



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escuela Técnica
Superior de Ingeniería
Informática

APLICACIÓN DE TÉCNICAS DE EXTRACCIÓN DE CONOCIMIENTO EN DATOS INMOBILIARIOS

Autor:

José Ramón Verdeguer Ferrero

Directores:

Cèsar Ferri Ramírez

Antonio Bella Sanjuán

Titulación:

Ingeniería Técnica en Informática de Sistemas

Fecha de entrega:

Septiembre 2012

Resumen:

Este proyecto trata de extraer conocimiento útil y novedoso de los datos de las viviendas almacenados en un portal web inmobiliario. Para lo cual, se preparan dichos datos de las viviendas y se realizan unos modelos de datos con ellos. Concretamente, se realizan tres modelos de datos, uno de ellos tiene como objetivo predecir la compra o la no compra de las viviendas. Los otros dos modelos tienen como objetivo predecir el precio de las viviendas vendidas. En uno de ellos no se tiene en cuenta los días que tarda en realizarse la venta y en el otro sí se tiene en cuenta tales días.

Por otra parte, cada uno de los tres modelos es evaluado con los diferentes algoritmos de la herramienta de minería de datos “WEKA”.

Palabras clave:

KDD, knowledge discovery in databases, descubrimiento de conocimiento en bases de datos, fase de preparación, vista minable, fase de minería de datos, modelos de datos, tarea predictiva, fase de evaluación, cross validation, validación cruzada, fase de interpretación y visualización, WEKA, Waikato Environment for Knowledge Analysis.

ÍNDICE TEMÁTICO

1. Introducción.....	4
2. Obtención de los datos.....	10
3. Aprendizaje de datos.....	58
4. Evaluación.....	131
5. Conclusiones.....	149
6. Referencias.....	153
Apéndice 1. La construcción del modelo 1.....	154
Apéndice 2. La construcción del modelo 2.....	170
Apéndice 3. La construcción del modelo 3.....	173
Apéndice 4. Programas propios.....	177
Apéndice 5. Organización de la carpeta del proyecto.....	181

1. Introducción

En este proyecto denominado “Aplicación de técnicas de extracción de conocimiento en datos inmobiliarios” se pretende extraer conocimiento útil y novedoso a partir de los datos de las viviendas almacenados en un portal web inmobiliario. Para lo cual utilizaremos una serie de técnicas (semi-)automáticas, que estudiarán y analizarán dichos datos, con la intención de obtener conclusiones que serían imposibles de percibir de forma manual, debido a la gran cantidad de datos. Cuantos más datos de viviendas tengamos a nuestra disposición, más elaborado y mejor podrá ser el análisis que se realice sobre ellos, siempre y cuando dichos datos sean fiables.

Del párrafo anterior una cuestión importante que se puede decir es que, la credibilidad del conocimiento que se pueda obtener de las viviendas, dependerá en gran parte de la calidad de los datos de las viviendas y también del uso que se haga de ellos. Ya que, si por ejemplo, los datos de las viviendas son de mala calidad y tienen incongruencias entre ellos, por muy buenas técnicas de análisis que se utilicen sobre ellos, sólo se puede esperar que se obtenga un conocimiento inadecuado acorde con esos malos datos. Por otro lado, si se hace un mal uso de los datos, aunque estos sean buenos tampoco se encontrará un conocimiento adecuado o por lo menos no todo lo bueno que se podría esperar.

En cuanto a la fuente de información de las viviendas, como hemos dicho utilizaremos un portal web inmobiliario que, evidentemente, habremos seleccionado. Sin embargo, lo ideal sería obtener los datos de la fuente primaria de la que se nutre el portal web, ya que, debe ser, seguramente, una base de datos o más raramente unos ficheros de datos. De esa forma la recogida de la información sería más cómoda y fácil, porque el contenido al que accederíamos estaría formado únicamente por los datos de las viviendas y no habría que preocuparse de encontrarlos entre las páginas del portal web, en las que además de los datos de las viviendas, aparece también otro tipo de información, como por ejemplo etiquetas “HTML”. Añadido a esto, puede ocurrir también que la interfaz (el portal web) cambie la estructura y organización que presenta, de tal forma, que se requieran variaciones para obtener los datos de las viviendas.

1.1 El proceso “KDD”

El proceso que se utilizará en este proyecto para realizar el análisis de datos será el denominado “KDD”, “Knowledge Discovery in Databases”, que en español viene a ser “Descubrimiento de Conocimiento en Bases de Datos”. El “KDD” es un proceso que

tiene como finalidad extraer conocimiento útil a partir de datos. El “KDD” está constituido por una serie de fases. La fase de preparación es la primera fase. En dicha fase se tratan los datos para que estén en óptimas condiciones para las siguientes fases del proceso. En esta fase se recopilan los datos, se unifican y se analiza el dominio de valores de cada uno de los datos. También se eliminan posibles errores en los datos, esto frecuentemente se llama limpiar los datos. Asimismo se eligen o seleccionan los datos eliminando los atributos no necesarios y los ejemplares que no se requieren. Además también puede hacer falta transformar los datos con la intención de que aparezcan con un formato distinto, sin que afecte por ello, al significado de los mismos.

En este proyecto, para almacenar la información de las viviendas y los diversos tratamientos que hagamos sobre ellas, utilizaremos ficheros de texto y no sobrescribiremos ninguno. Por lo tanto, no vamos a utilizar la tecnología de almacenes de datos, para almacenar y preparar la información de las viviendas. Los almacenes de datos ayudan a preparar los datos y a organizarlos de manera óptima según nuestras necesidades. Por ejemplo, disponibilidad de tener la información en forma más agregada o por el contrario de forma más disgregada. En definitiva los almacenes de datos ayudan a manejar los datos sobre todo cuando disponemos de gran cantidad de datos, sin embargo, como hemos dicho no los vamos a utilizar.

La fase de preparación de los datos termina cuando se obtiene la vista minable deseada. La vista minable se define como: el subconjunto de datos que se obtienen del tratamiento o preparación de un conjunto de datos preexistente.

La fase siguiente a la de preparación es la de minería de datos. En esta fase se construye el modelo a partir de los datos de la vista minable creada en la fase de preparación. El tipo de modelo a construir en la fase de minería de datos puede ser un modelo predictivo o un modelo descriptivo. Los modelos de datos de tipo predictivo tienen como finalidad encontrar patrones de comportamiento que nos ayuden a predecir datos desconocidos. En cambio, los modelos de datos de tipo descriptivo lo que buscan es encontrar patrones de comportamiento que resuman los datos existentes. Las tareas que se realizan en los modelos predictivos esencialmente son tareas de clasificación y de regresión y las tareas que se realizan en los modelos descriptivos son principalmente tareas de agrupamiento, de reglas de asociación y de correlaciones.

Una vez se tiene claro el tipo de modelo a construir se utiliza una herramienta de minería de datos como por ejemplo “WEKA” para construir el modelo. Se elige uno de los algoritmos de minería de datos que esté disponible en la herramienta que esté acorde con la tarea a realizar, es decir, que pueda realizar la tarea y así construir el modelo a partir de los datos. El algoritmo de forma automatizada construirá el modelo de datos con las técnicas que tenga a su disposición, por ejemplo, puede basarse en árboles de decisión, en redes neuronales, en la fórmula de bayes, en reglas o en vecinos más próximos entre otras técnicas. Sólo se tendrán que configurar los parámetros que posea el algoritmo antes de ejecutarlo, si es que no se desean los parámetros por defecto de que disponga el algoritmo, y éste automáticamente construirá el modelo de datos.

La siguiente fase del proceso “KDD” es la fase de evaluación. En esta fase se comprueba la validez del modelo. Se tienen en cuenta parámetros como la precisión del modelo, la comprensibilidad, la utilidad y la novedad del conocimiento extraído.

La fase de interpretación y visualización del modelo será la última de las fases. En dicha fase se interpretarán los patrones construidos.

Después de la última fase se obtiene la salida del proceso que es el conocimiento nuevo, que estará clarificado y comprobado por las fases de evaluación e interpretación por las que ha transitado. Una vez obtenido dicho conocimiento se puede explotar, es decir, difundirlo, utilizarlo y tomar decisiones a partir de él.

Por otra parte, comentar que el proceso “KDD” es un proceso iterativo lo que permite volver las veces que sea necesario a fases anteriores del proceso. Por ejemplo, después de extraer conocimiento y haberlo evaluado en la fase de evaluación se podría volver a las fases anteriores para volver a realizar el proceso o parte del proceso teniendo en cuenta el nuevo conocimiento extraído. O también, por ejemplo, volver a la fase de minería de datos, desde la fase de evaluación o desde de la fase interpretación y visualización, tantas veces como se quiera, para realizar un mismo modelo con otro algoritmo o con el mismo algoritmo pero con distintos parámetros de configuración y así obtener cada vez resultados distintos.

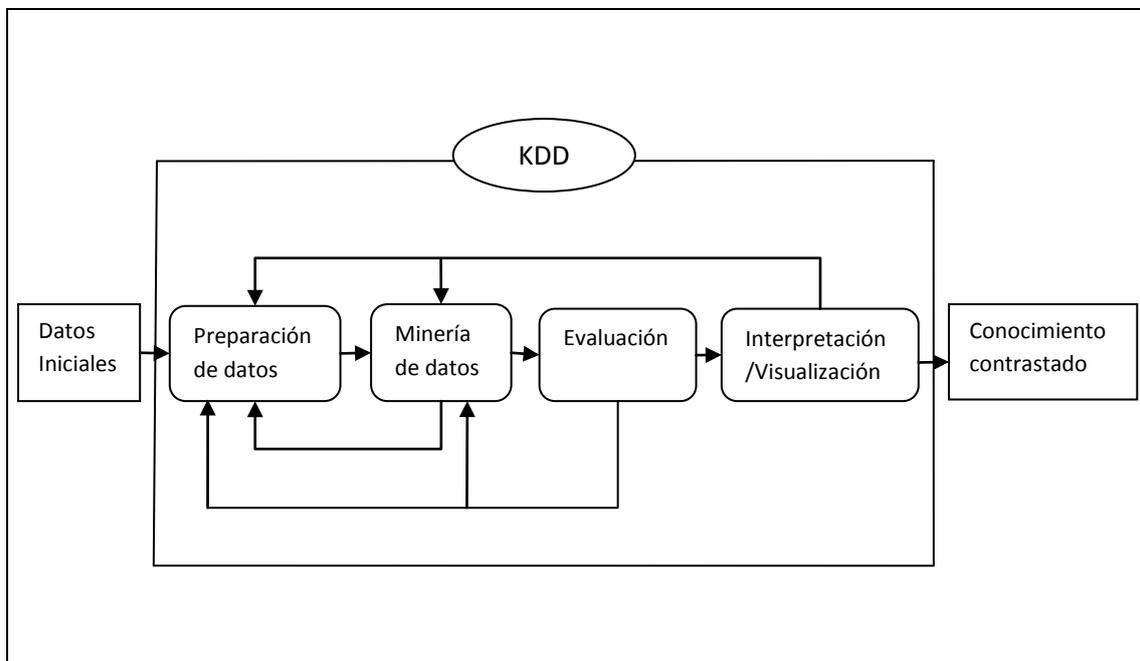


Diagrama 1.1: Las fases del proceso “KDD”.

1.2. La herramienta de análisis de datos “WEKA”

En cuanto a la herramienta de minería de datos que vamos a emplear en este proyecto, llamada “WEKA” (Waikato Environment for Knowledge Analysis, Entorno para Análisis de Conocimiento de la Universidad de Waikato) decir que es una herramienta de extracción de conocimiento de minería de datos y de aprendizaje automático. Dicha herramienta es de software libre y está desarrollada en el lenguaje de programación ‘Java’.

Cuando se ejecuta la herramienta “WEKA”, lo primero que suele captar la atención, es la imagen representativa de “WEKA”, que es un ave que se denomina del mismo modo que la herramienta ,es decir, se llama también “WEKA”, aunque de forma científica adopta el nombre de ‘Gallirallus australis’. Concretamente, dicho ave es un pájaro típico, aunque en peligro de extinción, del país Nueva Zelanda, país donde está ubicada la universidad ‘Waikato’. Además del ave también se aprecian cuatro aplicaciones o interfaces que se denominan ‘Explorer’, ‘Experimenter’, ‘KnowledgeFlow’ y ‘Simple CLI’.

En este proyecto de entre todas las aplicaciones disponibles de “WEKA” vamos a utilizar sólo la aplicación ‘Explorer’. La aplicación ‘Explorer’ presenta un total de seis pestañas llamadas: ‘Preprocess’, ‘Classify’, ‘Associate’, ‘Cluster’, ‘Selected Attributes’ y ‘Visualize’. En este proyecto vamos a utilizar principalmente la pestaña ‘Preprocess’ y la pestaña ‘Classify’.

La pestaña ‘Preprocess’ nos permite cargar y guardar los datos y realizar una serie de filtros a los datos, como filtros de selección de atributos y filtros de selección de instancias. También nos permite observar los valores de los datos.

En la pestaña ‘Classify’ tenemos los algoritmos de minería de datos para construir modelos de predicción de tareas de clasificación o de regresión.

La pestaña o panel ‘Associate’ nos ofrece la posibilidad de realizar reglas de asociación.

La pestaña ‘Cluster’ nos permite realizar agrupamientos. La pestaña ‘Selected attributes’ posee métodos que permiten reconocer que atributos tienen más importancia a la hora de predecir el valor de otro atributo utilizado como salida.

La pestaña ‘Visualize’ nos permite opciones de visualización que facilitan el análisis de los datos. En definitiva la aplicación ‘Explorer’ de “WEKA” nos va a facilitar el proceso de extracción de conocimiento.

Por último una característica importante que hay que resaltar de la herramienta “WEKA” es que el formato de fichero nativo que adopta es el denominado “arff” acrónimo de “Attribute-Relation File Format”. Este tipo de formato “arff” lo explicaremos más adelante en este proyecto, sólo en esta sección de introducción

señalar que, son ficheros de texto que tienen como peculiaridad que disponen de una sección en la que se describen como deben ser los datos que se introducen en ellos.

1.3. Modelos de las viviendas

En este proyecto, como se ha dicho, se quiere extraer conocimiento de los datos de las viviendas de un portal web inmobiliario. Para lo cual vamos a realizar tres modelos predictivos a partir de los datos que obtengamos del portal web inmobiliario. Para la realización de los tres modelos tendremos bien presente en todo momento el proceso “KDD”, y la aplicación de minería de datos “WEKA” que la utilizaremos para la realización de los modelos.

Los tres modelos de datos a realizar se citan seguidamente:

- Modelo1: Hallar la compra o la no compra de una vivienda.
- Modelo2: Precio por el que vender una vivienda independientemente del número de días.
- Modelo3: Precio por el que vender una vivienda teniendo en cuenta el número de días.

Ahora que hemos citado los tres modelos de datos vamos a decir que entendemos por un modelo y por un modelo de datos predictivo:

Un modelo es una representación de algún hecho o situación. En este proyecto, se intenta que los modelos sean lo más reales posible a lo que quieren emular.

Un modelo de datos predictivo es un modelo cuyo hecho o situación son los datos que se le suministran. El modelo de datos se realiza a partir de dichos datos e intenta encontrar unos patrones en tales datos que se le suministran, para posteriormente, sobre otros datos que se le pasarán al modelo obtener unas predicciones.

La creación de un modelo de datos predictivo consistirá en la construcción del modelo y la prueba del modelo:

- La construcción del modelo pretende implementar el modelo con los datos de entrada disponibles para tal cometido. Dichos datos de entrada del modelo se suelen denominar experiencias del conjunto de entrenamiento, debido a que sirven para entrenar al modelo. Cuando el modelo se haya implementado y por tanto se haya entrenado, éste habrá aprendido una función predictiva, por ejemplo, una clasificación o regresión, que servirá para predecir la salida de cualquier entrada que se le pase a la función. A la salida de la función se le denomina clase y a las entradas se les denomina variables explicativas. Un apunte que habría que decir es que, lógicamente, las experiencias disponibles para la construcción del modelo disponen de los datos de entrada y de los correspondientes datos que se consideran de salida.

- La prueba del modelo se utiliza para evaluar la credibilidad de la función predictiva creada en la construcción del modelo. Para ello se disponen de experiencias para el conjunto de prueba, que no necesariamente deben ser distintas a las del conjunto de entrenamiento. La cuestión es que para cada experiencia del conjunto de prueba se predecirá la clase. Y como se disponía de la clase real de cada una de las experiencias de prueba, entonces se puede realizar la comparación de la clase real con la predicha en cada una de las ellas. Con lo que, se podrá llegar a la finalidad de la prueba del modelo que es evaluarlo mediante el conjunto de prueba. Idealmente los valores de la clase real y los predichos deberían ser iguales.

1.4. Estructura de la memoria

A continuación vamos a comentar brevemente el contenido de la memoria:

En el apartado “2. Obtención de los datos”, estaremos en la fase primera del proceso “KDD”, preparación de datos. En este apartado obtendremos los datos del portal web inmobiliario y los trataremos para que sean lo más idóneos posibles para cada uno de los modelos de datos. Los trataremos de forma diferente hacia el final de la fase de preparación, según para el modelo de datos que se estén preparando los datos, teniendo al final tres vistas minables: una del modelo 1, una del modelo 2 y otra del modelo 3.

En el apartado “3. Aprendizaje de datos”, trataremos la fase de minería de datos del proceso “KDD”. En esta fase para cada uno de los tres modelos realizaremos un estudio de algoritmos, que consistirá en crear cada uno de los tres modelos varias veces, cada una de las veces, con un algoritmo distinto que nos ofrezca la aplicación “WEKA” para realizar el modelo correspondiente. Por otra parte, adicionalmente, utilizaremos el modelo 3, para tratar de predecir el precio por el que vender una misma vivienda en función de los días.

En el apartado “4. Evaluación”, llegaremos a la fase de evaluación del proceso “KDD”. En este apartado evaluaremos los algoritmos del estudio del modelo 1, del estudio del modelo 2 y del estudio del modelo 3.

Después aparecerá el apartado “5. Conclusiones” que aportará las conclusiones del proyecto y otro llamado “6. Referencias” que mostrará las referencias utilizadas.

Por último aparecerán cinco apéndices denominados “Apéndice 1. La construcción del modelo 1”, “Apéndice 2. La construcción del modelo 2”, “Apéndice 3. La construcción del modelo 3”, “Apéndice 4. Programas propios” y “Apéndice 5. Organización de la carpeta del proyecto”: Los tres primeros de estos apéndices como el nombre indica mostrarán la construcción elegida del modelo 1, del modelo 2 y del modelo 3 respectivamente. El apéndice cuarto tratará sobre los programas propios realizados y el apéndice cinco sobre la organización de la carpeta del proyecto.

2. Obtención de los datos

En esta sección vamos a tratar la fase de preparación de datos que es la primera del proceso “KDD”. Nos interesa obtener la información de las viviendas, para posteriormente realizar unos modelos de predicción sobre ellas. En este proyecto, la información de las viviendas la obtendremos de los datos de las viviendas que aparecen en el portal web en una fecha determinada primera, que denominaremos “fecha primera”. Lógicamente la fecha puede ser cualquier fecha. No obstante, como veremos en este mismo apartado además se precisará obtener los datos de las viviendas que aparecen en el portal web en otra fecha determinada segunda, “fecha segunda”. Esta última obtención de datos tendrá como finalidad, como veremos más adelante en este apartado, intentar resolver más información acerca de las viviendas que no habremos conseguido con la obtención de los datos en la “fecha primera” determinada. Todo esto lo explicaremos más adelante detalladamente.

En este apartado vamos a relatar el proceso de preparación y al final del apartado citaremos los datos concretos que recogeremos.

En la fase de preparación crearemos la vista minable particular de cada uno de los modelos de datos a tratar citados previamente en el apartado anterior. Para que la construcción de los modelos sea lo más adecuada posible se requiere que los datos que se utilicen estén lo más claros posibles y que se utilicen solamente los datos esencialmente necesarios. Esto será lo que desearemos alcanzar al llegar a la vista minable particular en cada uno de estos modelos. En el proceso para conseguirla en cada uno de los modelos, lo que realizamos en este proyecto, en primer lugar será alcanzar una vista de datos previa en común para los modelos que llamaremos vista de datos general. Cuando alcancemos lo que hemos dicho, la vista de datos general, seguiremos por separado el proceso de creación de la vista minable particular de cada uno de los modelos. Entonces, queda claro que lo primero que relataremos será como se llega a dicha vista de datos general común a todos los modelos. Para ello lo primero que tendremos en cuenta serán los datos o atributos que se necesiten y tengamos disponibles de las viviendas, cómo recopilar estos atributos, y las acciones a realizar para recogerlos, tratarlos y organizarlos. Y una vez llegados a la vista de datos general citaremos la preparación que realizaremos por separado de los datos para la creación de cada uno de los modelos.

2.1. Atributos disponibles

En primer lugar, vamos a citar los atributos disponibles que queremos para describir cada vivienda:

1. El identificador de la vivienda: a este dato o atributo lo llamaremos “identificador”, este atributo identifica inequívocamente a cada vivienda. Cada vivienda, por lo tanto, tiene un valor distinto para el atributo “identificador”. Este atributo es útil para diferenciar claramente una vivienda de otra.
2. El tipo de la vivienda: a este atributo lo denominaremos “tipo”.
3. El barrio donde está ubicada la vivienda: a este atributo lo llamaremos “barrio”.
4. El número de dormitorios de que dispone la vivienda: a este atributo lo llamaremos “dorm”.
5. La planta donde está instalada la vivienda: a este atributo lo llamaremos “planta”.
6. El número de metros cuadrados que tiene la vivienda: a este atributo lo llamaremos “metros”.
7. El precio del metro cuadrado de la vivienda: a este atributo lo denominaremos “euros/metro”.
8. El precio total de la vivienda: que llamaremos “euros”.
9. El estado en que se encuentra la vivienda: que lo nombraremos como “estado”.
10. La ciudad donde está la vivienda: que lo denominaremos “ciudad”.
11. La compra o no compra de la vivienda dependiendo de si ha sido vendida o no ha podido ser vendida la vivienda: a este atributo lo llamaremos “compra”.
12. El número de días que ha estado la vivienda en el portal web, hasta el momento de la venta o finalmente de la no venta: a este atributo lo denominaremos “días”.

Una vez sabemos los atributos disponibles que deseamos para describir las viviendas relatamos como encontrarlos en el portal web que disponemos como fuente de datos.

Al buscar en el portal web los atributos que requerimos encontramos la información que necesitamos referente a los atributos “identificador”, “tipo”, “barrio”, “dorm”, “planta”, “metros”, “euros/metros”, “euros” y “ciudad”, aunque no encontramos el atributo “compra”, ni el atributo “días” que hemos dicho que los queremos y que los tenemos disponibles. Estos dos atributos “compra”, “días” los tendremos a nuestra disposición no obteniéndolos directamente del portal web sino de otra forma distinta.

En cuanto al atributo “compra” para obtenerlo, en este proyecto, lo que realizaremos será realizar una segunda extracción de los datos con una diferencia entre ‘15’ y ‘20’

días entre la extracción primera. Precisamente esta segunda extracción de datos es la que se comentaba al principio de este apartado, y ahora ya sabemos porque se realiza. Esta segunda extracción de datos tiene como finalidad observar la evolución en el tiempo de los datos de las viviendas residentes en el portal web, es decir, percibir el dinamismo o las variaciones en los datos de las viviendas, entre la primera extracción y la segunda, para poder estimar tal atributo “compra”. Concretamente las variaciones que queremos observar serán las instancias de viviendas que: desaparecen o tienen un valor distinto respecto al atributo “euros” o que aparecen después de la primera extracción.

Además para obtener el atributo “compra” nos fijaremos también en otros datos que todavía no hemos mencionado. Uno de ellos será un atributo que aparece en la información de las viviendas en el portal web pero que aunque no lo hemos incluido en la información que hemos dicho que utilizaríamos en este proyecto para describirlas, sí que nos va hacer falta para poder estimar el atributo “compra”. Este atributo es la fecha de actualizado que lo llamaremos “fechaActualizado”. El otro dato que necesitaremos no aparece en el portal web y es la fecha en la que recopilamos los datos de las viviendas, a este dato lo denominamos “fechaExtracción” y se obtiene del computador.

Como decíamos al principio del apartado, a la primera vez en que extraemos la información de las viviendas la denominaremos “fecha primera” y a la segunda vez “fecha segunda”.

Dicho esto supondremos lo siguiente para estimar el atributo “compra”:

- Si una instancia de vivienda recogida en la “fecha primera”, tiene un valor referente al atributo “fechaActualizado”, que dista de seis meses o más respecto a la “fecha segunda” de extracción, concluimos que la vivienda no ha sido vendida, porque decimos que ha caducado.
- Si no, supondremos que si una instancia de vivienda vista en el portal web en la “fecha primera”, deja de ser vista en la “fecha segunda”, es porque ha sido vendida.
- Si en vez de lo anterior, apreciamos que la instancia de la vivienda permanece en la “fecha segunda”, pero resulta que el atributo “euros” presenta una variación, es decir el precio ha cambiado desde la “fecha primera” en que contemplamos la instancia de la vivienda, entonces concluimos que la instancia de vivienda vista en la “fecha primera” no ha sido vendida, porque ha cambiado el precio de la vivienda.
- Por último si la instancia de la vivienda permanece en el portal web en la “fecha segunda” y el atributo “euros” aparece sin variaciones, entonces concluimos que todavía no se ha declarado la vivienda como vendida o como no vendida, por lo que, no vamos a estimar el atributo “compra” en esta última situación.

Al respecto de la estimación de tal atributo “compra” destacar que: nos quedaremos con las instancias de viviendas a las que se pueda estimar el atributo “compra” con las suposiciones anteriores. Las instancias de viviendas a las que no se les ha podido estimar el atributo “compra” las desecharemos, por lo que habrá una tarea de selección.

Del párrafo anterior se puede deducir que las instancias de viviendas que nos interesan son las recogidas en la “fecha primera”, ya que, a las instancias de viviendas extraídas en la “fecha primera” es a las que intentamos estimar el atributo “compra”. Por el contrario, las instancias de viviendas extraídas en la “fecha segunda”, las requerimos, para compararlas con las instancias extraídas en la “fecha primera”, para así intentar estimar el atributo “compra” a éstas últimas. Sin embargo, al resto de instancias de viviendas que no se les ha podido estimar el atributo “compra” las denominaremos resto y las guardaremos aparte. Estas instancias de viviendas que llamaremos resto serán las siguientes:

- Las instancias de viviendas extraídas en la “fecha primera” que no se les haya podido estimar el atributo “compra”.
- Las extraídas en la “fecha segunda” que son nuevas ofertas de instancias de viviendas no caducadas almacenadas en la “fecha primera”.
- Las extraídas en la “fecha segunda” que no aparecen en la extracción realizada en la “fecha primera”.

Respecto a la situación referida a las viviendas que no se les ha podido obtener el atributo “compra” de la “primera fecha” hay que resaltar que es la situación más frecuente, debido a que suele ser bastante común que no aparezcan variaciones entre las instancias de viviendas vistas en la “fecha primera” y en la “fecha segunda”. No obstante, afortunadamente también aparece un gran número de instancias de viviendas a las que se les puede estimar el atributo “compra” y por tanto disponer de ellas para realizar los modelos.

Ahora vamos a citar respecto a las viviendas que se les ha obtenido el atributo “compra”, el momento en que suponemos que se ha realizado la compra si han sido vendidas, o el momento en el que se ha declarado la no compra de la vivienda si no lo han sido:

- Respecto al momento de la compra, si la vivienda tiene el atributo “compra”=’SÍ’ decimos que ha sido vendida entre la “fecha primera” que extraemos los datos y la “fecha segunda” que volvemos a extraer los datos. Y por tanto, el periodo que ha durado la venta de la vivienda es desde el atributo “fechaActualizado” que aporta la vivienda, hasta la declaración de la venta de la vivienda.

- También decimos que cualquier vivienda que tiene el atributo “compra”=’NO’ ha sido declarada como no vendida. En este caso si la no venta ha sido porque la vivienda ha caducado, decimos que la no venta se ha declarado, precisamente, seis meses después del atributo “fechaActualizado”. No obstante, si ha sido no vendida, no porque ha caducado, sino, porque ha variado el precio de la vivienda entre la “fecha primera” y la “fecha segunda”, entonces decimos que la fecha de no venta de la vivienda que su instancia aparece en la “fecha primera” es la fecha que aparece en el atributo “fechaActualizado” de la misma vivienda que aparece instanciada en la “fecha segunda”

con otra oferta. A semejanza del periodo que ha durado la venta de una vivienda vendida, el periodo que ha durado la no venta de una vivienda es desde el atributo “fechaActualizado” que aporta la vivienda, hasta la fecha de no venta de la vivienda.

Una pregunta que nos podemos hacer es: ¿por qué a las viviendas vendidas decimos que se han declarado como vendidas o no vendidas precisamente en la media de la fecha de extracción de la “fecha primera” y la “fecha segunda”?

Esto se ha hecho así porque la venta o no venta de dichas viviendas podría haberse provocado en cualquier día comprendido entre ambas fechas. Por lo tanto, lo más acertado es decir que es en la media de las dos fechas. Otra cuestión referente a la media es que es posible que sea divisible entre dos o que no sea divisible entre dos. Si no es divisible entre dos la media queda entre dos fechas, y como resolvemos la media en este caso es escogiendo la fecha del día inferior entre los dos días entre los que queda la media.

Por último, decir que el cálculo de este momento lo almacenaremos en las instancias de las viviendas que se les ha obtenido el atributo “compra” y lo nombraremos como atributo “FechaDeVenta/NoVenta”. Tal atributo remplazará al atributo “fechaExtracción”, que ya, no nos va a hacer falta para estas viviendas que se les ha podido estimar el atributo “compra”. No obstante, a las viviendas que no se les ha podido estimar el atributo “compra” no se les elimina el atributo “fechaExtracción”. Eso sí, todas estas viviendas que no tienen el atributo “compra” adoptarán como valor para el en el atributo “fechaExtracción” la “fecha segunda” de extracción, ya que, hasta dicha fecha no podemos decir nada de ellas respecto al atributo “compra”, y lógicamente, no tendrán el atributo “FechaDeVenta/NoVenta”, por no poder decir nada sobre la venta o no venta de tales viviendas al carecer del atributo “compra”.

Por último, volviendo a la estimación del atributo “compra”, vamos a relatar unas situaciones anormales que no las hemos citado antes porque son prácticamente inapreciables:

- La primera situación anormal es que de forma inesperada ocurra que a una instancia de una vivienda vista en la “fecha primera” que se le vaya a intentar estimar el atributo “compra” tenga el atributo “fechaActualizado” vacío, con lo que, no se puede saber desde cuando permanece la instancia de la vivienda en el portal web. Esto último provoca que no se pueda realizar íntegramente el proceso de intentar estimar el atributo “compra”, ya que, no se puede deducir la parte de si la vivienda ha caducado o no. Esto es porque al no saberse el atributo “fechaActualizado” no se puede determinar si la instancia de la vivienda vista en la “fecha primera” dista seis o más meses desde la “fechaActualizado” de la vivienda hasta la “fecha segunda” de extracción. Por lo tanto, si inesperadamente se aprecia alguna instancia de vivienda extraída en la “fecha primera” que posea el atributo “fechaActualizado” vacío se ignorará dicha instancia de vivienda.

- Otra situación anormal más extraña, y que no se ha observado afortunadamente, sería que en alguna instancia de vivienda vista en la “fecha primera” el atributo “identificador” apareciera vacío. Si esto ocurriera, entonces, esta situación también provocaría que no se pudiera realizar íntegramente el proceso de intentar estimar el atributo “compra”. Esta vez sería porque como el atributo “identificador” estaría vacío no se podría buscar la instancia de la vivienda en las instancias de viviendas extraídas en la “fecha segunda”, precisamente, porque se buscan mediante el atributo “identificador”. Por lo tanto si el atributo “identificador” de la instancia de la vivienda vista en la “fecha primera” apareciera vacío entonces se ignoraría.

- Por último otra anomalía que afortunadamente tampoco se ha observado sería que el atributo “euros” de una instancia de vivienda vista en la “fecha primera” apareciera vacío. De las instancias de viviendas de la “fecha segunda” se considera que no pueden tener ninguna anomalía, ya que, no van a afectar al proceso de comparación. Por lo tanto no se considerará como anomalía que pudieran tener el atributo “identificador” vacío o el atributo “fechaActualizado” vacío o que el atributo “euros” lo tuvieran vacío. En este último caso sí que podría pensarse que afectaría al proceso de comparación, no obstante, aunque no se ha visto ninguna instancia de vivienda de la “fecha segunda” con el atributo “euros” vacío, si esto ocurriera: al comparar el precio de la instancia de la vivienda de la “fecha segunda” con la misma instancia de vivienda de la “fecha primera” se consideraría que serían ofertas distintas de una misma vivienda y se concluiría que la vivienda de la “fecha primera” sería no vendida.

Ahora que hemos explicado como estimar el atributo “compra” vamos a decir cómo obtener el atributo que nos falta de los que hemos dicho que tenemos disponibles y queremos para describir las vivienda, que es el atributo “días”. Como hemos citado nos quedamos con las viviendas que tienen el atributo “compra” y el resto las dejamos apartadas, con lo que queremos averiguar el atributo “días” a dichas viviendas seleccionas.

El atributo “días” lo hallaremos, precisamente, a partir de los dos periodos que hemos citado con anterioridad cuando hablamos del periodo que dura la venta o el periodo que dura la no venta. Por lo tanto, para calcular el atributo días a las viviendas utilizaremos el atributo “fechaActualizado” y el atributo “FechaDeVenta/NoVenta”. La diferencia en días entre el atributo “fechaActualizado” y el atributo “FechaDeVenta/NoVenta” nos resolverá el atributo “días”, para las viviendas seleccionadas.

Con el cálculo del atributo “días” tendremos los atributos disponibles que deseamos para describir las viviendas.

2.2. Recopilación de datos desde el portal web

Una vez sabido como conseguir los atributos que necesitamos para describir las viviendas vamos a relatar precisamente el proceso de recopilación de dichos atributos de forma cronológica:

1. La primera tarea será recopilar las instancias de las viviendas. Para lo cual accederemos al portal web en la “fecha primera” para recoger la información de las viviendas referida a los atributos “identificador”, “tipo”, “barrio”, “dorm”, “planta”, “metros”, “euros/metros”, “euros”, “estado”, “fechaActualizado” y “ciudad”, de cara a almacenarla en un fichero de texto. Además, también guardaremos en cada una de las instancias de las viviendas el atributo “fechaExtracción” obtenido del computador. En cuanto al formato del contenido en el fichero de almacenamiento se utiliza para cada instancia de vivienda una línea para depositar los valores de sus atributos, es decir, cada vivienda será un registro del fichero de texto. Los valores de los atributos de cada vivienda aparecerán separados por un punto y coma, inclusive el valor del último atributo que aparezca, que también tendrá un punto y coma de separación. El orden en el que figurarán los valores de los atributos de cada vivienda en cada una de las líneas del fichero de texto creado para la extracción será el siguiente: en primer lugar el valor del atributo “identificador” de la vivienda y después los valores de los atributos de la vivienda “tipo”, “barrio”, “dorm”, “planta”, “metros”, “euros/metros”, “euros”, “estado”, “fechaActualizado”, “ciudad” y “fechaExtracción”.

El fichero de extracción se denominará “viviendasdd-mmm-yyyy.txt”, en donde, “vivienda” es una cadena constante que forma parte del nombre y “dd-mmm-yyyy” es la fecha en la que se crea el fichero, que por cierto será la misma fecha que la del atributo “fechaExtracción”. En tal fecha “dd” es el día, “mmm” el mes, formado por las tres primeras letras del mes en el idioma inglés, y “yyyy” el año. Dicha parte “dd-mmm-yyyy” nos ayudará a diferenciar un fichero “viviendad1d1-m1m1m1_-y1y1y1y1.txt” de otro “viviendad2d2-m2m2m2-y2y2y2y2 .txt” si han sido creados en fechas de tiempo distintas. La parte del nombre precedida por el punto es la extensión del fichero e indica que es un fichero de texto con formato “.txt”. El nombre del fichero es “viviendadd-mmm-yyyy.txt”, como podía haber sido otro, pero como hay que ponerle un nombre le hemos puesto este. Un fichero abstracto de este tipo de ficheros que almacenan las instancias de viviendas extraídas del portal web se aprecia seguidamente:

```
identificador1;tipo1;barrio1;dorm1;planta1;metros1;euros/metro1;euros1;estado1;fechaActualizado1;ciudad1;fechaExtracción1;
identificador2;tipo2;barrio2;dorm2;planta2;metros2;euros/metro2;euros2;estado2;fechaActualizado2;ciudad2;fechaExtracción2;
.....;
identificadorN;tipoN;barrioN;dormN;plantaN;metrosN;euros/metrosN;eurosN;estadoN;fechaActualizadoN;ciudadN;fechaExtracciónN;
```

Recuadro 2.1: Fichero abstracto “Viviendasdd-mmm-yyy.txt”, en este fichero aparece ‘n’ instancias de viviendas.

Un ejemplo de un posible contenido de este tipo de ficheros se aprecia a continuación en el recuadro 2.2.

```
VP0000001111122;pisos;L'eixample;3;7;123;3.008;370.000;obra nueva;2008-12-26 10:38:59.0;valencia;27-nov-2009;
VP0000001111444;pisos;Algirós;3;7;120;2.875;345.000;;2009-11-10 12:08:50.0;valencia;27-nov-2009;
.....;
VP0000002149013;pisos;eixample;3;4;94;8.312;781.315;a reformar;2009-10-24 11:32:45.0;barcelona;27-nov-2009;
VP0000002872954;pisos;sarrià-sant gervasi;5;5;275;10.000;2.750.000;;2009-11-19 18:59:37.0;barcelona;27-nov-2009;
```

Recuadro 2.2: Ejemplo de un fichero “Viviendasdd-mmm-yyyy.txt”, exactamente, el ejemplo nos muestra ‘4’ instancias de viviendas del fichero “Viviendas27-nov-2009.txt”, ya que, el atributo “fechaExtracción” de dichas viviendas es ‘27-nov-2009’.

En cuanto al recuadro 2.2 vamos a describir los valores que presenta la primera instancia de vivienda, que se muestra en la primera línea, que tiene como “identificador” =’VP0000001111122’, que es el que mostraba en el portal web dicha vivienda. Además es una vivienda del tipo ‘piso’ del barrio de ‘L’eixample’ de la ciudad de Valencia, dispone de ‘3’ dormitorios y está situada en la planta ‘7’. Tiene ‘123’ metros cuadrados de superficie. En cuanto al metro cuadrado vale ‘3.008 €’, y el precio total es ‘370.000 €’. Además el estado es ‘obra nueva’ que nos informa que se ha realizado no hace mucho tiempo algún tipo de obra en la vivienda. Y la oferta de la vivienda se introdujo en el portal web el ‘26 de diciembre de 2008’ y por último que la hemos extraído el ‘27 de noviembre de 2009’.

2. Respecto a la segunda tarea, después de un periodo de tiempo en una “fecha segunda” determinada se vuelve a acceder al portal web inmobiliario para volver a almacenar las instancias de las viviendas del portal web inmobiliario en otro fichero de texto. Este fichero lo llamaremos “viviendasd’d-m’m’m’-y’y’y’.txt”, es decir, lo nombraremos de la misma forma que el fichero explicado anteriormente que llamamos “viviendasdd-mmm-yyyy.txt”, pero con la parte del nombre perteneciente a la fecha distinta. Almacenará las instancias de las viviendas de la misma forma que el fichero de la tarea anterior “viviendasdd-mmm-yyy.txt”. Es decir, es idéntico al fichero “viviendasdd-mmm-yyy.txt”, pero la extracción de las instancias de las viviendas se realizará en un periodo de tiempo posterior, y por lo tanto la fecha “d’d’-m’m’m’-y’y’y’y’”, que referencia a la “fecha segunda” de extracción es mayor que la fecha “dd-mmm-yyy” que referencia a la “fecha primera” de extracción.

2.3. Estimación del atributo “compra”

3. La tercera tarea será estimar el atributo “compra” con las deducciones que se relataban anteriormente cuando hablábamos del atributo “compra”. Para realizar dichas deducciones deberemos comparar el fichero “viviendasdd-mmm-yyyy.txt” creado en la primera tarea, y el fichero “viviendasd’d’-m’m’m’-y’y’y’y’.txt” creado en la segunda tarea. Los dos ficheros tienen las instancias de las viviendas del portal web, pero en periodos de tiempo distintos. El fichero “viviendasdd-mmm-yyyy.txt” que actuará como “fichero base” lo compararemos respecto al fichero “viviendasd’d’-m’m’m’-y’y’y’y’.txt” que actuará como “fichero comparar”.

Con “fichero base” queremos decir: que en este fichero aparecen las instancias de las viviendas a las que queremos estimar el atributo “compra”, que son precisamente las extraídas en la “fecha primera”.

Y con “fichero comparar” queremos decir: que tal fichero contiene las instancias de las viviendas de la “fecha segunda” y lo utilizamos para comparar sus instancias de viviendas con las instancias de las viviendas del “fichero base”.

Las instancias de las viviendas del “fichero base” que se les haya podido estimar tal atributo “compra” irán al fichero “VSINOFecha1-Fecha2.txt” y a las que no se les haya podido estimar de tal fichero se destinarán a otro fichero llamado “VRestoFecha1-Fecha2.txt”. A este último fichero “VRestoFecha1-Fecha2.txt” también irán las instancias de las viviendas del “fichero comparar” que sean ofertas nuevas de instancias de viviendas no caducas del “fichero base”, y también las instancias de viviendas del “fichero comparar” no presentes en el “fichero base”.

Uno se podría preguntar: ¿el por qué en una de las condición se dice que las nuevas ofertas de viviendas del “fichero comparar” no estén caducadas en el “fichero base”?

Y la respuesta sería porque si una vivienda del “fichero base” está caducada no vamos a intentar buscar si aparece en el “fichero comprar” con un precio distinto, ya que, suponemos que no va a aparecer la instancia de la vivienda con un precio distinto en el “fichero comparar” respecto al que presenta en el “fichero base”, ya que, después de permanecer inmóvil la instancia de la vivienda seis o más meses en el portal web parece improbable que varíe de repente en el “fichero comparar”. Además de esto también decimos en este proyecto, que cuando tenemos constancia que una vivienda ha caducado se acaba el ciclo de la vivienda y ya no la intentamos buscar más veces en otras ofertas, y por ello, aunque supuestamente no apareciera en el “fichero comparar”, igualmente, sería declarada como no vendida.

Una vez contado que instancias de viviendas van al fichero “VSINOFecha1-Fecha2.txt” y cuales instancias de viviendas van al fichero “VRestoFecha1-

Fecha2.txt” vamos a explicar el porqué del nombre que les hemos puesto a estos dos ficheros denominados “VSINOFecha1-Fecha2.txt” y “VRestoFecha1-Fecha2.txt”:

-A los ficheros que poseen las instancias de las viviendas con el atributo “compra” se les nombra como “VSINOFecha1-Fecha2.txt” debido a que la letra “V” en el que empieza esta clase de ficheros se extrae de la letra primera de la cadena “viviendas” y “SINO” viene a decir que las viviendas tienen el atributo “compra” estimado con dos posibles valores ‘SÍ’ o ‘NO’. La parte del nombre “Fecha1” es la “fecha primera” que aparece en el fichero “viviendasdd-mmm-yyyy.txt” que se utiliza como “fichero base” y sin los guiones de separación entre día, mes y año, es decir, Fecha1=ddmmmyyyy, y “Fecha2” es la “fecha segunda” que aparece en el fichero “viviendasd’d-m’m’m’-y’y’y’y’.txt” que actúa como “fichero comparar” también sin los guiones de separación es decir, Fecha2= d’d’m’m’y’y’y’y’.

- El nombre de la clase de ficheros que tienen el resto de las instancias de las viviendas que no se les ha podido estimar el atributo “compra” que es “VRestoFecha1-Fecha2.txt” se les pone este nombre debido a que la letra “V” con la que empiezan estos ficheros se extrae también de la primera letra de la cadena “viviendas” y la cadena “Resto” viene a decir que son el resto de las viviendas a las que no se les ha podido hallar el atributo “compra”. En el nombre de esta clase de ficheros: “Fecha1” y “Fecha2” referencian respectivamente a la “fecha primera” y a la “fecha segunda” de extracción.

En cuanto a la forma en que se almacenará la información en el fichero donde van a parar las viviendas que se les ha estimado el atributo “compra”, es decir, “VSINOFecha1-Fecha2.txt”, será de la siguiente forma: se copiarán las instancias de las viviendas que se les haya podido estimar el atributo “compra”, y tal atributo “compra” se añadirá al final de la información de cada una de dichas instancias de viviendas. Concretamente, después del punto y coma correspondiente al último atributo que había hasta entonces en las instancias de las viviendas que era concretamente el atributo “fechaExtracción”. Después de añadir el valor del atributo “compra” se añade otro punto y coma de separación. Por otro lado decíamos que el último atributo previamente a insertar el atributo “compra” era el atributo “fechaExtracción” porque para estas instancias de viviendas que se les ha otorgado el atributo “compra”, desaparece y en su lugar introducimos el atributo mencionado anteriormente cuando hablábamos de cómo poder estimar el atributo “compra”, “FechaDeVenta/NoVenta”.

Respecto a la forma de almacenarse, el resto de instancias de viviendas, que no se les ha podido obtener el atributo “compra” en el fichero “VRestoFecha1-Fecha2.txt” decir, que las instancias de viviendas que vayan a tal fichero “VRestoFecha1-Fecha2.txt” procedentes del “fichero base” se copiarán tal cual salvo el atributo “fechaExtracción” que se pondrá como fecha de extracción la fecha de extracción segunda. En cambio las instancias de viviendas que procedan del “fichero comparar” a tal fichero “VRestoFecha1-Fecha2.txt” sí se copiarán tal cual.

Ahora, como resumen de las distintas situaciones del atributo “compra” pondremos en primer lugar una tabla 2.1 que las señale, y luego otra tabla 2.2 que informe de lo que sucede con tales situaciones. Esta separación en dos tablas es debida entre otras cosas a que no cabe toda la información en una sola.

Previamente se muestran algunas aclaraciones referentes a algunas frases y valores que aparecen en dichas tablas y que pueden necesitar una explicación:

-La frase ¿Aparece en fichero base? : Indica si la instancia de la vivienda aparece en el “fichero base”.

-La frase denominada ¿Fecha caducada en fichero base? : Indica si la vivienda del “fichero base” ha caducado.

- La frase ¿Aparece en fichero comparar? : Indica si la instancia de la vivienda aparece en el “fichero comparar”.

- La frase ¿Variación en precio? : Señala si el precio de la vivienda difiere en el “fichero base” respecto al precio de la misma vivienda en el “fichero comparar”.

¿Comprada vivienda base?: Indica si la vivienda del “fichero base” ha sido comprada.

¿Vivienda base almacenada en?: Señala donde se almacena la instancia de la vivienda del “fichero base”.

¿Vivienda comparar almacenada en?: Señala donde se almacena la instancia de la vivienda del “fichero comparar”.

- El valor ‘Situación [1-5]’ viene a significar: Las posibles situaciones de una vivienda en el “fichero base” y en el “fichero comparar”.

- El valor ‘(-)’ viene a significar: Imposible de calcular porqué faltan datos, por ejemplo, no existe la instancia de la vivienda en el “fichero base” o en el “fichero comparar” y no se puede estimar, por lo tanto, si la fecha de la vivienda está caducada o si ha habido variación en el precio. También indica que la instancia de la vivienda del “fichero base o comparar” es imposible almacenarla debido a que no existe.

- El valor ‘Indiferente’ viene a significar: No importa el valor que tenga para el resultado.

Vivienda	¿Aparece en “fichero base”?	¿Fecha caducada en “fichero base”?	¿Aparece en “fichero comparar”?	¿Variación en precio?
Situación 1	SÍ	SÍ	Indiferente	Indiferente
Situación 2	SÍ	NO	SÍ	SÍ
Situación 3	SÍ	NO	SÍ	NO
Situación 4	SÍ	NO	NO	-
Situación 5	NO	-	SÍ	-

Tabla 2.1: Indica las distintas situaciones de la comparación.

Vivienda	¿Comprada vivienda base?	¿Vivienda base almacenada en?	Vivienda comparar almacenada en
Situación 1	NO	VSINOFecha1-Fecha2.txt	ningún lugar
Situación 2	NO	VSINOFecha1-Fecha2.txt	VRestoFecha1-Fecha2.txt
Situación 3	?	VRestoFecha1-Fecha2.txt	ningún lugar
Situación 4	SÍ	VSINOFecha1-Fecha2.txt	-
Situación 5	-	-	VRestoFecha1-Fecha2.txt

Tabla 2.2: Indica que ocurre con las situaciones de la tabla 2.1 respecto al atributo “compra”. Y también donde van a parar las instancias de las viviendas del “fichero base” y del “fichero comparar”, si es que existen en la situación a tratar.

En la tabla 2.1 se han indicado las distintas situaciones de la comparación ignorando las situaciones anormales que pueden aparecer en el “fichero base” que comentamos con anterioridad en este proyecto, ya que, esas situaciones como dijimos se dan raramente y por lo tanto, no las hemos expuesto en la tabla 2.1. Como dijimos de dichas situaciones anormales, la única que se ha apreciado, pero muy infrecuentemente es en la que en las instancias de las viviendas del “fichero base” aparezca el atributo “fechaActualizado” vacío. En cambio las situaciones anormales de que en las instancias de las viviendas del “fichero base” aparezca el “identificador” vacío o el atributo “euros” vacío no se han encontrado. No obstante podemos poner una tabla que indique dichas situaciones defectuosas del “fichero base”.

Vivienda del fichero base	¿“FechaActualizado” de la vivienda base	¿“Identificador” de la vivienda base?	¿“Euros” de la vivienda base?	¿“Vivienda base” almacena en?
Situación anormal 1	Vacío	indiferente	indiferente	ningún lugar
Situación anormal 2	no vacío	vacío	indiferente	ningún lugar
Situación anormal 3	no vacío	no vacío	Vacío	ningún lugar

Tabla 2.3: Situaciones anormales de las instancias de las viviendas del “fichero base”.

A continuación veremos cómo se distribuyen las instancias de las viviendas de ejemplo de dos ficheros “Viviendasdd-mmm-yyyy.txt”. De ellos se muestran ‘4’ instancias de viviendas como ejemplo. Exactamente se observan las dos primeras instancias de viviendas y las dos últimas. El recuadro 2.3 presenta el fichero “Viviendas27-nov-2009.txt” que actuará de “fichero base”. El recuadro 2.4 presenta el fichero “Viviendas13-dic-2009.txt” que actuará como “fichero comparar” del fichero “Viviendas27-nov-2009.txt”, ya que, el fichero “Viviendas13-dic-2009.txt” está creado en una fecha posterior al fichero “Viviendas27-nov-2009.txt” y dentro del periodo permitido de diferencia de fechas. Ya que, como decíamos el periodo permitido debe ser entre ‘15’ y ‘20’ días y entre estos dos ficheros hay una diferencia de ‘16’ días.

La instancia de vivienda que aparece en la fila segunda en el fichero “Viviendas27-nov-2009.txt” no está presente en el fichero “Viviendas13-dic-2009.txt” (ni en las ‘4’ instancia de viviendas que se observan del fichero “Viviendas13-dic-2009.txt” en el recuadro 2.4, ni en el resto de instancias de viviendas de dicho fichero que no se

observan en dicho recuadro 2.4.) La instancia de vivienda de la segunda fila que se presenta en el fichero en “Viviendas13-dic-2009.txt” no aparece en ninguna instancia de vivienda del fichero “Viviendas27-nov-2009.txt”. Además se aprecia otra variación entre los dos ficheros, concretamente, en la instancia de la vivienda de la fila penúltima de ambos ficheros; la instancia de vivienda es la misma porque presenta el mismo identificador, pero la oferta ha cambiado en el “fichero comparar”. En cambio la instancia de la vivienda de la fila primera perteneciente a ambos ficheros aparece exactamente igual en los dos ficheros, ídem con la última instancia de vivienda de ambos ficheros.

```
VP0000001111122;pisos;L'eixample;3;7;123;3.008;370.000;obra nueva;2008-12-26 10:38:59.0;valencia;27-nov-2009;
VP0000001111444;pisos;Algirós;3;7;120;2.875;345.000;;2009-11-10 12:08:50.0;valencia;27-nov-2009;
.....;
VP0000002149013;pisos;eixample;3;4;94;8.312;781.315;a reformar;2009-10-24 11:32:45.0;barcelona;27-nov-2009;
VP0000002872954;pisos;sarrià-sant gervasi;5;5;275;10.000;2.750.000;;2009-11-19 18:59:37.0;barcelona;27-nov-2009;
```

Recuadro 2.4: Fichero que actúa como base denominado: “Viviendas27-nov-2009.txt”. De este fichero se observan ‘4’ instancias de viviendas, las dos primeras y las dos últimas.

```
VP0000001111122;pisos;L'eixample;3;7;123;3.008;370.000;obra nueva;2008-12-26 10:38:59.0;valencia;13-dic-2009;
VP0000004444433;pisos;L'eixample;4;7;100;2.908;290.800;obra nueva;2009-12-12 12:24:23.0;valencia;13-dic-2009;
.....;
VP0000002149013;pisos;eixample;3;4;94;7.954;747.676;a reformar;2009-12-10 12:25:38.0;barcelona;13-dic-2009;
VP0000002872954;pisos;sarrià-sant gervasi;5;5;275;10.000;2.750.000;;2009-11-19 18:59:37.0;barcelona;13-dic-2009;
```

Recuadro 2.5: Fichero que actúa como comparar denominado: “Viviendas12-dic-2009.txt”. De este fichero se observan las dos primeras instancias de viviendas y las dos últimas.

```
VP0000001111122;pisos;L'eixample;3;7;123;3.008;370.000;obra nueva;2008-12-26 10:38:59.0;valencia;26-jun-2009;NO;
VP0000001111444;pisos;Algirós;3;7;120;2.875;345.000;;2009-11-10 12:08:50.0;valencia;5-dic-2009;SÍ;
.....;
VP0000002149013;pisos;eixample;3;4;94;8.312;781.315;;2009-10-24 11:32:45.0;barcelona;10-dic-2009;NO
```

Recuadro 2.6: Fichero de las viviendas que se les ha podido estimar el atributo “compra” del fichero “Viviendas27-nov-2009.txt”, que ha sido comparado con el fichero “Viviendas12-dic-2009.txt”. Tal fichero se denomina “VSINO27nov2009-12dic2009.txt”. Se observan ‘3’ instancias de viviendas de tal fichero, exactamente, las dos primeras y la última.

```
.....;
VP0000002149013;pisos;eixample;3;4;94;7.954;747.676;a reformar;2009-12-10 12:25:38.0;barcelona;13-dic-2009;
VP0000002872954;pisos;sarrià-sant gervasi;5;5;275;10.000;2.750.000;;2009-11-19 18:59:37.0;barcelona;13-dic-2009;
VP0000004444433;pisos;L'eixample;4;7;100;2.908;290.800;obra nueva;2009-12-12 12:24:23.0;valencia;13-dic-2009;
.....;
```

Recuadro 2.7: Fichero denominado “VResto27nov2009-12dic2009.txt” que van a parar el resto de las instancias de viviendas que carecen del atributo “compra”. En tal fichero se observan ‘3’ instancias de viviendas, la primera y la última que se observan proviene del fichero “Viviendas12-dic-2009.txt” y la segunda que se observa del fichero “Viviendas27-nov-2009.txt”.

Una vez expuestos los recuadros vamos a explicar el proceso, en el cual se va a ignorar la comprobación de las situaciones anormales de la tabla 2.3.

En principio tenemos el fichero “Viviendas27-nov-2009.txt” que actuará como “fichero base” y el fichero “Viviendas13-dic-2009.txt” que lo hará como “fichero comparar”. Se trata de comparar los dos ficheros para intentar estimar el atributo “compra” a las instancias de las viviendas del “fichero base”. Y después almacenar las instancia de las viviendas del “fichero base” que se les haya podido estimar el atributo “compra” en el fichero “VSINO27nov2009-13dic2009.txt” junto al atributo “compra” estimado. Además también se desea almacenar el resto de las instancias de las viviendas en el fichero “VResto27nov2009-12dic2009.txt”. Todo este proceso se base en las suposiciones comentadas anteriormente en este mismo apartado sobre la estimación del atributo “compra”. Vamos a explicar, para las instancias de las viviendas que se encuentran en el recuadro 2.3 y en el recuadro 2.4 correspondientes al fichero “Viviendas27-nov-2009.txt” que actuará como base y al fichero “Viviendas13-dic-2009.txt” que actuará como comparar, dicho proceso.

- En primer lugar lo que deberemos hacer será obtener la primera instancia de vivienda del fichero que actúa como base: “Viviendas27-nov-2009.txt” que se observa en el recuadro 2.3. Según la estimación del atributo “compra” lo primero que deberemos apreciar es si la vivienda ha caducado. Como el atributo “fechaActualizado” que presenta la vivienda tiene el valor ‘2008-12-26 10:38:59.0’ que representa a la fecha ‘26 de diciembre de 2008’ y la “fecha segunda” de extracción presente en el “fichero comparar” es ‘13-dic-2009’ que viene a indicar la fecha ‘13 de diciembre de 2009’, concluimos que la vivienda ha caducado. La vivienda ha caducado, debido a que según las suposiciones, si hay una diferencia de seis meses o más entre el valor del atributo “fechaActualizado” y la “fecha segunda” de extracción, la vivienda se declara como caducada, y como faltan 13 días para que haya una diferencia de un año entre ellas se concluye que ha caducado la vivienda del “fichero base”. La instancia de dicha vivienda del “fichero base” se introducirá en el fichero “VSINO27nov2009-13dic2009.txt” con el atributo “compra”=‘NO’. Como el momento en que caduca la vivienda es a los seis meses desde la fecha correspondiente al valor de su atributo “fechaActualizado”, el atributo “FechaDeVenta/NoVenta”, que surge al declara a la vivienda como no vendida, adoptará como valor una fecha de seis meses posterior al valor de la fecha del atributo “fechaActualizado” que posee la vivienda. Al declararse la vivienda del “fichero base” como caducada no intentaremos buscarla en el fichero “Viviendas13-dic-2009.txt” que actúa como “fichero comparar” y pasaremos a la próxima instancia de vivienda del “fichero base”.
- Pasamos a la segunda instancia de vivienda del “fichero base”, “Viviendas27-nov-2009.txt”. En ésta vemos que no ha caducado la vivienda, porque no han pasado seis o más meses entre el atributo “fechaActualizado” que presenta ‘2009-11-10 12:08:50.0’ y la “fecha segunda” de extracción que es ‘13-dic-2009’. Por lo tanto, buscaremos la instancia de la vivienda en el “fichero comparar”, pero no la

encontraremos porque no aparece ni en las '4' instancias de viviendas que se observan del "fichero comparar" en el recuadro 2.4 ni tampoco en el resto de las instancias de las viviendas del "fichero comparar", con lo que, deducimos que ha sido vendida y la pondremos en el fichero "VSINO27nov2009-13dic2009.txt". Por otro lado al atributo "fechaDeVenta/Noventa" que surge al venderse la vivienda, le otorgaremos como valor, la media entre la fecha de extracción de la "primera fecha" y la fecha de extracción de la "segunda fecha" que será la fecha del '5 de diciembre de 2009'. Tras ello pasaremos a la siguiente instancia de vivienda del "fichero base" que se observa en el recuadro 2.3 que es la penúltima.

- Obtenemos la penúltima instancia de vivienda del "fichero base" y observamos que no ha caducado. La buscamos en el "fichero comparar" y la encontramos en la penúltima instancia de vivienda del "fichero comparar", ya que disponen del mismo identificador. No obstante, vemos, que aunque es la misma instancia de vivienda, ha variado el atributo "euros" en la instancia del "fichero comparar", por tanto concluimos que ha cambiado la oferta de la vivienda. Ahora echamos un vistazo al valor del atributo "fechaActualizado" de la vivienda y vemos que también ha variado entre los dos ficheros, ya que en el "fichero comparar" aparece con un valor comprendido entre las dos fechas de extracción "fecha primera" y "fecha segunda" y en el "fichero base" con una fecha anterior. Por lo que copiamos la instancia de la vivienda "penúltima" del "fichero base" en el fichero "VSINO27nov2009-13dic2009.txt", y le ponemos como valor al atributo "FechaDeVenta/NoVenta" el valor del atributo "fechaActualizado" que aparece en la nueva oferta de la vivienda en el "fichero comparar". Además la nueva oferta de la vivienda que aparece en el "fichero comparar" la copiamos en el fichero "VResto27nov2009-13dic2009.txt".
- Llegamos a la última instancia de vivienda del "fichero base" que se observa en el recuadro 2.3 y observamos que no está caducada. La buscamos en el "fichero comparar" y la encontramos en la última instancia de vivienda que se observa en el recuadro 2.4 del "fichero comparar". Nos fijamos en el valor del atributo "euros" de la instancia de la vivienda que aparece en el "fichero base" y también en el "fichero comparar" y vemos que adopta el mismo valor en ambos ficheros, por lo que, no podremos estimar el atributo "compra" de la vivienda. Con lo que, copiaremos dicha instancia de vivienda que aparece sin variaciones en el "fichero base" y en el "fichero comparar" en el fichero "VResto27nov2009-13dic2009.txt". Hay que resaltar que, en este proyecto, la copiaremos en el fichero "VResto27nov2009-13dic2009.txt" desde el "fichero base" cambiando el valor del atributo "fechaExtracción" de la vivienda por el valor de la "fecha segunda" de extracción.
- Como hemos tratado la última instancia de vivienda del "fichero base", ya no encontraremos más en tal fichero, por lo que, habremos terminado de estimar el atributo "compra" a las viviendas, ya que, esto lo realizamos sobre las instancias de viviendas del "fichero base". No obstante, ahora vamos a comprobar si alguna de las '4' instancia de viviendas del "fichero comparar" que se observan en el recuadro 2.4 no aparecen en el "fichero base".

instancia de vivienda de la última línea que ha sido no vendida y que han hecho falta ‘47’ días para decirlo.

2.5. Refinamiento de los datos hasta alcanzar la vista general

Con el cálculo del atributo “días” a las viviendas seleccionadas llegaremos a disponer de todos los atributos de que disponemos. Una vez estemos en esta situación continuaremos con el proceso de preparación de datos con la realización de tareas de cambio de formato, de selección de datos, de consistencia, de coherencia y de limpieza.

La primera tarea consistirá en una serie de alguna de esas tareas o acciones que hemos mencionado y que las englobaremos en una sola. La tarea central de la serie tratará sobre una tarea de cambio de formato que consistirá en transformar o cambiar el formato de las instancias de las viviendas a uno entendible por la aplicación “WEKA”. Por entendible para la aplicación “WEKA”, queremos decir que “WEKA” pueda cargar las instancias de las viviendas.

Otra tarea incluida en el conjunto de tareas consistirá en la eliminación del atributo “euros/metros”. Este atributo “euros/metros” es uno de los que habíamos dicho que utilizaríamos para describir las viviendas, sin embargo, es observable que dicho atributo tiene mucha relación con el atributo “euros” y con el atributo “metros”, ya que, si relacionamos el atributo “euros/metros” con el atributo “metros” y los multiplicamos (“euros/metros” X euros) nos informan precisamente del atributo “euros”. Con lo que aquí tendremos como hemos dicho una tarea de selección de atributos, aunque también se puede percibir desde otra perspectiva como una tarea de limpieza, ya que, se eliminan redundancias. Decimos que es de selección de atributos, porque eliminamos el atributo “euros/metros” y de limpieza de eliminación de redundancias, porque es una información que con la división del atributo “euros” por el atributo “metros” tenemos disponible, ya que (“euros” / “metros” = “euros/metros”).

Por otro lado, cambiando de tarea dentro de la serie de tareas relataremos una tarea de consistencia que consistirá en asegurarnos que el atributo “dorm” y el atributo “planta” adopten siempre un tipo de dato nominal en vez de numérico. Como el tipo de datos es una cuestión importante en la preparación de los datos vamos a hacer un paréntesis en la explicación de lo que hemos citado como serie de tareas englobadas en una sola tarea, para decir unos detalles sobre los tipos de datos y sobre los tipos de datos ligados a los atributos, que utilizamos para describir las viviendas. Estos detalles los relatamos seguidamente:

En análisis de datos una manera usual de contemplar o interpretar los datos ligados a los atributos es como datos nominales o datos numéricos, es decir, considerar a los atributos como de tipo de dato nominal o como de tipo de dato numérico. Nosotros vamos a considerar o interpretar los datos de esta forma y más importante la aplicación de

análisis de datos “WEKA”, que utilizaremos en este proyecto, también interpretará nuestros atributos como nominales o como numéricos. Aunque “WEKA” también abarca otros tipos de datos, como los tipos de datos “String” y “Date”.

Seguidamente vamos a aclarar que se entiende por atributos nominales y que por atributos numéricos:

Un atributo cualquiera se interpreta como de tipo nominal, si por lo menos uno de los valores ligados a él no contiene exclusivamente números. Por el contrario, un atributo cualquiera se interpreta como numérico, si cada uno de los valores que tiene está formado únicamente por caracteres numéricos.

Por la definición de los tipos de datos nominales y numéricos que hemos hecho parece que los datos adjuntos a un atributo son los que van a declarar si el atributo es nominal o numérico. Esto es lo que hace precisamente la aplicación “WEKA” cuando le pasamos los datos de los atributos sin el tipo de dato ligado a cada atributo, es decir, sin estar declarado explícitamente. Si ocurre esto, es decir, si no aparecen los atributos acompañados del tipo de dato, entonces, “WEKA” decide según los valores que posee cada atributo el tipo de datos de cada atributo. Concretamente, en nuestros atributos, los que sólo poseen valores numéricos “WEKA” los interpretará como atributos numéricos, y los atributos que no poseen valores numéricos o aun poseyendo valores numéricos no todos los valores son numéricos, “WEKA” los trata como atributos nominales. La aplicación “WEKA” realiza esto de dicha manera, porque se le pasan los valores de los atributos sin un tipo de dato declarado explícitamente. No obstante, si nosotros creáramos un tipo de fichero de datos que exigiera la definición de los tipos de datos para los atributos, entonces, sí se tendrían que restringir los valores de los atributos a los tipos de datos.

Por lo que, retomando la tarea dentro de la serie de tareas de que el atributo “dorm” y “planta” se interpreten siempre como nominales, sólo habrá que añadirles una letra en alguno de los valores que adopten para confirmar que sean siempre nominales. Exactamente, lo que haremos será añadir una letra delante de todos los valores pertenecientes al atributo “dorm” o al atributo “planta”, para que se interpreten siempre como nominales. Aunque, por lo dicho con anterioridad, bastaría con que un único valor en cada uno de dichos atributos se le añadiera una letra para que se interpretarán como nominales. No obstante, se prefiere que se añada la letra en todos los valores para que tengan una homogeneidad y así que en los valores del atributo “dorm” y “planta” se observe que todos sus valores son de naturaleza nominal. En cuanto a la letra que se les añade a los valores de tales atributos, decir que, la letra “p” de planta es la que se introduce delante de los valores del atributo “planta” y la letra “d” de dormitorio es la que se introduce delante de los valores del atributo “dorm”.

Por otra parte, después de haber comentado el tipo de datos que queremos para el atributo “dorm” y para el atributo “planta”, decir para el resto de atributos, que también queremos que sean siempre del tipo de dato que deseamos entre los dos tipos de datos que nos interesan, nominales o numéricos. Esto recibe especial importancia sobretodo

en el atributo “euros”, ya que, en este precisamos necesariamente que sea siempre numérico, porque como veremos en el apartado siguiente, “3.Aprendizaje”, se utilizará como atributo de salida de los modelos de regresión que realizaremos, y por lo tanto “WEKA” deberá interpretarlo siempre como un atributo del tipo numérico. A su vez el atributo “compra” es necesario que sea de tipo nominal, porque también se realiza un modelo de clasificación en que el atributo “compra” es la salida. Y como los modelos de clasificación tienen como salida valores nominales se necesita que el atributo “compra” sea nominal. La tabla 2.4 muestra los tipos de datos asociados a los atributos que se espera que se reflejen en la aplicación “WEKA” al cargarlos:

Atributo	Tipo de dato
Identificador	Nominal
Tipo	Nominal
Barrio	Nominal
Dorm	Nominal
Planta	Nominal
Metros	Numérico
Euros	Numérico
Estado	Nominal
Ciudad	Nominal
Compra	Nominal
Días	Numérico

Tabla 2.4: El tipo de los atributos descriptivos.

Como el atributo “euros/metros” lo eliminamos no lo incluimos en la tabla 2.4, aunque claramente en la aplicación “WEKA” se presentaría como un atributo “numérico”.

En la tabla 2.4 aparecen los tipos de datos que se esperan para los atributos. Como se ha relatado, son los tipos de datos que se esperan y los que se deberían esperar en futuras extracciones de datos, ya que es prácticamente imposible, que cambiarán en futuras extracciones, que se desearan realizar sobre el portal web, hasta el extremo de que “WEKA” los tenga que interpretar de forma numérica en vez de nominal o viceversa. Es decir, por ejemplo, sería muy raro (aunque sería el menos raro de todos los cambios en los tipo de datos de los atributos) que el atributo “identificador” sólo presentara valores numéricos en vez de presentar también letras. Si esto sucediera, entonces “WEKA” lo interpretaría como numérico en vez de como nominal que es la forma en que se presenta en las extracciones realizadas en este proyecto. Pero si esto pasara, aunque no sería idóneo sí que podríamos permitirnoslo, ya que realmente los únicos cambios de tipo que no podemos permitirnos serían que el atributo “euros” se convirtiera en atributo nominal y que el atributo “compra” se convirtiera en atributo numérico. Pero esto es prácticamente imposible de que ocurra para el atributo “euros” e imposible del todo para el atributo “compra”.

Decimos que es casi imposible que ocurra para el atributo “euros”, porque ver el precio de las viviendas en letras en vez de números es difícil de creer y poco adecuado, por lo que, es casi imposible que aparezca en letras en el portal web los valores referentes al atributo “euros”. No obstante, si por alguna circunstancia esto ocurriera y los precios de las viviendas aparecieran en letras, entonces habría que buscar el modo, necesariamente, de interpretar dichas letras y traducirlas a números de alguna forma.

Respecto al atributo “compra” decimos que estamos seguros que siempre adoptará el tipo de dato nominal, porque este atributo lo estimamos, no lo extraemos, y por tanto, sabemos que los valores que le atribuimos son nominales. De forma similar también sabemos que el atributo “días” será siempre numérico debido a que tampoco lo extraemos sino que lo calculamos.

Siguiendo con la misma idea volvemos a recalcar que es casi imposible que en otras extracciones futuras cambie el formato de los valores de tal manera que “WEKA” tenga que interpretarlos de forma diferente, ya que, por ejemplo, al igual que hemos dicho del atributo “euros” que es casi impensable, que adopte en el portal web un tipo de dato nominal, es igual de improbable, que el atributo “barrio” adopte valores numéricos en el portal web, ya que también sería imposible de creer. No obstante, si muy extrañamente cambiaran de tipo alguno de los atributos y se reflejará en “WEKA” con otro tipo distinto del indicado en la tabla 2.4 siempre se podrían cambiar al tipo que quisiéramos. En el caso de que un atributo apareciera hipotéticamente como numérico cuando deseáramos que apareciera como nominal, sólo deberíamos añadir una letra a los valores del atributo para transformarlo a nominal. Por otro lado en los atributos, que hipotéticamente, aparecieran como nominales y quisiéramos que fueran numéricos habría que realizar algún tipo de traducción para pasar los caracteres no numéricos a caracteres numéricos para que se interpretaran de forma numérica.

Una vez explicado el tipo de datos de cada uno de los atributos que hemos introducido debido a que relatábamos la tarea de que el atributo “dorm” y el atributo “planta” sean del tipo nominal, vamos a seguir con la serie de tareas y precisamente llegamos a una que también tiene que ver con los tipos de datos. Concretamente, se trata de comprobar que el atributo “euros” tenga todos los valores numéricos y no aparezca por error algún valor que no sea numérico. Aunque esto no se ha visto en ninguna de las extracciones que se han realizado en este proyecto, hay que comprobar si aparece algún valor no numérico en el atributo “euros” y entonces si se encuentran eliminar esas instancias de viviendas que tienen el atributo “euros” mal escrito. Esta comprobación de que el atributo “euros” presente todos los valores como numéricos es necesaria, ya que, como hemos dicho anteriormente, el atributo “euros” forma parte de la salida de los modelos de regresión que realizaremos y por lo cual debe ser siempre numérico. Sobre esta tarea de la serie de tareas, decir que intenta detectar y eliminar las instancias de las viviendas que supuestamente tuvieran valores no numéricos en el atributo “euros” por error en los datos que aparecieran en el portal web, no si por lo que fuera hubieran cambiado con el tiempo voluntariamente a letras los valores del atributo “euros” en el portal web.

En cuanto al atributo “compra” como sabemos que es siempre nominal no tenemos que comprobar si es del tipo nominal. Y respecto al resto de atributos tampoco vamos a comprobar si son del tipo que deseamos, porque hemos dicho anteriormente que no es indispensable que sean del tipo que deseamos para ellos, aunque si es recomendable. Por lo cual si por lo que fuera se presentarán con el tipo numérico en vez de nominal o viceversa alguno de los atributos lo muy deseable es que se realizaran las tareas necesarias para que adoptaran los tipos de datos que se reflejan en la tabla 2.4 para cada uno de estos atributos.

Por otro lado, la última tarea de la serie de tareas que vamos a comentar, tiene como objetivo que los valores numéricos de los atributos “euros” y “metros” se representen correctamente siempre en la aplicación “WEKA”, por lo que, será una tarea de coherencia de datos. Esta tarea se debe realizar, porque los valores de los atributos “euros” y “metros” en las extracciones tienen puntos de separación de millares y “WEKA” interpreta los puntos en los valores numéricos no como separación de millares, sino como puntos decimales, por lo que, interpretará de forma errónea los valores. (Si estos puntos desaparecieran de los valores del atributo “euros” o “metro” en futuras extracciones del portal web, entonces esta tarea se realizaría sin consecuencias.)

En el caso del atributo “euros”, en todos sus valores aparecen puntos de separación de millares y en el caso del atributo “metros”, aunque no es habitual, si que aparece algún valor que tiene un punto de separación de millares, ya que, hay algunos valores en el atributo “metros”, que alcanzan los cuatro dígitos. En el caso del atributo “euros” hay valores que pasan incluso de seis dígitos, por lo que habrá que eliminar por dos veces el punto en esos valores. Con esta última tarea se terminaría la serie de tareas.

Después de la serie de tareas tendremos las instancias de las viviendas disponibles para que “WEKA” las importe, sin embargo, antes de importarlas realizaremos otra tarea que puede que no sea necesaria de realizar, porque simplemente en las instancias de las viviendas extraídas no se dé el caso: se trata de eliminar las posibles instancias de viviendas repetidas. Es decir, eliminar las muy excepcionalmente instancias de viviendas que aparezcan más de una vez en las instancias extraídas del portal web. Decimos que puede ser una tarea innecesaria, porque puede que no aparezca ninguna instancias de vivienda más de una vez en las instancias de las viviendas del portal web, no obstante, hay que comprobar si hay instancias de viviendas duplicadas o ver incluso si aparecen más de dos veces, y entonces eliminarlas y quedarnos, sólo, con una. Esta tarea es de limpieza de datos ya que se eliminan las repeticiones de instancias de viviendas que enturbian la información que se dispone de las viviendas. Decir que en los datos concretos extraídos para la construcción de los modelos, que explicaremos más adelante en este mismo apartado, aparecían unas casi inapreciables instancias de viviendas duplicadas en el portal web, si tenemos en cuenta todas las que habían en el portal web. Sin embargo, puede que en otras extracciones, en el portal web, no aparezca ninguna instancias de vivienda más de una vez, ya que en el portal web inmobiliario desean que se observe una misma instancia de vivienda, lógicamente, una sola vez, ya que, no tiene sentido que se presente más de una vez.

Después de esta tarea de limpieza, sí que nos dispondremos a importar las instancias de las viviendas a la aplicación “WEKA” y observaremos el atributo “identificador” que deberá aparecer como distinto para todas las viviendas. Lo que nos confirmará, en efecto, que cada instancia de vivienda es distinta. Este atributo “identificador”, ya no lo necesitaremos para nada una vez carguemos las instancias de las viviendas en la aplicación “WEKA” y veamos que todas las viviendas son diferentes, y por lo tanto, lo eliminaremos. Esta eliminación se debe a que no nos va a ayudar en la construcción de los modelos, ya que, como tiene un valor diferente para cada vivienda es difícil que se pueda extraer algún patrón de comportamiento del atributo “identificador”, además dificultaría la construcción de los modelos, por disponer de una gran cantidad de valores nominales dicho atributo “identificador”, uno para cada vivienda.

Cuando realicemos esta tarea de eliminación del atributo “identificador” habremos llegado a la vista de datos general que habíamos dicho que es la parte común de preparación de los modelos que pretendemos construir. Una vez citadas estas tareas de cambio de formato, de limpieza, de selección, de consistencia y de coherencia que nos permiten alcanzar la vista de datos general vamos a enumerarlas y contar las acciones concretas que realizan:

5. La quinta tarea consistirá en la serie de tareas que hemos dicho que englobaríamos como una sola tarea. La tarea central como dijimos tiene como finalidad que las instancias de las viviendas puedan cargarse en la aplicación de análisis de datos “WEKA”. La idea será cargar los datos en la herramienta “WEKA” debido a que nos va a facilitar el proceso de preparación restante y a que queremos utilizar dicha herramienta también para construir los modelos de datos citados anteriormente.

Los datos de las viviendas que disponemos no se pueden cargar en “WEKA” debido a que “WEKA” no puede interpretarlos. La aplicación “WEKA” tiene como formato característico el formato “arff” como ya se comentó en el apartado de introducción y no tenemos las instancias de las viviendas en dicho formato, sin embargo, no es necesario que adaptemos las instancias de las viviendas al formato “arff”, ya que, “WEKA” puede importar otra serie de ficheros que no están en formato “arff” como el formato “csv”, o los datos de una base de datos. “WEKA” al importar un fichero lo transforma automáticamente en la carga al formato “arff”. Precisamente el formato “csv” es un formato de texto factible de adoptar desde el formato que tenemos en las instancias de las viviendas. Por lo que, se pretende convertir el fichero de las instancias de las viviendas que deseamos cargar que está en formato “txt”, al formato “csv” con lo que podremos importarlo a la aplicación “WEKA”.

Seguidamente comentamos unas diferencias entre los ficheros “txt” y “csv”:

Los ficheros “csv” son al igual que los “txt” ficheros de texto pero más específicos. Es decir tienen unas reglas más restrictivas que los ficheros de texto “txt”, debido a que los ficheros de texto en formato “csv” deben tener como separación de valores una coma y el último valor de cada línea no debe llevar ningún tipo de separación. Estas separaciones dividen el contenido en columnas. Un fichero “csv”, además,

tiene que tener una primera línea con los nombres de los atributos, que tienen que identificar a los valores, y estos deben estar también separados por una coma menos el último nombre de atributo que no debe llevar ningún tipo de separación. Añadido a esto cada nombre de atributo debe estar en la columna donde estén los valores que representa. Al contrario que los ficheros “csv” los ficheros “txt” no tienen estas restricciones, ni tampoco es necesario que haya ni siquiera atributos en ellos, pueden tener cualquier tipo de contenido, aunque es evidente, que en este proyecto, el contenido que presentamos en los ficheros “txt” sí que representan registros de viviendas. Y que cada uno de estos registros tiene los valores separados en columnas, ya que, cada valor de un registro de vivienda, de nuestros ficheros “txt”, está separado por un punto y coma.

Para realizar tal transformación lo que haremos será crear otro fichero con extensión “.csv” a partir del fichero de las instancias de las viviendas con el atributo “compra” y el atributo “días” que tiene la extensión “.txt”. Está como hemos dicho será la tarea central de la serie de tareas. Las otras tareas que realizaremos dentro de la serie de tareas será, como ya hemos dicho con anterioridad, eliminar el atributo “euros/metros”, que los atributos “dorm” y “planta” sean del tipo nominal, también eliminar las posibles instancias de viviendas que adoptarán por error un valor no numérico en el atributo “euros” y por último que los atributos “euros” y “metros” se interpreten de forma correcta. La serie de tareas a realizar, de la tarea quinta, la citamos seguidamente de forma cronológica:

- Primeramente, colocar una cabecera con el nombre de los atributos en la primera línea del fichero de salida de la serie de tareas englobadas en una tarea. Estos aparecerán separados por comas menos el último de los nombres de atributos que no llevará ningún tipo de separación. Dicha cabecera se muestra seguidamente:

Identificador,Tipo,Barrio,Dorm,Planta,Metros,Euros,Estado,Ciudad,Compra,Dias
--

Recuadro 2.9: Cabecera con el nombre de los atributos.

Se copiarán con algunas modificaciones cada una de las líneas, instancias de viviendas, del fichero de entrada a esta tarea. A no ser que por error el atributo “euros” de la vivienda sea no numérico, y entonces se pasaría a otra instancia de vivienda sin copiarla. Esto se realizará de la siguiente forma:

-Se modificarán los puntos y comas de separación por una coma, menos el último punto y coma que separa al último valor de cada instancia de vivienda, que lo eliminaremos sin remplazarlo por separador alguno.

-Cuando lleguemos al atributo “dorm” colocaremos la letra “d” delante del valor de dicho atributo. Esto se realizará siempre que el atributo “dorm” sea no vacío.

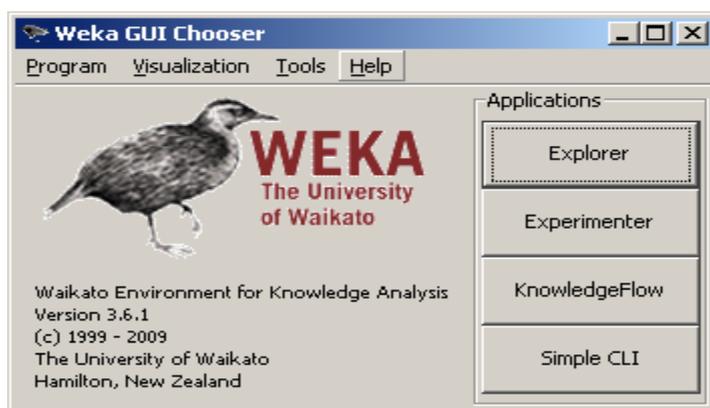


Figura 2.1: Ventana inicial de “WEKA”.

Cuando se pulse sobre la Aplicación ‘Explorer’, se mostrará la ventana gráfica de la pestaña llamada ‘Preprocess’, como se observa en la figura 2.2.

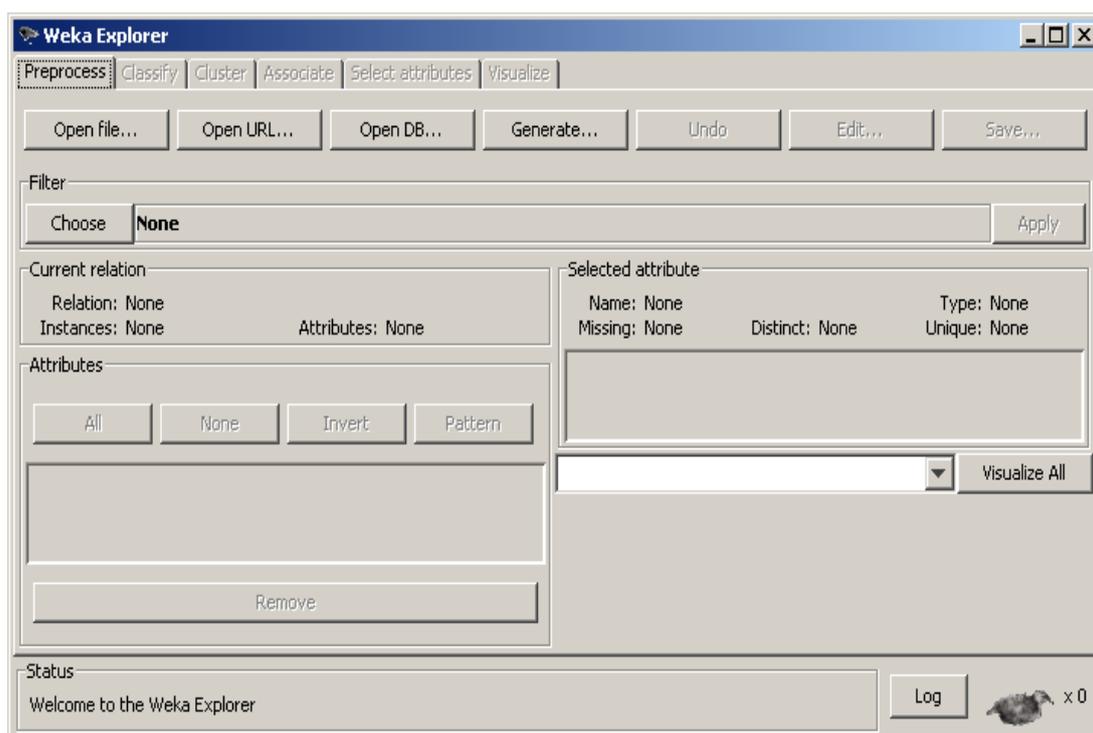


Figura 2.2: Ventana de la aplicación ‘Explorer’ de “WEKA”, ‘Preprocess’.

Finalmente, desde la ventana ‘Preprocess’ podremos realizar la importación que comentábamos. Para lo cual habrá que pulsar sobre el botón ‘Open file...’, mostrado en la parte superior izquierda de la figura 2.2, y aparecerá un cuadro de diálogo que nos permitirá abrir un fichero para “WEKA”. Como el fichero a abrir es un fichero con extensión “.csv”, elegiremos antes de seleccionar ningún fichero la opción abrir ficheros “.csv”. Con lo que si estamos en el directorio donde reside el fichero “.csv” que

queremos importar, entonces, sólo tendremos que seleccionarlo. Cuando se realiza la acción, “WEKA” importará el fichero con extensión “.csv” transformándolo durante el proceso al formato “.arff”, con lo que se visualizará en la ventana ‘Preprocess’ los datos de tal fichero. En la figura 2.3 se observa el resultado de la importación de un fichero de salida de la tarea sexta.

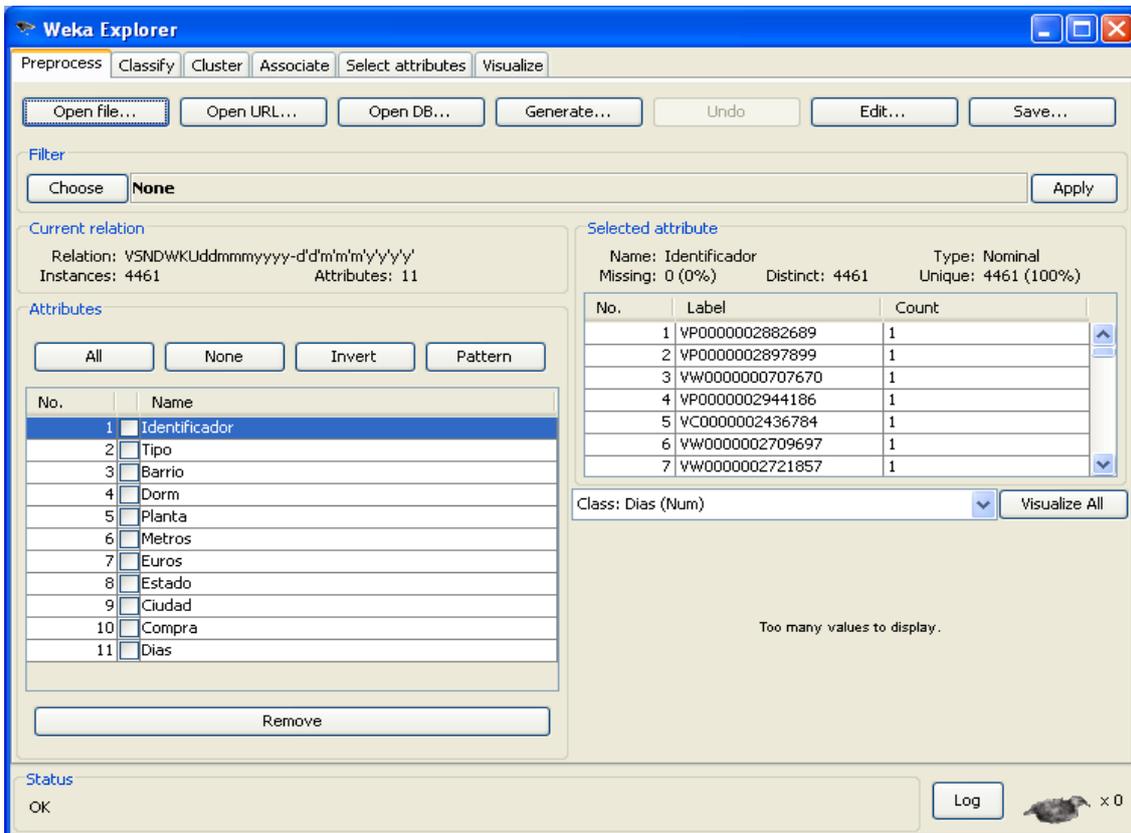


Figura 2.3: Fichero “VSNDWKUFecha1-Fecha2.csv” importado a la aplicación ‘Explorer’ de la herramienta “WEKA”.

En la ventana ‘Preprocess’ de la aplicación ‘Explorer’ de “WEKA” se observan cuatro secciones:

La sección ‘Filter’ es la que aparece más hacia arriba en forma alargada y se utiliza para realizar operaciones de tratamiento a los datos, como por ejemplo eliminar atributos o eliminar valores. En este proyecto, ya hemos hecho alguna de estas operaciones anteriormente, sin la ayuda de estos filtros, como al eliminar el atributo “euros/metros”. No obstante, utilizaremos los filtros que nos proporciona “WEKA” en la sección ‘Filter’ para seguir el proceso de preparación de los datos. Ya que precisamente los filtros transforman los datos de alguna u otra forma y esto nos sirve para seguir con la fase de preparación. A los filtros se puede acceder pulsando sobre el botón ‘Choose’ (Elegir) contenido en la sección ‘Filter’. Dichos filtros se presentan organizados según sean “supervisados” o “no supervisados” y a su vez según actúen sobre los atributos o sobre

las instancias. Los filtros supervisados realmente no son filtros corrientes, ya que, necesitan de algoritmos de minería de datos como los que se encuentran en la pestaña ‘Classifier’ (Clasificador) de “WEKA”, para ejecutarse. (La ventana ‘Classifier’ la explicaremos en el apartado posterior “3. Aprendizaje”.) Los filtros supervisados no los vamos a utilizar en este proyecto. Por contra los filtros no supervisados actúan directamente sobre los datos de entrada de que disponemos. Por ejemplo: “Add”, “Copy”, “Remove”, “RemoveWithValues”, “RemoveType”, “TimeSeriesTranslate” son algunos de los filtros presentes en la sección ‘Filter’.

La sección “current relation” que aparece inmediatamente debajo de la sección ‘Filter’ a la parte izquierda relata algunas características globales del fichero: como el nombre, el número de instancias que posee y el número de atributos que tiene el fichero.

La sección “Attributes” que aparece debajo de la sección “current relation” muestra un número que se le asigna a cada atributo y el nombre de cada atributo. Dicha información aparece en forma de tabla con dos columnas correspondientes: al número que se le asigna a cada atributo, y al nombre de cada atributo, esta última información aparece en un control gráfico del tipo ‘checkbox’. La columna titulada como “N” (número) indica la ordenación de los atributos y la columna “Name” el nombre de los atributos. Siempre aparece una fila resaltada de las demás. Los valores de esa fila resaltada aparecen detallados en otra sección situada a la derecha. En la sección ‘Attributes’, además, aparece un botón llamado “Remove”, el cual se utiliza para la operación de eliminar cualquier atributo que tenga marcada la casilla de verificación referente a su ‘checkbox’. Hay que apreciar que la opción de marcado de casilla no tiene nada que ver y es independiente con la de resaltar una fila. La operación de eliminar atributos es uno de los filtros de la sección ‘Filter’, pero al ser una operación usual, aquí aparece una manera más directa de realizarla.

Por otra parte, la sección que hemos dicho, que aparece más a la derecha se llama “selected attribute” y muestra la información del atributo resaltado en la sección ‘Attributes’. Al pulsar en la sección ‘Attributes’ sobre una fila cualquiera de los atributos disponibles hará que en la sección ‘select attribute’ se visualice la información correspondiente al atributo seleccionado. La información que aparece del atributo seleccionado es: el nombre, el tipo de datos, el número de valores desconocidos del atributo, el número de valores distintos que dispone el atributo en el fichero cargado y el orden de aparición y las veces que aparece cada uno de esos valores distintos. También se aprecia un control gráfico del tipo ‘comboBox’ y un botón ‘Visualize all’, en el que, esta vez el atributo indicado en ese ‘comboBox’ (que no el atributo seleccionado en la sección ‘Attribute’) aparecerá representado comparado con el resto de atributos. Por último más hacia abajo en la misma sección aparece una representación gráfica de los distintos valores del atributo seleccionado en la sección ‘Attribute’. Una vez descrita la ventana ‘Preprocess’, decir, que el proceso de preparación de los datos lo continuaremos desde dicha ventana.

Después de detallar la ventana ‘Preprocess’, vamos a relatar la última tarea antes de llegar a la vista de datos general. Esta tarea consistirá, precisamente, en la tarea que citábamos de eliminar el atributo “identificador”, para ello utilizaremos las opciones de la ventana ‘Preprocess’, seguidamente enunciamos tal tarea.

7. La tarea séptima va a consistir en eliminar el atributo “identificador”. Para eliminar dicho atributo “identificador” utilizaremos el filtro “Remove” (borrar), de la pestaña ‘Preprocess’ de “WEKA”. Dicha acción se puede hacer en la ventana ‘Preprocess’ de dos formas que pasamos a decir:

- Una forma es buscando el filtro en la sección ‘Filter’ de la ventana ‘Preprocess’. Efectivamente, al pulsar sobre el botón ‘Choose’ de la sección ‘Filter’ aparecerán los filtros de que se disponen en la aplicación ‘Explorer’ de “WEKA”. Se apreciarán dos tipos de filtros unos supervisados y otros no supervisados. Como dijimos sólo utilizamos los no supervisados y dentro de tales nos aparecerán otras dos clases de filtros: unos que actúa sobre los atributos y otros sobre las instancias, en este caso las instancias de viviendas que tenemos. Como queremos eliminar el atributo “identificador” seleccionaremos la clase que dispone de los filtros de atributos, con lo que se observarán los filtros disponibles de dicha clase y seleccionaremos el filtro que buscamos llamado “Remove”. Una vez lo seleccionamos nos aparecerá el filtro en el control gráfico ‘label’ (etiqueta) que tiene a su disposición la sección ‘Filter’. Ahora que lo tenemos seleccionado tendremos que configurarlo para que actúe sobre el atributo que deseamos, con lo que pulsaremos sobre el ‘label’ en donde aparece escrito el nombre del filtro y aparecerá un cuadro de dialogo disponible para configurar el filtro. Lo único que realizaremos será indicar en el campo ‘indexItem’ el índice número “1” correspondiente al atributo que queremos eliminar. Hecho esto ya tendremos el filtro “Remove” configurado según nuestras pretensiones con lo que pulsaremos sobre el botón aceptar para que se guarde la configuración. Cuando tenemos el filtro configurado solamente faltará realizar la acción de presionar el botón “apply” (aplicar) para que el atributo “identificador” se elimine.
- La opción anterior aunque es correcta ya decimos que no es la que utilizaremos en este proyecto cuando utilicemos el filtro “Remove”. La razón es porque como el filtro “Remove” es el filtro comúnmente más utilizado la ventana ‘Preprocess’ dispone de una forma más rápida de realizarlo. Dicha forma consiste en utilizar el botón “Remove” que aparece en la sección “Attributes” en la parte baja de la sección. Lo que deberemos hacer es seleccionar en la sección “Attributes” el ‘checkBox’ del atributo que queremos eliminar que es el atributo “identificador” y una vez seleccionado pulsar sobre dicho botón de la parte baja de la sección “Attributes” llamado “Remove”, una vez pulsado tal botón habremos eliminado el atributo “identificador”.

Una vez explicado los dos posibles caminos para utilizar el filtro “Remove” sobre el atributo “identificador”, decir que se puede utilizar la opción que se quiera, pero en este proyecto, como hemos dicho, utilizaremos siempre la segunda opción, porque es más rápida de realizar. Por lo que, para eliminar el atributo “identificador” utilizaremos los pasos de la segunda opción. En la figura 2.4 se muestra la selección del ‘checkbox’ correspondiente al atributo “identificador” antes de pulsar sobre el botón “Remove”. Y la figura 2.5 muestra el resultado de eliminar dicho atributo “identificador”.

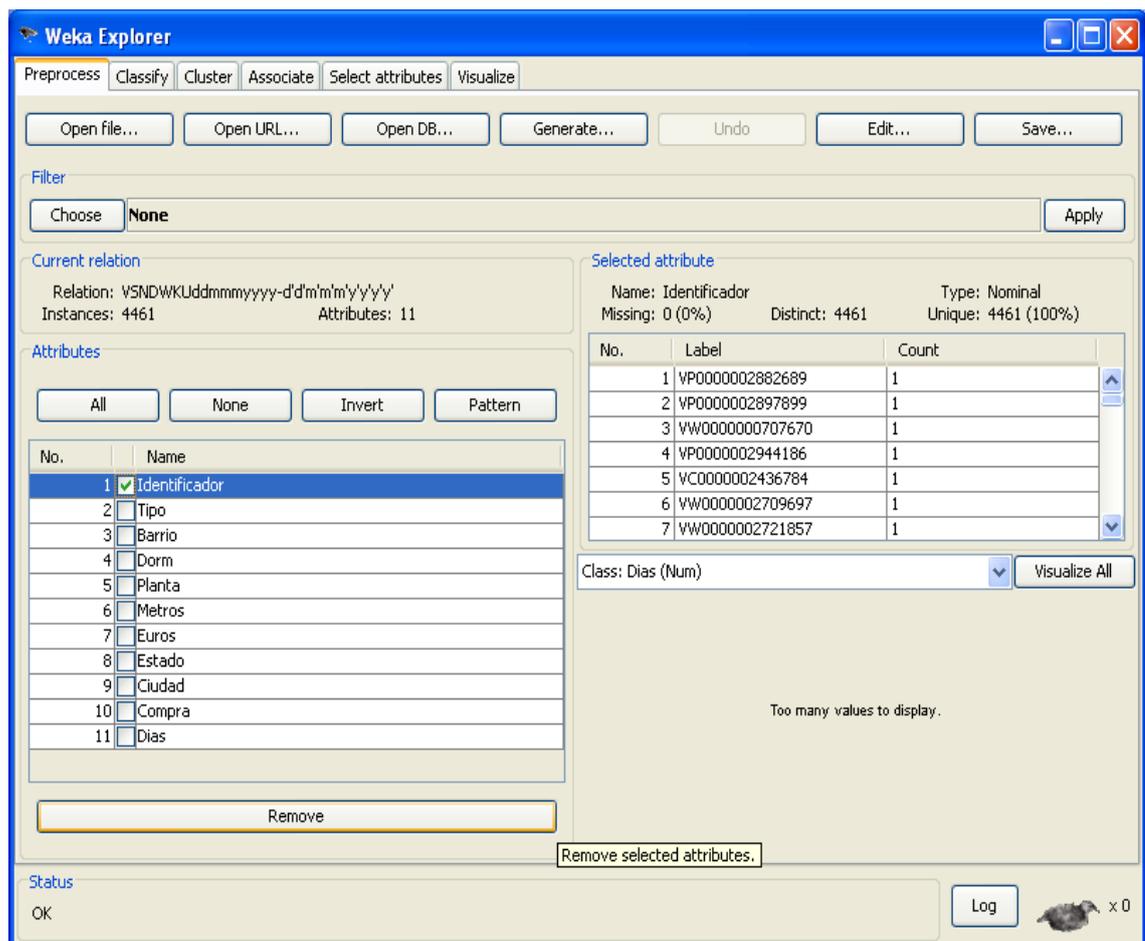


Figura 2.4: Se observa la selección del ‘checkbox’ correspondiente al atributo “identificador”. Con la finalidad de eliminar dicho atributo “identificador”, cuando presionemos sobre el botón “Remove”, situado en la parte inferior de la sección “Attributes”.

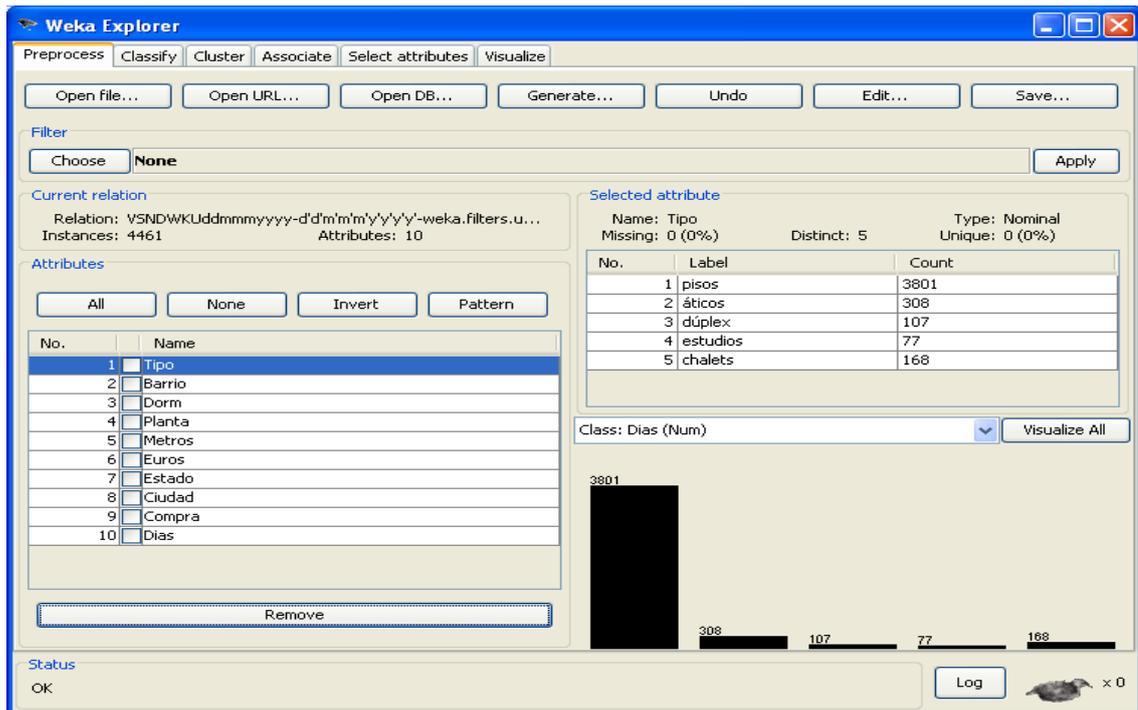


Figura 2.5: En esta figura se observa el resultado de pulsar el botón “Remove”. Como se aprecia tras haber pulsado dicho botón “Remove” el atributo “identificador” habrá sido eliminado.

Como se observa en la figura 2.5 al eliminar el atributo “identificador” se pasa de ‘10’ atributos a ‘11’ atributos. Y el número de asignación de los atributos ha sido disminuido en una unidad.

También como se puede apreciar en la figura 2.5 el nombre de la relación que aparece en la sección ‘Current relation’ ha variado al haberse realizado el filtro. Concretamente, se le concatena al nombre de la relación, el comando del filtro utilizado. Como en la figura 2.5 no se observa todo el nombre de la relation debido a que es muy largo se introduce en el recuadro 2.11.

```
relation: VSNDWKUddmmmyyyy-d'd'm'm'y'y'y'-WEKA.filters.unsupervised.attribute.Remove-R1
```

Recuadro 2.11: Muestra el nombre de la relación después de aplicar el filtro “Remove” al atributo “identificador”. Al nombre original “VSNDWKUFecha1-Fecha2”, siendo, Fecha1=ddmmmyyyy y Fecha2=d’d’m’m’y’y’y’y’ se le concatena “-WEKA.filters.unsupervised.attribute.Remove-R1”, que se compone de la ruta del filtro, del nombre del filtro y de la opción “R1”, está ultima opción indica que se actuó sobre el atributo que tenía asignado el número “1”.

Siguiendo con el nombre de la relación en la aplicación “WEKA”, decir, que al igual que ha cambiado con el resultado de la aplicación del filtro “Remove”, cambia con la aplicación de cualquier otro filtro. Es más cada vez que se realiza un filtro se añade al nombre de la relación la ruta del filtro y las opciones con las que se ha aplicado, por lo que, el nombre cada vez se hace más y más largo. Decimos esto porque en este proyecto, como veremos, vamos a aplicar más filtros y el nombre de

la relación, lógicamente, cambiara como hemos indicado, sin embargo, aunque esto ocurra como es siempre la misma idea, ya no la vamos a comentar más y dejaremos que cambie el nombre sin más comentarios sobre la modificación del nombre de la relación.

Una vez aplicado el filtro “Remove” lo único que nos faltará será guardar el fichero resultante de dicha operación. Para ello pulsaremos el botón ‘Save’ (Guardar) ubicado en el conjunto de botones superiores de la ventana ‘Preprocess’. En el cuadro de dialogo que aparece tras pulsar el botón ‘Save’ guardaremos tal fichero con el formato característico de “WEKA”. El nombre que le pondremos será “ViviendasVistaGeneralFecha1-Fecha2.arff”.

Aunque decíamos que no íbamos a hablar más de la modificación del nombre de la relación, una cuestión importante que sí que hay que decir sobre dicho nombre, es que aunque guardemos los ficheros con distintos nombres, al cargarlos, en el nombre de la relación, en la sección ‘Current relation’ aparecerá el nombre del fichero que se cargo por primera vez antes de haberlo guardado por primera vez, concatenado con los filtros que se hayan realizado después a lo largo del tiempo. Esto es debido, a que cuando guardamos un fichero con el botón “Save” de la ventana ‘Preprocess’ de “WEKA” le podemos poner el nombre que queramos que, lógicamente, se le asignará, sin embargo, esto no afectará al nombre de la relación, lo que si que afecta al nombre de la relación es como decíamos con anterioridad que se realicen filtros sobre la información del fichero. Por ejemplo, el fichero de la vista de datos general que al guardarlo se le ha asignado el nombre “ViviendasVista_GeneralFecha1-Fecha2.arff” tendrá asignado en el campo ‘relation’ el nombre “VSNDWKUFecha1-Fecha2”, concatenado con el comando del filtro “Remove” utilizado. Dicho nombre precisamente es el que se observa en el recuadro 2.11.

Con la eliminación del atributo “identificador” llegaremos a la vista de datos general. Una vez se alcanza la vista de datos general terminará el proceso de preparación común para los modelos, como ya indicamos con anterioridad. Como el fichero de la vista de datos general es el primero que dispondremos con formato “arff” vamos a analizar con una cierta profundidad dicho formato “arff”, antes de continuar con el proceso de preparación.

2.6. El formato de fichero “.arff”

El formato de archivos “.arff”, acrónimo de “Attribute-Relation File Format”, es el formato característico de los ficheros en la aplicación de minería de datos “WEKA”. Los archivos “.arff” son archivos de texto también y tienen las siguientes características:

-Se reflejan dos secciones en un fichero “.arff”. Estas secciones son: “sección de cabecera” y “sección de datos”.

-La “sección de cabecera” a su vez está compuesta por dos secciones: La sección correspondiente al nombre de la relación y la sección de declaración de atributos”.

-La sección correspondiente al nombre de la relación es la sección en donde se declara el nombre de la relación. La forma de la declaración es: *@relation [IdentificadorRelación]*.

Donde: “[IdentificadorRelación]” se sustituye por el nombre de la relación.

-En la sección de declaración de atributos aparecen los metadatos de los atributos. Los atributos se definen de la siguiente forma:

@attribute [IdentificadorAtributo] [tipo].

Donde: “[IdentificadorAtributo]” se sustituye por el nombre del atributo, y “[tipo]” se sustituye por uno de los tipos siguientes: ‘numérico’, ‘entero’, ‘fecha’, ‘cadena’ o nominal.

-Los tipos de datos más comunes y los que utilizamos en esta memoria son el tipo ‘numérico’ y el tipo nominal. El tipo ‘numérico’ lo utilizaremos precisamente para los atributos numéricos de naturaleza continua como el atributo “euros”, por ejemplo. El tipo nominal de naturaleza discreta lo utilizaremos para atributos que tengan un conjunto de posibles valores como el atributo “compra”. Los atributos numéricos se definen de la siguiente forma:

@attribute [IdentificadorAtributo] numeric.

Y los atributos nominales de la siguiente forma:

@attribute [IdentificadorAtributo] {[Valor1],..., [ValorN]}.

En los dos casos [IdentificadorAtributo] se sustituye por el nombre del atributo que corresponda. En el caso de los atributos nominales los valores {[Valor1],..., [ValorN]} se sustituyen por los posibles valores que pueda adoptar el atributo.

-En la “sección datos” aparecen los datos correspondientes a los atributos definidos en la sección de declaración, es decir, las instancias. En cada instancia los valores aparecen separados por una coma, menos el valor correspondiente al último atributo que está

separado por un salto de línea. Esta “sección de datos” empieza con la palabra clave “@data”. Tras dicha palabra clave aparecen las instancias.

```
@relation ViviendasSNDWKUddmmmyyy-d'd'm'm'yyyy
@attribute Tipo {pisos,áticos,dúplex,estudios,chalets}
@attribute Barrio {Extramurs, Patraix ... , 'El pla del real' }
@attribute Dorm {d3,d5, ..., d8 }
@attribute Planta {p7, p3,..., p5 }
@attribute Metros numeric
@attribute Euros numeric
@attribute Estado {'obra nueva','a reformar'}
@attribute Ciudad {valencia,madrid,barcelona}
@attribute Compra {NO,SÍ}
@attribute Días numeric

@data
pisos,Extramurs,d5,p4,158,310000,'a reformar',valencia,SÍ,21
pisos,L'eixample,d2,p3,65,129000,?,valencia,NO,13
.
.
.
pisos,Rascanya,d2,p2,60,65000,'a reformar',valencia,NO,27
```

Recuadro 2.12: Ejemplo de fichero “ViviendasVistaGeneralFecha1-Fecha2.arff”.

2.7. Vistas minables de los modelos

Prosiguiendo con el proceso de preparación llegamos al momento en que hemos dicho que seguiríamos por separado la preparación de los datos para cada uno de los modelos a tratar. Lo que desearemos será alcanzar la vista minable particular de esos modelos.

Primero explicaremos la preparación individual que se le deben hacer a los datos que se suministren al modelo 1, después los del modelo 2 y por último los del modelo 3. Aunque aparecen algunas similitudes entre la preparación específica de los modelos, decir que vamos a presentar la preparación de cada uno de los tres modelos por separado, porque de esta forma el proceso se aprecia de forma más comprensible y clara. Precisamente estas similitudes aparecen en la preparación específica del modelo 1 y del modelo 2 cuando deseamos eliminar el atributo “días”. Y también en la preparación específica del modelo 2 y del modelo 3 en el momento que queremos eliminar, las instancias de viviendas que adoptan el valor ‘NO’ en el atributo “compra”.

2.7.1. Vista minable del modelo 1

Vamos a empezar por la preparación específica de los datos para la realización del modelo 1, que estará comprendida entre la vista de datos general común a todos los

modelos y la deseada vista minable particular del modelo 1. Lo primero que vamos a hacer es volver a enunciar tal modelo:

-Modelo 1: Hallar la compra o la no compra de una vivienda.

Como el enunciado del modelo indica se trata de hacer predicciones sobre el atributo “compra”. Por lo cual, de entre los atributos que tenemos está claro que el atributo “compra” lo queremos. Además, también queremos los atributos “tipo”, “barrio”, “dorm”, “planta”, “metros”, “euros”, “estado” y “ciudad”, ya que estos atributos nos ayudarán en la predicción del modelo. Sin embargo, hay un atributo que no hemos nombrado de entre todos los que disponemos en la vista de datos general, que es el atributo “días”, que no lo vamos a incluir en la vista minable del modelo 1, porque es una consecuencia del atributo “compra”. Una vez eliminado el atributo “días” de las viviendas de la vista de datos general alcanzaremos la vista minable particular del modelo que estamos explicando. Seguidamente vamos a describir dicha tarea de eliminar el atributo “días”.

8. En esta tarea desearemos crear un nuevo fichero “.arff” a partir del fichero de la vista de datos general sin el atributo “días”. Lo primero que haremos será cargar la vista de datos general en la aplicación ‘Explorer’ de “WEKA”. Una vez cargada en la sección “Attributes” de la ventana ‘Preprocess’ seleccionaremos el ‘checkbox’ correspondiente al atributo “días” y luego presionaremos sobre el botón “Remove” de la parte inferior de dicha sección para así eliminar el atributo “días”. En la figura 2.6 se observa la selección del atributo “días” para ser eliminado y en la figura 2.7 se observa que dicho atributo ha sido eliminado, después de presionar el botón “Remove”.

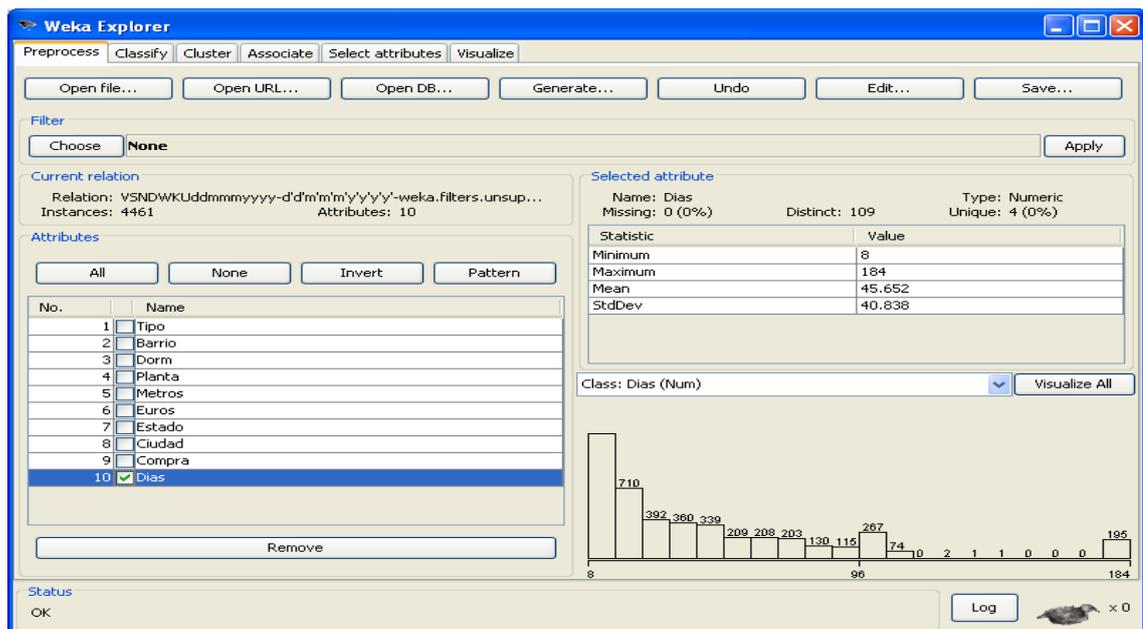


Figura 2.6: Seleccionamos el ‘checkbox’ ligado al atributo “días”, para eliminar dicho atributo al pulsar sobre el botón “Remove”.

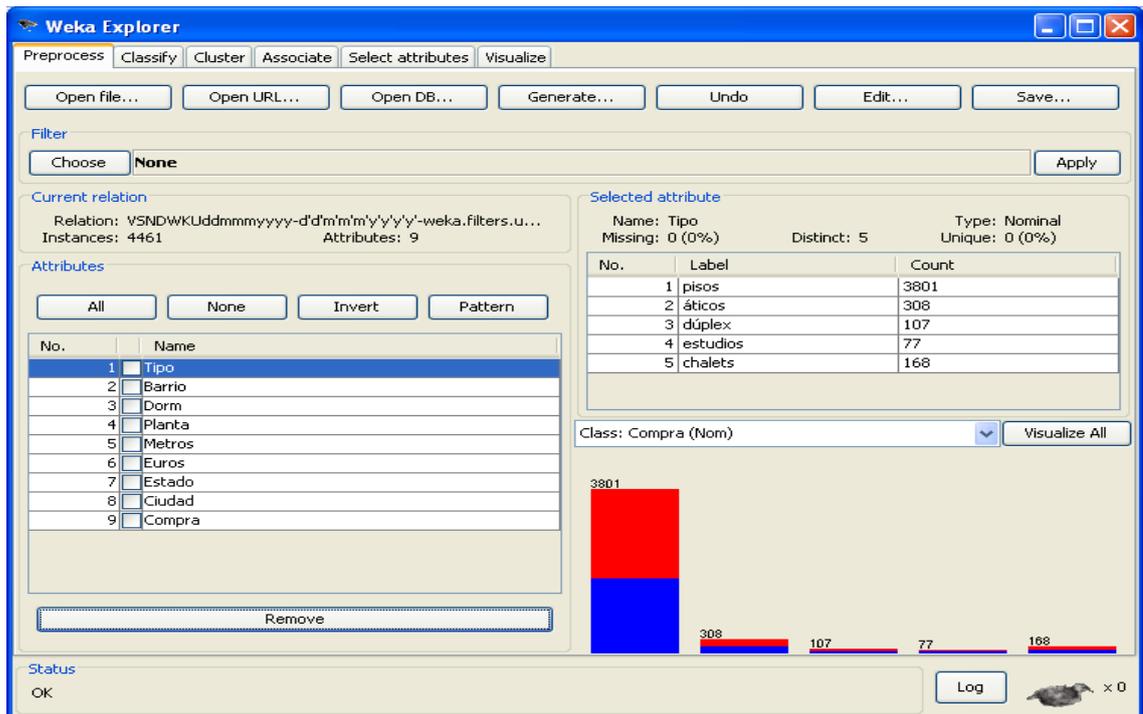


Figura 2.7: En esta figura se observa el resultado de pulsar el botón “Remove”. Previamente a pulsar dicho botón “Remove”, el ‘checkbox’ perteneciente al atributo “días”, estaba seleccionado para eliminar dicho atributo con la pulsación de tal botón.

Una vez eliminado el atributo “días” tendremos la vista minable particular del modelo 1. El nombre con el que guardaremos dicho fichero que contiene la vista minable será “ViviendasModelo1VistaMinableFecha1-Fecha2.arff”.

2.7.2. Vista minable del modelo 2

Después de crear la vista minable particular del modelo 1 nos centraremos en crear la vista minable particular del modelo 2. Para realizar la vista minable del modelo 2 daremos un retroceso y volveremos a la vista de datos general de las viviendas. Ya que la última tarea realizada estaba indicada con exclusividad para la preparación de los datos del modelo 1, en ningún caso para la del modelo 2 o la del modelo 3. Por lo tanto, para la preparación de los datos en el modelo 2 ignoramos la tarea octava que se ha realizado específicamente para los datos destinados al modelo 1 aun sabiendo que dicha tarea realizada para alcanzar la vista minable propia del modelo 1 vale para la del modelo 2, ya que, también se elimina el atributo “días”, en este modelo. Pero como hemos dicho, previamente, no se van a reusar las tareas específicas hechas para la preparación de otro modelo. Por lo que, volveremos a hacer entre otras acciones, para alcanzar la vista minable particular de este modelo 2, la tarea de eliminar el atributo “días”. Lo primero que vamos a hacer es volver a enunciar este modelo.

-Modelo2: Precio por el que vender una vivienda independientemente del número de días.

Como se cita en el texto que describe el modelo 2 un atributo que se requiere es el atributo “euros”, que es el que se desea predecir. Los atributos “tipo”, “barrio”, “dorm”, “planta”, “metros”, “estado”, “ciudad” también se necesitan porque todos aportan información relevante de cara a explicar el atributo “euros”. Sin embargo, como ya hemos comentado el atributo “días” lo vamos a eliminar, debido a que el enunciado del modelo señala que es independiente de los días. En cuanto a la eliminación de atributos, con eliminar el atributo “días”, habremos eliminado todos los atributos que no se requieren para este modelo de los presentes en la vista de datos general, que es la entrada de la primera tarea especificada de este modelo al igual que en los otros dos modelos. No obstante, no habremos llegado a la vista minable particular eliminando el atributo “días”, ya que, para este modelo se necesita otra operación de preparación de los datos. La otra tarea a realizar es debido a que el enunciado hace mención al precio de las viviendas vendidas, y no al de las no vendidas, por esta razón debemos eliminar las instancias de viviendas no vendidas que son las que tienen el atributo “compra”=’NO’. Seguidamente vamos a introducir como se han realizado las tareas específicas del modelo 2, en la tarea novena y décima.

9. En esta tarea vamos a crear un fichero para copiar las instancias de las viviendas de la vista de datos general sin el atributo “días”. Esta tarea como se puede apreciar es la misma que la tarea octava enunciada anteriormente, sin embargo, el fichero en el que guardaremos el resultado va a tener un nombre diferente al del fichero creado en la tarea octava, ya que, se va a nombrar como “ViviendasPreviaModelo2Vista_MinableFecha1-Fecha2.arff”. El fichero creado tendrá el mismo contenido que el del fichero de la vista minable del modelo 1, porque se habrá realizado la misma acción a partir de la misma vista de datos general de las viviendas. Sin embargo, se podría haber realizado antes que eliminar el atributo “días”, la eliminación de las instancia de viviendas que tienen el atributo “compra”=’NO’ y entonces el fichero sería distinto que el que este en la vista minable del modelo 1. Respecto a la acción, no vamos a volver a explicar el procedimiento de la eliminación del atributo “días”, ya que, es el mismo que el que se ha realizado en la tarea octava, y el resultado de la tarea tampoco vamos a visualizarlo en ninguna figura porque es el mismo que se observa en la figura 2.7.

10. Después de eliminar el atributo “días” falta eliminar las instancias de viviendas que no han sido vendidas. Para lo cual, en esta tarea décima crearemos un nuevo fichero, a partir del fichero “ViviendasPreviaModelo2VistaMinableFecha1-Fecha2.arff” en el cual se eliminaba el atributo “días” a las viviendas. Ahora para eliminar las instancias de viviendas no vendidas del fichero “ViviendasPreviaModelo2Vista_

MinableFecha1-Fecha2.arff” la aplicación “WEKA” tiene implementado un filtro de eliminación de instancias denominado “RemoveWithValues” (eliminar con valores) .

Lo primero que deberemos hacer será cargar en la aplicación ‘Explorer’ de “WEKA” el fichero “ViviendasPreviaModelo2VistaMinableFecha1-Fecha2.arff”. Luego pulsaremos sobre el botón ‘Chosee’ de la sección ‘Filter’ para buscar el filtro “RemoveWithValues”. Para encontrar dicho filtro seleccionaremos la carpeta de los filtros no supervisados y luego la carpeta de los filtros de instancias. Una vez seleccionado el nombre de filtro aparecerá en el ‘label’ de la sección ‘Filter’. Al pulsar sobre tal label nos aparecerá un cuadro de diálogo en el que podremos configurar el filtro.

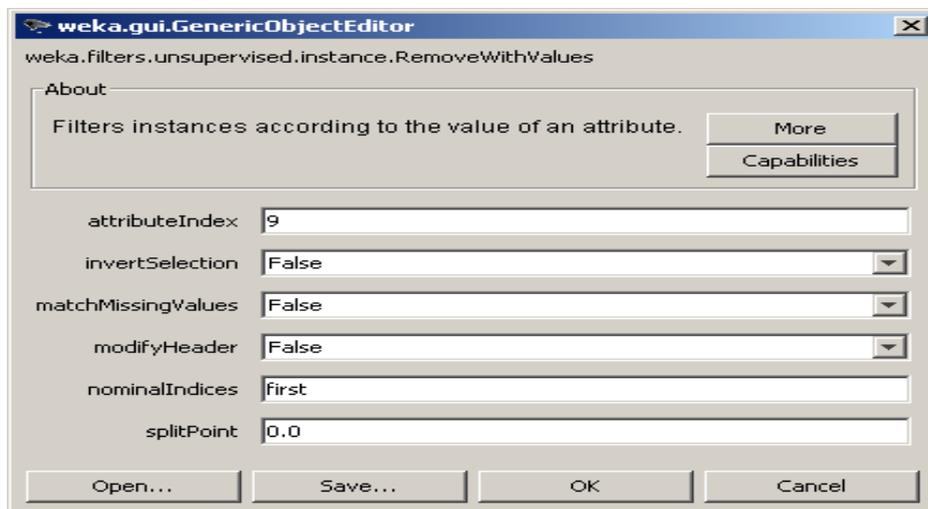


Figura 2.8: Cuadro de diálogo perteneciente a la configuración del filtro “RemoveWithValues”. Con la configuración correcta para eliminar las instancias que deseamos eliminar.

En la figura 2.8, se muestra el cuadro de dialogo, en el que se puede configurar el filtro “RemoveWithValues”. En tal figura 2.8 ya se ha configurado el filtro con nuestras necesidades y por lo tanto, vamos a explicar los valores que aparecen en cada de las opciones, desde la opción “attributeIndex”, hasta la última opción que se observa llamada “splitPoint”.

Lo primero que vamos a indicar de las opciones es que hay dos que son exclusivas para atributos nominales y otra que es exclusiva para atributos numéricos. Las dos opciones que actúan sólo en atributos nominales son “modifyHeader” y “nominalIndices” y la que sólo actúa con atributos numéricos es la última que aparece llamada “splitPoint. Por lo tanto, las opciones “modifyHeader” y “nominalIndices” se ignoran cuando el atributo sobre el que se quiere hacer la operación es numérico y la opción “splitPoint” se ignora si el atributo es nominal. Dicho esto vamos a explicar cada una de las opciones en orden de aparición:

- “attributeIndex”: A esta opción se le debe indicar el atributo al que se le quiere aplicar el filtro. Para indicar un atributo determinado hay que insertar en el ‘textBox’ de la opción el número que tiene asignado el atributo en la sección ‘Attributes’. Esta opción evidentemente no puede ser ignorada. En este caso en la figura 2.8 se ha introducido en tal opción el número ‘9’, porque dicho número es el que hace referencia al atributo “compra” que es el que queremos filtrar.
- “invertSelection”: Esta opción se puede establecer a ‘true’ o ‘false’. El valor ‘true’ indica que se invierte la tarea de filtrado. El valor ‘false’ indica que no se invierte la tarea de filtrado. Esta opción se tiene en cuenta siempre. En la figura 2.8 se presenta con el valor ‘false’, porque no se quiere invertir la tarea de filtrado.
- “matchMissingValue”: Puede adoptar también solamente el valor ‘true’ o el valor ‘false’. Si tiene el valor ‘true’, entonces también se borran las instancias en las que el atributo seleccionado no tenga valor alguno, es decir, este perdido. Si posee el valor ‘false’ no se borran las instancias en que el atributo seleccionado no tenga valor alguno. Esta opción tampoco se ignora nunca. En este caso presenta el valor ‘false’, aunque no importaría que tuviera el valor ‘true’ debido a que el atributo “compra” no posee valores perdidos en los valores que se observan en la ventana ‘Preprocess’.
- “modifyHeader”: Esta opción puede tener dos posibles valores que son también ‘true’ o ‘false’. Si el valor que adopta es ‘true’, entonces además de eliminar las instancias del atributo seleccionado que tengan unos determinados valores, se elimina también las etiquetas que referencian a esos valores. Por el contrario, si el valor que tiene es ‘false’ no se eliminan las etiquetas que referencian a los valores eliminados. Al no eliminarse las etiquetas que referencia a los valores eliminados éstas indican que hay cero valores con esas características. En la configuración que estamos analizando presenta el valor ‘false’ porque no queremos eliminar la etiqueta ‘NO’ en este caso. Una cuestión que hay que tener en cuenta de dicha opción “modifyHeader” es que cuando actúa con la opción ‘true’ elimina los valores que se desean eliminar además de en las instancias que se observan, también en la cabecera del fichero “.arff” que se esté manipulando en “WEKA”.
- “nominalIndices”: En el ‘textBox’ adyacente a tal opción se deben indicar los valores nominales que se quieren eliminar del atributo al que se le aplica el filtro. En este caso sólo queremos eliminar las instancias de viviendas que tienen el valor ‘NO’ en el atributo “compra”. Por lo que, indicamos el valor “first” o el valor numérico ‘1’, ya que el atributo “compra” presenta el valor ‘NO’ en la primera posición. Si quisiéramos eliminar el valor ‘SÍ’, entonces introduciríamos el valor “last” (último) o el valor numérico ‘2’ porque el valor ‘SÍ’ está en la última posición detrás del valor ‘NO’.

- “splitPoint”: Está es la última opción de la configuración del filtro “RemoveWithValues”. Esta opción elimina los valores del atributo numérico al que se le aplica el filtro que sean menores al que aparezca en su ‘textbox’. Como en este caso el atributo es nominal la opción “splitPoint” se ignora. Una cuestión que hay que tener en cuenta es que como hemos dicho la opción “invertSelection” invierte la selección si se establece al valor ‘true’, por lo tanto, si la opción “invertSelection” tuviera el valor ‘true’, entonces, las opciones “nominalIndice” y “splitPoint” actuarían a la inversa.

Cuando dispongamos del filtro correctamente configurado presionaremos sobre el botón “OK”, con lo que desaparecerá el cuadro de dialogo y la configuración del filtro se habrá guardado y sólo faltará pulsar el botón ‘Apply’ que aparece en la parte derecha de la sección ‘Filter’, para eliminar los ejemplares cuyo atributo “compra” tengan el valor ‘NO’. La figura 2.9 muestra el resultado de la eliminación de las instancias de viviendas declaradas como no vendidas.

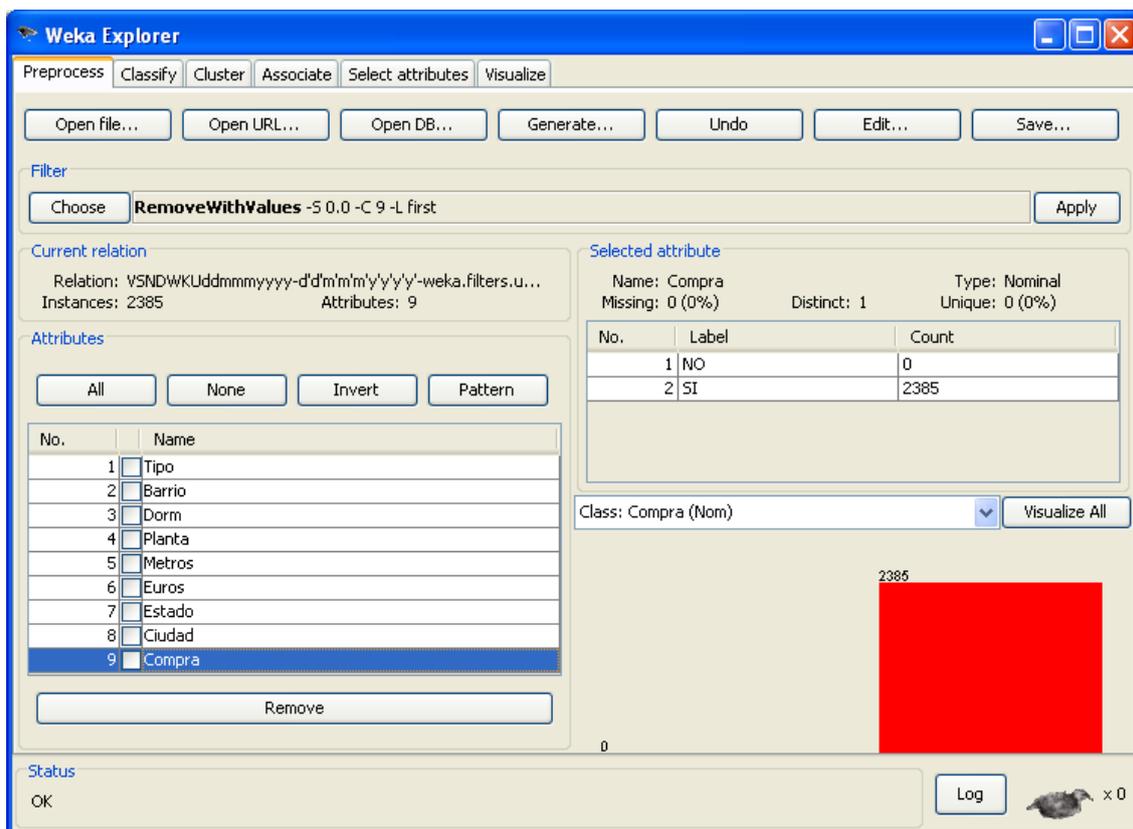


Figura 2.9: Muestra el resultado de la eliminación de las instancias de viviendas no vendidas. En la sección ‘Filter’ se observa el filtro “RemoveWithValues” con la configuración que se ha utilizado: “RemoveWithValues –S 0.0 –C 9 –L first”, ‘-C 9’, indica que se le pasa al filtro el atributo que tiene asignado el número ‘9’, ‘-L first’, indica que se aplica sobre el primero valor del atributo que tiene asignado el número ‘9’, ‘-S 0.0’ se ignora debido a que dicho argumento pertenece a la opción ‘splitPoint’.

2.7.3. Vista minable del modelo 3

Después de alcanzar la vista minable del modelo 1 y del modelo 2 nos faltará llegar a la vista minable del modelo 3. Para lo cual, al igual que hicimos con la preparación particular del modelo 2 retrocederemos a la vista de datos general, ya que, las últimas tareas realizadas han tenido como finalidad o alcanzar la vista minable particular del modelo 1 o del modelo 2. Seguidamente volvemos a enunciar el modelo 3:

-Modelo3: Precio por el que vender una vivienda teniendo en cuenta el número de días.

El enunciado del modelo 3 es muy parecido al del modelo 2 como se puede apreciar si leemos el enunciado de ambos modelos. De hecho la diferencia entre ambos modelos reside en que en éste se tiene en cuenta además el atributo “días” y en el modelo 2 no. Por lo tanto, en este modelo se requiere el atributo “días” como un atributo predictivo más para predecir el atributo “euros”, además del resto de atributos de la vista de datos general. Con lo que no se va a eliminar ninguno de los atributos disponibles en la vista de datos general de las viviendas para alcanzar la vista minable particular del modelo 3, ya que se necesitan todos. Lo que sí se debe realizar para alcanzar la vista minable particular de este modelo es eliminar las viviendas que no se han podido vender que aparecen en la vista de datos general de las viviendas. Esta tarea es idéntica a la segunda tarea que se ha realizado en el modelo 2. Sin embargo, la volveremos a realizar con el atributo “días” como parte de los atributos de las viviendas, por lo que, esta vez tendremos en el resultado final las instancias de las viviendas vendidas con el atributo “días”. Con esta tarea alcanzaremos la vista minable particular del modelo 3. Seguidamente relatamos la tarea específica del modelo 3, que viene a ser la tarea undécima:

11. En la tarea undécima volveremos a cargar en la aplicación ‘Explorer’ de “WEKA” el fichero “ViviendasVistaGeneralFecha1-Fecha2.arff” que contiene como hemos dicho varias veces la vista de datos general. Se trata de eliminar las instancias de las viviendas que tengan el valor ‘NO’ en el atributo “compra”, pero a diferencia del modelo 2 en las viviendas se dispone del atributo “días” en esta tarea. Al ser la misma tarea que en la tarea décima volveremos a seleccionar el filtro “RemoveWithValues”. Como el atributo “compra” se referenciará con el mismo número que en la tarea décima la configuración del filtro es exactamente idéntica, y no la vamos a volver a detallar. Simplemente decir, que en la figura 2.10 se muestra el resultado de la tarea undécima. Con la realización de esta tarea se alcanza la vista minable particular del modelo 3. Debido a esta circunstancia se guardará el fichero resultante de esta tarea con el nombre “ViviendasModelo3VistaMinableFecha1-Fecha2.arff”.

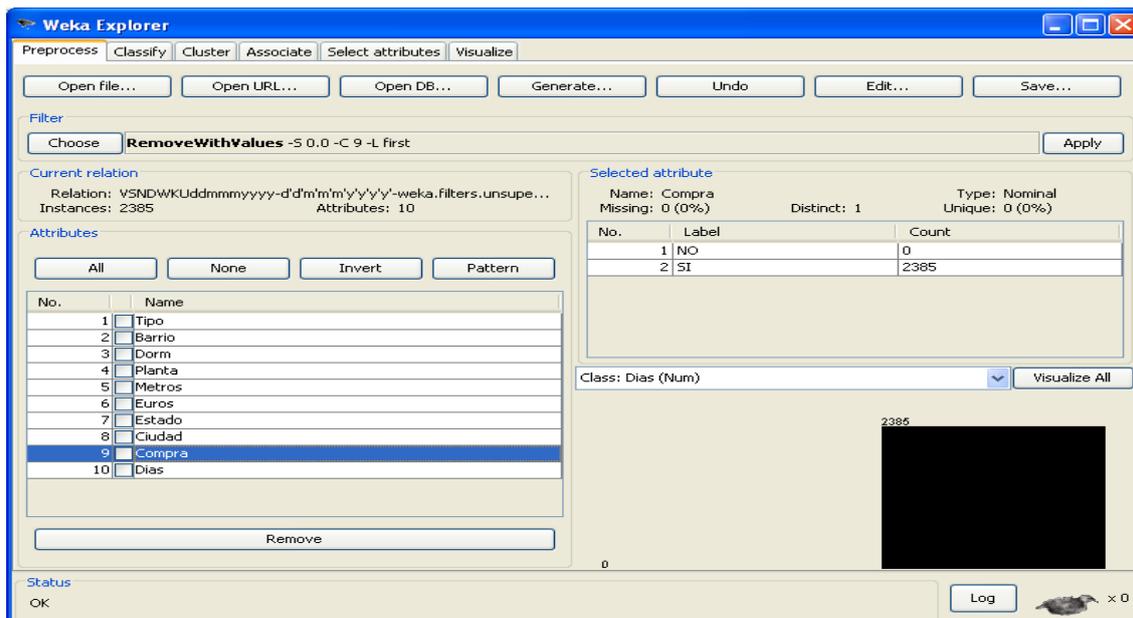


Figura 2.10: En este caso también se eliminan las instancias de viviendas que tienen el valor 'NO' en el atributo "compra". Sin embargo, el resultado no es el mismo que el resultado del filtro "RemoveWithValues" en el modelo 2. En efecto, el resultado es distinto debido a que en este caso en las viviendas también se observa el atributo "días".

Una vez tenemos la vista minable particular de cada modelo habremos concluido la fase de preparación del proceso "KDD". Aunque como el proceso "KDD" es un proceso iterativo podría darse el caso de tener que volver a la fase de preparación con posterioridad, no obstante, en este proyecto, no vamos a realizar más de una pasada o iteración en la fase de preparación.

2.8. Datos concretos

Ahora lo que haremos para finalizar este apartado será mencionar los datos concretos a utilizar para la construcción de los modelos de datos (que no son los de ejemplo que se han utilizado solamente para explicar el proceso de preparación) estos datos que se van a utilizar para los modelos son datos de las viviendas extraídos del portal web inmobiliario de las ciudades de "Valencia", "Madrid" y "Barcelona". Por lo que, las instancias de viviendas de otras ciudades no las hemos extraído.

Como se ha comentado en el proceso de preparación se requiere la extracción de los datos del portal web, en dos fechas de tiempo distintas. Nosotros concretamente hemos realizado la primera extracción de los datos de las viviendas 'el 26 de enero de 2010' y la segunda extracción 'el 12 de febrero de 2010'. Precisamente vamos a estudiar dichas extracciones, ya que contienen los datos que una vez preparados se utilizarán para crear los modelos.

Lo primero que vamos a realizar consistirá en introducir una tabla con la información de entrada y de salida de las tareas relativas a la preparación de dichos datos concretos. Después se expondrá otra tabla formada por el nombre de los ficheros de salida de las tareas, que informará del número de atributos que posea cada fichero y de la talla de cada fichero. Esta tabla se expone con la finalidad de saber cuánto contenido posee cada uno de los ficheros de salida. Dicha tabla indicará la tarea en la que el fichero es creado también. Una percepción que hay que tener en cuenta es que en cuanto a los atributos de cada fichero hay algunos que se utilizan, como se explica en la preparación, para obtener otros. Por otra parte, por último, se analizarán los valores de los atributos para ello mostraremos esta vez en “WEKA” dichos valores. Para ello utilizaremos el fichero de la vista de datos general ya que en dicha vista aparecen todas las viviendas y atributos definitivos de éstas, que se utilizan en alguno de los modelos.

La tabla de la información de entrada y de salida de las tareas de la preparación de los datos concretos se expone seguidamente:

Tarea	Información de entrada que recibe la tarea	Fichero/s de salida después de realizar la tarea
1ª	(Las instancias de viviendas almacenadas en el portal web “el 26 de enero de 2010”).	Viviendas26-ene-2010.txt
2ª	(Las instancias de viviendas almacenadas en el portal web “el 12 de febrero de 2010”).	Viviendas12-feb-2010.txt
3ª	1) Viviendas26-ene-2010.txt (“fichero base”) 2) Viviendas12-feb-2010.txt (“fichero comparar”)	1) VSINO26ene2010-12feb2010.txt (seleccionado) 2) VResto26ene2010-12feb2010.txt (descartado)
4ª	El fichero: VSINO26ene2010-12feb2010.txt	VSNDias26ene2010-12feb2010.txt
5ª	El fichero: VSNDias26ene2010-12feb2010.txt	VSNDWK26ene2010-12feb2010.csv
6ª	El fichero: VSNDWK26ene2010-12feb2010.csv	VSNDWКУ26ene2010-12feb2010.csv
7ª	El fichero: VSNDWКУ26ene2010-12feb2010.csv	ViviendasVistaGeneral26ene2010-12feb2010.arff
8ª	El fichero: ViviendasVistaGeneral26ene2010_-12feb2010.arff	ViviendasModelo1VistaMinable26ene2010_-12feb2010.arff
9ª	El fichero: ViviendasVistaGeneral26ene2010_-12feb2010.arff	ViviendasPreviaModelo2VistaMinable26ene2010_-12feb2010.arff
10ª	El fichero: ViviendasPreviaModelo2Vista_Minable26ene2010-12feb2010.arff	ViviendasModelo2VistaMinable26ene2010_-12feb2010.arff
11ª	El fichero: ViviendasVistaGeneral26ene2010_-12feb2010.arff	ViviendasModelo3VistaMinable26ene2010_-12feb 2010.arff

Tabla 2.5: Información de entrada y de salida de las tareas de preparación de los datos concretos.

Después de introducir la tabla 2.5, se introducirá la tabla 2.6 correspondiente a la información referente al número de atributos y a la talla (número de instancias) de cada uno de los ficheros de salida, de las tareas del proceso de preparación concreto, realizado en este proyecto, en cada fila se vuelve a indicar también en que tarea se crea cada uno de los ficheros:

Nombre del fichero	Creado en la tarea número	Número de atributos	Talla
Viviendas26-ene-2010.txt	1	12	28338
Viviendas12-feb-2010.txt	2	12	28979
VSINO26ene2010-12feb2010.txt	3	13	4593
VResto26ene2010-12feb2010.txt (descartado)	3	12	28555
VSNDias26ene2010-12feb2010.txt	4	12	4593
VSNDWK26ene2010-12feb2010.csv	5	11	4593
VSNDWKU26ene2010-12feb2010.csv	6	11	4583
ViviendasVistaGeneral26ene2010-12feb2010.arff	7	10	4583
ViviendasModelo1VistaMinable26ene2010-12feb2010.arff	8	9	4583
ViviendasPreviaModelo2VistaMinable26ene2010-12feb2010.arff	9	9	4583
ViviendasModelo2VistaMinable26ene2010-12feb2010.arff	10	9	2448
ViviendasModelo3VistaMinable26ene2010-12feb2010.arff	11	10	2448

Tabla 2.6: Información del número de atributos y de la talla de cada fichero de salida, también se vuelve a indicar la tarea en la que se crea cada uno de dichos ficheros de salida.

Una vez introducida la tabla 2.5 y la tabla 2.6 nos centraremos como dijimos en los valores de los atributos de los datos concretos, para ello observaremos los datos del fichero “ViviendasVistaGeneral26ene2010-12feb2010.arff”, ya que, dispone de todos los atributos disponibles que se utilizarán en los modelos. En las figuras siguientes observamos dichos valores con la ayuda de la ventana ‘Preprocess’ de la aplicación “WEKA”.

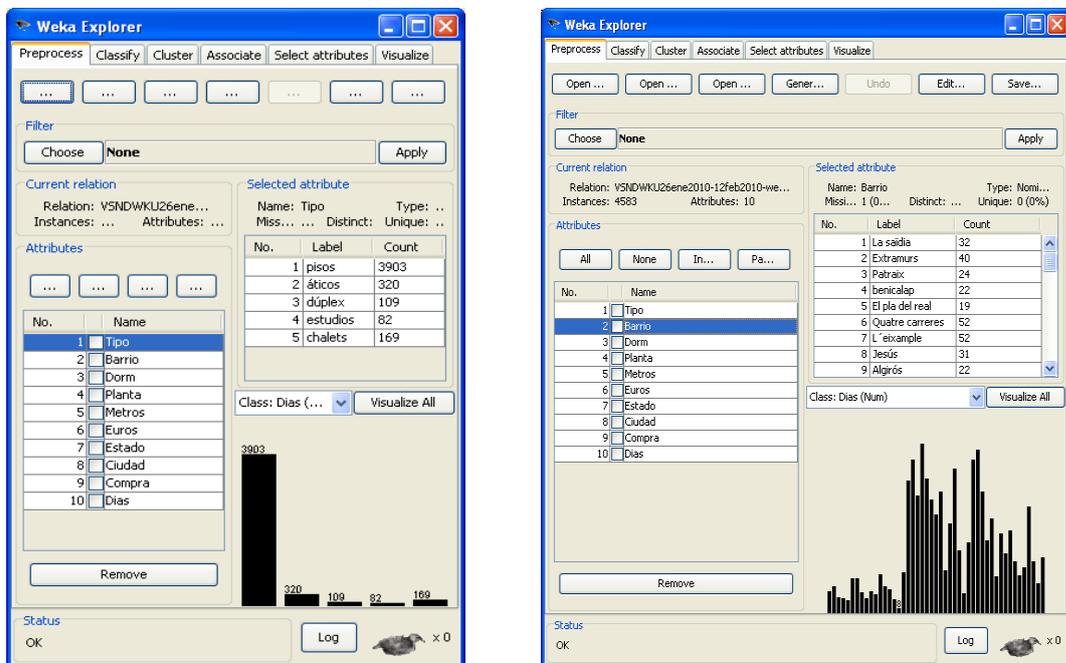


Figura 2.11: Se observan los valores del atributo “tipo” y los primeros valores del atributo “barrio”.

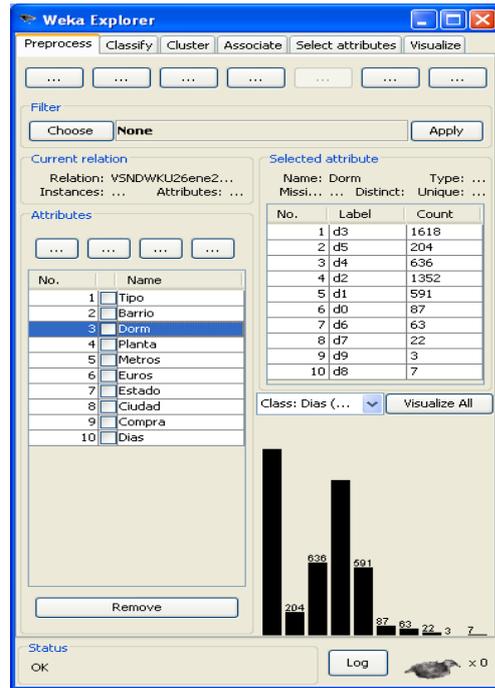
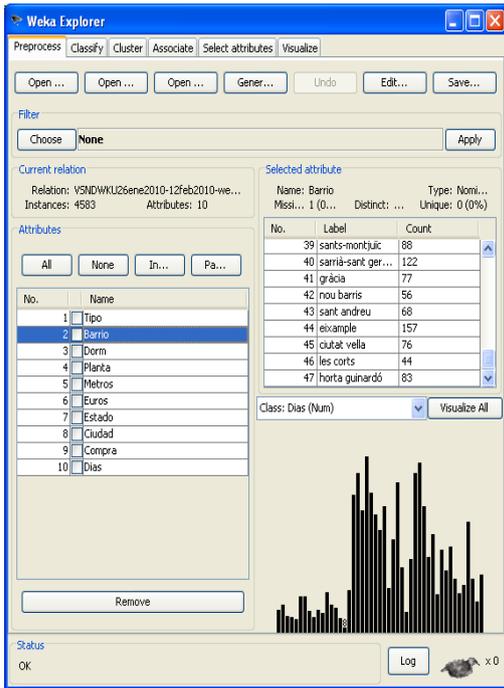


Figura 2.12: Se observan los últimos valores del atributo "barrio" y los valores del atributo "dorm".

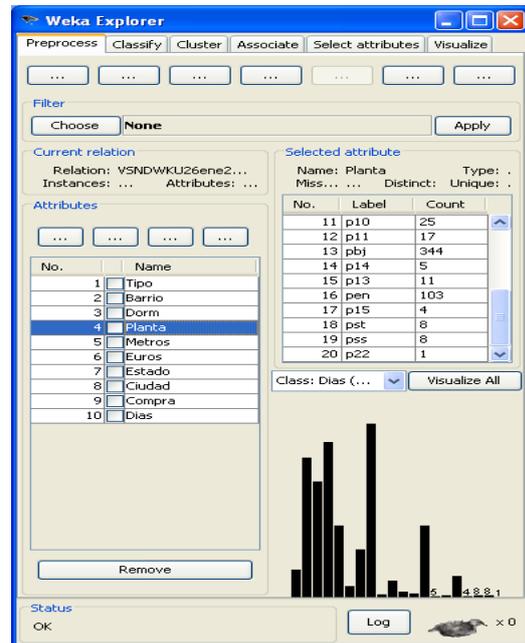
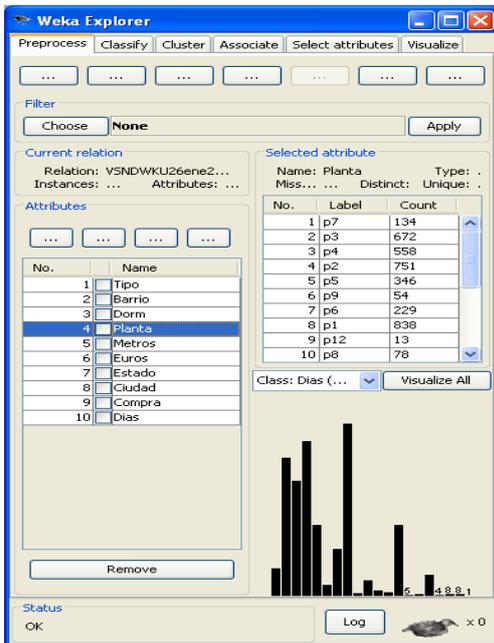


Figura 2.13: Se muestran los valores del atributo "planta".

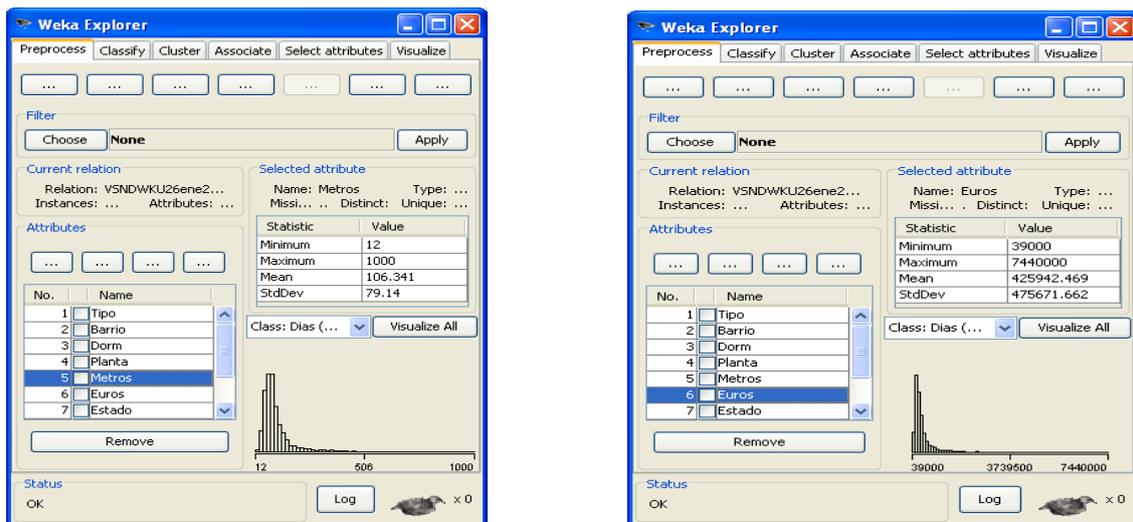


Figura 2.14: Se observan los valores del atributo “metros” y del atributo “euros”.

Una cuestión que hay que mencionar es que como ya se dijo con anterioridad “WEKA” utiliza para diferenciar la parte fraccionaria de la parte entera de un número, el punto (‘.’) siguiendo la notación anglosajona, como se puede apreciar en los valores estadísticos que se observan en la figura 2.14 en los atributos numéricos “metros” y “euros”. Además como también se observa, en dicha figura 2.14 no se aprecia ningún símbolo que separe los millares en los números. Al respecto indicar, que en esta memoria cuando la intención sea mostrar la información tal cual se visualiza en “WEKA” vamos a dejar los números con la notación que se presenta en “WEKA”. No obstante, cuando la intención sea mostrar algún valor para describirlo, realizar algún cálculo o concluir algún valor pondremos una coma para separar la parte fraccionaria de la entera. Y para evitar confusiones ya no separaremos mediante un punto los millares.

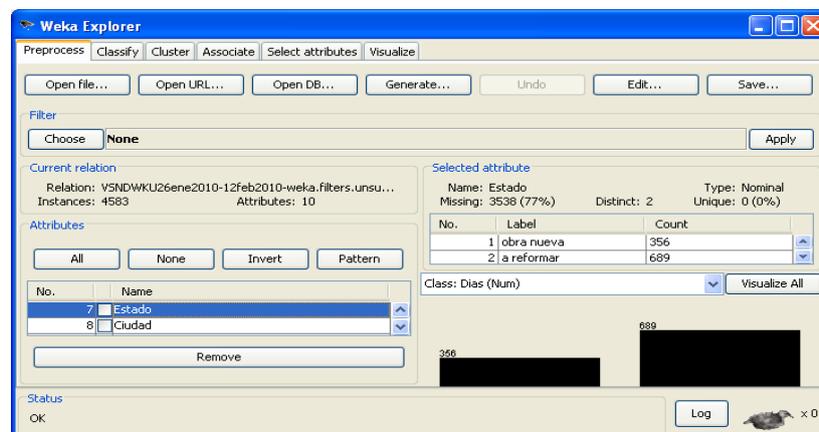


Figura 2.15: Se muestran los valores del atributo “estado”.

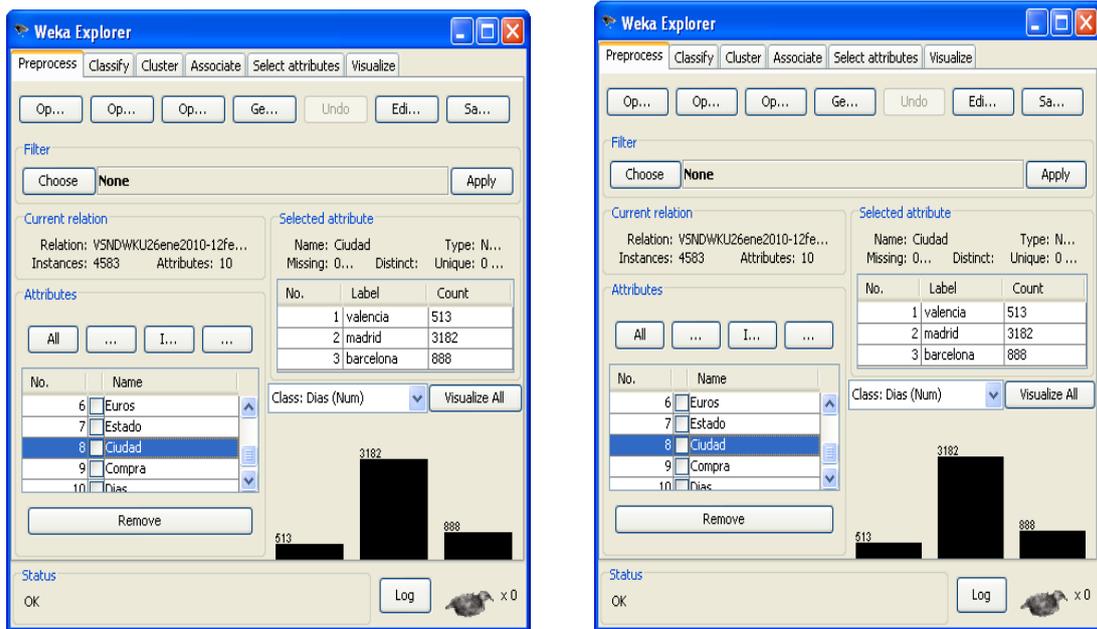


Figura 2.16: Valores del atributo “ciudad” y del atributo “compra”.

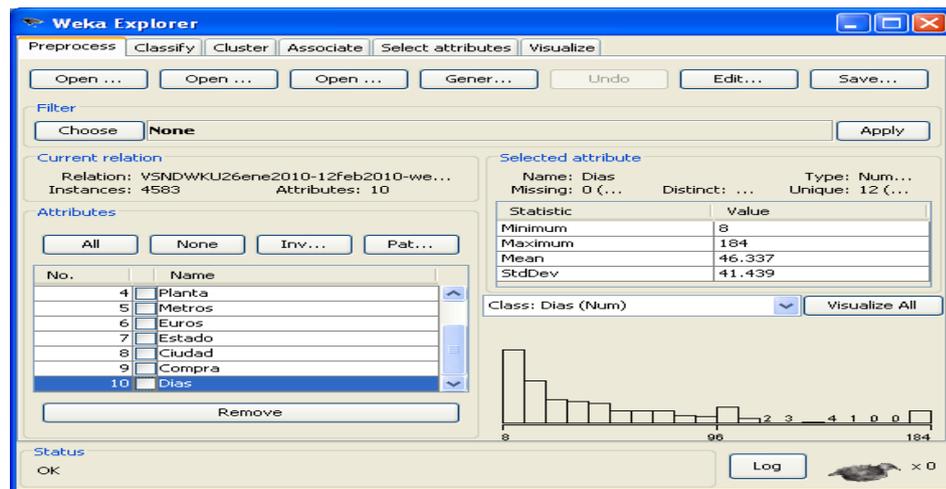


Figura 2.17: Se muestran los valores atributo “días”.

En estas imágenes que se han introducido referentes a los valores de los atributos de los datos concretos, se observa que en los atributos nominales se muestran los distintos valores junto con el número de viviendas que los poseen, mientras que, en los atributos numéricos se observa cuatro medidas estadísticas: el valor mínimo, el valor máximo, la media y la desviación estándar. Además en cada uno de los atributos ya sea nominal o numérico se percibe una representación gráfica de los valores de cada atributo. No obstante, del atributo “barrio” no hemos mostrado todos los valores que posee debido a que son un total de “47” barrios y no se pueden observar todos en las dos imágenes utilizadas para mostrar dichos valores del atributo “barrio”, por lo que, se va a mostrar seguidamente un recuadro con todos los valores de dicho atributo.

'La saïdia', 'Extramurs', 'Patraix', 'benicalap', 'El pla del real', 'Quatre carreres', 'L'eixemple', 'Jesús', 'Algirós', 'campanar-beni...', 'Poblats marítims', 'Ciutat vella', 'Camins al grau', 'Rascanya', 'L'olivereta', 'Benimaclet', 'villa de vallecas', 'puente de vallecas', 'carabanchel', 'chamberí', 'centro', 'chamartín', 'latina', 'moncloa', 'hortaleza', 'moratalaz', 'tetuán', 'arganzuela', 'villaverde', 'vicálvaro', 'usera', 'ciudad lineal', 'salamanca', 'san blas', 'retiro', 'fuencarral', 'barajas', 'sant martí', 'sants-montjuïc', 'sarrià-sant gervasi', 'gràcia', 'nou barris', 'sant andreu', 'eixample', 'ciutat vella', 'les corts', 'horta guinardó'

Recuadro 2.13: Se aprecian los "47" valores distintos del atributo "barrio".

Seguidamente, detallaremos de forma general los valores de los atributos que se observan en las figuras que van desde la 2.11 a la 2.17 y también los que se aprecian en el recuadro 2.13:

- En cuanto al atributo "tipo" como se percibe en la imagen izquierda de la figura 2.11 adopta los valores 'pisos', 'áticos', 'dúplex', 'estudios' y 'chalets'.
- Respecto al atributo "barrio" tiene los valores que se observan en el recuadro 2.13 y que también se observan algunos en la imagen derecha de la figura 2.11 y en la imagen izquierda de la figura 2.12. El primer valor que se observa es el del barrio llamado 'La saïdia' y el último corresponde a la 'horta guinardó'. De entre todos los valores de barrio hay uno que puede llevar a confusión porque lleva tres puntos suspensivos dicho valor es 'campanar-beni...', los tres puntos suspensivos en este caso forman parte del nombre. Y hay otros dos barrios que pueden llevar a confusión debido a que tienen el mismo nombre, aunque el portal web para diferenciarlos ha hecho que uno empiece por la letra mayúscula y el otro por la letra minúscula. Estos dos barrios son 'Ciutat vella' y 'ciutat vella', y aunque sólo se diferencia en la letra mayúscula el primero que se ha denominado 'Ciutat vella', es un barrio de la ciudad de Valencia y el segundo 'ciutat vella', de la ciudad de Barcelona.
- Respecto al atributo "dorm" los valores están comprendidos entre 'd0'...'d9', es decir, aparecen viviendas que tienen desde cero dormitorios hasta nueve dormitorios, esto se aprecia en la imagen derecha de la figura 2.12. En el atributo "planta" como se observa en las imágenes de la figura 2.13 y 2.14 se aprecian los valores entre 'p1'...'p15', y además los valores 'p22', 'pbj', 'pen', 'pst' y 'pss'. Estos valores vienen a indicar que las instancias de viviendas recopiladas están entre la planta primera y la planta decimoquinta o están en la planta veintidosava o también en la planta baja debido al valor 'pbj' o en el entresuelo por el valor 'pen' o en la planta del sótano por el valor 'pst' o por último en el semisótano debido al valor 'pss'.
- En cuanto al atributo "metros" como se observa en la imagen izquierda de la figura 2.14 los valores oscilan entre el valor '12' metros y el valor '1000' metros, siendo la media '106,345' metros. En cuanto a los valores del atributo "euros" se observa en la imagen derecha de la figura 2.14 que abarcan entre '39000 €' y '7440000 €', con una media de '425976,503 €'.

- Por lo que hace al atributo “estado” se percibe en la figura 2.15 que los valores que se observan son ‘obra nueva’ que indica que se ha hecho una obra nueva y ‘a reformar’, que referencia a que está para reformar.
- En referencia al atributo “ciudad” como se observa en la imagen izquierda de la figura 2.16 y como se dijo con anterioridad se observan valores de las ciudades de Valencia, Madrid y Barcelona.
- En cuanto al atributo “compra” está claro que como hemos dicho varias veces aparecen dos valores ‘SÍ’ y ‘NO’.
- Respecto al atributo “días” como se observa en la figura 2.17 aparecen valores entre ‘8’ días y ‘184’ días, con una media de ‘46,339’ días.

Por otra parte, lo último que vamos a indicar de los datos concretos son los valores perdidos que aparecen. Para lo cual introducimos la tabla 2.7 indicando dichos valores perdidos (missing) que son valores desconocidos que aparecen vacíos.

Atributo	Nº de valores perdidos (Missing)
Tipo	0
Barrio	1
Dorm	0
Planta	384 (8 %)
Metros	0
Euros	0
Estado	3537 (77%)
Ciudad	0
Compra	0
Días	0

Tabla 2.7: Valores perdidos.

Respecto a la tabla 2.7, hay que resaltar que por la preparación realizada: los atributos “euros”, “compra” y “días” es imposible que tengan valores perdidos, ya que, no se permite que estos tres atributos tengan algún valor perdido. En los otros atributos es muy raro que aparezca algún valor vacío, salvo el atributo “estado” que tiene muchos valores vacíos como se observa en dicha tabla. En cuanto al atributo “planta” que posee algunos valores vacíos, cabe decir, que muchos de ellos se deben a las viviendas del tipo ‘chalet’, ya que estas en el portal web no presentaban valor para tal atributo “planta”. Por último, también se aprecia que el atributo “barrio” posee un valor perdido.

Con lo acabado de relatar, es decir, la descripción de los valores de las viviendas que se van a utilizar para construir los modelos concluimos este apartado.

3. Aprendizaje de datos

En este apartado “Aprendizaje de datos” llegamos a la fase denominada minería de datos del proceso “KDD”. En este proyecto, en dicha fase vamos a realizar el aprendizaje de nuestros tres modelos de datos de los que obtuvimos sus respectivas vistas minables en el apartado anterior. Dichos modelos los volvemos a citar seguidamente:

Modelo1: Hallar la compra o la no compra de una vivienda.

Modelo2: Precio por el que vender una vivienda, independientemente de los días.

Modelo3: Precio por el que vender una vivienda, teniendo en cuenta los días.

Para cada uno de los tres modelos, lo que haremos será un estudio de algoritmos. Cada modelo lo construiremos tantas veces como algoritmos de minería de datos se tengan disponibles en la aplicación “WEKA” para realizarlo. La finalidad será seleccionar en cada uno de los estudios una de las construcciones o aprendizajes descartando el resto. Tal construcción seleccionada en cada estudio la analizaremos y le extraeremos patrones y conclusiones. Además adicionalmente utilizaremos el modelo 3 para predecir los euros por los que vender una misma vivienda en función de los días.

3.1. El aprendizaje del modelo 1

En el “modelo 1: hallar la compra o la no compra de una vivienda”, se quieren descubrir patrones que puedan predecir el atributo “compra” de nuevos ejemplares de viviendas de los cuales no se sabe el valor de dicho atributo. La tarea a realizar en este modelo de datos será una tarea predictiva y dentro de las tareas predictivas será una tarea de clasificación, ya que, se trata de clasificar ofertas de viviendas como vendidas, atributo “compra”=‘SÍ’, o como no vendidas, atributo “compra”=‘NO’. La tarea se sabe que es de clasificación porque los valores del atributo “compra” a predecir son discretos, ya que, el atributo “compra” es de tipo nominal. Como la clase a predecir sólo puede adoptar dos posibles valores, es decir, el dominio del atributo “compra” es {‘SÍ’, ‘NO’} la clasificación a realizar se dice que es una clasificación binaria. Por lo tanto, la clasificación consistirá en formar dos grupos: uno con las instancias de viviendas que se clasifiquen como vendidas (“compra”=‘SÍ’) y otro grupo con las que se clasifiquen como no vendidas (“compra”=‘NO’.)

En cuanto a los datos a utilizar para construir el modelo serán los almacenados en el fichero “Modelo1VistaMinable26ene2010-12feb2010.arff”, que contiene los datos concretos de las viviendas en la forma de la vista minable del modelo 1.

Vivienda	Tipo	Barrio	Dorm	Planta	Metros	Euros	Estado	Ciudad	Compra
1	Pisos	La saïda	d3	P7	125	345000	obra nueva	Valencia	NO
2	Pisos	El pla del real	d4	p2	92	270000	a reformar	Valencia	NO
3	Pisos	Latina	d2	p1	60	130000	a reformar	Madrid	Sí
4	Pisos	Moncloa	d3	p5	126	534999		Madrid	Sí
5	dúplex	sants_ -montjuic	d3	p2	102	378000	obra nueva	Barcelona	Sí

Tabla 3.1: Parte del fichero “Modelo1VistaMinable26ene2010-12feb2010.arff” del cual se muestran ‘5’ instancias de viviendas de un total de ‘4583’ de que dispone tal fichero.

Los atributos “tipo”, “barrio”, “dorm”, “planta”, “metros”, “euros”, “estado” y “ciudad” actuarán como variables explicativas y el atributo “compra”, como es el atributo a predecir, lo hará como variable de salida o clase. En la tabla 3.1 se observan dichos atributos y alguno de los valores adjuntos que disponen. Se observan dos instancias de viviendas con el atributo “compra”=’NO’ y tres con el atributo “compra”=’SÍ’. En total el fichero posee ‘4583’ instancias de viviendas, de las cuales adoptan el atributo “compra”=’SÍ’, ‘2448’ y el atributo “compra”=’NO’, ‘2135’.

Una vez conocida la tarea a realizar y mostradas algunas de las instancias de viviendas que se le suministrarán al modelo, vamos a construir el modelo de datos cada vez con cada uno de los algoritmos que puedan realizarlo de los disponibles en “WEKA”, es decir, vamos a tratar el estudio de algoritmos del modelo 1. Estos algoritmos están organizados por familias, cada una de ellas posee unas técnicas para construirlo. El primer requisito que necesitaremos para realiza tal estudio será disponer de las instancias de las viviendas de entrada del modelo sobre la aplicación ‘Explorer’ de “WEKA”. Para ello abriremos en dicha aplicación el fichero “Modelo1VistaMinable26ene2010-12feb2010.arff”, que aparecerá con los datos reflejados en la ventana ‘Preprocess’.

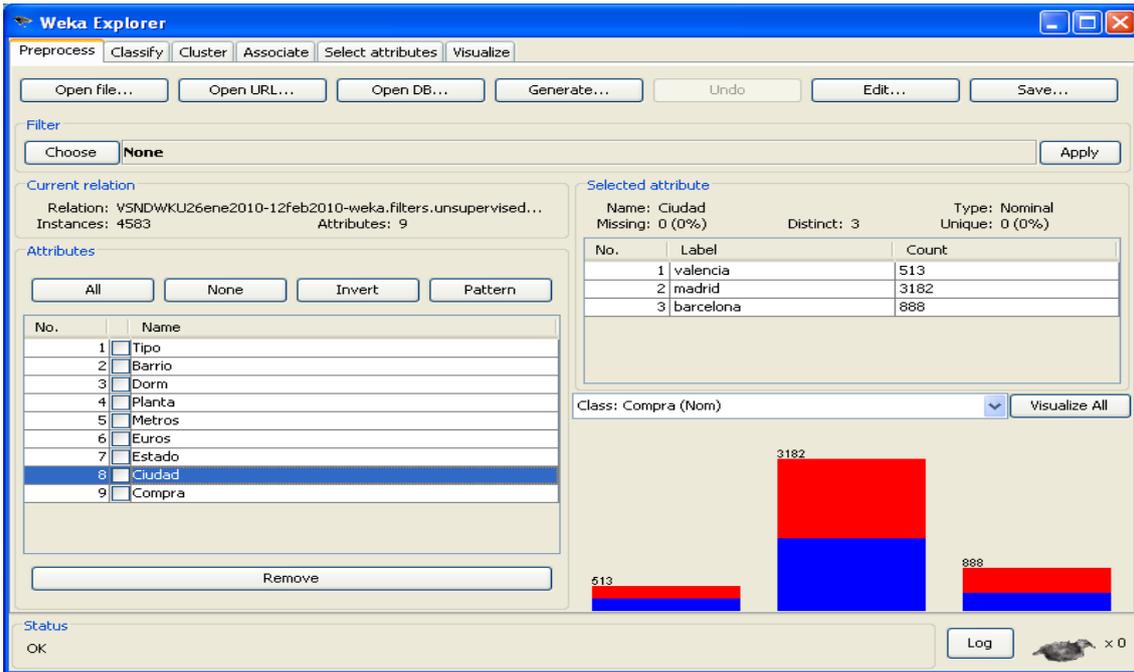


Figura 3.1: Se observa la pestaña 'Preprocess' de la aplicación, 'Explorer' de "WEKA", con los datos del fichero "Modelo1VistaMinable26ene2010-12feb2010.arff".

Ahora que tendremos los datos de entrada habrá que indicarle a "WEKA" que el atributo "compra" es la clase del modelo, también escoger el método que se utilizará para la prueba del modelo y disponer de los algoritmos para la serie de construcciones. La ventana 'Classify' es la que nos ofrece todas estas opciones, ya que, dispone de un control "comboBox" para indicar la clase, de unos "radioButton" para seleccionar el método de evaluación y de los algoritmos de predicción de minería de datos que tiene implementados la aplicación "WEKA". Por lo que, transitaremos a tal ventana desde la ventana 'Preprocess'. La figura 3.2 muestra la ventana 'Classify'.

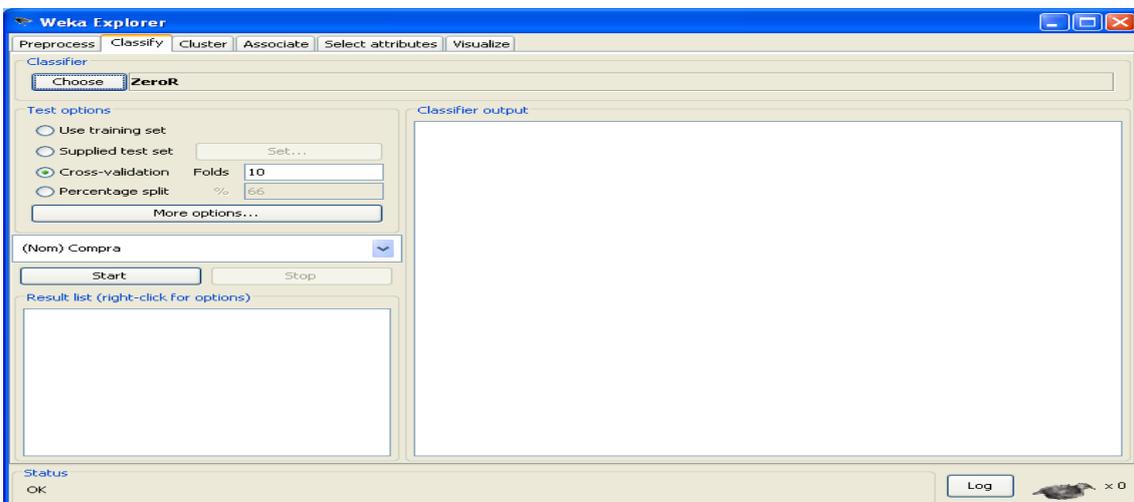


Figura 3.2: Ventana 'Classify'.

En esta ventana debemos seleccionar como clase a predecir, el atributo “compra”, la clase del modelo se selecciona en el control ‘comboBox’ situado debajo de la sección ‘Test options’. En dicho ‘comboBox’ se observa que el atributo “compra”, ya aparece como clase a predecir por lo que, ya tendremos escogida por defecto la clase del modelo. También debemos escoger el método de validación que aparecen en la sección ‘Test options’. Como deseamos indicar el método de “Validación cruzada 10 X 10” y también aparece seleccionado por defecto, tampoco habrá que seleccionarlo, porque es el método de validación escogido por defecto. El método de validación lo explicaremos en el apartado “4.Evaluación”.

Ahora, una vez tenemos la clase a predecir y el método de validación podremos realizar el estudio de algoritmos. Cada vez que se utilice un algoritmo para construir el modelo habremos pasado por la fase de minería de datos y por la fase de evaluación una vez, por lo que haremos, estas dos fases tantas veces de manera conjunta en el estudio del modelo como algoritmos distintos utilicemos. Los algoritmos disponibles en “WEKA”, que utilizaremos cada uno, para realizar una construcción distinta del modelo 1, los mostramos seguidamente ordenados por familias o tipo de algoritmos:

- ❖ De la familia “*Bayes*”:
 - “*BayesNet*”, “*NaiveBayes*”, “*NaiveBayesSimple*” y “*NaiveBayesUpdateable*”.
- ❖ De la familia “*Functions*”:
 - “*Logistic*”, “*RBFNetwork*”, “*SimpleLogistic*”, “*SMO*” y “*VotedPerceptron*”.
- ❖ De la familia “*Lazy*”:
 - “*IB1*”, “*IBK*”, “*Kstar*” y “*LWL*”.
- ❖ De la familia “*Metalearned*”:
 - “*AdaBoostM1*”, “*AttributeSelectedClassifier*”, “*Bagging*”,
“*ClassificationViaClustering*”, “*ClassificationViaRegression*”,
“*CVParameterSelection*”, “*Dagging*”, “*Decorate*”, “*END*”,
“*EnsembleSelection*”, “*FilteredClassifier*”, “*Grading*”, “*LogitBoost*”,
“*MultiBoostAB*”, “*MultiClassClassifier*”, “*MultiScheme*”,
“*OrdinalClassClassifier*”, “*RacedIncrementalLogitBoost*”,
“*RandomSubSpace*”, “*RotationForest*”, “*Stacking*”, “*StackingC*”,
“*ThresholdSelector*” y “*Vote*”.

- ❖ De la familia, “*NestedDichotomies*” dentro de la familia, “*Metalearned*”:
 - “*ClassBalancedNDm*”, “*DataNearBalancedND*” y “*ND*”.
- ❖ De la familia, “*Miscellaniuous*”:
 - “*HyperPipes*” y “*VFI*”.
- ❖ De la familia, “*Rules*”:
 - “*ConjunctiveRule*”, “*DecisionTable*”, “*DTNB*”, “*Jrip*”, “*OneR*”, “*PART*”, “*Ridor*” y “*ZeroR*”.
- ❖ Y por último de la familia, “*Trees*”:
 - “*RandomForest*”, “*RandomTree*”, “*RepTree*” y “*SimpleCart*”.

Aunque “WEKA” tiene más algoritmos de predicción, los acabados de listar son los únicos que pueden realizar la tarea de clasificación que se requiere en el modelo 1. Hay que resaltar, como ya se dijo en la introducción, que cada familia de algoritmos se valdrá de unas técnicas para realizar la construcción del modelo, con lo que los algoritmos de cada familia tendrán una forma distinta de obtener la construcción del modelo y con ello de mostrar los patrones de dicho modelo. Por ejemplo, los algoritmos de la familia “bayes” se basan en la fórmula de bayes, los algoritmos de la familia “tree” son en forma de árboles de decisión y los de la familia “rules” construyen el modelo mediante reglas. Otra peculiaridad de cada familia de algoritmos se observa en el tiempo que requieren para construir el modelo que tienen asignado, por ejemplo, los algoritmos de todas las familias excepto los algoritmos de la familia “lazy” tienen un tiempo para construir el modelo antes de probarlo, esto es así, ya que, dichos algoritmos de la familia “lazy” retardan el aprendizaje a la prueba. En esta parte del apartado vamos a centrarnos precisamente en la construcción del modelo 1 con los diferentes algoritmos. Una vez citados los algoritmos realizaremos el estudio, al respecto decir, que iremos seleccionando cada uno de los algoritmos desde la sección ‘Classifier’ de la ventana ‘Classify’.

En dicha sección ‘Classifier’ aparece otro botón de los denominados ‘Choose’, que tanto afloran en la aplicación “WEKA”, esta vez para seleccionar los distintos algoritmos de minería de datos de que dispone. Al pulsar sobre tal botón se observan los algoritmos clasificados por familias y por lo tanto, como nosotros también los hemos citado por familias será bastante fácil ir escogiendo cada vez, uno a uno, en orden los algoritmos enumerados con anterioridad:

-Cada vez que seleccionemos uno de los algoritmos del estudio en la sección ‘Classifier’, se mostrará el nombre de tal en el control alargado ‘label’ de tal sección. Si pulsamos tal control titulado con el nombre del algoritmo escogido se nos mostrará una

ventana emergente en donde aparecerán los parámetros por defecto con los que se ejecutará el algoritmo. Estos parámetros si se desea se pueden modificar, sin embargo, en este proyecto utilizaremos siempre los parámetros por defecto. Una vez tengamos un algoritmo seleccionado sólo faltará pulsar el botón ‘Start’ (empezar) que aparece debajo del ‘comboBox’ que indica la clase a predecir.

Recapitulando para realizar el estudio de algoritmos debemos disponer de lo siguiente en la aplicación “WEKA”:

- Los datos de entrada suministrados.
- El método validación cruzada 10 X 10.
- La clase a predecir atributo “compra”, con lo que el resto de atributos: “tipo”, “barrio”, “dorm”, “planta”, “metros”, “euros” y “ciudad” serán atributos de carácter explicativos.
- Seleccionar el primer algoritmo citado para construir el modelo. Pulsar el botón ‘Start’ para construir el modelo con el algoritmo. Ídem con el resto de algoritmos citados.

Una vez pulsado el botón ‘Start’ para construir el modelo con uno de los algoritmos del estudio, el nombre del algoritmo con el que se esté ejecutando el modelo se observará en el recuadro situado en la parte inferior izquierda denominado ‘Result list’ (lista de resultados.) De forma simultánea, en el gran recuadro ‘textarea’ que aparece a la derecha, perteneciente a la sección ‘Classifier output’ es en donde se mostrará la salida que “WEKA” nos vaya ofreciendo de la ejecución o construcción del modelo. Lo primero que “WEKA” mostrará será la información que se tiene como entrada para construir el modelo, esta información la muestra de forma inmediata y consta de: el nombre del algoritmo con su configuración, los datos de entrada del fichero cargado, la clase y el método de evaluación. Esta información es titulada por “WEKA” como “Run information” (información de ejecución.) Con esta información llamada “Run information”, “WEKA” construye el modelo.

La construcción del modelo es una tarea generalmente costosa, este coste depende primordialmente del número de datos de entrada que haya, del algoritmo que se utilice y de la tarea que haya que realizar. Cuando se empieza a elaborar el modelo el símbolo del pájaro “WEKA” que aparece en la parte inferior derecha de la ventana ‘Classifier’ cambia de reposo a ponerse en movimiento girando de izquierda a derecha indicando que se está procesando el modelo. Cuando aparece la información llamada ‘Classifier Model’ el modelo habrá sido construido y se mostrará la salida de la construcción. Con la construcción del modelo todavía no se ha acabado el proceso de ejecución, ya que, falta probar el modelo. Inmediatamente después de la obtención del modelo se empieza a probar el modelo, decir que, el pájaro de “WEKA” no habrá parado de dar cambios de posición de izquierda a derecha entre la transición entre la construcción del modelo y la prueba del modelo.

La prueba del modelo también es un proceso que suele tardar bastante tiempo, dependiendo también de la información de entrada, en este caso referente al número de instancias de viviendas a probar, al método de evaluación elegido. Como la prueba del modelo corresponde a la parte de la evaluación del modelo lo explicaremos en el apartado, “4 evaluación”. Sólo diremos que cuando concluye la evaluación del modelo, finaliza la ejecución del modelo. El pájaro representativo de “WEKA” simboliza tal hecho al sentarse y dejar de moverse.

Una cuestión importante es que los algoritmos de la familia “lazy”, como retardan la construcción del modelo transitarán de forma inmediata a la prueba del modelo, saltando la construcción previa del modelo, esto lo refleja “WEKA” informando de que la construcción previa del modelo tarda “0” segundos de tiempo.

Una vez la ejecución del modelo concluye, en la correspondiente instancia que aparece en la sección “Result-list” referida al modelo ejecutado, podremos ver pulsando el botón derecho del ratón, las opciones que nos aporta el modelo construido, como por ejemplo, los resultados de la prueba. También una vez terminada la ejecución del modelo, podremos seleccionar otro algoritmo del estudio, para realizar otra ejecución diferente del modelo, con lo que ahora en la sección ‘Output Classifier’ se visualizará otra construcción del modelo. Las instancias que creadas se mantienen en la sección ‘Result-list’ hasta que se eliminen o se cierre la herramienta ‘Explorer’ de “WEKA”. Por lo que, siempre se puede volver a observar la construcción de un modelo al pulsar sobre cualquier instancia formada por el nombre del algoritmo y la hora de ejecución. Otro detalle importante es que si se quisiera podría ejecutar más de una vez el modelo con el mismo algoritmo, ya que la instancia sería distinta ya que la hora de ejecución sería una posterior. No obstante, lógicamente, los resultados de la construcción del modelo y de la evaluación serían los mismos, ya que, se habría utilizado exactamente la misma información de entrada. Lo que sí podría dar resultados distintos sería si un mismo algoritmo se ejecutará con distintos parámetros de configuración, situación que como hemos afirmado no se realiza en este proyecto. En el estudio de algoritmos almacenaremos en ficheros de texto separados la información de salida de las diversas ejecuciones del modelo para poder analizarla siempre que queramos.

En el estudio de los algoritmos del modelo 1 empezaremos por la ejecución del modelo con el primer algoritmo de la lista que citábamos anteriormente “BayesNet” y seguiremos en orden por la ejecución del modelo cada vez con uno del resto de algoritmos de la lista hasta llegar a la ejecución del modelo con el último algoritmo de la lista, en este caso el algoritmo “SimpleCLI”. Por lo que, cuando haya concluido la ejecución del modelo mediante el algoritmo, “SimpleCLI”, habrá terminado el estudio de algoritmos.

Seguidamente vamos a mostrar algunas imágenes del estudio del modelo 1, aunque no vamos a mostrar imágenes de la parte de la evaluación, ya que dichas imágenes

referentes a la evaluación de los algoritmos del estudio las mostraremos en el apartado siguiente “4.Evaluación”.

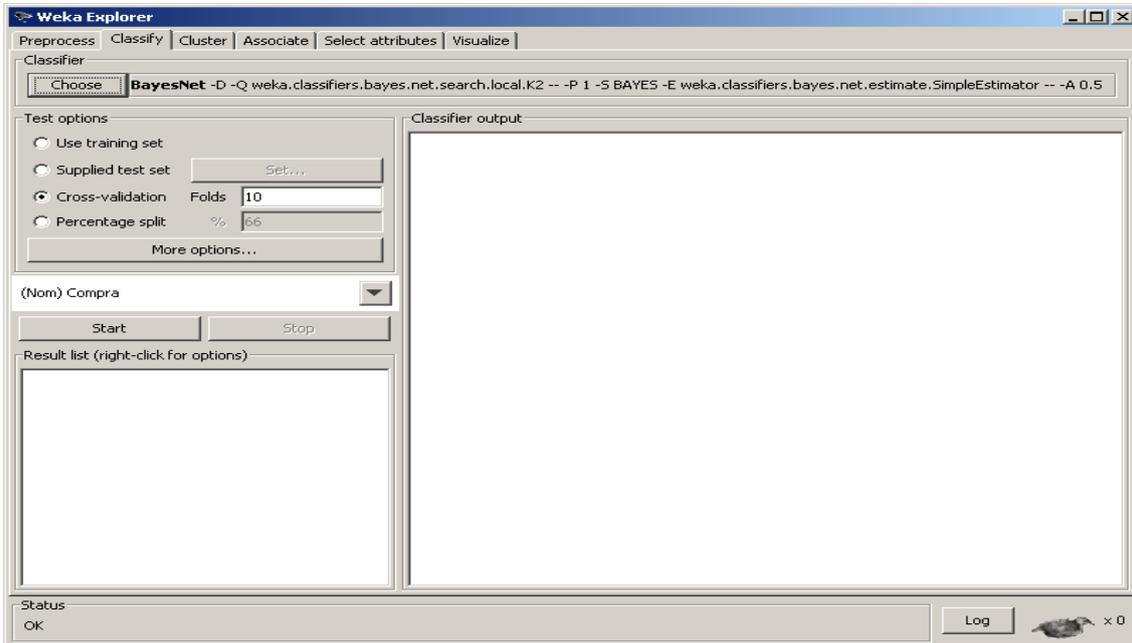


Figura 3.3: En la sección ‘Classifier’ se ha seleccionado el algoritmo “BayesNet”, que es el algoritmo primero del estudio de algoritmos del modelo 1.

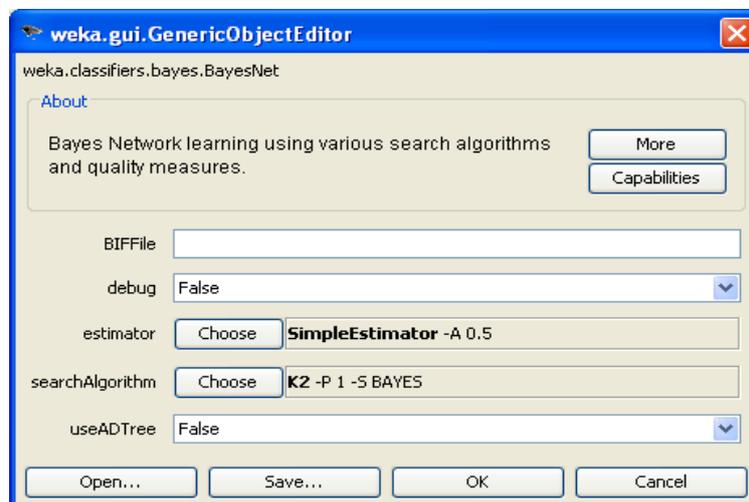


Figura 3.4: Cuadro de dialogo correspondiente a los parámetros de configuración por defecto que se utilizarán en el algoritmo “BayesNet”. A los parámetros de configuración de un algoritmo se accede pulsando sobre el ‘label’ en el que aparece el nombre del algoritmo junto a la configuración de que dispone.

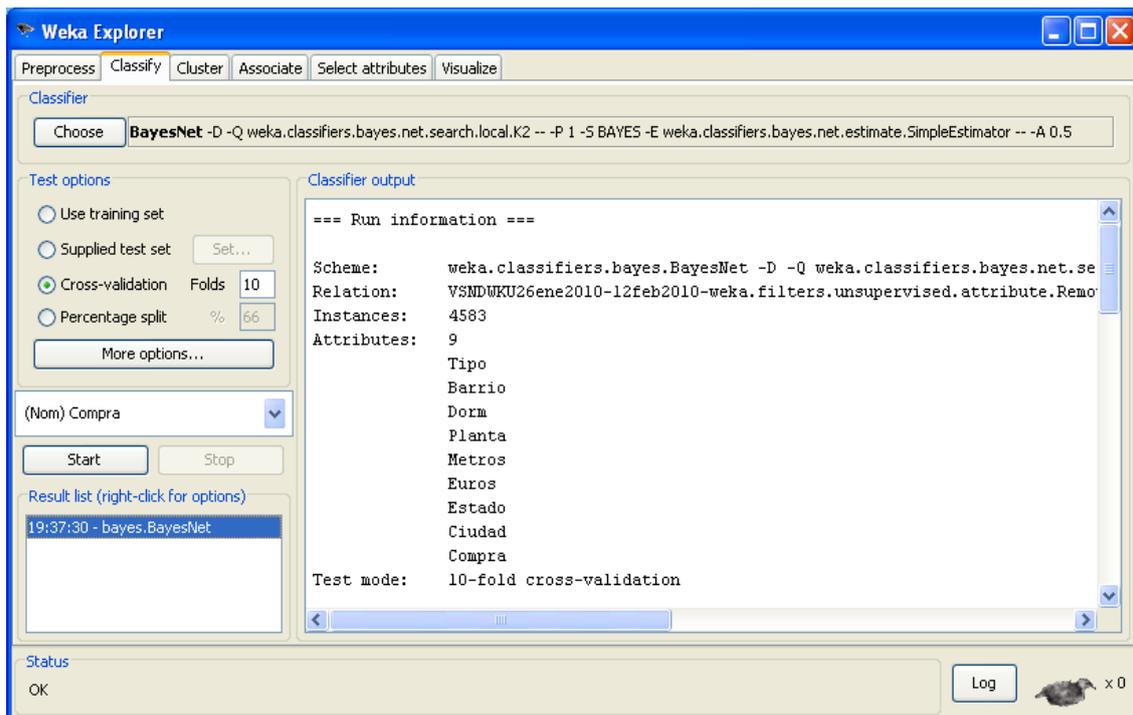


Figura 3.5: Se observa la sección 'Run información' después de haber ejecutado el modelo con el algoritmo "BayesNet".

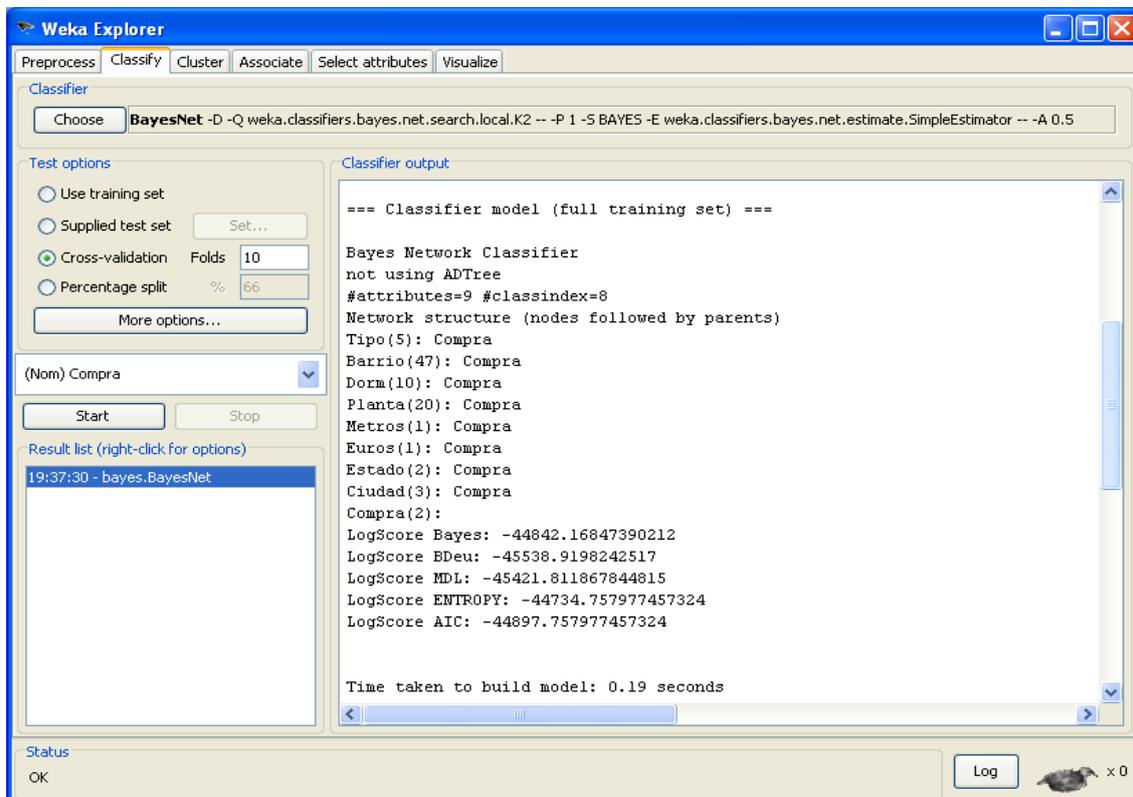


Figura 3.6: Muestra la salida, 'Classifier model', que nos ofrece "WEKA" de la construcción del modelo con el algoritmo "BayesNet".

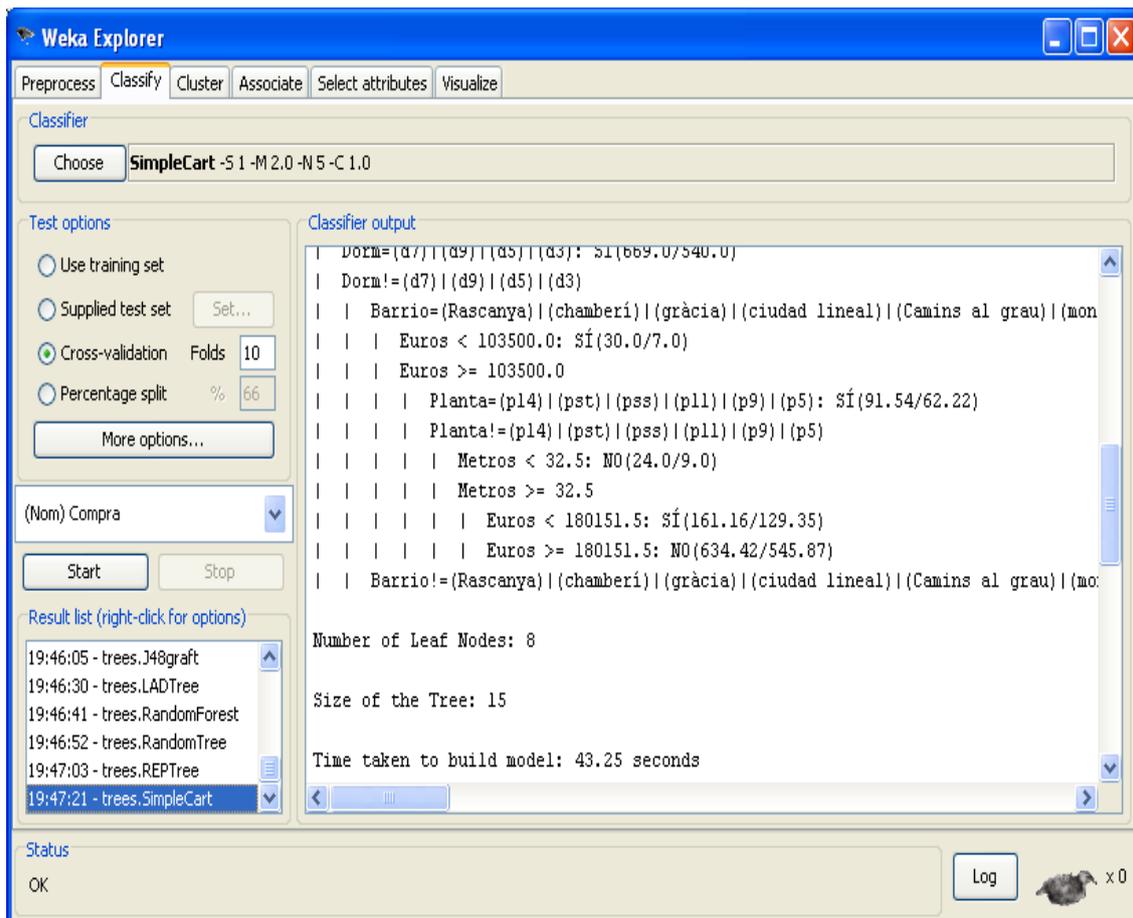


Figura 3.7: Con la ejecución del algoritmo “SimpleCart” llegamos a la última construcción del modelo 1.

Por otro lado, las imágenes relativas a la evaluación de los algoritmos del estudio, las mostraremos en el apartado “4. Evaluación”, no obstante, solamente diremos respecto a la evaluación que la construcción elegida del modelo 1 es la creada con el algoritmo “RotationForest”. Esto lo decimos porque, como hemos dicho con anterioridad en este apartado, queremos saber la construcción elegida en el estudio de algoritmos del modelo 1 para examinar los patrones de tal construcción seleccionada y así poder extraer conclusiones sobre la venta o la no venta de las viviendas.

Como la construcción realizada mediante el algoritmo “RotationForest” es la seleccionada se incluyen las figuras 3.8, 3.9, 3.10 y 3.11 que muestran aspectos de la ejecución del modelo con tal algoritmo.

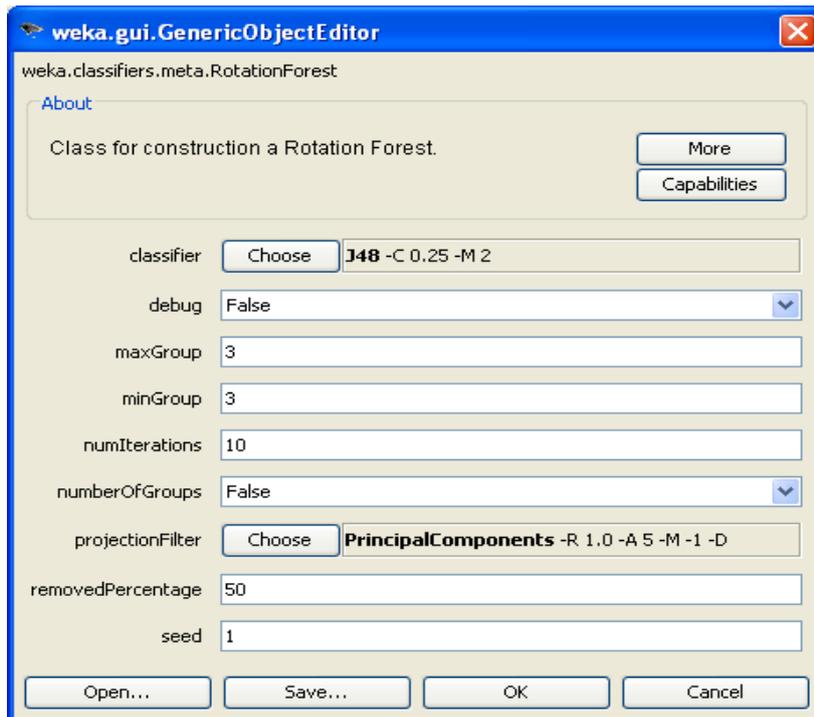


Figura 3.8: Se observan los parámetros de configuración que utilizamos en el algoritmo “RotationForest”. Además dichos parámetros de configuración son los que tiene por defecto el algoritmo.

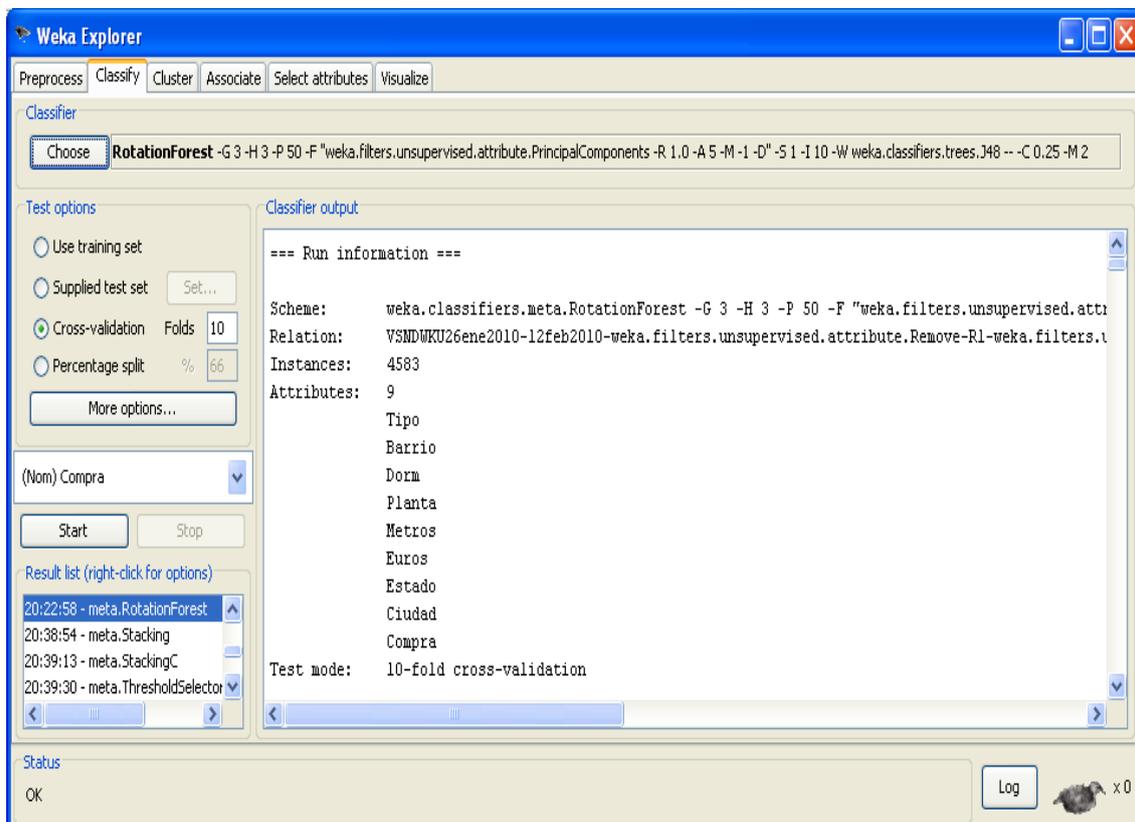


Figura 3.9: Se muestra la información de ejecución del modelo con el algoritmo “RotationForest”.

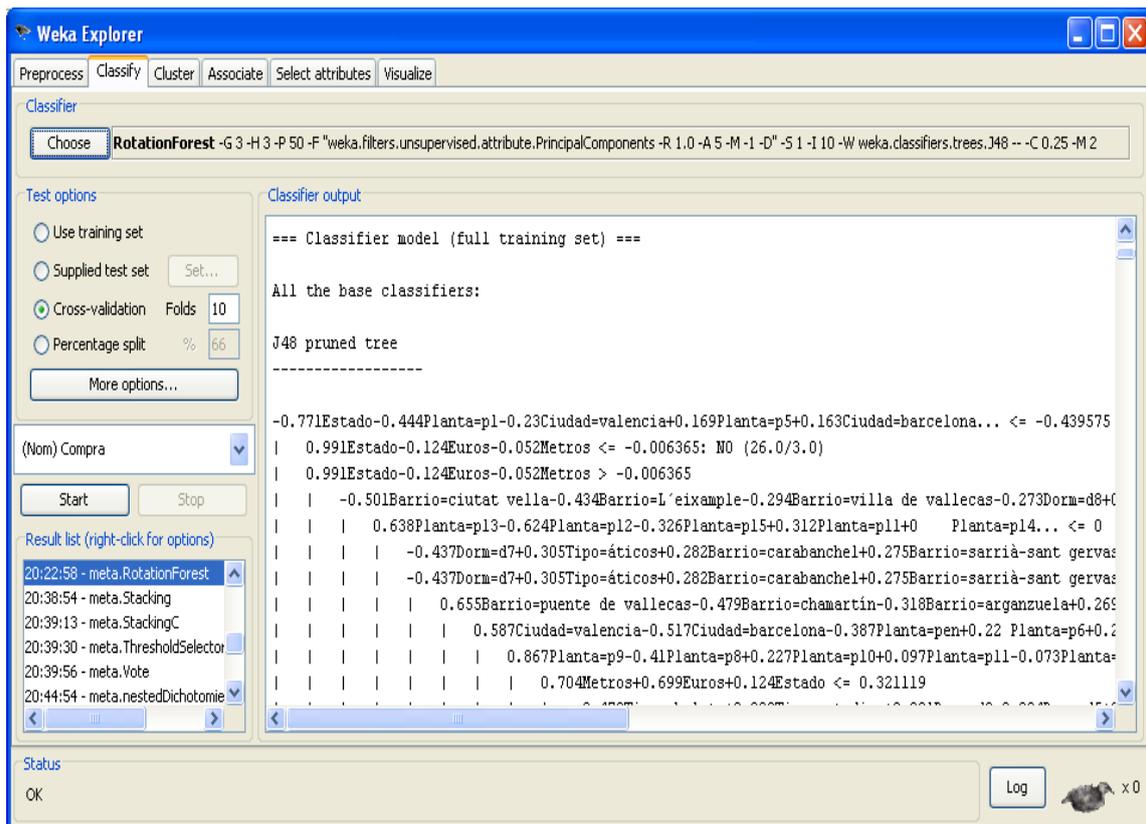


Figura 3.10: Se observa el principio de la salida que nos ofrece “WEKA” de la construcción del modelo con el algoritmo “RotationForest”.

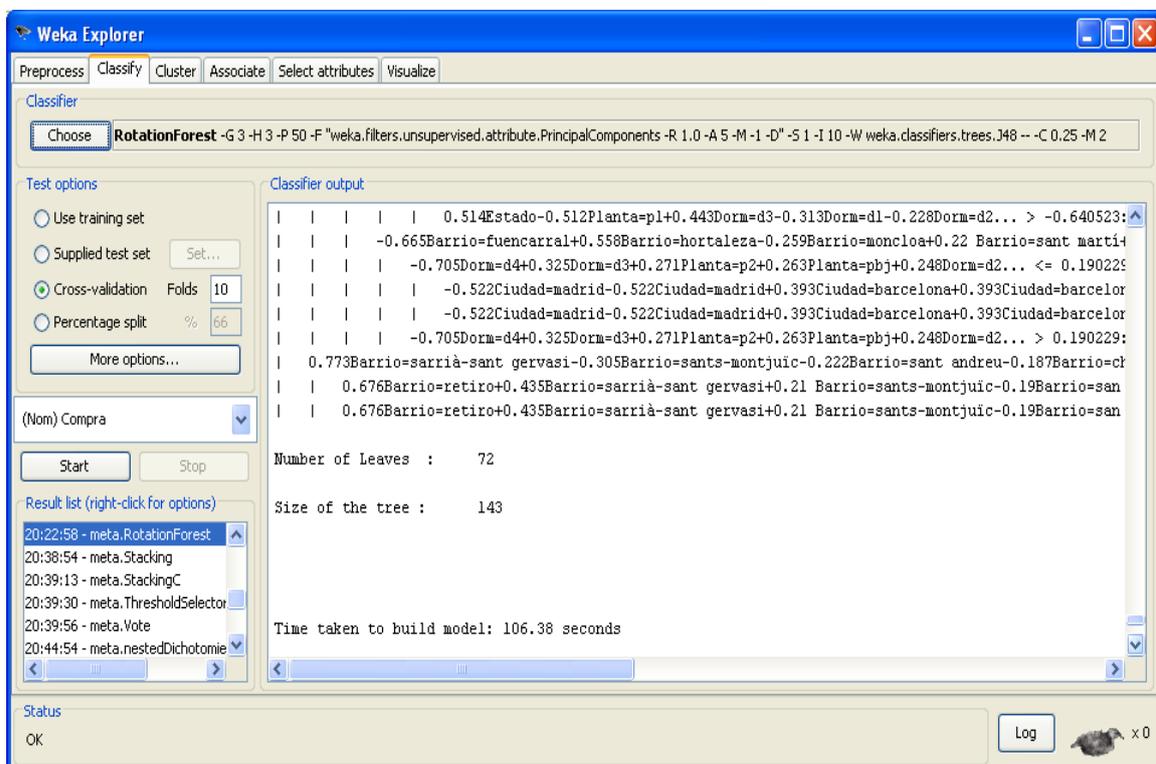


Figura 3.11: Se observa el final de la salida que nos ofrece “WEKA” de la construcción del modelo con el algoritmo “RotationForest”.

En las figuras 3.10 y 3.11 que se muestra parte de la salida de la construcción escogida se observan trozos de árboles de clasificación. Particularmente la figura 3.10 muestra partes del primer árbol y la figura 3.11 el final del último árbol. La figura 3.11, además, muestra información general de dicho último árbol relativa al número de nodos hojas que presenta '72' y al tamaño '143'. Estos conceptos referentes a los nodos de los árboles los explicaremos un poco más adelante, cuando contemos algunos conceptos sobre los árboles de clasificación. Aunque en las figuras 3.10 y 3.11 sólo se observa un trozo del modelo hay que resaltar que el modelo completo que ofrece "WEKA" consta de un total de '10' árboles de clasificación.

Una cuestión que se puede preguntar alguien es: ¿por qué si el algoritmo "RotationForest" es de la familia "metaLearned" la salida que muestra "WEKA" son árboles de clasificación?

Y la respuesta se puede observar en la figura 3.8 que muestra los parámetros de configuración del algoritmo seleccionado para construir el modelo "RotationForest" (que en este proyecto siempre son los de por defecto.) Porque en los parámetros de configuración de dicho algoritmo aparece un parámetro denominado, "classifier", (clasificador) al que se le debe indicar un algoritmo para implementar precisamente el algoritmo "RotationForest" y como se aprecia en tal figura es el algoritmo "J48". Por consiguiente, como el algoritmo "RotationForest" se implementa con el "J48" que es de la familia "trees" (árboles), la salida muestra árboles construidos con "J48".

Seguidamente, vamos a analizar la salida del modelo 1 realiza con el algoritmo seleccionado "RotationForest", pero antes de nada, vamos a mostrar la forma algorítmica que persiguen los modelos de clasificación, ya que, el modelo 1 como bien se sabe es de clasificación. Y también vamos a decir unos conceptos sobre las estructuras de datos en forma de árboles clasificación, porque el modelo seleccionado se presenta en forma de árboles de clasificación.

El recuadro 3.1 siguiente muestra la forma algorítmica general de los modelos de clasificación.

```

Si variableExplicativa1 operador Valor1 [Y ...[ Y variableExplicativaN operador Valor N ] entonces
  [Si ....
    .
    .
    asignadaClasificación = valorClasificación1
  ]
FinSi
[ Sino Si variableExplicativa1 ...
  [ Si...
    .
    .
  ]
  asignadaClasificación = valorClasificaciónM
  [ Sino Si ...
  [ Si...
    .
    .
  ]
  asignadaClasificación = valorClasificaciónN
]
]]

```

Recuadro 3.1: Forma algorítmica que persiguen los modelos de clasificación.

En nuestro caso como se ha explicado el algoritmo “RotationForest” muestra por la salida que ofrece “WEKA” el modelo en forma de árboles de clasificación o también conocidos como árboles de decisión. La idea, para cada uno de dichos árboles es la misma que la que se presenta en el recuadro 3.1, clasificar la clase de salida (en el recuadro 3.1, “asignadaClasificación”) a partir de los valores de los atributos explicativos (en el recuadro 3.1, “variableExplicativa1”,...,“variable ExplicativaN” .) En el caso del modelo 1 “asignadaClasificación” sería el atributo “compra” y sólo habría dos valores “valorClasificación1” y “valorClasificación2” que serían ‘SÍ’ y ‘NO’. Por último, como también se sabe, los atributos explicativos en el modelo 1, es decir, los que se utilizan para predecir la clase, atributo “compra”, serían el resto de atributos “tipo”, “barrio”, “dorm”, “planta”, “euros”, “estado”, “ciudad”.

Ahora, como decíamos, una vez expuesta la forma algorítmica general de los modelos de clasificación vamos a comentar los conceptos más destacados sobre las estructuras de datos en forma de árboles de clasificación:

Las estructuras de datos en forma de árboles de clasificación están formadas por nodos enlazados mediante condiciones. Se pueden distinguir tres clases distintas de nodos:

- Una de estas clases está formada por un sólo nodo que es la entrada del árbol. Dicho nodo se llama raíz.
- Otra clase de nodos esta formada por los nodos llamados hojas que se presentan al final del árbol y concluyen el valor de la clasificación.
- La clase de nodos que falta se llaman nodos internos y aparecen entre las otras dos clases de nodos que hemos comentado, es decir, entre el nodo raíz y los nodos hoja.

También se puede hacer mención a cada nodo como nodo padre o como nodo hijo. Un nodo se puede referenciar como padre de otro nodo, si presenta un enlace a otro nodo de nivel inferior. Para esto el nodo padre y el nodo hijo deben distar en un nivel. Por las características de los nodos padre y los hijos, se deduce que el nodo raíz no tiene padre y los nodos hoja no tienen hijos.

También podríamos decir, que los árboles están constituidos por ramas en un nivel de abstracción superior. Las ramas como son partes del árbol están formadas, por tanto, por nodos. Todas las ramas de un árbol constituyen el árbol y cada una es un camino distinto del árbol desde el nodo raíz a un nodo hoja, por lo dicho, todas las ramas de un árbol surgen del nodo raíz del árbol y finalizan en un nodo hoja. El único nodo que no puede ser compartido entre las ramas de un árbol es el nodo hoja correspondiente a cada una de las ramas. Por lo que hay tantas ramas como nodos hoja.

Para seguir una rama se deben verificar las condiciones adyacentes a tal que hacen que se pase de un nodo superior de la misma a otro inferior de ella hasta alcanzar el nodo hoja de la rama. Si esto ocurre, entonces, se clasificará la instancia con el valor del nodo hoja.

El tamaño de un árbol se calcula con el número de nodos que posee el árbol.

Según el número de condiciones que puedan surgir de un nodo diremos que el árbol, es binario si pueden surgir hasta dos condiciones de cada nodo, ternario si pueden surgir hasta tres condiciones de cada nodo y n-ario si pueden surgir “n” condiciones de cada nodo. Al respecto del número de nodos que surgen de un árbol hay que resaltar que cualquier árbol no binario se puede transformar en uno equivalente binario.

Seguidamente, para explicar la construcción seleccionada realizada con el algoritmo “RotationForest” que está formada por los ‘10’ árboles de clasificación vamos a introducir dos recuadros, el recuadro 3.2 y el recuadro 3.3 (completando las figuras 3.11 y 3.12 que también mostraban parte del primer y último árbol de la construcción seleccionada.) Después de mostrar los recuadros 3.2 y 3.3 y de explicarlos introduciremos una tabla 3.5 que indicará el tamaño y el número de nodos hojas de cada uno de los ‘10’ árboles con lo que reflejaremos el número de nodos y el número de ramas de cada árbol.

```

J48 pruned tree
-----
-0.771Estado=0.444Planta=p1-0.23Ciudad=valencia+0.169Planta=p5+0.163Ciudad=barcelona... <= -0.439575
| 0.991Estado=0.124Euros=0.052Metros <= -0.006365: NO (26.0/3.0)
| 0.991Estado=0.124Euros=0.052Metros > -0.006365
.
.
| | | | | -0.656Ciudad=madrid+0.485Ciudad=barcelona+0.357Ciudad=valencia-0.273Planta=p1+0.186Estado... > -0.460829: Sí (3.0)
| | -0.501Barrio=ciutat vella-0.434Barrio=L'eixample-0.294Barrio=villa de vallecas-0.273Dorm=d8+0.216Barrio=sant andreu... > 0.152513
| | 0.45 Barrio=ciudad lineal-0.443Barrio=centro-0.3Barrio=hortaleza+0.287Barrio=pueblo de vallecas-0.276Barrio=latina... <= 0.058445: Sí (157.0/40.0)
| | 0.45 Barrio=ciudad lineal-0.443Barrio=centro-0.3Barrio=hortaleza+0.287Barrio=pueblo de vallecas-0.276Barrio=latina... > 0.058445: NO (9.0)
-0.771Estado=0.444Planta=p1-0.23Ciudad=valencia+0.169Planta=p5+0.163Ciudad=barcelona... > -0.439575
| -0.514Dorm=d6+0.44 Barrio=sarrià-sant gervasi+0.348Dorm=d5-0.259Barrio=salamanca-0.259Barrio=fuencarral... <= 0.0523
| -0.438Barrio=L'olivereta-0.424Barrio=El pla del real+0.348Barrio=Extramurs-0.261Barrio=Algirós+0.255Barrio=les corts... <= -0.155687
| | -0.645Barrio=sant martí+0.307Barrio=Camins al grau+0.301Barrio=Extramurs+0.262Barrio=moratalaz+0.197Barrio=sants-montjuïc... <= 0.143901: Sí (6.0)
| | -0.645Barrio=sant martí+0.307Barrio=Camins al grau+0.301Barrio=Extramurs+0.262Barrio=moratalaz+0.197Barrio=sants-montjuïc... > 0.143901: NO (5.0/1.0)
.
.
| | | | | -0.699Planta=p3+0.638Planta=p4-0.219Planta=p2+0.134Planta=pen+0.108Planta=p1... > 0.109594: Sí (2.0)
| | -0.437Dorm=d7+0.305Tipo=áticos+0.282Barrio=carabanchel+0.275Barrio=sarrià-sant gervasi-0.25Dorm=d8... > 0.115324: NO (6.0)
| | 0.655Barrio=pueblo de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... > 0.037055: Sí (12.0)
Number of Leaves : 46

Size of the tree : 91

```

Recuadro 3.2: Se observa el principio y el final y el entremedio del primer árbol del modelo 1 construido con el algoritmo "RotationForest".

```

J48 pruned tree
-----
-0.757Estado=0.341Planta=p1-0.307Dorm=d1-0.222Dorm=d0+0.166Dorm=d5... <= -0.436024
| 0.514Estado=0.512Planta=p1+0.443Dorm=d3-0.313Dorm=d1-0.228Dorm=d2... <= -0.640523
| 0.838Barrio=ciudad lineal-0.368Barrio=chamartín-0.241Barrio=centro+0.199Barrio=salamanca-0.16Barrio=carabanchel... <= 0.202706
.
.
| | | | | -0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.045422: NO (13.0)
| | | | | -0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.044922: Sí (281.0/134.0)
| | | | | 0.703Barrio=Patraix-0.689Barrio=Benimaclet+0.176Barrio=El pla del real-0.008Barrio=L'olivereta-0.008Barrio=campanar - beni..... > -0.001383: Sí (874.0/353.0)
| | | | | -0.553Dorm=d7+0.429Planta=pss-0.338Dorm=d0+0.284Planta=p7+0.268Planta=p8... > -0.043858: Sí (2180.0/1069.0)
-0.757Estado=0.341Planta=p1-0.307Dorm=d1-0.222Dorm=d0+0.166Dorm=d5... > -0.436024
| 0.773Barrio=sarrià-sant gervasi-0.305Barrio=sants-montjuïc-0.222Barrio=sant andreu-0.187Barrio=chamberí-0.165Barrio=retiro... <= 0.143972
| 0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... <= -0.16556: Sí (12.0/1.0)
| 0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... > -0.16556
| | -0.665Barrio=fuencarral+0.558Barrio=hortaleza-0.259Barrio=moncloa+0.22 Barrio=sant martí+0.208Barrio=latina... <= 0.565388
.
.
| 0.773Barrio=sarrià-sant gervasi-0.305Barrio=sants-montjuïc-0.222Barrio=sant andreu-0.187Barrio=chamberí-0.165Barrio=retiro... > 0.143972
| 0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... <= -0.104739: NO (2.0)
| 0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... > -0.104739: Sí (15.0)

Number of Leaves : 72

Size of the tree : 143

```

Recuadro 3.3: Se observa el principio y el final y el entremedio del último árbol del modelo 1 construido con el algoritmo "RotationForest".

“WEKA” nos ofrece de la salida de cada árbol de clasificación del modelo seleccionado: las condiciones y los nodos que lo forman. Como se puede apreciar, por ejemplo, en parte de la salida del primer árbol en el recuadro 3.2 o en parte de la salida del último árbol en el recuadro 3.3. Los nodos que no son nodo hoja aparecen en la parte izquierda de cada condición, dicha parte izquierda es la parte que está a la izquierda del operador de la condición. Por eso, las condiciones que surgen de un mismo nodo están formadas por la misma parte izquierda. Los nodos hojas por su parte aparecen, al lado de las condiciones que conducen a ellos, de forma independiente etiquetados como ‘SÍ’ o ‘NO’.

Hay que resaltar también, que en la salida del modelo que nos ofrece “WEKA” se observan unas líneas verticales delante de las condiciones, a excepción de las condiciones que surgen del nodo raíz que no tienen dichas líneas. Cuantas más líneas verticales tiene delante una condición, está en un nivel más inferior el nodo del que surge dicha condición. Las condiciones que tienen el mismo número de líneas son del mismo nivel. Una cuestión muy importante de dichas líneas discontinuas es que si se sigue la línea más externa que empieza debajo de una condición nos conduce a la condición contraria, siempre y cuando, la contraria esté descrita más hacia abajo. Por contra, si la contraria aparece antes en la salida descrita, entonces dicha línea más externa nos conduce al final de la descripción del árbol. De esta frase última se deduce que dichos árboles son binarios. Esto se observa claramente en las dos condiciones resaltadas en negrita en el árbol primero (recuadro 3.2) y en el árbol último (recuadro 3.3.) Dichas dos condiciones resaltadas del árbol primero salen del nodo raíz correspondiente del árbol, ídem con las del árbol del recuadro 3.3. Por lo que, si la primera condición del nodo raíz no se cumple en dichos árboles, entonces se cumple la segunda condición que surge del nodo raíz y viceversa. Se puede observar que dichas condiciones resaltadas surgen del nodo raíz por la jerarquía que poseen, ya que ninguna de estas condiciones aparece precedida por una línea vertical.

En los árboles binarios se suele identificar a las dos condiciones que surgen de un nodo padre como condición izquierda y derecha, la izquierda suele ser la primera y la derecha la segunda condición. Por eso “WEKA” muestra la salida del árbol de izquierda a derecha. La condición de la izquierda de un nodo padre aparece con el operador “menor o igual que” y la de la derecha con el operador contrario “mayor que”.

Después de esta explicación, informamos del número de nodos hoja y por tanto, del número de ramas de cada uno de los ‘10’ árboles después de haber introducido parte del primer y último árbol en los recuadros 3.2 y 3.3.

Árbol del modelo	Número de nodos hoja (Number of Leaves)	Tamaño del árbol (Size of the tree)
1	46	91
2	40	79
3	199	397
4	68	135
5	54	107
6	34	67
7	60	119
8	78	155
9	59	117
10	72	143

Tabla 3.2: Muestra el número de nodos hojas “Number of Leaves” y el tamaño del árbol “Size of tree” de cada uno de los ‘10’ árboles que forman el modelo.

A continuación en forma de explicación de los patrones de las viviendas que se pueden encontrar en los ‘10’ árboles que forman el modelo seleccionado: vamos a analizar ‘4’ ramas del primer de tal modelo seleccionado. Después representaremos una pequeña porción de dicho árbol, concretamente, hasta el nivel de profundidad tercero. En dicha representación precisamente sólo una de las ramas será expuesta por completo por ser de nivel tercero. Una vez quede analizada cualquiera de las ‘4’ ramas podremos entender cualquier rama de la salida del modelo.

Para analizar nuestra primera rama que llamaremos “rama 1ª” vamos a mostrar las condiciones que se deben verificar hasta alcanzar el nodo hoja de tal rama. Dicho nodo hoja será clasificado con el valor ‘SÍ’, vivienda vendida o con el valor ‘NO’, vivienda no vendida.

Seguidamente mostramos el contenido de la “rama 1ª”.

(Nodo 1º o raíz) (Condición 1ª) -0.771Estado-0.444Planta=p1-0.23Ciudad=valencia+0.169Planta=p5+0.163Ciudad=barcelona... <= -0.439575
(Nodo 2º) (Condición 3ª) 0.991Estado-0.124Euros-0.052Metros <= -0.006365
(Nodo 4º hoja) : NO (26.0/3.0)

Recuadro 3.4: Rama 1ª del árbol primero.

La “rama 1ª” que estamos analizando clarifica todo el modelo, ya que el conjunto completo de este modelo está formado por árboles y cada árbol está formado por ramas como por ejemplo ésta. La “rama 1ª” perteneciente al árbol primero está formada por tres nodos y por dos condiciones. La “condición 1ª” surge del nodo raíz del árbol, que es común para todas las ramas del árbol. Si esta “condición 1ª” se verifica se transita al “nodo 2º” de la “rama 1ª”.

Seguidamente para explicar que debe suceder para transitar del “nodo 1º” al “nodo 2º” de la “rama 1ª” vamos a detallar las partes que se observan de una condición con la intención de poder tratarlas:

Como ya dijimos con anterioridad la parte izquierda de una condición es el nodo del que surge la condición. Dicha parte izquierda (el nodo) está formada en cada nodo por una suma de parte de las variables explicativas del modelo. Dichas variables explicativas son el resultado del análisis realizado a los valores de los atributos explicativos que se suministraron al modelo. Por lo que también se pueden denominar en vez de variables explicativas, atributos explicativos.

Dichas variables explicativas se observa que están precedidas por un peso y que suelen estar igualadas a un valor, aunque también de forma menos habitual aparecen sin estar igualadas a ningún valor.

Cuando una variable explicativa aparece igualada a un valor viene a indicar que se sustituye por su peso adyacente, siempre que el valor de la misma variable explicativa en este caso perteneciente a la vivienda que se esté analizando posea el mismo valor. Por el contrario, si contiene un valor distinto, entonces, se sustituye por el valor '0'. En el caso en el que la variable explicativa no esté igualada a ningún valor, entonces siempre se sustituye por el peso que posee, siempre que el ejemplar que se esté analizando no tenga un valor vacío (perdido) correspondiente a dicha variable explicativa, porque, entonces, se sustituiría también por el valor '0'.

En cada nodo, las variables explicativas aparecen ordenadas de mayor a menor peso indiferentemente de que tengan un signo positivo o negativo y por tanto, de más a menos importancia o significado en la condición. Dichas variables explicativas pueden aparecer varias veces en un mismo nodo. Al final de la suma de las variables explicativas suelen aparecer unos puntos suspensivos, estos puntos suspensivos indican que el resto de la suma carece de importancia o de peso como para tenerse en cuenta.

La parte central de una condición esta compuesta por el operador de la condición.

La parte derecha de una condición está compuesta por el valor numérico que se comparará con la suma de las variables explicativas.

Apreciando la “condición 1^a” de la “rama 1^a” que surge del “nodo 1^o” (raíz) se observa que en la parte de las variables explicativas entra en escena en primer lugar el atributo explicativo “estado” una vez, después el atributo explicativo “planta” por partida doble intercalado por el atributo explicativo “ciudad”, y finalmente vuelve a aparecer el atributo “ciudad”, por lo que, éste último aparece dos veces también. Mirando de forma global dicha suma de las variables explicativas (parte izquierda) junto con el operador (parte central) y el término constituido por un valor (parte derecha) se aprecia lo que debe pasar para que se cumpla la condición. Al estudiar esta condición se podrá entender cualquier otra condición.

Como vamos a analizar la “condición 1^a” de la “rama 1^a” la volvemos a mostrar seguidamente:

“Condición 1^a” de la “rama 1^a” que surge del “nodo 1^o”.

`-0.771Estado-0.444Planta=p1-0.23Ciudad=valencia+0.169Planta=p5+0.163Ciudad=barcelona... <= -0.439575`

Lo primero que hay que decir es que como decíamos tal “condición 1^a” de la “rama 1^a” enlaza el “nodo 1^o” o raíz con el “nodo 2^o” de dicha rama. Vamos a relatar cuando se verifica la condición y también cuando no se verifica. Para lo cual en primer lugar indicamos como afecta a tal “condición 1^a” cada uno de los miembros de que dispone:

Al querer analizar los miembros nos encontramos con la parte izquierda o nodo y el primero que apreciamos es el miembro, “-0.771Estado”. Que indica que si la instancia de la vivienda ingresada tiene un valor, es decir no está vacío (perdido) se sustituye por el valor ‘-0,771’. Si por el contrario la instancia de la vivienda ingresada tiene perdido

el valor referente al atributo “estado” entonces “-0.771Estado” se sustituye por el valor ‘0’. Por otro lado, el miembro que aparece inmediatamente a la derecha que contiene al atributo explicativo “planta”, “-0.444Planta=p1” viene a decirnos que si el atributo “planta” tiene asignado el valor ‘p1’ (planta primera) se sustituirá dicho miembro por el valor ‘-0,444’. Sin embargo, si el atributo planta posee cualquier otro valor distinto de ‘p1’, entonces se sustituye el miembro “-0.444Planta=p1” por el valor ‘0’. Respecto al siguiente miembro “-0.23Ciudad=valencia” viene a indicar que se sustituye por el valor ‘-0,23’, siempre y cuando el atributo “ciudad” de la instancia de la vivienda ingresada tenga como ciudad el valor Valencia, en otro caso se sustituirá por el valor ‘0’. Referente al siguiente miembro “+0.169Planta=p5” se sustituirá por ‘0,169’ si la planta de la instancia de la vivienda ingresada es ‘p5’, en otro caso se sustituirá por el valor ‘0’. Finalmente, en cuanto al último miembro que aparecen en la suma “+0.163Ciudad=barcelona” se sustituirá por ‘+0,163’ si la instancia de la vivienda ingresada tiene como ciudad Barcelona, en otro caso se sustituirá por el valor ‘0’.

La “condición 1^o” se cumplirá si el resultado de la suma de las variables explicativas de la condición es menor o igual a “-0,439575”.

En cuanto a que valores debe tener el ejemplar ingresado para que se cumpla la “condición 1^a” de la “rama 1^a” vamos a poner una tabla 3.3 con las combinaciones de los valores que debe tener el atributo “estado”, “planta” y “ciudad

Estado	Planta	Ciudad	Condición 1 ^a : Se cumple
no perdido	(Indiferente)	(Indiferente)	Sí
(Indiferente)	p1	no Barcelona	Sí

Tabla 3.3: Muestra las combinaciones de los valores de los atributos “planta”, “estado” y “euros” que hacen que se cumpla la “condición 1^a” de la “rama 1^a”.

Ningunos otros valores hacen que se cumpla la condición. Por tanto, según la tabla 3.3, para que la condición se cumpla el atributo explicativo “estado” debe ser no perdido como se muestra en la fila primera o como se aprecia en la segunda fila el atributo “planta” debe tener el valor “p1” que viene a referenciar a la planta primera y la ciudad no ser de Barcelona.

Ahora, en la tabla 3.4 ponemos el valor que alcanzan las sumas de las posibles combinaciones de valores de la fila primera de la tabla 3.3.

-0.771Estado	-0.444Planta=p1	-0.23Ciudad=valencia	+0.169Planta=p5	+0.163Ciudad=barcelona	Suma
no perdido (se da)	'p1' (se da)	'valencia' (se da)	'p1' (no se da)	'valencia' (No se da)	-1,445
no perdido (se da)	'p1' (se da)	'barcelona' (no se da)	'p1' (no se da)	'barcelona' (Se da)	-1,052
no perdido (se da)	'p1' (se da)	no se da	'p1' no se da	no se da	-1,215
no perdido (se da)	'p5' (no se da)	'valencia' (se da)	'p5' (se da)	'valencia' (No se da)	-0,832
no perdido (se da)	'p5' (no se da)	'barcelona' (no se da)	'p5' (se da)	'barcelona' (Se da)	-0,439
no perdido (se da)	'p5' (no se da)	no se da	'p5' (se da)	no se da	-0,602
no perdido (se da)	no se da	'valencia' (se da)	no se da	'valencia' (No se da)	-1,001
no perdido (se da)	no se da	'barcelona' (no se da)	no se da	'barcelona' (Se da)	-0,608
no perdido (se da)	no se da	no se da	no se da	no se da	-0,771

Tabla 3.4: Se observan las posibles combinaciones y la suma de estas cuando el valor del atributo “estado” es no perdido. La suma de cada combinación hace que se cumpla la condición que se está estudiando.

Efectivamente las sumas de las filas de la tabla 3.4 hacen que se cumpla la condición. La única suma de las que se observan en la tabla 3.4 que puede llevar a confusión es la de la fila número ‘5’, que aparece resaltada en negrita, en la cual, la suma alcanza el valor ‘-0,439’. Esta suma aparentemente se podría considerar mayor que ‘-0,439575’ si se tienen en cuenta los tres últimos decimales del valor del término independiente de la parte derecha de la condición. Sin embargo, como en las variables explicativas “WEKA” no visualiza más de tres decimales: vamos a comparar la suma que resulte de dichas variables, teniendo en cuenta de la parte decimal del término independiente únicamente los tres primeros decimales que posee. Esto lo hacemos así porque en las variables explicativas se trunca el resto de valores decimales a partir del tercer decimal. Y por lo tanto, claramente la suma de la fila ‘5’ también hace que se cumpla la condición ya que, “-0,439 <= -0,439” y por lo tanto como “-0,439 = -0,439” la condición se cumple.

En cuanto a la comprobación de que la “condición 1ª” se cumple también con los valores de la fila segunda de la tabla 3.3, hay que resaltar: que si la planta es ‘p1’ el miembro “-0.444Planta=p1” se sustituye por ‘-0,444’ y si la ciudad no es Barcelona el miembro “+0.163Ciudad=barcelona”, se sustituye por el valor ‘0’ y entonces, la condición sería: “-0,444 <= -0,439” que se cumple debido a que “-0,444 < -0,439”.

Respecto a los valores que hacen que no se cumpla la “condición 1ª” porque la suma de las variables explicativas que se tienen en cuenta en dicha condición sería mayor que ‘-0,439’, las mostramos en la tabla siguiente 3.5.

Estado	planta	Ciudad	Condición 1ª: Se cumple
Perdido	no p1	(Indiferente)	NO
Perdido	(indiferente)	Barcelona	NO

Tabla 3.5: Cada fila muestra las combinaciones que hacen que no se cumpla la condición.

Con las combinaciones que hacen que se cumpla la condición y con las que hacen que no se cumpla queda explicada la “condición 1ª”. En estos momentos supondremos que la “condición 1ª” se verifica para así llegar al “nodo 2º” de nuestra rama. Si también se cumple la “condición 2ª” de nuestra rama (etiquetada como “condición 3ª”, en relación con el árbol) se transita al “nodo 3º” de la rama y último por ser el nodo hoja de la rama. Dicho nodo hoja está clasificado como ‘NO’. Por eso, si una instancia de vivienda cualquiera sigue esta rama será declarada como no vendida.

La expresión (26.0/3.0) que aparece de forma adyacente al nodo hoja indica que ‘26’ instancias de viviendas utilizadas en la construcción del modelo han llegado a este nodo hoja y ‘3’ de ellos han sido incorrectamente clasificados.

En cuanto a los patrones que se pueden observar en la “rama 1ª”, decir que: si el árbol que se tiene en cuenta es el primero del modelo (que es donde está ubicada la “rama 1ª”) se concluye que las viviendas cuyas instancias ingresaran para predecirles el atributo “compra”, que siguieran los valores de tabla 3.6, para los atributos “planta”, “ciudad” “estado”, “metros” y “euros” se predecirían como no vendidas, ya que alcanzarían el nodo hoja de la “rama 1ª”.

Planta	Ciudad	Estado	metros	Euros
p1	no barcelona	vacío	cualquier valor no vacío	(indiferente)
p1	no barcelona	vacío	(indiferente)	cualquier valor no vacío

Tabla 3.6: Patrones que deben darse para una vez estamos en el árbol primero que se cumplan todas las condiciones de la “rama 1ª” y se alcance el nodo hoja de dicha rama.

Con la tabla 3.6 concluimos la explicación de la “rama 1ª”. Seguidamente mostramos el contenido de la “rama 2ª” en el recuadro 3.5:

(Nodo 1º o raíz) (Condición 1ª) -0.771Estado-0.444Planta=p1-0.23Ciudad=valencia+0.169Planta=p5+0.163Ciudad=barcelona... <= -0.439575
(Nodo 2º) (Condición 4ª) 0.991Estado-0.124Euros-0.052Metros > -0.006365
(Nodo 5º) (Condición) -0.501Barrio=ciutat vella-0.434Barrio=L´eixample-0.294Barrio=villa de vallecas-0.273Dorm=d8+0.216Barrio=sant andreu... <= 0.152513
(Nodo) (Condición) 0.638Planta=p13-0.624Planta=p12-0.326Planta=p15+0.312Planta=p11+0 Planta=p14... <= 0
(Nodo) (Condición) -0.437Dorm=d7+0.305Tipo=áticos+0.282Barrio=carabanchel+0.275Barrio=sarrià-sant gervasi-0.25Dorm=d8... <= -0.211321
Nodo hoja: SÍ (787.0/304.0)

Recuadro 3.5: “Rama 2ª” del árbol primero.

La “rama 2ª” está también en el primer árbol del modelo. En la “rama 2ª” se observa nuevamente la “condición 1ª” de la “rama 1ª”, por lo que, el “nodo 2º” de ambas ramas debe ser el mismo. No obstante, como se observa la “condición 2ª” de la “rama 2ª” (“condición 4ª” respecto al árbol) es la contraria que la “condición 2ª” de la “rama 1ª”, ya que para continuar por el “nodo 3º” de la “rama 2ª” (“nodo 5º” del árbol) la condición que se debe cumplir es exactamente la contraria que la “condición 2ª” de la “rama 1ª”. Por lo que la “rama 2ª” a partir de su “condición 2ª” se diferencia de la “rama 1ª”. En cuanto a la explicación general de la “rama 2ª” hay que recalcar que la rama está formada por ‘6’ nodos y por ‘5’ condiciones. En este caso el nodo hoja de la rama está etiquetado con el valor ‘SÍ’. En cuanto a la construcción de la rama ‘787’ instancias de viviendas utilizadas en la construcción del modelo han sido las que han llegado a tal conclusión y ‘304’ lo han hecho erróneamente.

Respecto a los patrones que se observan de la “rama 2ª”: para que una instancia de vivienda alcance el nodo hoja de la rama debe poseer los valores que se observan en la tabla 3.7 para los atributos “planta”, “estado”, “euros” y “barrio”. Si esto ocurre, la vivienda ingresada al llegar al nodo hoja de la rama se predecirá como vendida.

Estado	Ciudad	Barrio	Planta	Dorm	Metros	Euros
cualquier valor no vacío	(indiferente)	no sant andreu	no p13 y no p11	d7 o d8	(indiferente)	(indiferente)
(indiferente)	no Barcelona	cualquiera no de Barcelona	no p13 y no p11	d7 o d8	no vacío	vacío

La tabla 3.7: Patrones que deben darse para una vez estamos en el árbol primero que se cumplan todas las condiciones de la “rama 2ª” y se alcance el nodo hoja de la “rama 2ª”.

Seguidamente mostramos el contenido de la “rama 3ª” en el recuadro 3.6.

(Nodo 1º) (Condición 2ª)
 -0.771Estado-0.444Planta=p1-0.23Ciudad=valencia+0.169Planta=p5+0.163Ciudad=barcelona... > -0.439575

(Nodo 3º) (Condición 5ª)
 -0.514Dorm=d6+0.44 Barrio=sarrià-sant gervasi+0.348Dorm=d5-0.259Barrio=salamanca-0.259Barrio=fuencarral... <= 0.0523

(Nodo 6º) (Condición)
 -0.438Barrio=L´olivereta-0.424Barrio=El pla del real+0.348Barrio=Extramurs-0.261Barrio=Algirós+0.255Barrio=les corts... <= -0.155687

(Nodo) (Condición)
 -0.645Barrio=sant martí+0.307Barrio=Camins al grau+0.301Barrio=Extramurs+0.262Barrio=mortalaz+ 0.197Barrio=sants-montjuïc... <= 0.143901

(Nodo hoja): Sí (6.0)

Recuadro 3.6: “Rama 3ª” del primer árbol.

Al igual que la “rama 1ª” y la “rama 2ª”, la “rama 3ª” reside en el primer árbol del modelo 1 seleccionado en el estudio. La “rama 3ª” está formada por ‘5’ nodos y ‘4’ condiciones. El nodo hoja de la rama esta etiquetado como “SÍ”. En cuanto a la construcción de la rama: ‘6’ viviendas de las utilizadas en la construcción del modelo han sido las que han llegado a tal conclusión. Una cuestión de la “rama 3ª” es que como se aprecia la “condición 1ª” de dicha rama (“condición 2ª” respecto al árbol) es la contraria a la “condición 1ª” de la “rama 1ª” y a la de la “rama 2ª”, por lo que, a partir de ahí se diferencia de las otras dos ramas estudiadas con anterioridad.

Respecto a los patrones de la “rama 3ª” se observa en la tabla 3.8 los valores respecto al atributo explicativo “estado”, “planta”, “dorm”, “ciudad” y “barrio” que debe tener una instancia de una vivienda ingresa en el árbol primero para alcanzar el nodo hoja de la “rama 3ª” y concluir que dicha vivienda ingresada es no vendida

Estado	Planta	Dorm	Ciudad	Barrio
vacío	no p1	no d5	valencia	L´olivereta o El pla del real o Algirós

Tabla 3.8: Patrones que deben darse para una vez estamos en el árbol primero que se cumplan todas las condiciones de la “rama 3ª” y se alcance el nodo hoja de dicha rama.

A continuación mostramos el contenido de la “rama 4^a”:

(Nodo 1 ^o) (Condición 2 ^a) -0.771Estado=0.444Planta=p1-0.23Ciudad=valencia+0.169Planta=p5+0.163Ciudad=barcelona... > -0.439575
(Nodo 3 ^o) (Condición 6 ^a) -0.514Dorm=d6+0.44 Barrio=sarrià-sant gervasi+0.348Dorm=d5-0.259Barrio=salamanca-0.259Barrio=fuencarral... > 0.0523
(Nodo 7 ^a) (Condición) 0.655Barrio=pueblo de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=Chamberí +0.208Barrio=eixample... <= 0.037055
(Nodo) (Condición) 0.587Ciudad=valencia-0.517Ciudad=barcelona-0.387Planta=pen+0.22 Planta=p6+0.205Planta=p7... <= -0.500504
(Nodo hoja): NO (7.0)

Recuadro 3.7: La “rama 4^a” del árbol primero del modelo 1.

La “rama 4^a” está en el árbol primero del modelo al igual que la “rama 1^a” que la “rama 2^a” y que la “rama 3^a”. La “rama 4^a” tiene “5” nodos y “4” condiciones. El nodo hoja está etiquetado como “NO”. En la construcción del modelo, ‘7’ instancias de las viviendas disponibles para construir el modelo han seguido la rama. En este caso está rama se diferencia de la “rama 3^a” estudiada con anterioridad a partir de la “condición 2^a” de la rama (“condición 6^a” del árbol.)

En referencia a los patrones de la “rama 4^a” se debe cumplir que el atributo “estado” sea perdido que el atributo “ciudad” sea Barcelona y que el barrio sea ‘sarrià-sant gervasi’ para alcanzar así el nodo hoja de la “rama 4^a” y concluir que la vivienda ingresa es no vendida.

Estado	Ciudad	Barrio
Perdido	Barcelona	sarrià-sant gervasi

Tabla 3.9: Patrones que deben darse para que una vez estemos en el primer árbol se cumplan todas las condiciones de la “rama 4^a” y se alcance el nodo hoja de tal rama.

Ahora, para seguir con la explicación del modelo vamos a representar los tres primeros niveles del primer árbol del modelo en el diagrama siguiente:

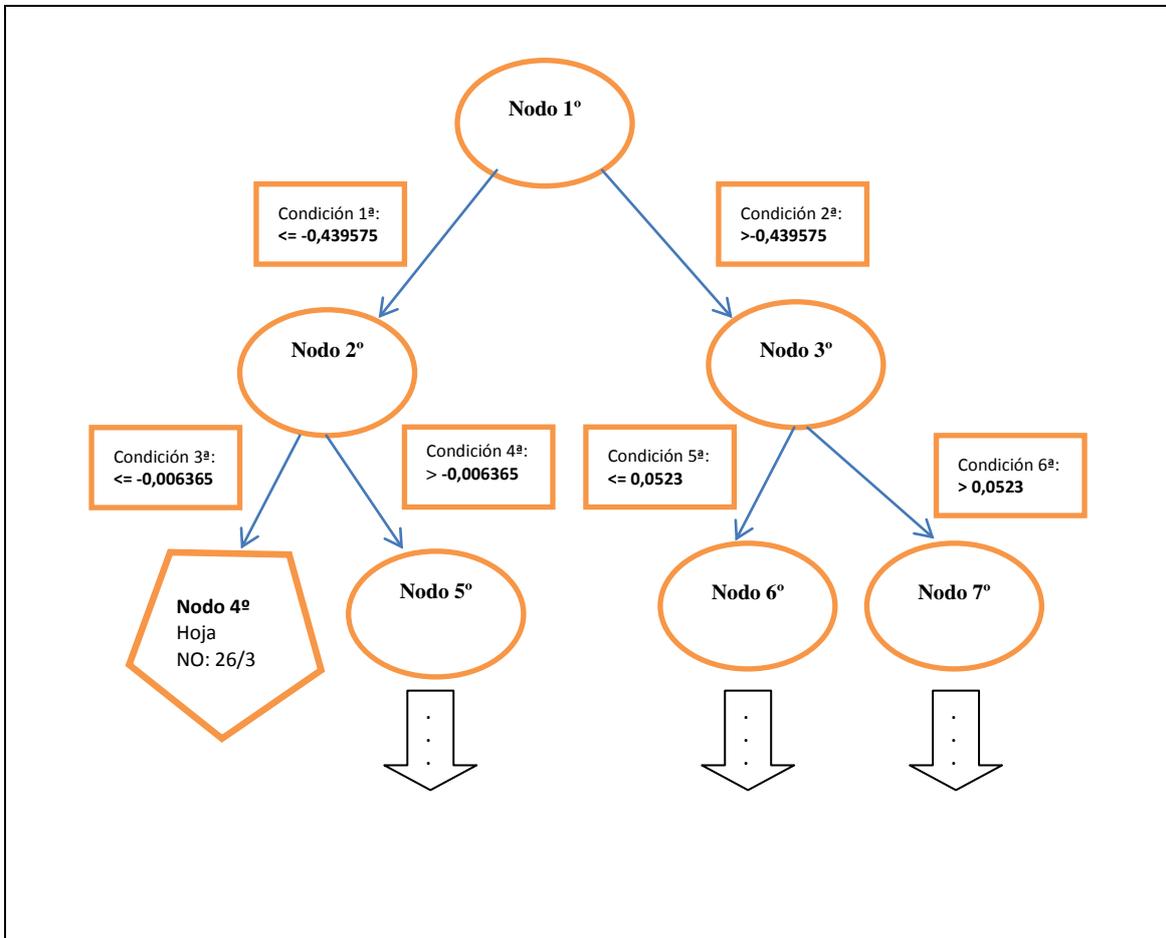


Diagrama 3.1: Diagrama del árbol primero, en el que se observa la rama 1ª íntegramente y parte de las otras ramas, entre ellas parte de la “rama 2ª”, de la “rama 3ª” y de la “rama 4ª”.

En el diagrama 3.1 se observa parte de todas las ramas del árbol primero del modelo, y por tanto, se debe apreciar parte de las ‘4’ ramas que explicamos del primer árbol. La “rama 1ª” que explicamos es la única que se observa íntegramente de las explicadas. La “rama 1ª” se observa completamente, porque, está constituida por tres niveles, es decir, su nodo hoja está en el nivel tercero y como la representación del diagrama 3.1 contempla hasta el nivel tercero, entonces, aparece íntegramente. Sin embargo, como las otras tres ramas, “rama 2ª”, “rama 3ª” y “rama 4ª” se alargan más allá del nivel tercero se muestra la porción de dichas ramas que llega hasta el nivel tercero. La “rama 1ª” es la que se observa más a la izquierda formada por los “nodos: 1º, 2º, 4º”. El “nodo 4º” es el nodo hoja de dicha “rama 1ª”. La parte de la “rama 2ª” que se observa está formada por los “nodos: 1º, 2º, 5º”. La parte de la “rama 3ª” que se aprecia está constituida por los

“nodos: 1º, 3º, 6º”. Y por último, la parte de la “rama 4ª” que se visualiza está formada por los “nodos: 1º, 3º, 7º”. Una observación que se aprecia es que cada una de las ‘4’ ramas va por un camino distinto de la porción del árbol que se observa en la diagrama 3.1. Es importante resaltar que el único nodo exclusivo de una rama que se observa en el diagrama 3.1 es el nodo hoja que se observa de la “rama 1ª”. Por lo tanto, cada una de las ‘46’ ramas de que dispone el árbol primero pasa por algunos de los nodos que se observan en el diagrama 3.1.

Seguidamente, mostramos una tabla que nos informe de la localización de las ‘4’ ramas estudiadas entre las ‘46’ ramas de dicho árbol primero.

Orden de aparición	Rama	Clasificada como	Viviendas del modelo que la construyen	Viviendas del modelo que la construyen erróneamente
1	La rama 1ª	NO	26	3
2	La rama 2ª	SÍ	787	304
3	Rama Siguiente	SÍ	3	1
4	Rama Siguiente	NO	2	0
...
22	La rama 3ª	SÍ	6	0
23	Rama Siguiente	SÍ	5	1
24	Rama Siguiente	NO	30	0
...
36	La rama 4ª	NO	7	0
37	Rama Siguiente	NO	2	0
38	Rama Siguiente	SÍ	35	4
...
44	Rama Siguiente	SÍ	2	0
45	Rama Siguiente	NO	6	0
46	Rama Siguiente	SÍ	12	0

Tabla 3.10: Se observan algunas de las “46” ramas del árbol primero del modelo. Entre ellas las ramas estudiadas: “rama 1ª”, “rama 2ª”, “rama 3ª” y “rama 4ª”.

Al igual que hemos visto con ‘4’ ramas del primer árbol de clasificación del modelo, decir que, cada una de las ramas de los ‘10’ árboles de clasificación (de forma similar a las ‘4’ ramas vistas) descubre unos patrones útiles para explicar la compra o no compra de las viviendas. Por otro lado, decir que, cualquier vivienda a clasificar será más acertadamente clasificada si tiene asignado en el atributo “ciudad” la ciudad de Valencia, Barcelona o Madrid, al igual que si tiene asignado en el atributo “barrio” uno de los barrios estudiados. Al igual, si el valor del atributo “tipo” de la vivienda es uno de los estudiados que son ‘pisos’, ‘ático’, ‘estudio’, ‘chalets’ o ‘dúplex’, tendrá una mejor clasificación. Y también si el atributo “estado” aparece vacío (perdido) o adopta los valores ‘a reformar’ o ‘obra nueva’ la clasificación será mejor.

El modelo con los ‘10’ árboles de clasificación se muestra completamente en el “apéndice 1: La construcción del modelo 1”.

En cuanto a las construcciones del modelo descartadas en el estudio están en la ruta del proyecto “NombreAlumno/anexos/modelos/modelo1/resto” en archivos de texto separados, uno para cada construcción del modelo. Además en “NombreAlumno/anexos/modelos/modelo1/seleccionado” aparece un archivo de texto con la construcción del modelo seleccionada con los ‘10’ árboles, al igual que en el apéndice 1 que hemos dicho que también aparece dicha construcción.

3.2. El aprendizaje del modelo 2

Ahora vamos a describir el aprendizaje del “modelo 2: Precio por el que vender una vivienda, independientemente de los días.” El modelo de datos pretende encontrar patrones en las viviendas para predecir el atributo “euros” sin tener en cuenta el atributo “días”. La finalidad será predecir dicho atributo “euros” a viviendas no analizadas en las cuales se desconozca el atributo “euros”. La tarea que se requiere para realizarlo es también una tarea de predicción, ya que, se quiere predecir el atributo “euros” a instancias no estudiadas a partir de los datos de las viviendas de que disponemos para este modelo. La tarea de predicción en este caso será una regresión, ya que, el atributo a predecir “euros” es un atributo de tipo numérico y por lo tanto de naturaleza continua. Entonces, por lo dicho, habrá que aprender una función de regresión. El atributo “euros” será la variable de salida o clase y el resto de atributos serán las variables de entrada o explicativas, que nos permitirán predecirlo.

En cuanto a los datos que se van a utilizar para construir este modelo son los datos de las viviendas del portal web que están almacenadas en el fichero “Modelo2Vista_Minable26ene2010-12feb2010.arff”, que tiene la vista minable sobre la que se va a construir el modelo. La tabla 3.11 muestra unos ejemplos de dicho fichero.

Vivienda	Tipo	Barrio	Dorm	Planta	Metros	Euros	Estado	Ciudad	Compra
1	pisos	extramurs	d5	p4	158	310000	a reformar	Valencia	SÍ
2	pisos	latina	d2	p1	60	130000	a reformar	Madrid	SÍ
3	pisos	ciudad lineal	d2	Pbj	58	220000	Obra nueva	Madrid	SÍ
4	pisos	Sant martí	d3	p2	87	310000		Barcelona	SÍ
5	áticos	Horta guinardó	d3	p4	70	164000	a reformar	Barcelona	SÍ

Tabla 3.11: Muestra de ‘5’ ejemplares del fichero “Modelo2_26ene2010-12feb2010.arff”, este fichero tiene un total de “2448” ejemplares.

El atributo “euros” es la clase y el resto de atributos que se observan en la tabla 3.11 “tipo”, “barrio”, “dorm”, “planta”, “metros”, “estado”, “ciudad” y “compra” son los atributos explicativos. El atributo “compra” en el fichero que estamos analizando adopta el valor ‘SÍ’ en todos los ejemplares. Por lo cual en la tabla 3.11 no se observa el valor

‘NO’ en el atributo “compra”, ni por lo mismo tampoco aparecerá en ninguna vivienda del fichero “Modelo2VistaMinable26ene2010-12feb2010.arff”.

Otra cuestión que se puede apreciar en las viviendas es que en el fichero que tiene la vista minable del modelo 2 aparecen ejemplares que también están en el fichero “Modelo1VistaMinable26ene2010-12feb2010.arff”, que está en la forma de la vista minable del modelo 1, y que analizamos cuando hablábamos del modelo 1. Esto explica que la tabla 3.11 que contiene “5” ejemplares del fichero “Modelo2VistaMinable26ene2010-12feb2010.arff” tenga la instancia de la vivienda de la fila primera idéntica a la instancia de la vivienda de la fila primera de la tabla 3.1, que contiene ‘5’ ejemplares del fichero “Modelo1VistaMinable26ene2010-12feb2010.arff”. Todo esto es debido a que los dos ficheros parten del mismo origen de datos, pero finalmente han tenido diferencias en la fase de preparación. Exactamente la única diferencia que tiene el fichero “Modelo2VistaMinable26ene2010-12feb2010.arff” es que se han eliminado los ejemplares cuyo atributo “compra”=‘NO’, mientras que en el fichero “Modelo1VistaMinable26ene2010-12feb2010.arff” no se han eliminado. Por lo que se puede apreciar que el fichero “Modelo1VistaMinable26ene2010-12feb2010.arff” es un conjunto que abarca los ejemplares del fichero “Modelo2VistaMinable26ene2010-12feb2010.arff”, ya que, contiene todas las viviendas del fichero “Modelo2VistaMinable26ene2010-12feb2010.arff” más las que tienen el atributo “compra” igual a ‘NO’, que han sido rechazados por el fichero “Modelo2VistaMinable_26ene2010-12feb2010.arff”.

Una vez conocida la tarea y los datos vamos a realizar al igual que hemos en el modelo 1, un estudio para el modelo 2 para construir el modelo con cada uno de los algoritmos que “WEKA” proporciona que puedan construirlo. En este caso, por lo tanto, al igual que hemos hecho en el modelo 1, se realizarán una serie de iteraciones constituidas por la fase de minería de datos y la fase de evaluación. Habrá tantas iteraciones conjuntas de la fase de minería de datos y de evaluación como construcciones se realicen del modelo 2. Como se realizará una construcción por algoritmo disponible, se realizarán tantas como algoritmos diferentes haya disponibles para construir el modelo. Al igual que en el modelo 1, una vez realizado el estudio, nos quedaremos con la construcción que mejor evaluación nos aporte. Lo primero que haremos será abrir en la aplicación ‘Explorer’ de la herramienta “WEKA” el fichero “Modelo2VistaMinable26ene2010-12feb2010.arff”. Una vez abierto tal fichero se apreciarán los atributos y valores del fichero cargado.

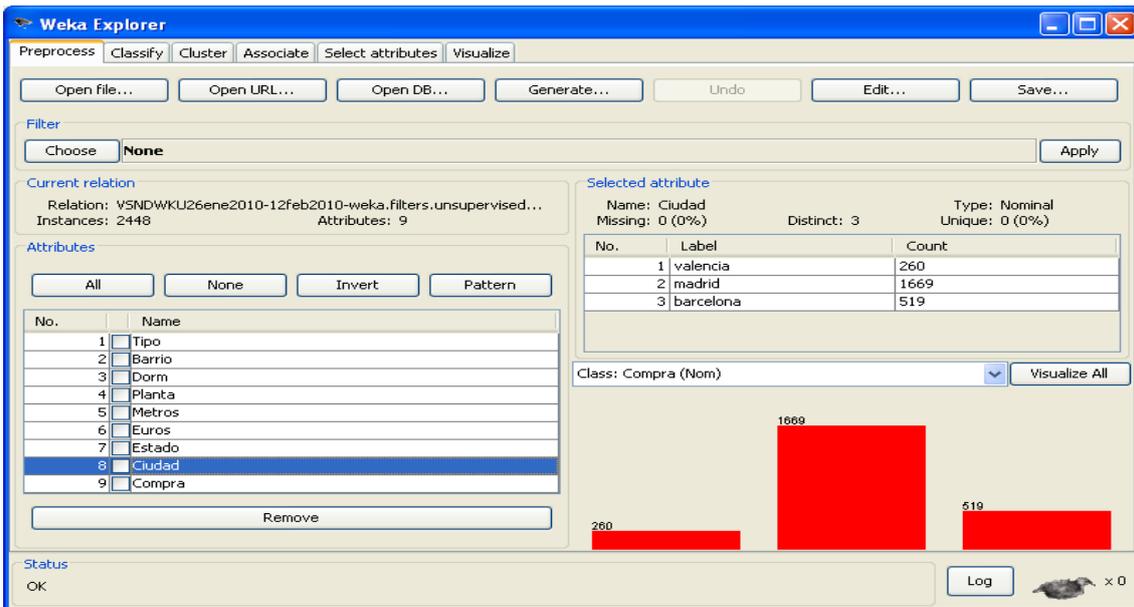


Figura 3.12: Se observa la pestaña 'Preprocess' de la aplicación 'Explorer' de "WEKA" con los datos del fichero "Modelo2VistaMinable26ene2010-12feb2010.arff".

Seguidamente, transitaremos a la ventana 'Classify' para realizar el estudio de algoritmos. En tal ventana seleccionaremos como clase a predecir el atributo "euros" y como método de validación dejaremos el que aparece por defecto, "Validación cruzada 10 X 10". Con lo que sólo nos faltará ir seleccionando los algoritmos disponibles en la sección 'Classifier' que puedan realizar la tarea de regresión del modelo 2 para así realizar el estudio. La figura 3.13 muestra la ventana 'Classify' una vez hemos puesto como clase el atributo "euros", ya que, es el atributo a predecir en modelo 2.

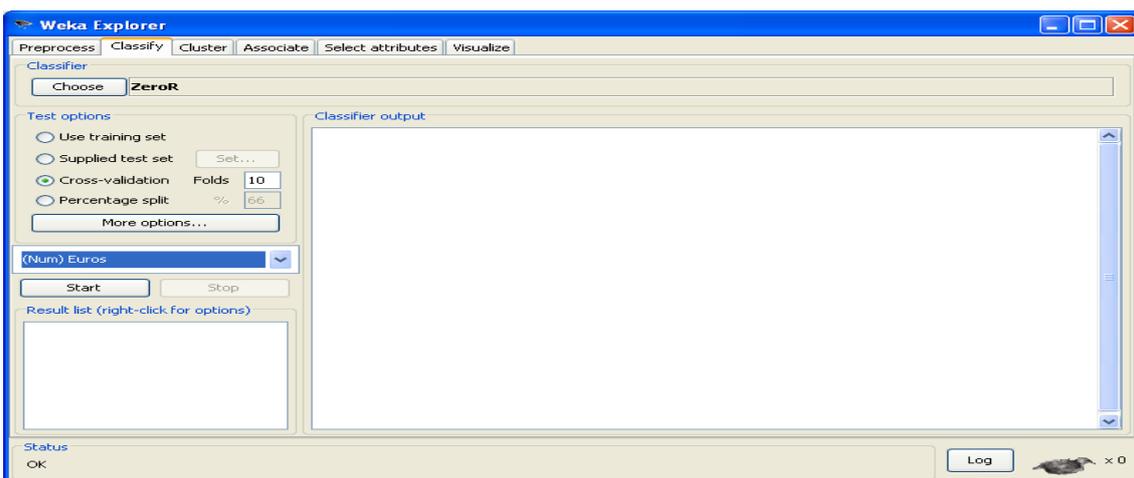


Figura 3.13: Se muestra la ventana 'Classify' después de seleccionar como clase el atributo "euros".

Es importante observar como hemos dicho con anterioridad que la tarea es de regresión, porque la clase seleccionada es numérica, debido a que el atributo "euros" es numérico,

si la clase fuera nominal, entonces realizaríamos una clasificación al igual que en el modelo 1. Los algoritmos que “WEKA” tiene disponible y que pueden realizar el modelo de regresión son los siguientes:

- ❖ De la familia “*Functions*” los algoritmos:
 - “*LeastMedSq*”, “*LinialRegression*” y “*RBFNetwork*”.
- ❖ De la familia “*Lazy*”:
 - “*IBK*”, “*Kstart*” y “*LWL*”.
- ❖ De la familia de algoritmos “*MetaLearned*”:
 - “*AdditiveRegression*”, “*Bagging*”, “*CVParameterSelection*”, “*MultiScheme*”, “*RandomSubSpace*”, “*RegressionByDiscretization*”, “*Stacking*” y “*Vote*”.
- ❖ De la familia “*Rules*”:
 - “*ConjunctiveRule*”, “*DecisionTable*”, “*M5Rules*” y “*ZeroR*”.
- ❖ De la familia “*Tree*”:
 - “*DecisionStump*”, “*M5P*” y “*REPTree*”.

Una vez indicada la lista de algoritmos es importante reseñar que hay algoritmos en esta lista que también se utilizaron en el modelo 1 que era de clasificación. Por lo que, uno se puede preguntar: ¿cómo es posible que un mismo algoritmo sirva para una tarea de clasificación y sirva también para una tarea de regresión? La respuesta sería que hay algoritmos que sirven para realizar varias tareas distintas y en este caso son capaces de realizar una tarea de clasificación o una tarea de regresión. Con lo que aun siendo tareas claramente de naturaleza distinta hay algoritmos que se moldean y pueden realizar las dos tareas de clasificación y de regresión, por consiguiente, tales algoritmos están disponibles en el estudio del modelo 1 y en el estudio del modelo 2.

Para el estudio realizaremos el modelo con cada uno de los algoritmos citados en la lista anterior. En cada ejecución del modelo “WEKA” mostrará de forma inmediata la información de entrada para construir el modelo ‘Run Information’ y a la vez el pájaro simbólico de “WEKA” que aparece en la parte inferior derecha empezará a moverse indicando que ha empezado la ejecución y comenzará a construirse la regresión con el algoritmo que se esté ejecutando del estudio. En el momento en que se termine la construcción del modelo se visualizará la regresión y se procederá con el procedimiento de evaluación de la regresión con el método validación cruzada 10 X 10. Tras la

visualización del resultado de la evaluación la ejecución se habrá terminado y el pájaro de “WEKA” parará de moverse indicando que se ha finalizado la ejecución al igual que en el modelo 1. A medida que se vaya realizando el estudio en la sección “Result-list” se mostrarán las instancias que se vayan creando.

El estudio de algoritmos lo empezaremos con la construcción del modelo 2 mediante el primer algoritmo de la lista “LeastMedSq” y lo terminaremos con la prueba de la construcción del último algoritmo de la lista denominado “REPTree”.

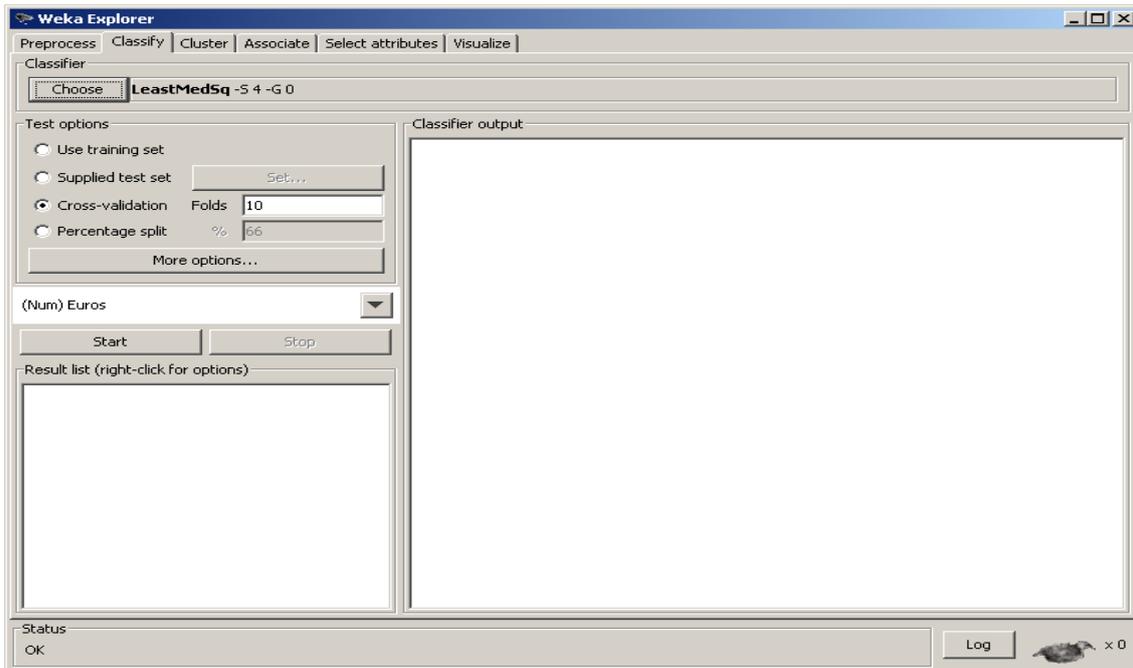


Figura 3.14: Se ha seleccionado en la sección ‘Classifier’ el algoritmo “LeastMedSq”, que es el algoritmo primero a utilizar en el estudio de algoritmos del modelo 2.

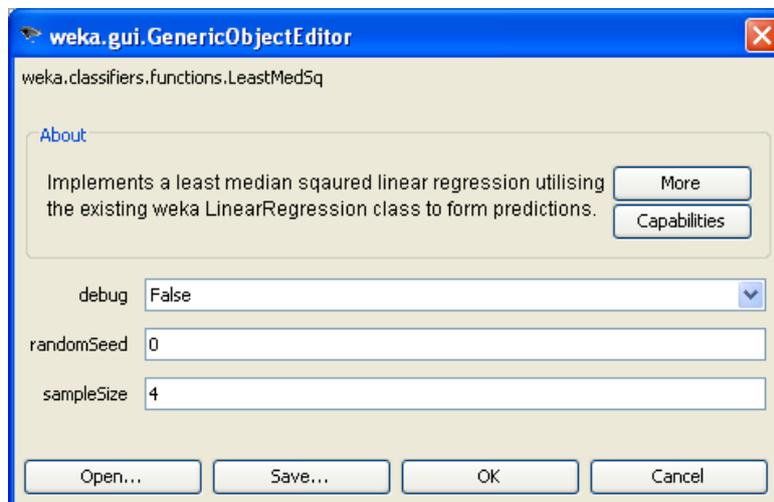


Figura 3.15: Parámetros de configuración por defecto con los que se ejecutará el algoritmo “LeastMedSq”.

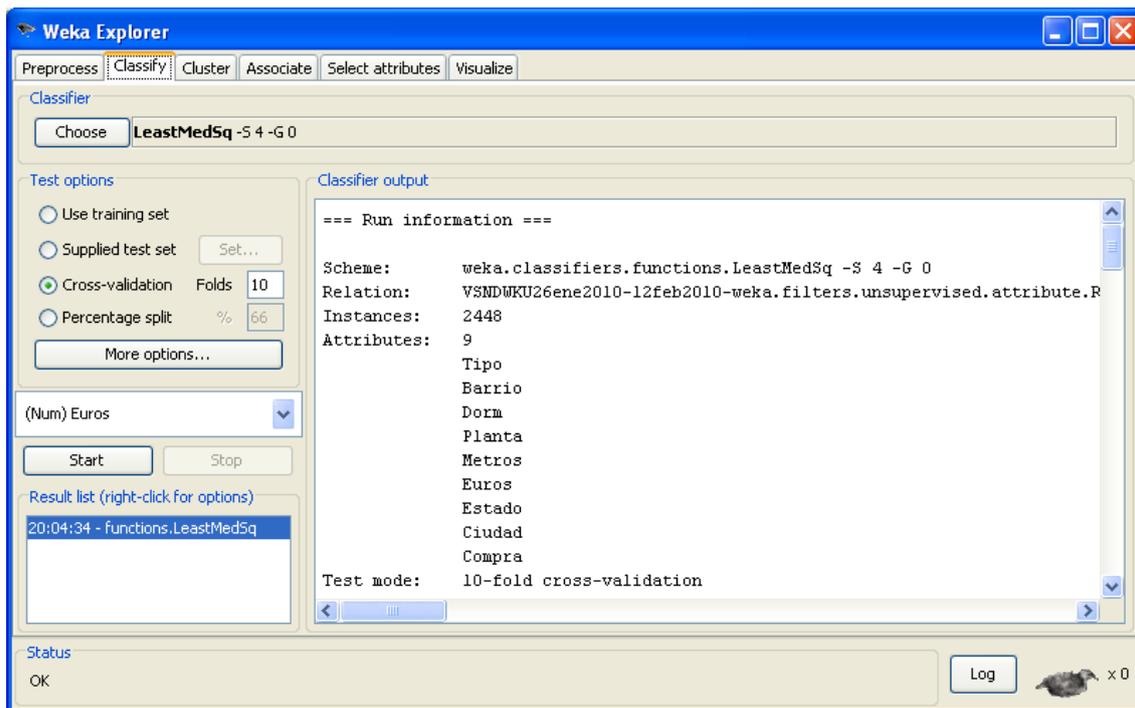


Figura 3.16: Muestra la información de ejecución del algoritmo "LeastMedSq"

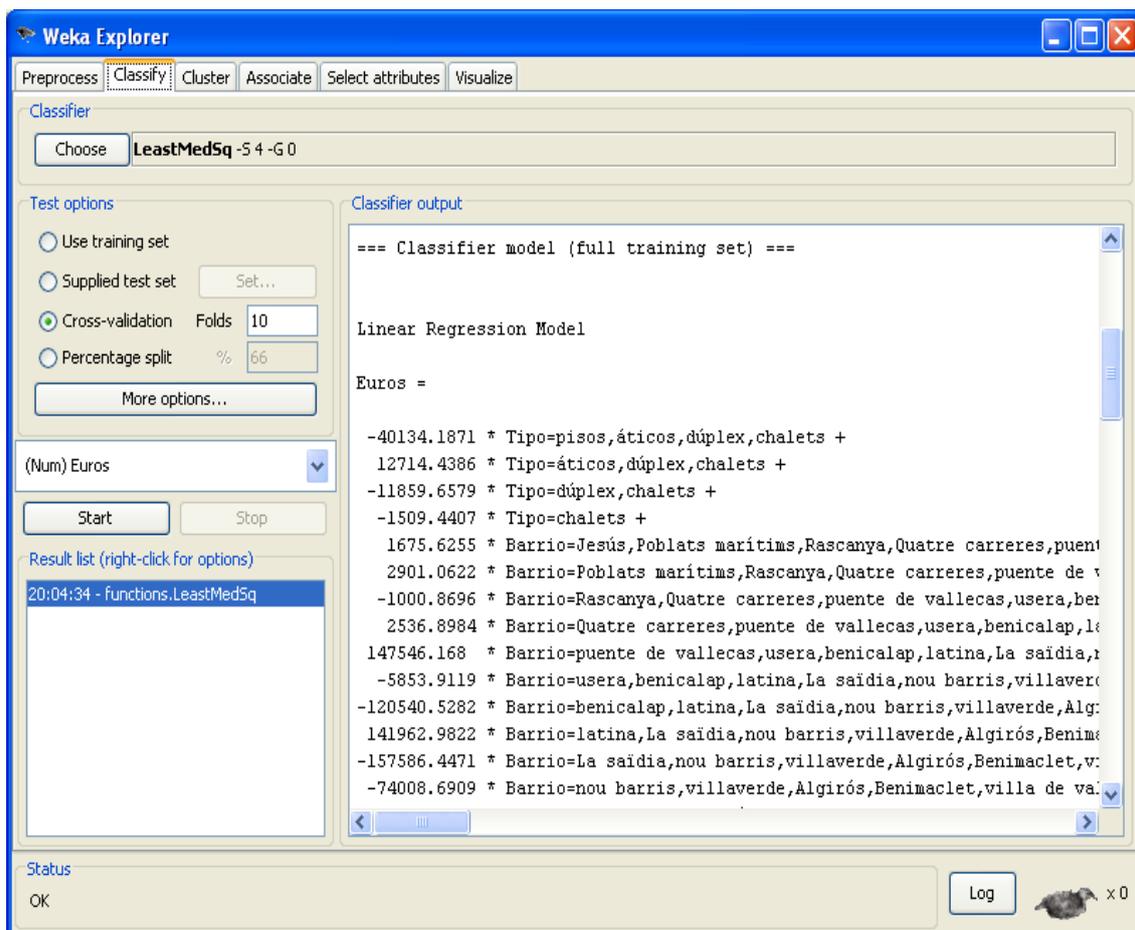


Figura 3.17: Muestra el principio de la salida de la construcción del modelo 2 con el algoritmo "LeastMedSq".

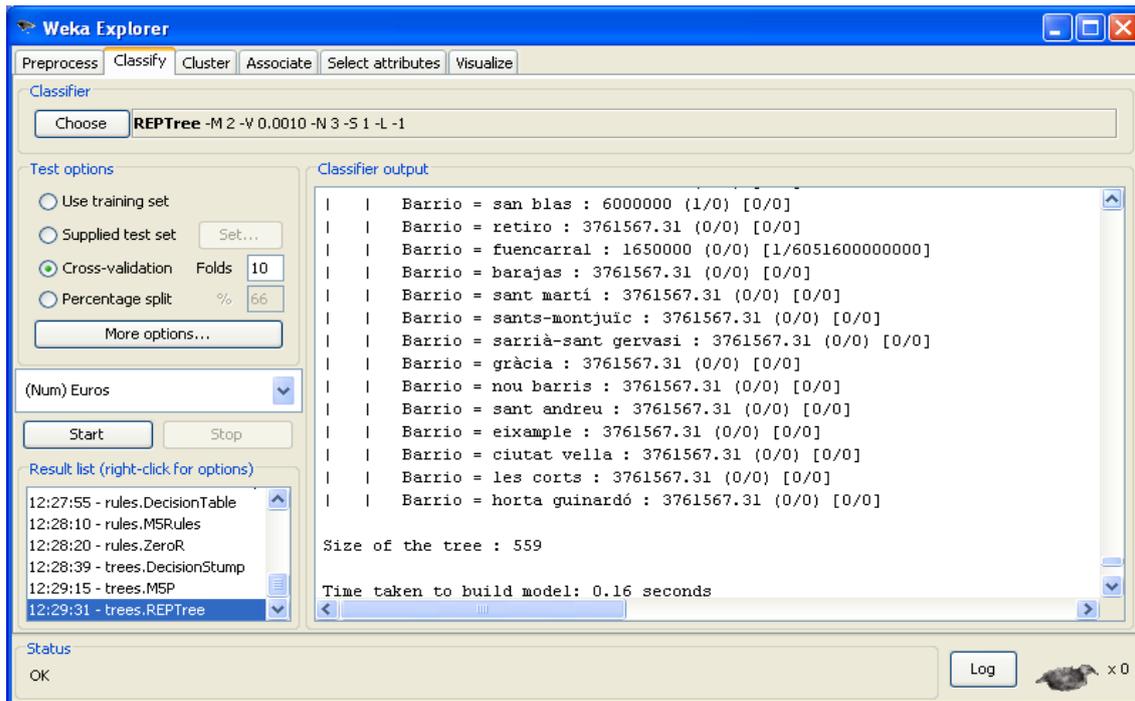


Figura 3.18: Muestra la parte final de la información de la construcción del modelo 2 con el algoritmo “REPTree”, que es el algoritmo último del estudio del modelo 2.

Una vez realizado el estudio se guardará la salida de cada uno de los algoritmos.

Aunque la parte de la evaluación de los algoritmos la detallaremos en el apartado posterior, sólo diremos que la construcción seleccionada del modelo 2 es la realizada mediante el algoritmo “M5P” de la familia “tree”. Esto lo decimos porque vamos a querer analizar dicha construcción que presenta la mejor evaluación, debido a que es la que seleccionamos para realizar el modelo descartando el resto.



Figura 3.19: Parámetros de configuración por defecto utilizados en el algoritmo “M5P”.

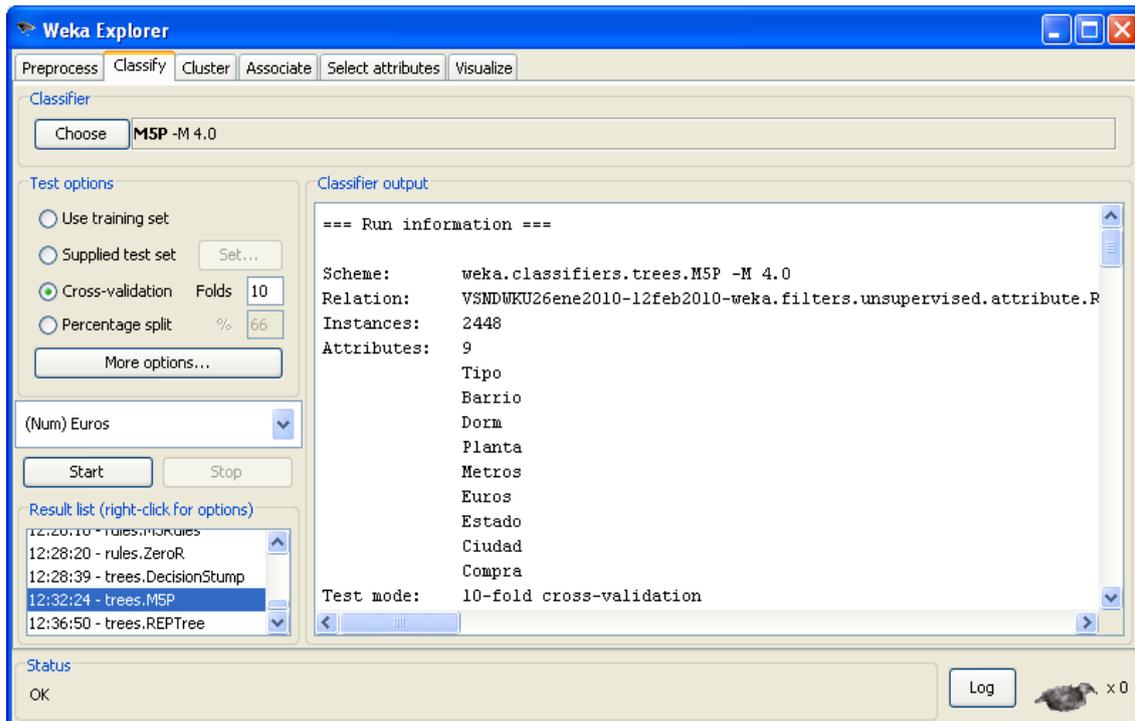


Figura 3.20: Muestra la información de entrada de la ejecución del modelo 2 con el algoritmo "M5P".

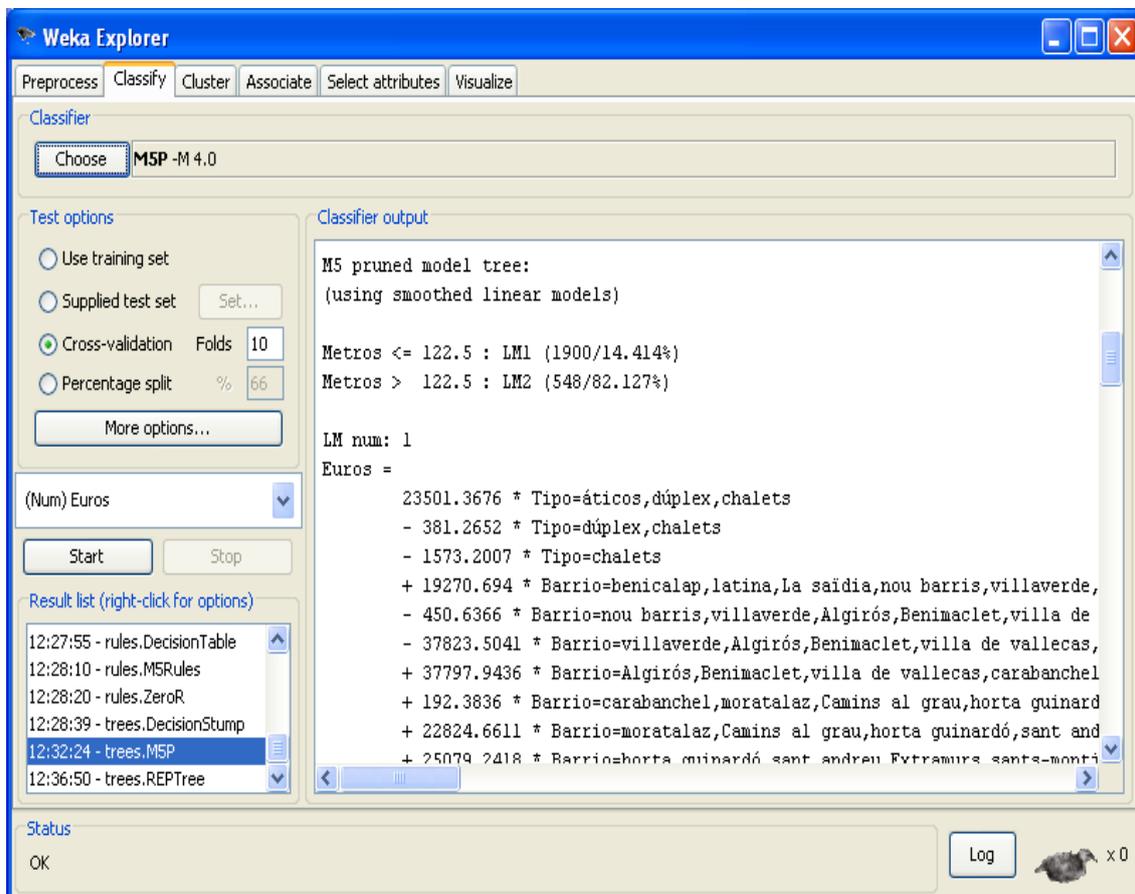


Figura 3.21: Se observa la parte inicial de la construcción del modelo 2, con el algoritmo "M5P".

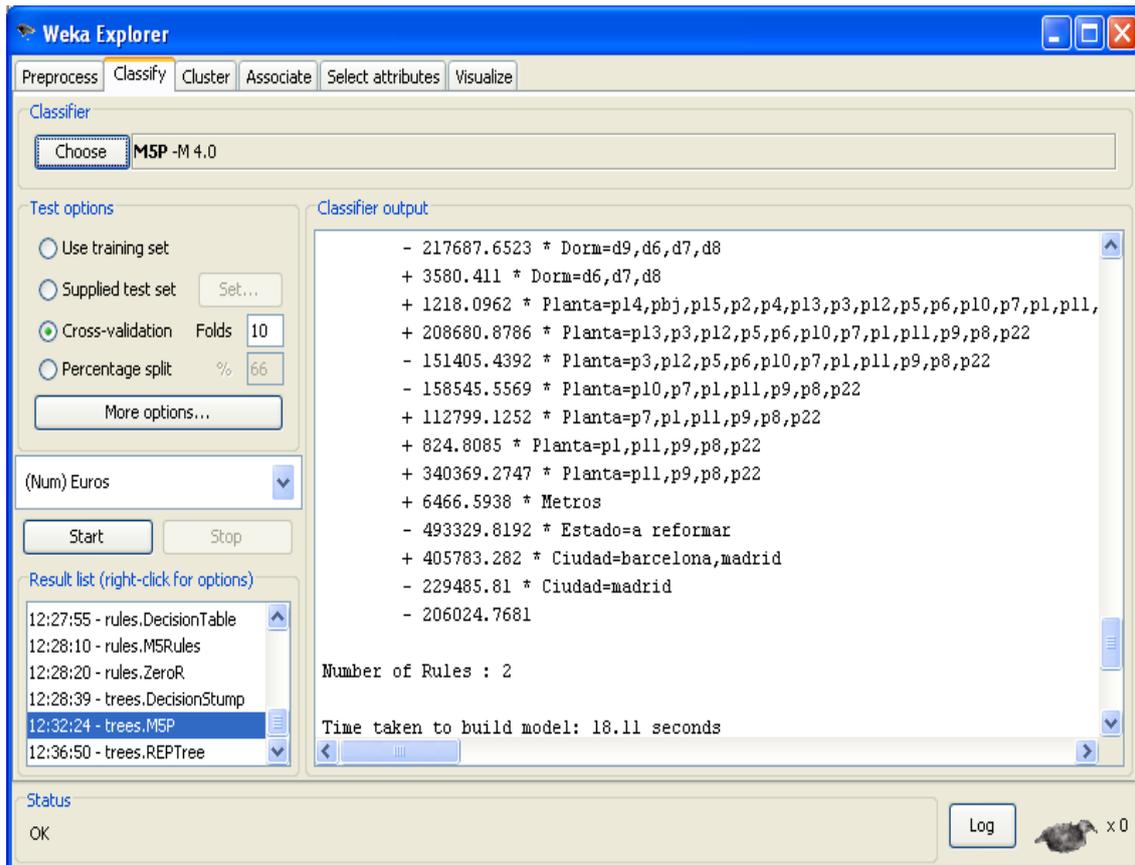


Figura 3.22: Se observa la parte final de la construcción del modelo 2, con el algoritmo “M5P”.

La salida del modelo que nos ofrece “WEKA” con el algoritmo elegido en el estudio, “M5P”, es en forma de regresión, exactamente, se observan dos regresiones ‘LM1’ y ‘LM2’. Como se observa en la parte superior de la figura 3.21 el modelo ‘LM1’ se utiliza para las viviendas a predecir el atributo “euros” que tengan en el atributo “metros” un valor menor o igual a ‘122,5 metros’ y el ‘LM2’ para las que tengan un valor mayor en el atributo “metros” a ‘122,5 metros’. Una cuestión que hay que resaltar es que a pesar de que el algoritmo “M5P” es de la familia “tree” (árbol) la salida ofrecida no es en forma de árbol o árboles sino directamente se muestra en forma de regresión, concretamente, como hemos indicado en forma de dos regresiones.

Las regresiones en general y las dos del modelo en particular tienen una variable a la que se le asigna una combinación del resto de las variables de la regresión. A la variable que se le asigna la combinación lineal se la conoce como variable de salida o clase y a las variables que forman parte de la combinación lineal se las denomina variables explicativas. En este caso el atributo “euros”, lógicamente, como es el atributo a predecir en el modelo 2: es la variable de salida en el ‘LM1’ y también en el ‘LM2’.

Por otro lado, decimos que la combinación es lineal debido a que todos los miembros que forman parte de ella aparecen precisamente de forma lineal, es decir, en ningún caso hay alguno elevado al cuadrado o al cubo o elevado a alguna otra potencia. Por lo que,

‘LM1’ y ‘LM2’ son recursiones lineales o como también se denominan, modelos lineales. De ahí, que ‘LM1’ signifique “Lineal Model 1” y ‘LM2’ signifique “Lineal Model 2”. Además de lineal cada una de las dos recursiones se dice que es múltiple, porque hay más de un atributo explicativo nominal.

Por lo que respecta, al análisis de patrones vamos a empezar explicando la forma de calcular cada miembro explicativo de una combinación lineal según posean un atributo explicativo nominal o un atributo explicativo numérico:

Un miembro de un atributo explicativo nominal se sustituye por el valor “1”, si la instancia de la vivienda ingresada, posee para ese atributo explicativo, uno de los valores a los que está asignado el atributo explicativo en el miembro y por el valor “0” cuando esto no sucede. Después ese valor “1” o valor “0” se multiplicará por el peso adyacente del miembro.

Un atributo explicativo nominal suele aparecer varias veces en varios miembros de una combinación lineal, esto se debe a que es la única forma de que dispone para proporcionar un valor numérico diferente en la combinación lineal a los valores nominales que poseen.

Por otro parte, un atributo explicativo numérico sólo aparece en uno de los miembros de una combinación lineal, debido a que, se sustituye de forma directa por el valor que presente de dicho atributo numérico la vivienda ingresada. Después el valor que resulte en la sustitución del atributo numérico se multiplicará por el peso adyacente del miembro.

Complementando a las figuras 3.21 y 3.22 vamos a poner la estructura general de los dos modelos lineales ‘LM1’ y ‘LM2’ que estamos analizando construidos con el algoritmo seleccionado para realizar el modelo 2 “M5P”. Concretamente, introducimos las primeras y las últimas líneas de cada uno de los dos modelos lineales:

```
Euros =
23559.2063 * Tipo=áticos,dúplex,chalets
- 381.3376 * Tipo=dúplex,chalets
- 1574.3677 * Tipo=chalets
+ 18905.9671 * Barrio=benicalap,latina,...
.
.
.
- 24606.224 * Ciudad=madrid
- 118773.6269
```

Recuadro 3.8: Modelo Lineal 1 ‘LM1’ de la construcción elegida del modelo 2.

```

Euros =
-1296.4125 * Tipo=dúplex,chalets
- 284874.2403 * Tipo=chalets
- 1535.1619 * Barrio=nou barris,villaverde,...
.
.
- 229483.4927 * Ciudad=Madrid
- 206028.7702

```

Recuadro 3.9: Modelo Lineal 2 'LM2' de la construcción elegida del modelo 2.

En los recuadros 3.8 y 3.9 se observa que el primer atributo explicativo que aparece en los miembros es el atributo "tipo" el siguiente el atributo explicativo "barrio" y el último el atributo explicativo "ciudad". Hay que resaltar que como decíamos los atributos explicativos si son nominales como los que se observan en el recuadro 3.8 y en el recuadro 3.9 es habitual que aparezcan cada uno de ellos en más de un miembro, y además, lógicamente, que aparezcan juntos, uno detrás de otro, todos los miembros referentes a cualquier atributo nominal. Por eso se observa que las tres primeras líneas del modelo 'LM1' en el recuadro 3.8 y las dos primeras del modelo 'LM2' en el recuadro 3.9 pertenecen al atributo "tipo". También al respecto, que el atributo "barrio" es el que aparece en más miembros del 'LM1' o del 'LM2', aunque sólo se observe en el recuadro 3.8 o en el recuadro 3.9 un solo miembro del atributo "barrio". Además dicho atributo "barrio", también es el que en algunos de sus miembros aparece asignado a las series de valores más numerosa. Todo esto se debe a que el atributo "barrio" posee muchos valores diferentes. El último miembro en cada uno de los modelos lineales aparece sin ningún atributo explicativo que le acompañe, esto es debido, a que los modelos lineales tienen un término independiente, es decir, un miembro que siempre se le asigna a la salida.

Respecto a todos los atributos explicativos que forman parte de los modelos decir que además de los atributos explicativos "tipo", "barrio" y "ciudad" que se observan en los recuadros 3.8 y 3.9, también aparecen en los dos modelos 'LM1' y 'LM2', aunque no se aprecian en dichos recuadros, los atributos explicativos "dorm", "planta", "metros" y "estado". Y de entre ellos el atributo explicativo "metros" es el único numérico.

Sin embargo, el atributo explicativo "compra" que también se pasa como atributo explicativo al modelo, no aparece en la combinación lineal 'LM1', ni en la combinación lineal 'LM2'. Esto se debe a que el atributo explicativo "compra" tiene el valor 'SÍ' en todas las instancias que se han utilizado para construir el modelo 'LM1' y el modelo 'LM2'.

Lógicamente, todas las instancias de viviendas nuevas ingresadas a predecir el atributo "euros" deben tener el atributo explicativo "compra" igual a SÍ, ya que en este modelo se pretende predecir por cuanto vender una vivienda, por lo que, se da por hecho que la vivienda es vendida, y hay que descubrir por cuanto venderla.

Para entender de forma más profunda el modelo y los patrones detallaremos el modelo lineal “LM1” y una vez entendido éste, podremos entender todo el modelo 2 formado por el modelo lineal ‘LM1’ y el modelo lineal ‘LM2’. Lo primero que haremos será mostrar el recuadro 3.10 de la parte inicial de la combinación lineal del ‘LM1’ del modelo 2. Posteriormente, se introducirán en el recuadro 3.11, algunos de los miembros de cada uno de los atributos explicativos del modelo lineal 1 de forma reducida. Por último, nos centraremos de forma separada en los miembros referentes a cada uno de los atributos explicativos, para poder obtener conclusiones de cómo afectan al atributo “euros” cada uno de los atributos explicativos y así obtener patrones a partir de ellos.

En primer lugar mostramos el recuadro 3.10 con los ‘5’ primeros miembros del ‘LM1’ introducidos de manera íntegra.

```

23501.3676 * Tipo=áticos,dúplex,chalets
- 381.2652 * Tipo=dúplex,chalets
- 1573.2007 * Tipo=chalets
+ 19270.694 * Barrio=benicalap,latina,La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,
carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc, vicálvaro,ciutat vella,
sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella, L'eixample,centro,
El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi, hortaleza,chamartín,moncloa
- 450.6366 * Barrio=nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau ,horta
guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,
tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,
les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
.
.
.

```

Recuadro 3.10: Parte inicial de la combinación lineal del ‘LM1’, del cual se muestran los cinco primeros miembros.

Seguidamente en el recuadro 3.11, observaremos algunos de los miembros pertenecientes a los atributos explicativos de la combinación lineal del ‘LM1’.

```

23501.3676 * Tipo=áticos,dúplex,chalets
- 381.2652* Tipo=dúplex,chalets
- 1573.2007* Tipo=chalets
+ 19270.694* Barrio=benicalap,latina,La saïdia,...
.
.
.
- 131663.0512* Barrio=hortaleza,chamartín,moncloa
+ 52639.3656* Barrio=chamartín,moncloa
- 60006.1702 * Barrio=Moncloa
- 14509.5593* Dorm=d2,d3,d4,...
- 33871.7504* Dorm=d3,d4,d5,...
- 21790.0035* Dorm=d4,d5,d9,...
- 2164.3337* Dorm=d9,d6,d7,d8
+ 1052.6221* Dorm=d6,d7,d8
+ 358.1139 * Planta=p14,pbj,p15,...
+ 12906.7895 * Planta=p15,p2,p4,...
+ 10635.365 * Planta=p4,p13,p3,...
+ 65524.7747 * Planta=p13,p3,p12,...
- 75613.7002 * Planta=p3,p12,p5,...
- 468.7697* Planta=p10,p7,p1,...
+ 201.1062 * Planta=p7,p1,p11,...
+ 242.4894 * Planta=p1,p11,p9,p8,p22
+ 31717.1932* Planta=p11,p9,p8,p22
+ 4044.4422 * Metros
- 58917.4469 * Estado=a reformar
+ 100105.6718 * Ciudad=barcelona,madrid
- 23843.2991 * Ciudad=madrid
- 118415.6469

```

Recuadro 3.11: Estructura general de la combinación lineal del ‘LM1’. Aparecen presentes todos los atributos explicativos. No obstante se han omitido muchos de los miembros del atributo explicativo “barrio”. Y los miembros que poseen muchos valores han sido reducidos a los primeros valores.

Ahora vamos a detallar por separado las partes de la combinación lineal del ‘LM1’ referidas a cada uno de los atributos explicativos. Empezaremos por el orden en el que aparecen en ‘LM1’. Por lo cual como se puede observar en el recuadro 3.11: primero empezaremos por el atributo explicativo “tipo”, y después por los atributos explicativos “barrio”, “dorm”, “planta”, “metros”, “estado” y “ciudad”. Además por último nos centraremos en el término independiente del ‘LM1’.

Como hemos dicho vamos a empezar por el atributo explicativo “tipo”. En el recuadro 3.12 se observan los miembros pertenecientes al atributo explicativo “tipo” del ‘LM1’ del modelo 2 seleccionado.

```

23501.3676* Tipo=áticos,dúplex,chalets
- 381.2652* Tipo=dúplex,chalets
- 1573.2007* Tipo=chalets

```

Recuadro 3.12: Miembros del atributo “tipo” en el ‘LM1’.

En cuanto a los patrones que se observan del atributo explicativo “tipo” en ‘LM1’, decir que: el valor ‘áticos’ es el tipo más caro. Esto se debe a que el valor nominal “áticos” es el único valor que aparece en el primero de los miembros del atributo “tipo” en donde el peso es positivo, y no aparece en los dos restantes miembros, los cuales disponen de un peso negativo. Por contra los valores ‘dúplex’ y ‘chalets’ no pueden ser los más caros, porque además de aparecer en el único miembro con peso positivo también aparecen en el miembro segundo del atributo “tipo” en donde el peso es negativo. Además respecto al valor nominal ‘chalets’, también aparece en el tercer y último miembro de dicho atributo “tipo”. A continuación calcularemos cada uno de los tres miembros.

El miembro “1º” del atributo “tipo”: “23501.3676* Tipo=áticos,dúplex,chalets” se sustituye por: “23501,3676 X 1 = 23501,3676 €”.

El miembro “2º” del atributo “tipo”: “- 381.2652* Tipo=dúplex,chalets” se sustituye por: “- 381,2652 X 0 = 0 €”.

Y el miembro “3º”: “- 1573.2007* Tipo=chalets” se sustituye por: “-1573,2007 X 0= 0 €”.

Con lo que el valor ‘áticos’ aporta un total de:

Total suma valor ‘áticos’ = miembro1TipoParaÁticos + miembro2TipoÁticos + miembro3TipoÁticos = 23501,3676 + 0 + 0 = 23501,3676 €

En cuanto al valor ‘dúplex’ aporta un total de:

Total suma valor ‘dúplex’ = miembro1TipoParaDúplex + miembro2TipoDúplex + miembro3TipoDúplex = 23501,3676 - 381,2652 + 0 = 23120,1024 €

Respecto al valor ‘chalets’ aporta un total de:

Total suma valor ‘chalets’ = miembro1TipoParaChalets + miembro2TipoChalets + miembro3TipoChalets = 23501,3676 - 381,2652 - 1573,2007 = 21547,9017 €

Respecto al atributo “tipo” hay que explicar que además del valor ‘áticos’, ‘dúplex’ y ‘chalets’, en las viviendas que se han utilizado para entrenar el ‘LM1’, también hay viviendas que adoptan el valor referente al atributo “tipo” denominado ‘pisos’ y ‘estudios’. Precisamente, el tipo ‘pisos’ es el más frecuente en los datos con los que se ha construido el modelo por lo que es un valor muy importante. Respecto del valor ‘pisos’ y del valor ‘estudios’ como no aparecen estos valores asignados a los miembros del atributo “tipo” en el ‘LM1’ añaden ‘0 € al atributo “euros”, es decir, nada. De esto se concluye que los valores “pisos” y “estudios” son los más baratos. Además, si la vivienda a predecir el atributo “euros”, tuviera otro valor distinto en el atributo “tipo”, a los valores del atributo “tipo” utilizados para construir el modelo dicho valor también añadirían ‘0 € al atributo “euros”. No obstante, en el caso de que fuera un valor distinto

a los estudiados la predicción del atributo “tipo” carecería de interés, ya que al no haber formado parte de la construcción del modelo sería un valor que no ha sido entrenado para predecir el atributo “euros”. Por lo que, si los datos de la vivienda a predecir no han sido entrenados la predicción del atributo “euros” no va a resultar útil.

Ahora pasaremos a detallar los miembros del atributo “barrio”. El recuadro 3.13 se refiere a dichos miembros del atributo “barrio”:

```
+ 19270.694 * Barrio=benicalap, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 450.6366 * Barrio=nou barrís, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 37823.5041 * Barrio=villaverde, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 37797.9436 * Barrio=Algirós, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 192.3836 * Barrio=carabanchel, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 22824.6611 * Barrio=mortalaz, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 25079.2418 * Barrio=horta guinardó, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 17602.9658 * Barrio=sant andreu, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 13421.874 * Barrio=Extramurs, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 36543.1412 * Barrio=vicálvaro, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 61052.4143 * Barrio=ciutat vella, ...,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 19072.5601 * Barrio=sant martí, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 694.2626 * Barrio=campanar - beni, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 1039.2879 * Barrio=tetuán, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 216.7501 * Barrio=Patraix, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 32115.2884 * Barrio=gràcia, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 63124.2232 * Barrio=san blas, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 28981.4835 * Barrio=barajas, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 452.6147 * Barrio=ciudad lineal, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 43271.1579 * Barrio=Ciutat vella, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 35427.188 * Barrio=L' eixample, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 51953.4764 * Barrio=centro, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 11922.0161 * Barrio=El pla del real, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 705.1037 * Barrio=les corts, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 19008.2381 * Barrio=eixample, ...,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 34677.7022 * Barrio=fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 63540.3739 * Barrio=salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 23096.4833 * Barrio=chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 62510.1766 * Barrio=sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 131663.0512 * Barrio=hortaleza,chamartín,moncloa
+ 52639.3656 * Barrio=chamartín,moncloa
- 60006.1702 * Barrio=Moncloa
```

Recuadro 3.13: Miembros del atributo “barrio” en el ‘LM1’.

En el recuadro 3.13 se han introducido ordenadamente los miembros del modelo lineal 1, ‘LM1’ que estamos analizando en los que está presente el atributo “barrio”. Cada uno de los miembros muestra como máximo siete de los valores de barrio a los que está igualado. Por lo que, respecto a los miembros que poseen más de siete barrios, decir que, aparece de la serie de barrios que tienen, el primero de los barrios y los seis últimos barrios que tienen asignados. En cuanto a la metodología utilizada para la asignación de los barrios hay que resaltar:

- Los miembros del atributo “barrio” están ordenados de tal forma: que el primero de los miembros es el que está igualado a más barrios y el último el que está igualado a

menos barrios. En todos los miembros del atributo “barrio” se sigue la misma dinámica, es decir, los miembros que aparecen antes están igualados a más valores que los que aparecen después. Además, no aparecen nuevos valores en un miembro posterior a otro, por consiguiente, cada miembro del atributo explicativo “barrio” que aparece posterior a otro, está igualado a un subconjunto de los barrios del anterior.

- Los valores que desaparecen de un miembro del atributo “barrio”: son los que aparecen al principio de la serie de barrios del miembro anterior. Respecto al número de valores que desaparecen hay que resaltar que no se sigue una pauta concreta, por lo que, siempre no desaparece el mismo número de valores entre un miembro y el siguiente. Habitualmente se eliminan, en un miembro del atributo “barrio” posterior, el primero de los valores de barrio, o los dos o tres primeros valores de barrios que tiene el miembro anterior. Esta regla se sigue hasta llegar al último de los miembros referentes al atributo “barrio”, el cual ya no tiene posterior. Lógicamente, de esta regla queda excluido el miembro del atributo “barrio” que aparezca en primer lugar, por no disponer de un miembro anterior del atributo “barrio”.

Estas características que se han relatado de las pautas que siguen los valores asignados a los miembros del atributo “barrio”: también se encuentran en los otros atributos explicativos. Por lo que, en el recuadro 3.12 en el que se observaban los miembros del atributo “tipo”, también se seguía dicha reducción y no inserción de nuevos valores de un miembro del atributo “tipo” respecto a otro posterior referente también al atributo “tipo”.

Una vez sabemos las características que se han mencionado vamos a buscar patrones para el atributo “barrio”:

Para encontrar patrones de dicho atributo “barrio” vamos a insertar una tabla indicando el barrio o los barrios más caros, si es que hay más de uno y también el barrio menos caro o los menos caros. Para ello, lo que vamos a hacer será realizar la suma de los pesos de los miembros del atributo “barrio” del ‘LM1’ que los tenemos todos en el recuadro 3.13.

Cuando la suma alcance mayor valor, entonces, estaremos ante el barrio o los barrios más caros si es que aparece más de uno, siempre que la suma sea mayor que cero. Por el contrario, cuando la suma alcance el menor valor estaremos ante el barrio o los barrios más baratos, siempre que la suma sea negativa. Decimos siempre que la suma sea mayor que cero en el caso de encontrar el barrio o los barrios más caros, porque si la mayor suma es menor que cero, entonces, un barrio que no apareciera reflejado en los miembros del atributo “barrio” sería más caro, ya que por lo menos, no disminuiría el atributo “euros”.

Por otra parte, para el barrio más barato, decimos siempre que la suma mínima sea menor que cero, porque de lo contrario, un barrio que no apareciera en los miembros del atributo “barrio” del modelo sería más barato.

De forma similar, si la suma mayor es igual a cero o la suma menor es igual a cero, entonces, los barrios que no aparecen en los miembros también serían los más caros o los más baratos dependiendo de si la suma mayor es cero o la suma menor es cero, eso sí pero esta vez junto con los valores que han dado valor cero en la suma mayor o menor.

Miembro barrio	Peso (€)	Suma acumulada (€)	Primer barrio (omitir el resto)
1º	19270,694	19270,694	benicalap
2º	- 450,6366	18820,0574	Nou barris
3º	- 37823,5041	-19003,4467	villaverde
4º	37797,9436	18794,4969	Algirós
5º	192,3836	18986,8805	carabanchel
6º	22824,6611	41811,5416	moratalaz
7º	25079,2418	66890,7834	horta guinardó
8º	- 17602,9658	49287,8176	sant andreu
9º	13421,874	62709,6916	Extramurs
10º	- 36543,1412	26166,5504	vicálvaro
11º	61052,4143	87218,9647	ciutat vella (Barcelona)
12º	- 19072,5601	68146,4046	sant martí
13º	- 694,2626	67452,142	campanar-beni...
14º	1039,2879	68491,4299	tetuán
15º	- 216,7501	68274,6798	Patraix
16º	32115,2884	100389,9682	gràcia
17º	- 63124,2232	37265,745	san blas
18º	28981,4835	66247,2285	barajas
19º	452,6147	66699,8432	ciudad lineal
20º	43271,1579	109971,0011	Ciutat vella (Valencia)
21º	- 35427,188	74543,8131	L'eixample
22º	51953,4764	126497,2895	Centro
23º	11922,0161	138419,3056	El pla del real
24º	705,1037	139124,4093	les corts
25º	- 19008,2381	120116,1712	eixample
26º	- 34677,7022	85438,469	fuencarral
27º	63540,3739	148978,8429	salamanca
28º	23096,4833	172075,3262	chamberí
29º	62510,1766	234585,5028	sarrià-sant gervasi
30º	- 131663,0512	102922,4516	hortaleza
31º	52639,3656	155561,8172	chamartín
32º	- 60006,1702	95555,647	moncloa

Tabla 3.12: Muestra la suma acumulada de los miembros del atributo “barrio” del ‘LM1’.

Como se aprecia en la tabla 3.12 la acumulación de la suma de los miembros del atributo “barrio” hasta el miembro “29º” inclusive obtiene la suma más elevada, ya que aunque todavía hay ‘3’ miembros más después del miembro “29º”, estos no hacen que la suma aumente. Por lo tanto, los barrios que aparezcan hasta el miembro “29º” y no vuelvan a aparecer son los más caros. Entre ellos está el barrio ‘sarrià-sant gervasi’, ya

que como hemos relatado anteriormente, entre un miembro anterior a otro posterior, por lo menos, se elimina en el en el miembro posterior, el primero de los valores que aparece en el miembro anterior. Sabemos que el barrio ‘sarrià-sant gervasi’ es uno de los más caros, porque en la tabla 3.12 nos informa que el primer barrio del miembro “29º” es ‘sarrià-sant gervasi’. Pero como no sabemos si se ha habrán eliminado más barrios en el miembro “30º”, vamos a poner otra tabla, esta vez con los primeros valores del miembro “29º” y con los primeros valores del miembro “30º” para ver si hay algún otro barrio del principio de la serie de barrios asignados al miembro “29º” que no aparece en el miembro “30º”.

Miembro barrio	Barrios
29º	sarrià-sant gervasi, hortaleza, chamartín, moncloa
30º	hortaleza, chamartín, moncloa

Tabla 3.13: Muestra parte de los barrios presentes en el miembro “29º” y en el miembro “30º”.

Como vemos el único barrio que aparece en el miembro “29º” y no aparece en el miembro “30º” es ‘sarrià-sant gervasi’, por lo que es el barrio más caro en el modelo ‘LM1’, de forma exclusiva.

Respecto al barrio más barato, en el miembro “3º” referente al atributo “barrios” se observa la menor suma, ya que es la única suma acumulada que tiene un valor negativo. Por lo que, vamos a mostrar también una tabla con los primeros barrios asignados en el miembro “3º” y los primeros asignados en el miembro “4º”. Los que no aparezcan en el miembro “4º” y si en el miembro “3º” serán los más baratos.

Miembro barrio	Barrios
3º	villaverde, Algirós, Benimaclet, villa de vallecas
4º	Algirós, Benimaclet, villa de vallecas

Tabla 3.14: Muestra los barrios presentes en el miembro “3º” y en el miembro “4º”.

Como se observa el único barrio que aparece en el miembro “3º” y no aparece en el miembro “4º” es ‘villaverde’, por lo que, es el barrio más barato sin haber algún otro igual de barato.

Con el barrio más caro y más barato dejamos de analizar los patrones del atributo “barrio” del modelo lineal ‘LM1’ que estamos estudiando, ya que, de la misma forma se puede averiguar cómo afecta al atributo “euros” cualquiera de los barrios. Y si hay alguno que no aparece en los miembros, entonces, también sabemos que afecta en ‘0 € al atributo “euros”, por tanto, podemos saber el valor de cualquiera de los barrios con los que se ha construido el modelo.

Una vez estudiado el atributo explicativo “barrio” en el modelo, pasamos al siguiente atributo explicativo que aparece en la combinación lineal y en este caso es el atributo “dorm”. Los miembros del atributo “dorm” del ‘LM1’ se muestran en el recuadro 3.14

- 14566.8688 * Dorm=d2,d3,d4,d5,d9,d6,d7,d8
- 33643.894 * Dorm=d3,d4,d5,d9,d6,d7,d8
- 21789.0152 * Dorm=d4,d5,d9,d6,d7,d8
- 2163.8796 * Dorm=d9,d6,d7,d8
+ 1051.6329 * Dorm=d6,d7,d8

Recuadro 3.14: Se muestran los miembros pertenecientes al atributo “dorm” del ‘LM1’.

Se aprecian en el recuadro 3.14, ‘5’ miembros correspondientes al atributo “dorm”. Aquí también vamos a decir, el número de dormitorios que se referencia con el atributo “dorm” que más aumenta el atributo “euros” y el número de dormitorios que menos aumenta el atributo “euros”. Echando un vistazo a los ‘5’ miembros vemos que el número de dormitorios más caros son precisamente los que no aparecen en valores de los miembros, ya que el resultado de la suma de los miembros del atributo “dorm” en ningún momento es mayor que cero. Esto es debido a que ‘4’ de los pesos de los ‘5’ miembros son negativos y el último que es positivo, no es bastante importante como para que la suma acumulada sea positiva. Por eso, los valores más caros son los que no aparecen en ninguno de los ‘5’ miembros referidos al atributo “dorm” y añaden un total de ‘0 € al precio del atributo “euros”.

En los datos con los que se ha construido el modelo los valores del atributo “dorm” que no aparecen en los ‘5’ miembros son ‘d0’ y ‘d1’ que vienen a indicar un total de ‘0’ dormitorios y un total de ‘1’ dormitorio respectivamente. No obstante, si la instancia de la vivienda ingresada a predecir adoptará otra valor distinto para el atributo “dorm”, a los utilizados en las instancias de las viviendas de entrada, sería también el más caro, ya que también añadiría ‘0 € al atributo “euros”, es decir, no añadiría tampoco nada. Pero a semejanza que con el atributo “tipo”, un valor distinto para el atributo “dorm”, de los que se han utilizado para construir el modelo en las viviendas ingresadas a predecir el atributo “euros”, es una predicción inadecuada del atributo “euros”.

Por otro lado, los valores más baratos están en el miembro penúltimo que se observa en el recuadro 3.14, ya que, es el último que presenta un peso negativo. Por lo que, se puede apreciar que el valor ‘d9’, que indica un total de ‘9’ dormitorios es el más barato. Para realizar el cálculo de cuanto abarata el precio, hay que realizar la suma desde el primer miembro relativo al atributo “dorm” hasta el penúltimo con lo que:

Suma Mínima atributo “dorm” = SumaMiembro1 + SumaMiembro2 + SumaMiembro3 + SumaMiembro4;

Suma Mínima atributo “dorm”= -14566,8688 - 33643,894 - 21789,0152 - 2163,8796 = -72163,6576 €

Por lo tanto el valor ‘d9’ disminuye el precio del atributo “euros” en 72163,6576 €

Por último, resaltar que todos los valores que aparecen en los miembros abaratan el precio del atributo “euros” siendo el valor ‘d9’ como hemos dicho el que más lo abarata.

A continuación llegamos al siguiente atributo explicativo: atributo “planta”. El recuadro siguiente muestra los miembros en los que está presente el atributo “planta” en el ‘LM1’:

+ 357.7121 * Planta=p14,pbj,p15,p2,p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
+ 12710.4458 * Planta=p15,p2,p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
+ 10848.9735 * Planta=p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
+ 64789.9535 * Planta=p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
- 74861.403 * Planta=p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
- 468.8344 * Planta=p10,p7,p1,p11,p9,p8,p22
+ 201.3856 * Planta=p7,p1,p11,p9,p8,p22
+ 242.3274 * Planta=p1,p11,p9,p8,p22
+ 31768.6154 * Planta=p11,p9,p8,p22

Recuadro 3.15: Miembros del atributo “planta” en el ‘LM1’.

Para el atributo explicativo “planta” también vamos a querer encontrar el máximo y el mínimo valor o valores si hay más de uno. Al igual que hemos hecho con el atributo “barrio” vamos a insertar una tabla para ver la suma acumulada, ya que aparecen bastantes miembros.

Miembro planta	Peso (€)	Suma acumulada (€)	planta (la primera en aparecer)
1º	+ 357,7121	357,7121	p14
2º	+ 12710,4458	13068,1579	p15
3º	+ 10848,9735	23917,1314	p4
4º	+ 64789,9535	88707,0849	p13
5º	- 74861,403	13845,6819	p3
6º	- 468,8344	13376,8475	p10
7º	+ 201,3856	13578,2331	p7
8º	+ 242,3274	13820,5605	p1
9º	+ 31768,6154	45589,1759	p11

Tabla 3.15: Muestra la suma acumulada de los miembros del atributo “planta”.

Como se observa la suma acumulada desde el miembro “1º” hasta el miembro “4º” es la máxima ya que en ninguna de las acumulaciones siguientes se supera dicha cantidad. Por lo que, los valores que aparezcan en el miembro “4º” y no aparezcan en el miembro “5º” son los más caros para el atributo “planta”.

Miembro “planta”	Plantas
4º	p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
5º	p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22

Tabla 3.16: Miembro “4º” y miembro “5º”.

En este caso sólo se observa que el valor ‘p13’ apareciendo en el miembro “4^o” no aparece en el miembro “5^o”, por lo tanto, la planta ‘p13’, decimotercera es la más cara. Por lo que, respecto a la planta más barata como la suma en ningún momento es negativa son los valores de la planta que no aparecen en ninguno de los miembros. Dichos valores aportarían ‘0 € al atributo de salida “euros”, y por lo tanto serían menores que cualquiera de las sumas acumuladas que siempre son mayores que cero. Los valores referentes al atributo “planta” con los que se ha construido el modelo y no aparecen en ninguno de los miembros del ‘LM1’ son ‘p2’, ‘p4’, ‘p14’, ‘p15’, ‘pen’, ‘pst’, ‘pss’, ‘pbj’. Y por tanto son las plantas más baratas, al igual que cualquier planta distinta a las que se han utilizado para construir el modelo, ya que también aportaría ‘0 € al precio total y también serían las plantas más baratas, pero igual, que se ha dicho para el atributo “tipo” y para el atributo “dorm” un valor distinto de los entrenados para el atributo “planta” en una instancia de vivienda ingresada a predecir el atributo “euros” provocaría una predicción inadecuada del atributo “euros en cuanto al atributo “planta”.

Pasando al siguiente atributo explicativo llegamos al atributo “metros” cuyo valor como decíamos hace que la instancia de la vivienda a predecir ingrese en el modelo ‘LM1’ si el valor es menor o igual a ‘122,5 metros’ o en el ‘LM2’ si el valor es mayor a ‘122,5 metros’. En este caso como se está detallando el modelo ‘LM1’, suponemos que la instancia de la vivienda ingresada a predecir el atributo “euros” tendría un valor menor o igual a ‘122,5 metros’. Pero además el atributo “metros” también forma parte de los dos modelos lineales ‘LM1’ y ‘LM2’. Como estamos detallando el ‘LM1’ indicamos un recuadro 3.16 con el miembro del atributo “metros” referente al ‘LM1’.

+ 4041.6995 * Metros

Recuadro 3.16: El miembro del atributo “metros” perteneciente al ‘LM1’.

Debido a que el atributo “metros” es numérico sólo aparece un miembro para tal atributo” metros”. Esto es así porque el valor que posea la instancia de la vivienda ingresada en el atributo “metros”, que como estamos en el ‘LM1’ debe ser menor o igual a ‘122,5 metros’ se sustituye directamente en el atributo “metros”. Es decir, como el atributo explicativo en este caso es numérico, se sustituye directamente (el valor referido al atributo “metros” que posee la vivienda a predecir el atributo “euros”) en dicho miembro y se multiplica por el peso. Así por ejemplo si la instancia de la vivienda a ingresar tuviera un valor de ‘100 metros’ se resolvería de la siguiente forma el miembro relativo al atributo “metros” en el ‘LM1’:

“+ 4041.6995 * Metros”; Se sustituiría el atributo explicativo por el valor ‘100’ y así quedaría
 $+ 4041,6995 * 100 = 404169,95 \text{ €}$

Por lo que, generalizando se concluye que el atributo “euros” aumenta en ‘4041,6995 € por cada metro.

Después de ver el miembro correspondiente al atributo “metros” apreciamos un único miembro del atributo explicativo “estado”. En el recuadro siguiente se muestra dicho miembro.

- 59041.7828 * Estado=a reformar

Recuadro 3.17: Miembro referido al atributo “estado” en el ‘LM1’.

Dicho miembro nos informa de que si el estado es “a reformar”: que viene a indicar que la vivienda está para hacerle una reforma, el atributo “euros” se abarata en ‘- 59041,7828 €’. Por contra, si el atributo “estado” adopta otro valor con el que se ha construido el modelo que no aparece en el miembro como ‘obra nueva’ o aparece vacío (perdido) como aparece frecuentemente en las viviendas con las que se ha construido el modelo, el atributo “euros” no se vería afectado ya que se añadirían ‘0 €’. Si la instancia de la vivienda ingresada a predecir adopta otro valor para el atributo “estado” distinto de los utilizados para construir el modelo también añadiría ‘0 €’ pero en este caso no sería una predicción adecuada para el atributo “euros” en cuanto al atributo “estado”.

Seguidamente, llegamos al último atributo explicativo que se observa en el ‘LM1’, que es el atributo “ciudad”. Del atributo “ciudad” se observan dos miembros en el ‘LM1’, dichos miembros se reflejan en el recuadro 3.18.

+ 101891.4869 * Ciudad=barcelona,madrid
- 24606.224 * Ciudad=madrid

Recuadro 3.18: Los dos miembros pertenecientes al atributo “ciudad” en el ‘LM1’.

Como se aprecia en el recuadro 3.18 aparecen los valores ‘barcelona y ‘madrid’ referentes a las ciudades de Barcelona y Madrid. Aunque no aparece la ciudad de Valencia que a pesar de que también ha sido utilizada por viviendas que han construido el modelo, no aparece en ninguno de los dos miembros. No obstante, la ciudad de Valencia, sí que aparece indirectamente en miembros del atributo “barrio”, que tienen asignados barrios de la ciudad de Valencia.

Según los dos miembros que se observan en el recuadro 3.18, la ciudad más cara es Barcelona. Esto es así, debido a que es el único valor que aparece en el primer miembro del atributo “ciudad” que posee un peso positivo y no en el segundo miembro del atributo “ciudad” que está acompañado en este caso por un peso negativo. Por tanto, se tiene que la ciudad de Barcelona es la más cara y aumenta el atributo “euros” en ‘101891,4869 €’.

Respecto a la ciudad de Madrid como aparece en los dos miembros resulta que aumenta el precio en:

$101891,4869 - 24606,224 = 77285,2629 \text{ €}$.

Con lo que, la ciudad de Madrid incrementa el atributo “euros” en ‘77.285,2629 €’.

Por lo que, de todo esto se deduce que: la ciudad de Barcelona es la más cara, después la de Madrid y la más barata es la de Valencia. Cualquier otra ciudad que se presentara en la instancia de una vivienda ingresada a predecir el atributo “euros”, también añadiría un total de ‘0 €’ al atributo “euros”, sin embargo, no sería una predicción adecuada del atributo “euros”.

Por último llegamos al término independiente del ‘LM1’ que se muestra en el recuadro 3.19.

- 118773.6269

Recuadro 3.19: El término independiente del ‘LM1’.

En cuanto a este término es una constante que se añade al valor del atributo “euros” en este caso es una constante que disminuye el precio de dicho atributo “euros” en un total de ‘-118773,6269 €’.

Con el término independiente llegamos al final del modelo lineal ‘LM1’. Por lo que, hemos explicado el ‘LM1’ y sus patrones. Además volvemos a indicar que una vez explicado el ‘LM1’ queda explicado también el ‘LM2’, ya que, todas las observaciones hechas para el ‘LM1’ valen para el ‘LM2’, ya que la combinación lineal de ambos modelos es similar en cuanto a organización y estructura.

Este modelo 2 seleccionado en el estudio de algoritmos con el ‘LM1’ y el ‘LM2’ que hemos detallado, realizado con el algoritmo “M5P” de la familia “tree”, aparece íntegramente en el “apéndice 2”. Por otra parte, en la carpeta del proyecto “NombreAlumno/anexos/modelos2/descartados/” están los modelos realizados del modelo 2 descartados. Y en “NombreAlumno/anexos /modelo2/seleccionado/” está el modelo 2 seleccionado realizado con el algoritmo “M5P” que hemos estado explicando, y que como hemos dicho aparece también íntegramente en el apéndice 2.

3.3. El aprendizaje del modelo 3

Ahora vamos hablar del último modelo de datos a realizar llamado “modelo 3: ¿Precio por el que vender una vivienda, teniendo en cuenta los días?” En este modelo de datos se quieren descubrir patrones para predecir el atributo “euros”. La diferencia entre este modelo y el modelo 2 estriba en que mientras en este modelo se va a tener en cuenta el atributo “días”, como una variable explicativa más para poder predecir el atributo “euros”, en el modelo 2 no se ha tenido en cuenta dicho atributo “días”. Por lo demás el modelo es similar: el resto de variables explicativas del modelo 3 pertenecen también al modelo 2 y las viviendas que se dispone para construir el modelo 3 son las mismas viviendas que las del modelo 2, salvo que en las viviendas del modelo 3 se dispone además del atributo “días”. Como el modelo 3 también trata de predecir el atributo “euros” la tarea a realizar también es la misma que en el modelo 2: una tarea predictiva de regresión. Por consiguiente, evidentemente, la clase a predecir será el atributo “euros” al igual que en el modelo 2. En cuanto a los datos que se van a utilizar para construir el modelo 3 son los datos almacenados en el fichero denominado “ViviendasModelo3VistaMinable26ene2010-12feb2010.arff” que contiene la vista minable del modelo 3. La siguiente tabla 3.17 muestra ‘5’ instancias de viviendas de tal fichero.

Vivienda	Tipo	Barrio	Dorm	Planta	Metros	Euros	Estado	Ciudad	Compra	Días
1	pisos	Extramurs	d5	p4	158	310000	a reformar	valencia	Sí	21
2	pisos	Latina	d2	p1	60	130000	a reformar	madrid	Sí	11
3	pisos	ciudad lineal	d2	pbj	58	220000	obra nueva	madrid	Sí	30
4	pisos	Sant martí	d3	p2	87	310000		barcelona	Sí	11
5	áticos	Horta guinardó	d3	p4	70	164000	a reformar	barcelona	Sí	29

Tabla 3.17: En la que se observan ‘5’ ejemplares del fichero “Modelo3VistaMinable26ene_2010-12feb2010.arff” de ‘2448’ ejemplares que tiene dicho fichero.

En la tabla 3.17, se pueden ver los atributos explicativos del modelo 3 que son los atributos: “tipo”, “barrio”, “dorm”, “planta”, “metros”, “estado”, “ciudad”, “compra” y “días” y el atributo de salida o atributo clase, que es el atributo “euros”. Para el modelo 3 también deseamos realizar un estudio con los algoritmos disponibles en la herramienta de minería de datos ‘Explorer’ de “WEKA” que sean capaces de construir el modelo, que como hemos dicho es una regresión. Por lo que, habrá que cargar el fichero que contiene los datos óptimos para construir el modelo 3, es decir, la vista minable del modelo 3 en la aplicación ‘Explorer’ de “WEKA”. Por consiguiente, abriremos el fichero denominado “ViviendasModelo3VistaMinable26ene2010-12feb2010.arff” en la aplicación ‘Explorer’ de “WEKA”. En la figura 3.23 se observa el fichero en la ventana ‘Preprocess’.

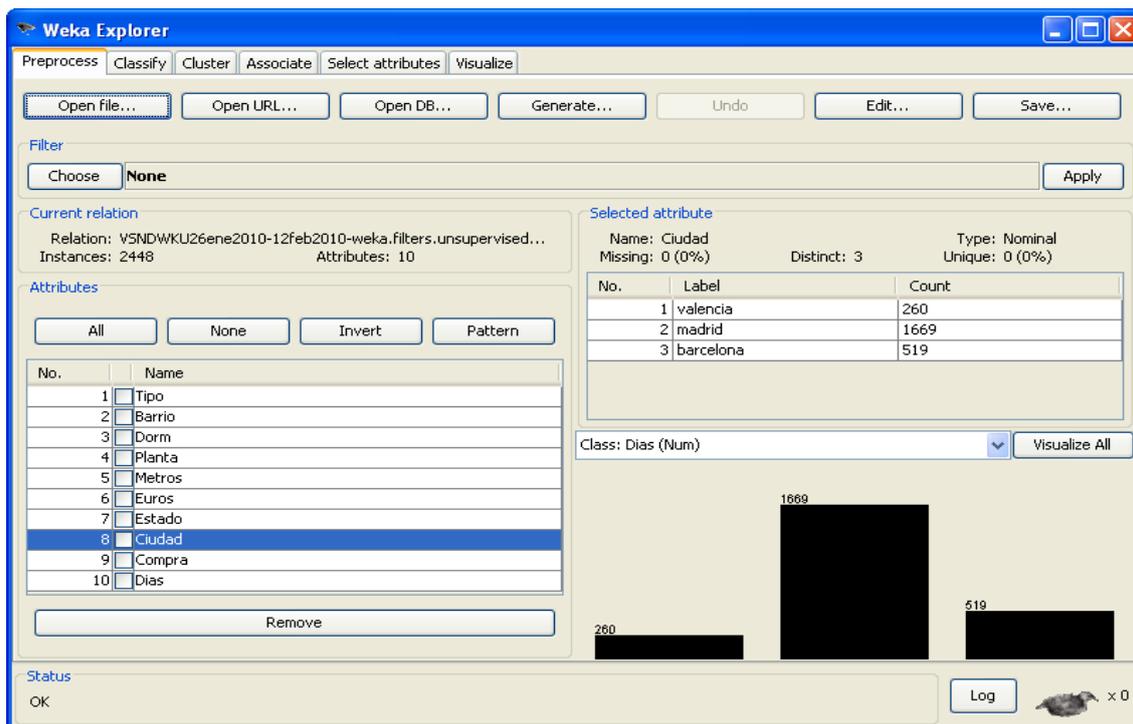


Figura 3.23: Se observa la pestaña 'Preprocess' de la aplicación, 'Explorer' de "WEKA", con los datos del fichero "ViviendasModelo3VistaMinable26ene2010-12feb2010.arff".

Al cargar el fichero se muestra la ventana 'Preprocess', pero como los algoritmos de minería de datos están en la ventana 'Classify' deberemos transitar a tal ventana denominada 'Classify'.

En cuanto a las opciones para construir el modelo que aparecen en la ventana 'Classify', decir que: como clase a predecir pondremos el atributo "euros" y como método de validación, dejaremos como anteriormente hemos hecho en el estudio de algoritmos del modelo 1 o del modelo 3, el que aparece seleccionado, "validación cruzada 10 X 10".

Ahora podremos realizar el estudio con los algoritmos que puedan construir el modelo 3, precisamente, en cuanto a tales algoritmos disponibles de "WEKA" para realizar el modelo 3, decir, que son exactamente los mismos que los que había disponibles para realizar el modelo 2. Esto es debido a que como el modelo 3 es un modelo de regresión, la aplicación "WEKA" ofrece los mismos algoritmos para poder construir el modelo 3, que los que ofrecían para construir el modelo 2. Y por consiguiente como listamos tales algoritmos cuando tratábamos el estudio del modelo 2, no los vamos a volver a citar, ya que son los mismos algoritmos. Como los algoritmos del modelo 3 utilizables son los mismos que los del modelo 2, el número de iteraciones constituidas por la fase de minería de datos y por la fase de evaluación serán el mismo número que en el modelo 2.

El estudio se realizará en el mismo orden que el que se realizó en el modelo 2, por lo que el primer algoritmo que se utilizará para construir el modelo será “LeastMedSq” y el último será el algoritmo “REPTree” con el que terminará el estudio. A continuación se muestran las figuras relativas al estudio referentes a la construcción del modelo con el primer algoritmo utilizado figura 3.24 y figura 3.25 y a la construcción del modelo con el último algoritmo figura 3.26. Una cuestión que hay que remarcar es que la construcción que hagamos con cada algoritmo disponible será con los parámetros de configuración por defecto que posea cada algoritmo al igual que en el modelo 1 o en el modelo 2.

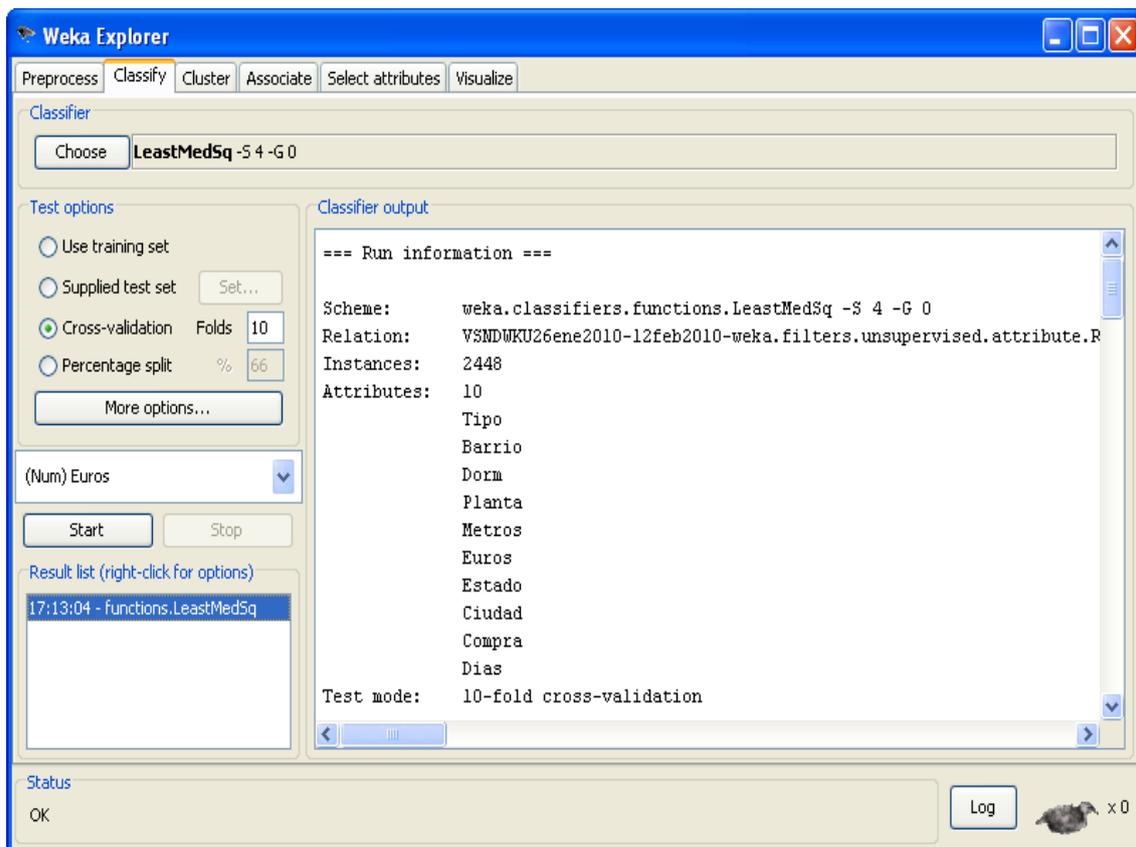


Figura 3.24: En esta figura se visualiza la información de ejecución del algoritmo “LeastMedSq”, que es el algoritmo primero del estudio del modelo 3.

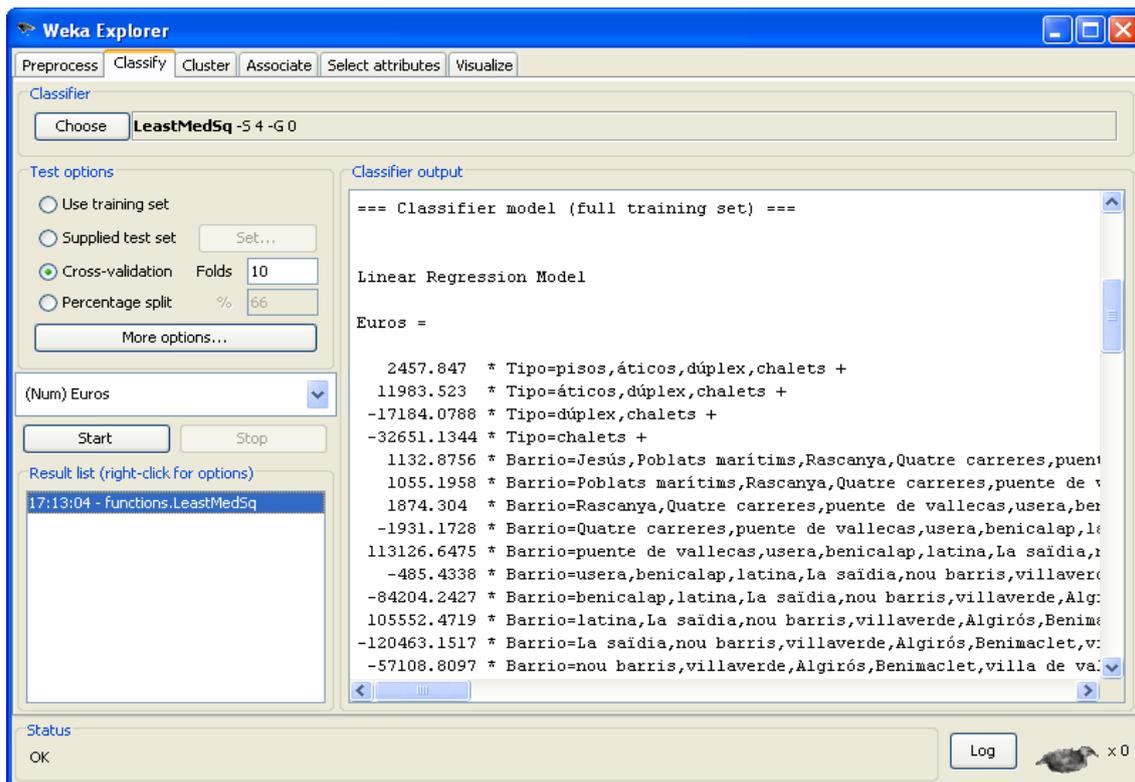


Figura 3.25: Se aprecia la parte inicial de la salida de la construcción que nos ofrece “WEKA” del algoritmo “LeastMedSq”.

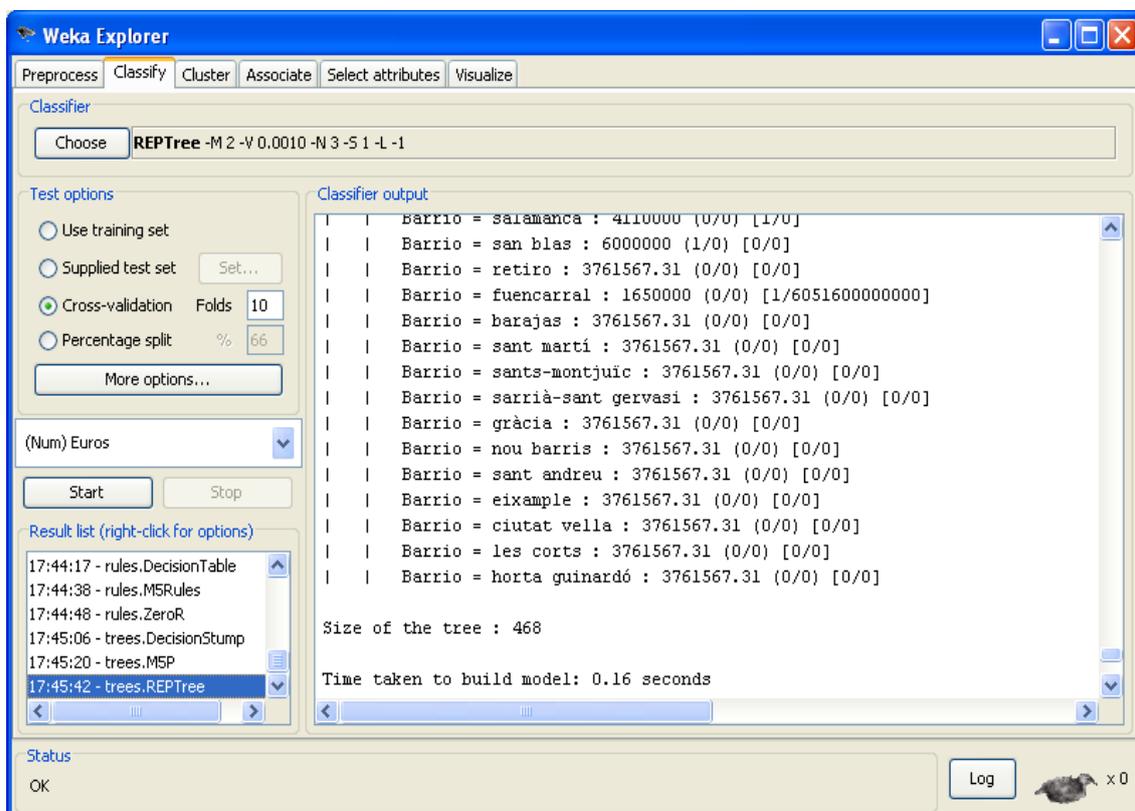


Figura 3.26: Se muestra la parte final de la salida de la construcción que nos ofrece “WEKA” del algoritmo “REPTree”, que es el algoritmo último del estudio del modelo 3.

Una vez realizado el estudio de algoritmos del modelo 3, sólo diremos en relación a la evaluación, que la construcción realizada con el algoritmo “M5P” es la elegida en el estudio del modelo 3, al igual que en el estudio de algoritmos del modelo 2, que también se seleccionó la construcción realizada con el algoritmo “M5P”. En las figuras 3.27, 3.28 y 3.29 se observa información relativa a la construcción del modelo 3 con el algoritmo “M5P”. Concretamente en la figura 3.27 se observa los datos de ejecución del algoritmo. En la figura 3.28 se muestra el principio de la salida de la construcción elegida y en la figura 3.29 el final de la salida de la construcción elegida. En cuanto a los parámetros de configuración con los que se ha construido el algoritmo “M5P”, no los vamos a mostrar, ya que, son los mismos que observamos en la figura 3.19 referente al modelo 2. Ya que, en este proyecto, siempre se utilizan los parámetros por defecto de los algoritmos para realizar las construcciones y como el algoritmo es el mismo, por lo tanto, esta claro que tiene los mismos parámetros por defecto.

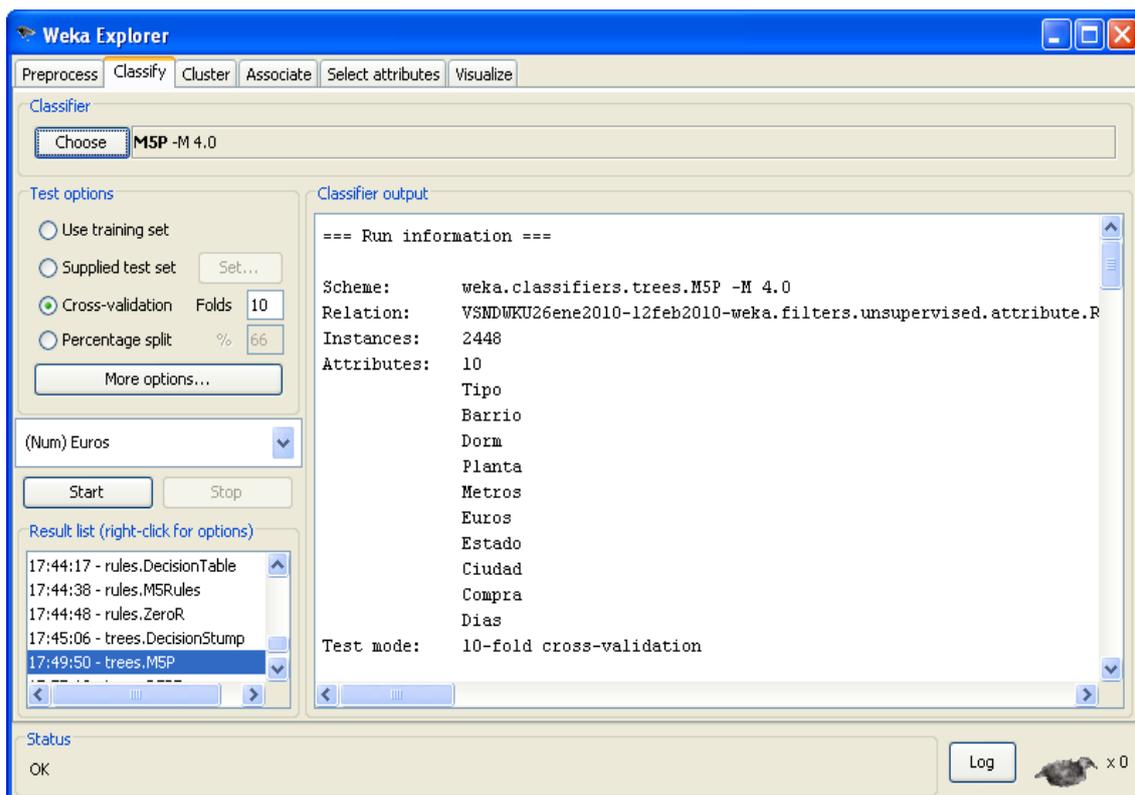


Figura 3.27: Muestra la información de entrada de la ejecución del modelo 3 con el algoritmo “M5P”.

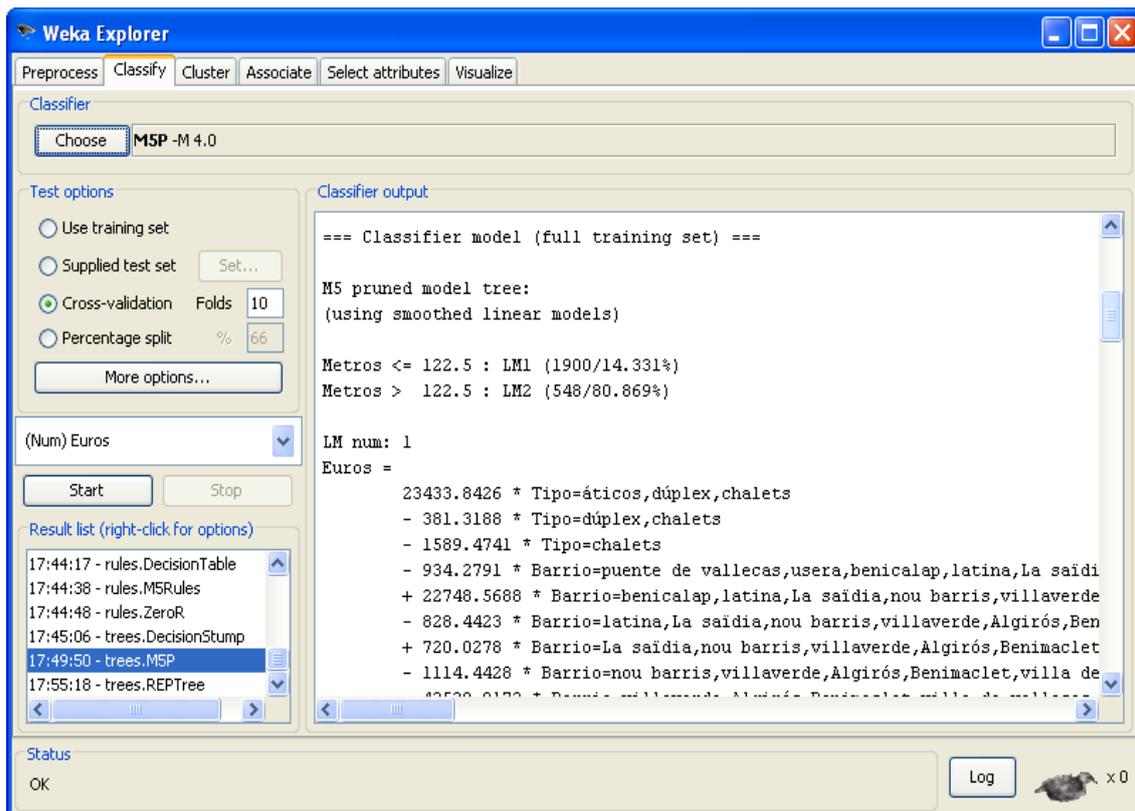


Figura 3.29: Se observa la parte inicial de la construcción del modelo 3, con el algoritmo “M5P”.

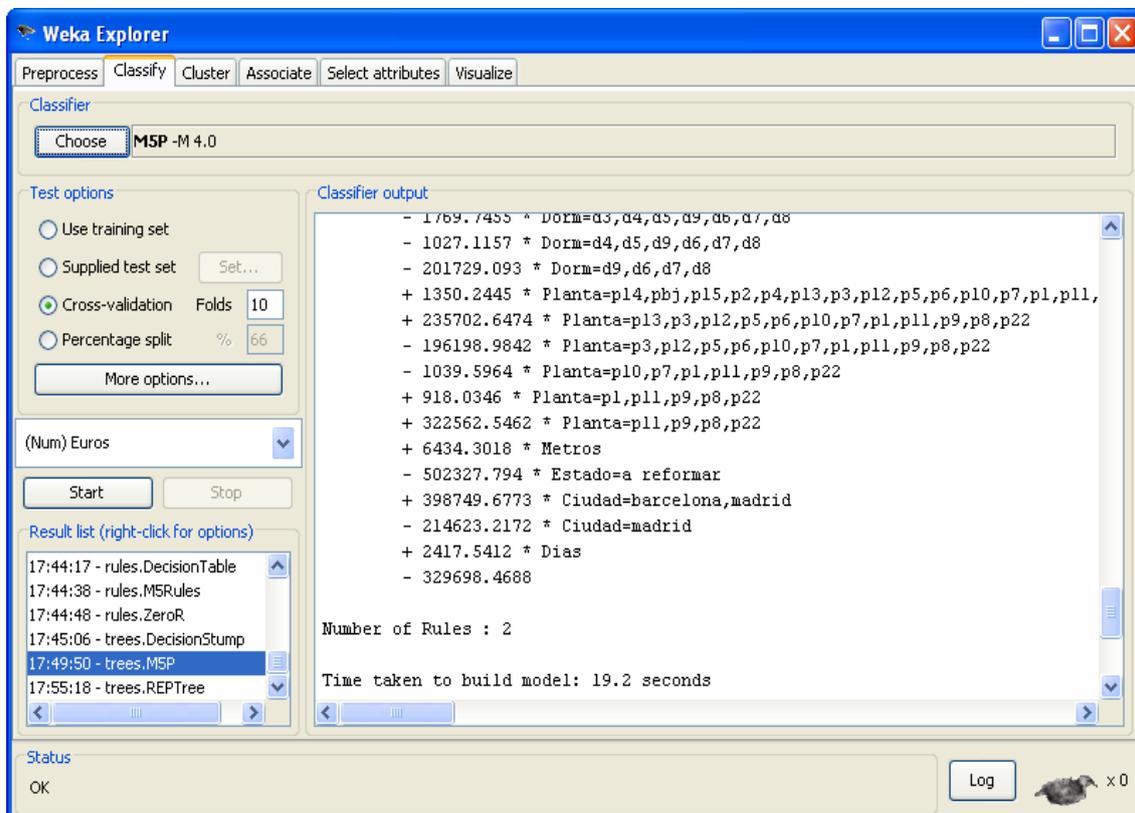


Figura 3.30: Se observa la parte final de la construcción del modelo 3, con el algoritmo “M5P”.

En cuanto a los patrones del modelo 3 tendremos en cuenta los de la construcción elegida, que es la realizada con el algoritmo “M5P”.

Como se ha recalcado varias veces el modelo 3 es un modelo de regresión, y además al igual que el modelo 2 es un modelo de regresión lineal múltiple, la regresión es lineal y múltiple por las mismas razones que se explicaron en el modelo 2 referentes a que la regresión del modelo 2 era lineal y múltiple. La salida en “WEKA” del modelo seleccionado del modelo 3 se asemeja en cuanto a la sintaxis, al modelo regresivo del modelo 2 seleccionado. Por lo que, las partes de la salida de la construcción seleccionada del modelo 2 explicadas cuando tratábamos el modelo 2 sirven para entender la salida de la construcción seleccionada del modelo 3. El modelo seleccionado en el modelo 3 como se aprecia en las figuras anteriores 3.30 y 3.31 está constituido por dos modelos recursivos: el modelo lineal 1 ‘LM1’ y el modelo lineal 2 ‘LM2’.

Al ‘LM1’ se accede si la vivienda tiene en el atributo “metros” un valor menor o igual a ‘122,5 metros’ y al modelo lineal 2 si el atributo “metros” tiene un valor mayor a ‘122,5 metros’. Curiosamente, en la construcción del modelo 2 también ocurría exactamente lo mismo. Ya que el modelo 2 seleccionado también está formado por dos modelos lineales ‘LM1’ y ‘LM2’, y para acceder al ‘LM1’ la instancia de la vivienda ingresada tiene que tener un valor menor o igual a ‘122,5 metros’ y para acceder ‘LM2’ un valor mayor a ‘122,5 metros’. No obstante, los modelos lineales además de diferenciarse en que, en los dos modelos lineales del modelo 3 aparece el atributo explicativo “días”, se puede apreciar también si se compara el ‘LM1’ del modelo 2 con el ‘LM1’ del modelo 3 o el ‘LM2’ del modelo 2 con el ‘LM2’ del modelo 3 que los miembros que aparecen son distintos. Ya que se observan diferencias en cuanto a los pesos, en cuanto a las igualdades y en cuanto al número de miembros de cada atributo explicativo, aunque, claramente se observa que dichas diferencias, en los miembros de los atributos explicativos, a excepción del atributo “días” que no aparece en el modelo 2 y si en el modelo 3, no son muy grandes y los pesos, los valores asignados y el número de miembros de cada atributo explicativo poseen bastantes similitudes.

Es importante destacar que: en cuanto a que el número de modelos lineales aparecidos es el mismo en ambos modelos: modelo 2 y modelo 3, hay que resaltar que no debería porque ser así y podría haber salido un número distinto de modelos lineales en el modelo 3 que en el modelo 2, pero aun así seguirían habiendo similitudes, ya que, el modelo 2 y el modelo 3 poseen en común, menos el atributo explicativo “días”, el resto de atributos explicativos y también la clase a predecir. Esto último, es decir, que tengan la misma clase a predecir, hace como hemos dicho varias veces que la tarea en ambos modelos sea la misma, una regresión porque dicha clase es numérica. A esto hay que añadirle que el algoritmo utilizado en la construcción elegida de los dos modelos es el mismo el algoritmo “M5P”, por lo que, al querer explicar la construcción elegida del modelo 3 y los patrones de tal construcción nos aparecen muchas semejanzas con la explicación que en el modelo 2.

En cuanto al atributo explicativo “días” que es lo que diferencia claramente el modelo 3 del modelo 2, decir que, como es numérico sólo aparece un miembro referente a tal atributo en cada uno de los dos modelos lineales del modelo 3. Esto es debido a que, como ya se dijo con anterioridad, los atributos numéricos se sustituyen directamente por el valor que posee en ellos la vivienda ingresada, y luego se multiplica dicho valor por el peso. En concreto, en el modelo lineal 1 ‘LM1’ aparece el miembro que se observa en el recuadro 3.20 relativo al atributo “días” como parte de la combinación lineal del ‘LM1’:

+ 250.493 * Dias

Recuadro 3.20: Muestra el miembro perteneciente al atributo “días” de la combinación lineal del ‘LM1’ del modelo 3 seleccionado.

Por lo tanto, la instancia de la vivienda ingresada en el ‘LM1’ para predecirle el atributo “euros”, en la parte referente al atributo “días”, se le multiplicará el valor del atributo “días” que posea por el peso que acompaña al miembro referente al atributo “días” en el modelo, que es ‘250,493’. Y por lo tanto, por cada día se le añadirá ‘250,493 € al precio del atributo “euros”. Esto viene a decir que cuantos más “días” hayan pasado para vender la vivienda, mayor será el precio del atributo “euros”.

Una cuestión importante que hay que resaltar es que la vivienda a predecir el atributo “euros”, al igual que en el modelo 2, debe tener el atributo “compra”=’SÍ’ y en ningún caso el atributo “compra”=’NO’. Además la vivienda a predecir el atributo “días” seguirá dicho patrón referente al atributo “días” del ‘LM1’, si el atributo “metros” es menor o igual a ‘122,5 metros’. En caso de que sea mayor a ‘122,5 metros’ se seguirá el ‘LM2’ del modelo, y por lo tanto, el patrón referente al atributo “días” del ‘LM2’. En el ‘LM2’ aparece el atributo “días” de la forma siguiente:

+ 2417.5412 * Dias

Recuadro 3.21: Muestra el miembro perteneciente al atributo “días” de la combinación lineal del ‘LM2’ del modelo 3 seleccionado.

El miembro del atributo “días” del ‘LM2’ viene a indicar que el precio del atributo “euros” se incrementará en ‘2417,5412 € por día. Por lo tanto, se puede apreciar que en el ‘LM2’, el incremento del atributo “euros” debido al atributo “días” es más elevado que en el ‘LM1’. La diferencia entre el peso del atributo “días” del ‘LM2’ y el ‘LM1’ es:

DiferenciaPesos = Peso atributo “días” (LM2) – Peso atributo “días” (LM1);
DiferenciaPesos = 2417,5412 – 250,493 = 2167,0482 €

Por otra parte, como hemos dicho el resto de atributos explicativos de los modelos lineales del modelo 3 seleccionado aparecen en forma similar (aproximada) en los modelos lineales del modelo 2 seleccionado. Por lo cual no vamos a volver a explicarlos en este caso.

Una cuestión que se podría preguntar alguien en relación al modelo 2 y al modelo 3 sería:

¿Por qué si en ambos modelos los datos de entrada son los mismos salvo los datos del atributo “días” que sólo aparece en el modelo 3 y la construcción seleccionada en ambos modelos es la realiza con el algoritmo “M5P”, los miembros resultantes referidos al atributo “barrio”, por ejemplo, presentan ligeras diferencias en el modelo 2 y en el modelo 3, ídem con el resto de atributos en común?

La razón sería porque como el modelo 3 tiene en los datos de entrada un atributo más, que es el atributo “días”, aparece algún tipo de modificación en el entrenamiento del algoritmo “M5P” que hace que los miembros referidos, por ejemplo, al atributo “barrio” en ambos modelos tengan algunas diferencias, en cuanto a pesos, valores asignados a al atributo y número de miembros del atributo. Ídem con el resto de atributos en común.

Una vez detallado el atributo explicativo “días” y comentada está última cuestión concluimos la explicación del modelo 3.

Dicha construcción elegida del modelo de datos está introducida íntegramente en el “apéndice 3”.

Las construcciones descartas del estudio del modelo 3 aparecen en ficheros separados en la carpeta: “NombreAlumno/anexos/modelos/modelo3/descartados”.

Y la construcción seleccionada aparece también en la carpeta “NombreAlumno/anexos /modelos/ modelo3/seleccionado” en un fichero separado, además de como hemos dicho en el “apéndice 3”.

3.3.1. Predecir las réplicas de una vivienda en función de los días

Ahora utilizaremos el modelo 3 para realizar otro estudio que consiste en: “Predecir el precio de una misma vivienda en función del valor de los días que los haremos variar”.

La idea es la siguiente: se trata de seleccionar una instancia de vivienda cualquiera de las que se le suministran al modelo 3 y predecirle el valor del atributo “euros” en función del valor del atributo “días” que será el único valor que irá variando.

Para lo cual el atributo “días” adoptará distintos valores en varias réplicas de la misma instancia vivienda seleccionada. El resto de atributos explicativos mantendrán siempre el valor original de que disponían, por lo que se mantendrán constantes.

Para ello habrá que construir nuevamente el modelo 3 (debido a que no lo tenemos implementado en ninguna aplicación) pero esta vez con un conjunto de prueba o test diferente: en este caso el conjunto de test serán las réplicas de la instancia de vivienda seleccionada. La cuestión será fijarse, una vez construido el modelo, en la predicción del atributo “euros” en cada uno de las réplicas de la instancia de la vivienda seleccionada.

Para realizar la predicción de las réplicas de una instancia de vivienda en función de los días, construiremos el modelo con el algoritmo “M5P”, ya que dicho algoritmo como hemos relatado con anterioridad es el elegido para realizar el modelo 3. Es importante señalar que en este caso la finalidad del conjunto de prueba, es decir, de las réplicas de la instancia de la vivienda, no es probar el modelo 3, sino predecir el atributo “euros” a tales y así observar el patrón referente al atributo “días” que como hemos dicho es el único que irá variando. También decir que, como hemos indicado debemos volver a construir el modelo, esto es debido a que no tenemos la construcción del modelo 3 implantada para su explotación en ninguna aplicación propia.

Una vez explicada la idea vamos a llevarla a cabo. Para ello primero crearemos un fichero denominado “ReplicasViviendaModelo3VariandoDias26ene201012feb2010.arff” en el que replicaremos una de las instancias de viviendas del fichero, “ViviendasModelo3VistaMinable26ene2010-12feb2010.arff”. Dicha instancia de vivienda seleccionada la replicaremos un total de “n – veces”. En donde pondremos un valor distinto para el atributo “días” en cada una de ellas y también asignaremos al atributo “euros” el valor “?” (desconocido) en ellas. Utilizamos el valor “?” debido a que, la herramienta “WEKA” representa con “?” los valores desconocidos.

Hay que resaltar que escogemos una de las instancias de vivienda del fichero “ViviendasModelo3VistaMinable26ene2010-12feb2010.arff”, porque sabemos que la vivienda es válida, debido a que, se utiliza en el modelo 3. No obstante, podríamos utilizar otra válida de otro lugar. Con válida queremos decir que tenga los atributos que utilizamos en el modelo, en el orden adecuado, y que el atributo “compra” posea el valor ‘SÍ’.

En cuanto a la variación del atributo “días”, resaltar que, no vamos a poner en ninguna réplica el valor real del atributo “días”, que aparece en la instancia de la vivienda seleccionada. Esto lo hacemos porque si dejamos en una réplica el valor real del atributo “días” dejaría de ser para dicha réplica desconocido el atributo “euros” y es una cuestión que no deseamos, ya que, queremos que todas las réplicas tengan desconocido el atributo “euros”.

En cuanto a la vivienda seleccionada para el fichero “ReplicasViviendaModelo3_VariandoDias26ene2010-12feb2010.arff” se ha elegido como vivienda al azar: una

vivienda de la ciudad de Valencia del barrio de ‘Extramurs’, que tiene ‘5’ dormitorios y que como todas las que se destinan al modelo 3 (ya que es una imposición del modelo) está vendida. Estas son algunas de las características que tiene la vivienda. La instancia de dicha vivienda escogida la replicaremos exactamente “15 – veces”, aunque el número de replications podría haber sido distinto.

Tipo	Barrio	Dorm	Planta	Metros	Euros	Estado	Ciudad	Compra	Días
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	1
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	14
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	27
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	40
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	53
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	66
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	79
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	92
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	105
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	118
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	131
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	144
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	157
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	170
pisos	Extramurs	d5	p4	158	?	a reformar	Valencia	SÍ	183

Tabla 3.18: Replicaciones de la instancia de una misma vivienda cambiando el atributo “euros”. Dichas réplicas las insertaremos en “ReplicasViviendaModelo3VariandoDias26ene2010_12feb2010.arff”.

Como el fichero de las réplicas de una instancia de la misma vivienda tendrá el formato “arff”, obligatoriamente debe tener una cabecera con el nombre de la relación y debe disponer de la información de los atributos, metadatos. Como nombre de la relación a al fichero se le pondrá el nombre que adopta el fichero.

Debido a que, la parte de la cabecera referida a los atributos en el fichero de las réplicas de una instancia de una misma vivienda debe ser idéntica a la adopta en el fichero de las viviendas que se le suministran al modelo 3, lo más idóneo será copiar dicha parte. Una porción del fichero de las réplicas de una misma vivienda, una vez se ha creado con la cabecera (mostrando sólo el primer y último valor de cada atributo nominal) se muestra seguidamente.

```

@relation ReplicasViviendaModelo3VariandoDias26ene2010-12feb2010.arff
@attribute Tipo {pisos,áticos,dúplex,estudios,chalets}
@attribute Barrio {'La saïdia',..., 'horta guinardó' }
@attribute Dorm {d3,..., d8 }
@attribute Planta {p7,..., p22}
@attribute Metros numeric
@attribute Euros numeric
@attribute Estado {'obra nueva','a reformar'}
@attribute Ciudad {valencia,madrid,barcelona}
@attribute Compra {NO,Sí}
@attribute Días numeric

@data
pisos,Extramurs,d5,p4,158,?, 'a reformar',valencia,Sí,1
.
.
.
pisos,Extramurs,d5,p4,158,?, 'a reformar',valencia,Sí,183

```

Recuadro 3.22: Fichero “ReplicasViviendaModelo3VariandoDias26ene2010-12feb2010.arff”.

Después de tener a nuestra disposición el fichero “ReplicasViviendaModelo3Variando_Dias26ene2010-12feb2010.arff” procederemos a construir otra vez el modelo 3 con el algoritmo seleccionado en el estudio “M5P”. Aunque esta vez, como el conjunto de prueba va a ser el fichero “ReplicasViviendaModelo3VariandoDias26ene2010_12feb2010.arff”, debemos seleccionar como método de validación la opción denominada, “Supplied test set” (conjunto de prueba suministrado) en vez de “validación cruzada 10 X10”, para poder escoger un fichero distinto del de entrada para la prueba. Para lo cual lo primero que haremos será abrir en la aplicación ‘Explorer’ de “WEKA” el fichero “ViviendasModelo3VistaMinable26ene2010-12feb2010.arff” para disponer de las instancias de viviendas de entrada para realizar el modelo. Seguidamente seleccionaremos en el panel ‘Classify’ de la aplicación ‘Explorer’ de “WEKA” como clase el atributo “euros”. Y en la sección ‘Test options’ seleccionaremos como método de validación “Supplied test set”.

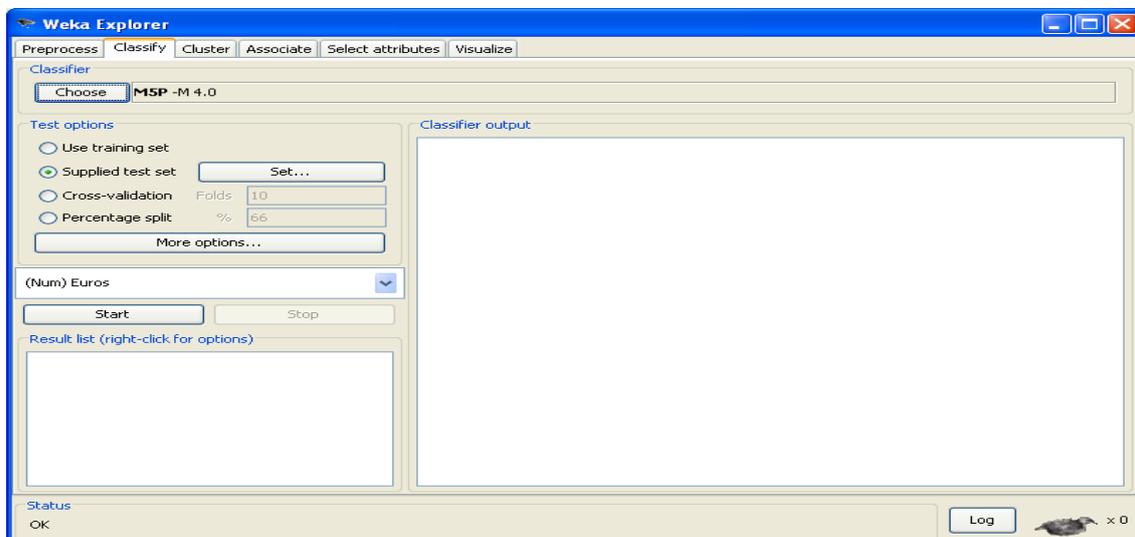


Figura 3.31: Seleccionamos ‘Supplied test set’ en la sección ‘Test options’.

Una vez seleccionado el método de validación ‘Supplied test set’ presionaremos sobre el botón “Set...” perteneciente a tal método de validación, para cargar el fichero a utilizar como conjunto de prueba. Al presionar tal botón aparecerá la ventana emergente de la figura 3.32 que mostramos seguidamente.

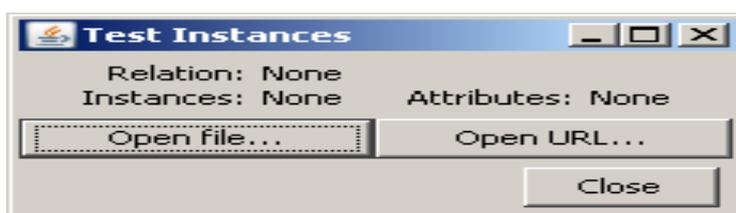


Figura 3.32: Después de pulsar el botón ‘Set...’ surge la ventana emergente titulada ‘Test Instances’.

En la ventana de la figura 3.32 titulada ‘Test Instances’ nos aparecen dos botones “Open file...” y “Open URL...” además, aparecen en la parte superior tres campos “Relation”, “Instances” y “Attributes” con valor “None” en cada uno de los tres. Este valor para estos tres campos nos viene a indicar que todavía no se ha abierto ningún fichero de prueba. Lo que vamos a hacer será pulsar sobre el botón “Open file...” que aparece a la izquierda, ya que deseamos abrir un fichero y no una “URL”. Al pulsar sobre el botón “Open file...” nos aparecerá el habitual cuadro de dialogo en donde seleccionar el fichero que deseamos que en este caso es el de las réplicas de una instancia de una misma vivienda. Tras seleccionar el fichero aparecerá otra vez la ventana emergente ‘Test Instances’ como muestra la figura 3.33.

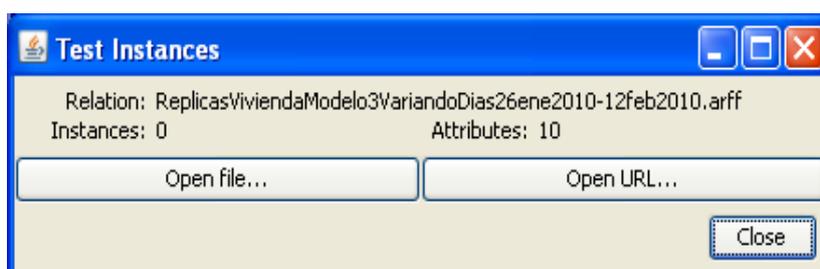


Figura 3.33: Muestra la ventana ‘Test Instances’ después de abrir el fichero de las réplicas de la vivienda.

Esta vez se observa que los campos “Relation”, “Instances” y “Attributes”, ya no aparecen asignados a “None”, debido a que, se ha abierto el fichero con las réplicas como conjunto de prueba. El campo “Relation” (relación) posee el nombre de la relación del fichero abierto como conjunto de prueba. El campo “Instances” está asignado al valor ‘0’, lo que viene a decir, que hay ‘0’ instancias para evaluar. Respecto al campo “Attributes”, a la derecha de la opción “Instances”, posee un valor igual a ‘10’ indicando que hay un total de ‘10’ atributos. Este valor debe coincidir con el número de

atributos de las viviendas disponibles para construir el modelo 3, que efectivamente también es ‘10’. Si no coincidiera el número total de atributos de las viviendas para construir el modelo y el número total de atributos de las réplicas del conjunto de prueba se produciría un error de compatibilidad y no se podría realizar la construcción del modelo.

Seguidamente cerraremos tal ventana ‘Test Instances’ y sólo nos faltará seleccionar el algoritmo “M5P” de la familia “tree”. Como dicho algoritmo es de la familia “tree” deberemos buscarlo en los algoritmos destinados a la familia “tree” tras pulsar sobre el botón ‘Choose’ en la sección ‘Classifier’. Luego pulsaremos sobre el botón “Start” para construir el modelo, con los parámetros de configuración del algoritmo por defecto como siempre en este proyecto. Una vez pulsamos sobre el botón “Start” observamos la información de ejecución en la sección “Run Information” y se aprecia que el método de evaluación en este caso es “Supplied test set” como hemos indicado en la sección “Test options” referente a los métodos de evaluación.

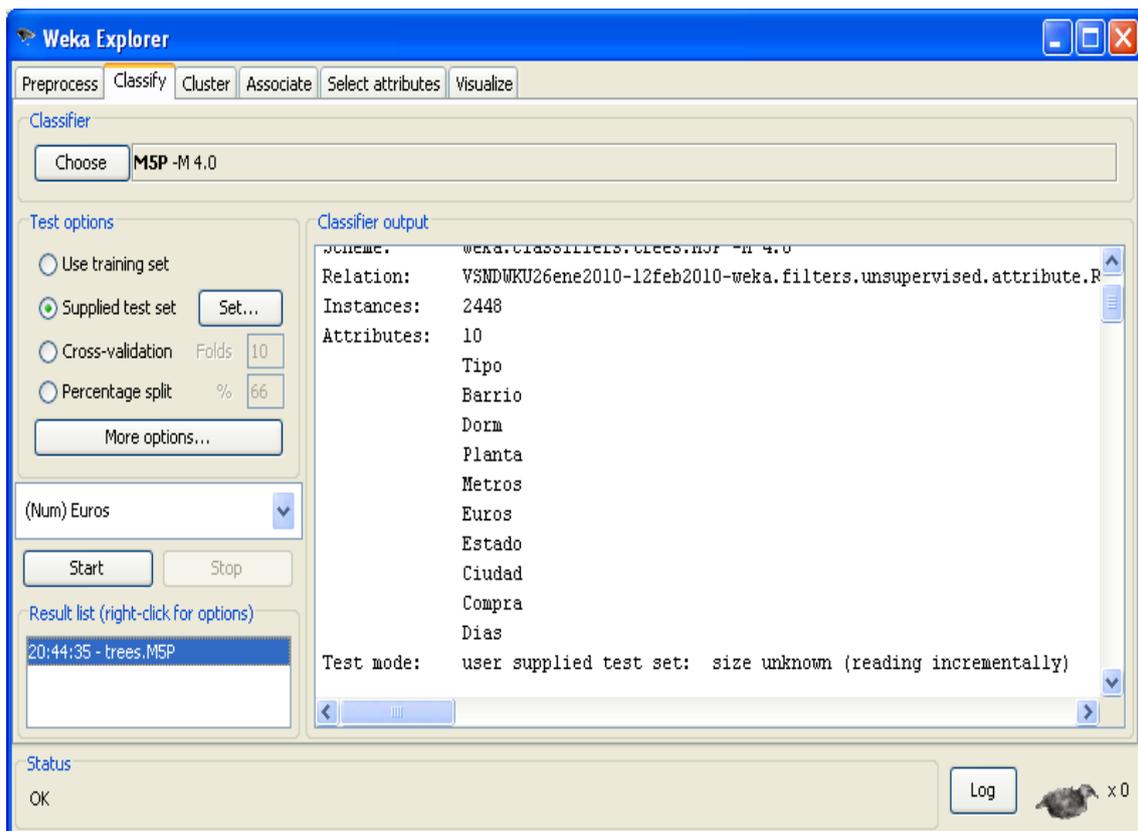


Figura 3.34: Se vuelve a construir el modelo 3 con el algoritmo “M5P” que fue el seleccionado en el estudio. Esta vez, no obstante, se utiliza “Supplied test set” como método de evaluación, ya que, se desea predecir el atributo “euros” a las réplicas que se pasen como conjunto de prueba, debido a que se desconocen.

En cuanto a la construcción del modelo decir que, cabe esperar que se cree la misma construcción que se observaba del algoritmo “M5P” en el estudio de algoritmos realizado. Esto se debe a que el modelo 3 se crea otra vez con el algoritmo “M5P” ejecutado con los mismos parámetros que son los de por defecto y que las instancias de viviendas disponibles de entrada son las mismas. La figura 3.35 muestra el principio de la construcción del modelo. La figura 3.36 muestran entre otra información la parte de la construcción en donde aparece el miembro referente al atributo “días” del ‘LM1’. Igualmente la figura 3.37 muestra además de otra información el miembro perteneciente al atributo “días” del ‘LM2’ y la parte final de la construcción.

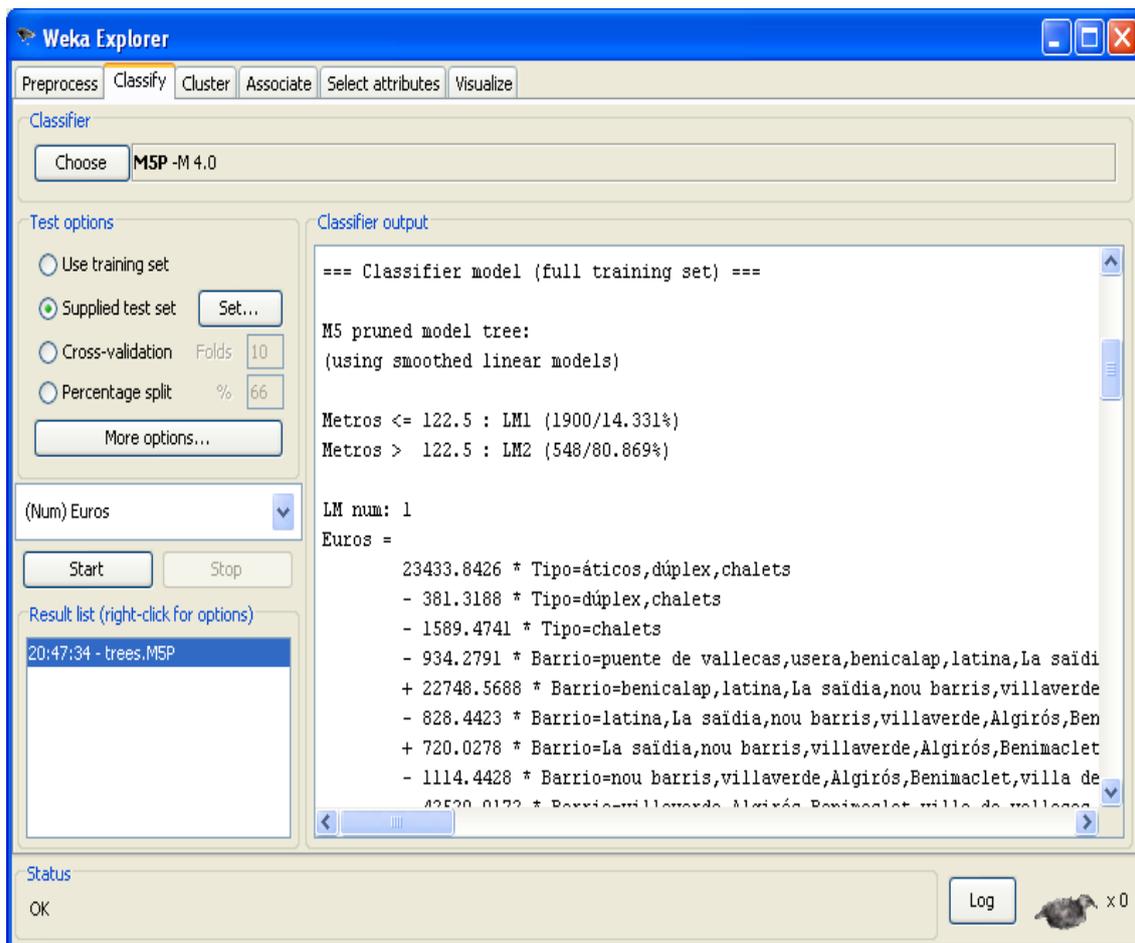


Figura 3.35: Al volverse a construir el modelo 3 seleccionado con los mismos datos y con las mismas opciones salvo el método de evaluación que es distinto se observa que la parte inicial del modelo es la misma.

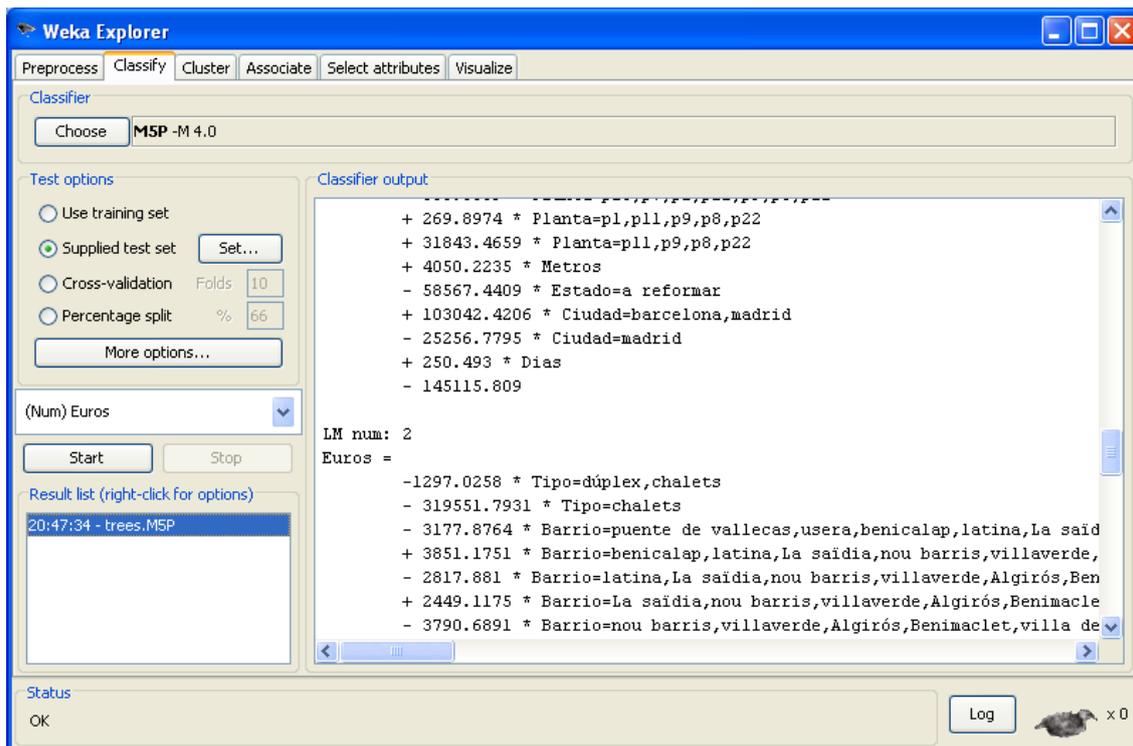


Figura 3.36: Se muestra entre otra información la parte del 'LM1' perteneciente al atributo "días", que aparece exactamente igual que en la construcción seleccionada en el estudio, debido a que, precisamente, es la misma construcción.

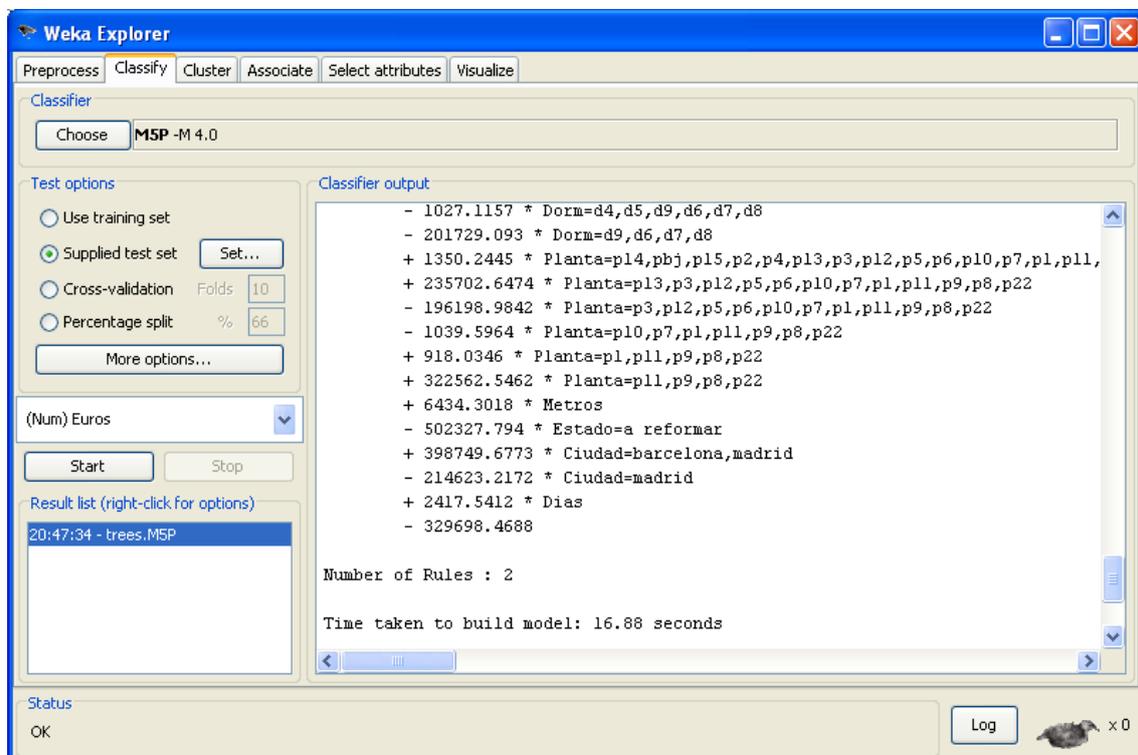


Figura 3.37: Se muestra entre otra información la parte del 'LM2' perteneciente al atributo "días", que aparece exactamente igual que en la construcción seleccionada en el estudio, debido a que, precisamente, es la misma construcción.

Sin embargo, la evaluación si será diferente, ya que, se ha escogido otro método de evaluación y además el conjunto de prueba es distinto.

Seguidamente mostramos una figura referida a la evaluación del modelo, que en este caso, como la finalidad no es probar el modelo, sino predecir el atributo “euros” de las réplicas, sí que la vamos a mostrar en este apartado.

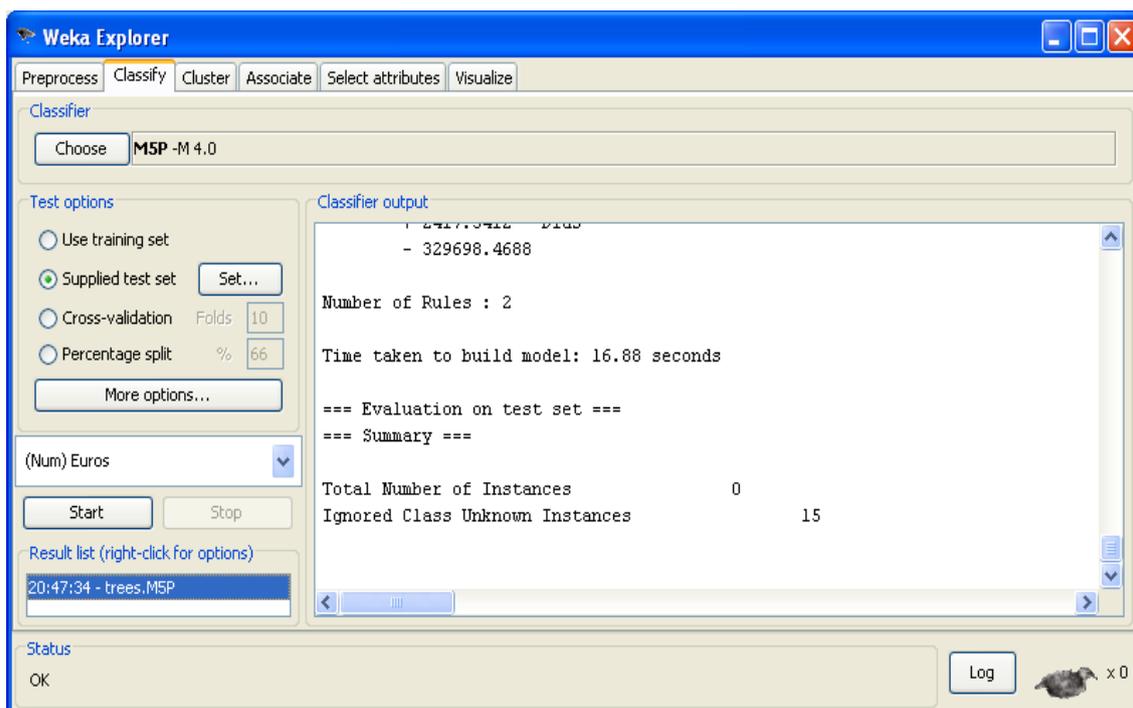


Figura 3.38: Resultado de la ejecución del algoritmo “M5P” con la opción “Supplied test set”, conjunto de prueba suministrado.

Aunque la figura 3.38 forma parte de la evaluación como hemos dicho, la intención no es evaluar el modelo que no se puede, sino predecir el atributo “euros”. Precisamente, la información que se nos ofrece de la evaluación nos refleja que no se ha podido probar el modelo. Esto se observa en el campo “Total Number of Instances” (numero total de instancias) y en el campo “Ignored Class UnKnown Instances” (Instancias de clase desconocida ignoradas.) El campo “Total Number of Instances” está igualado a “0” ya que no aparece el valor de la clase atributo “euros” en ninguna de las réplicas del conjunto de prueba. Asimismo la opción “Ignored Class UnKnown Instances” indica que aparecen ‘15’ instancias con la clase desconocida. Todo esto se resume diciendo que se ha predicho el atributo “euros” de las réplicas de la instancia de la vivienda, ya que, en la prueba se predice la clase, y que no se ha podido evaluar las réplicas de la instancia de la vivienda, porque, el valor real del atributo “euros” de cada réplica es desconocido. Esto, precisamente, es lo que deseábamos, es decir, obtener la predicción

del atributo “euros” en las réplicas y no probar el modelo, debido, a que ya lo probamos anteriormente en el estudio de algoritmos.

En las figuras 3.36 y 3.37 de la construcción se puede apreciar como afecta el atributo “días” al ‘LM1’ y al ‘LM2’, esto ya lo explicamos con anterioridad, cuando tratábamos la misma construcción del modelo 3, en la parte dedicada a detallar el modelo seleccionado. Lo primero a lo que debemos prestar atención es: ¿a cuál de los dos modelos lineales ‘LM1’ que se observa parte en la figura 3.35 y 3.36 y ‘LM2’ que se observa parte en la figura 3.37 ha ingresado cada uno de las réplicas? Y la respuesta, como seguramente ya se conoce al haber explicado con anterioridad la construcción, es que todas las réplicas ingresan en el ‘LM2’ ya que si el atributo “metros” es menor que ‘122,5 metros’ se utiliza el ‘LM1’ y si es mayor a ‘122,5 metros’ se utiliza el ‘LM2’. En este caso, como el atributo “metros” de las réplicas es mayor a ‘122,5 metros’, ya que las réplicas tienen el valor para el atributo metros igual a ‘158 metros’, ingresan todas en el ‘LM2’. Una vez sabemos que ingresan en el ‘LM2’ nos fijamos en la parte de la combinación lineal del ‘LM2’ referente al atributo “días”:

2417.5412 * Dias

Recuadro 3.23: Muestra el patrón del atributo “días” seguido en las réplicas.

El contenido del recuadro 3.23 ya lo analizamos anteriormente en este apartado, ya que, la construcción como hemos dicho es la misma que la construcción elegida en el estudio de algoritmos tratado. Como se observa y también dijimos con anterioridad cuando analizábamos la construcción elegida, ya que, es la misma, por cada día la predicción aumenta en ‘2417,5412 €’. Por tanto, viendo la diferencia entre los días de dos de nuestras réplicas se podría calcular en cuanto difieren respecto al atributo “euros”. No obstante, todo lo que hemos dicho, ha sido observando el modelo, por lo que, ahora queremos ver la predicción realizada por el modelo sobre las réplicas. Para lo cual, como la ejecución ya se ha terminado, vamos a buscar en las opciones de la instancia ejecutada, que aparece en la sección titulada ‘Result list’ de la ventana ‘Classify’, los resultados de la prueba del modelo. Buscaremos los resultados de la prueba del modelo, en la cual, como hemos dicho, no se habrá probado el modelo, aunque si se habrá realizado la predicción. Dicha sección ‘Result list’, en donde aparecen las opciones de la instancia ejecutada, como se sabe está en la zona baja y derecha de la ventana ‘Classify’. Para navegar sobre las opciones de la instancia en la sección ‘Result list’ hay que presionar con el botón derecho, sobre la instancia y nos aparecerá un listado o menú contextual en el que seleccionaremos la opción que buscábamos llamada ‘Visualize classifier errors’ (Errores del clasificador visualizados.)

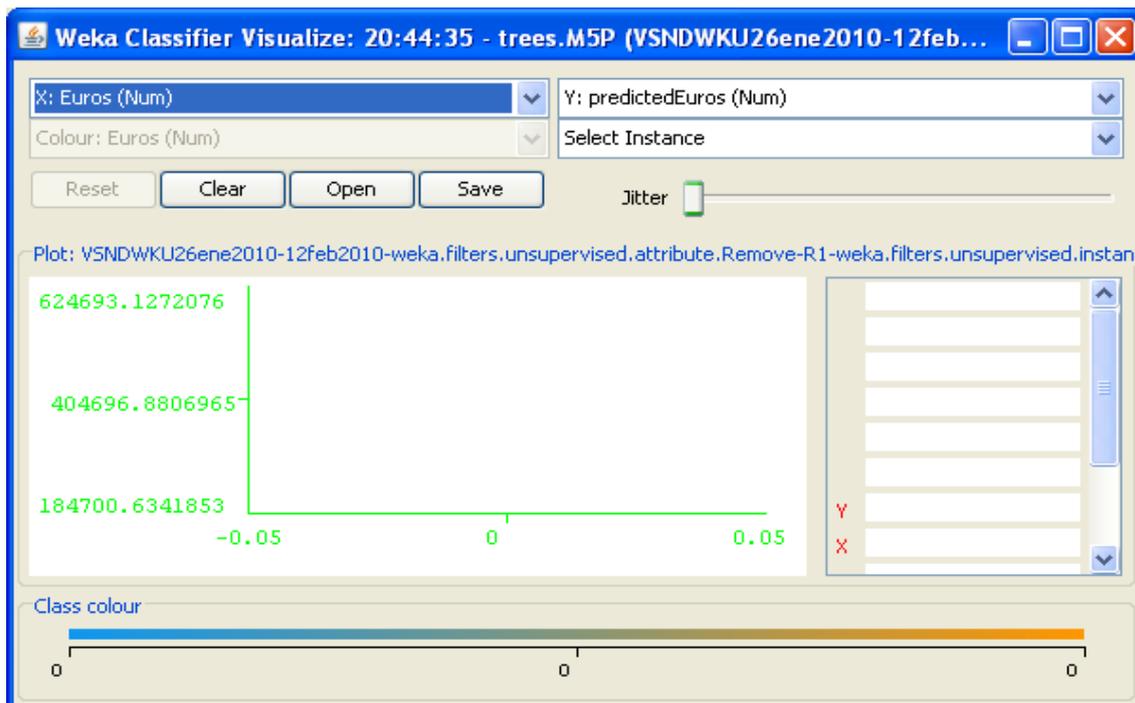


Figura 3.39: Ventana 'Classifier Visualize', en la cual no se observa gráfica alguna.

En la ventana de la figura 3.39, en el 'comboBox' de la parte superior izquierda aparece seleccionado el atributo "euros", que corresponde a la variable "x". Y en el 'comboBox' de la parte superior derecha aparece seleccionada la predicción "predictedEuros", el cual corresponde a la variable "y". Como se aprecia en principio no se observa nada, ya que, la gráfica desea mostrar el punto de comparación entre el valor real del atributo "euros" y la predicción del atributo "euros" de cada réplica, debido a que, se tiene seleccionado el atributo "euros" como variable independiente, y la predicción del atributo "euros" como dependiente. Y no se observa la representación de la gráfica, precisamente, porque el atributo "euros" es desconocido. Al no apreciarse nada en la gráfica se verifica que el modelo no ha podido ser probado. Lo que deseamos es ver la predicción, por lo que, seleccionando cualquiera de los atributos explicativos en el 'comboBox' de la izquierda correspondiente a la variable "x", se apreciará la predicción en función del atributo explicativo seleccionado. Entonces, como lo que deseamos es ver la predicción en función del atributo "días", seleccionaremos en el 'comboBox' de la variable "x" el atributo "días" con lo que nos aparecerá el resultado que queríamos ver en forma de gráfica, como se aprecia en la figura 3.40.

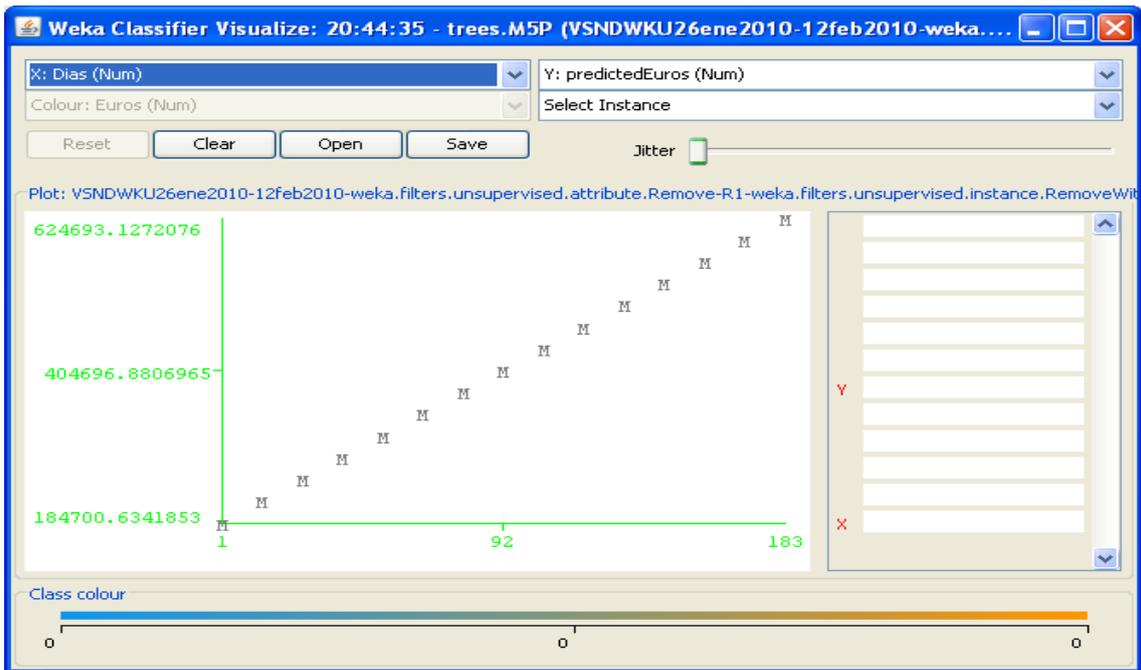


Figura 3.40: Esta vez se ha puesto como variable “x” el atributo “días” y ahora se observa la relación entre los días, y los euros que se han predicho.

Ahora guardaremos la información de la predicción, para lo cual, pulsaremos sobre el botón ‘Save’ que se observa en la parte central de la figura 3.40 al lado del botón “Open”. El fichero lo guardaremos con el nombre “PrediccionReplicasVivienda_Modelo3VariandoDias26ene2010-12feb2010.arff”. Como se observa guardamos el fichero con la extensión “.arff” con lo que podemos abrirlo en “WEKA”. En la figura 3.41 se muestra el fichero abierto en “WEKA”.

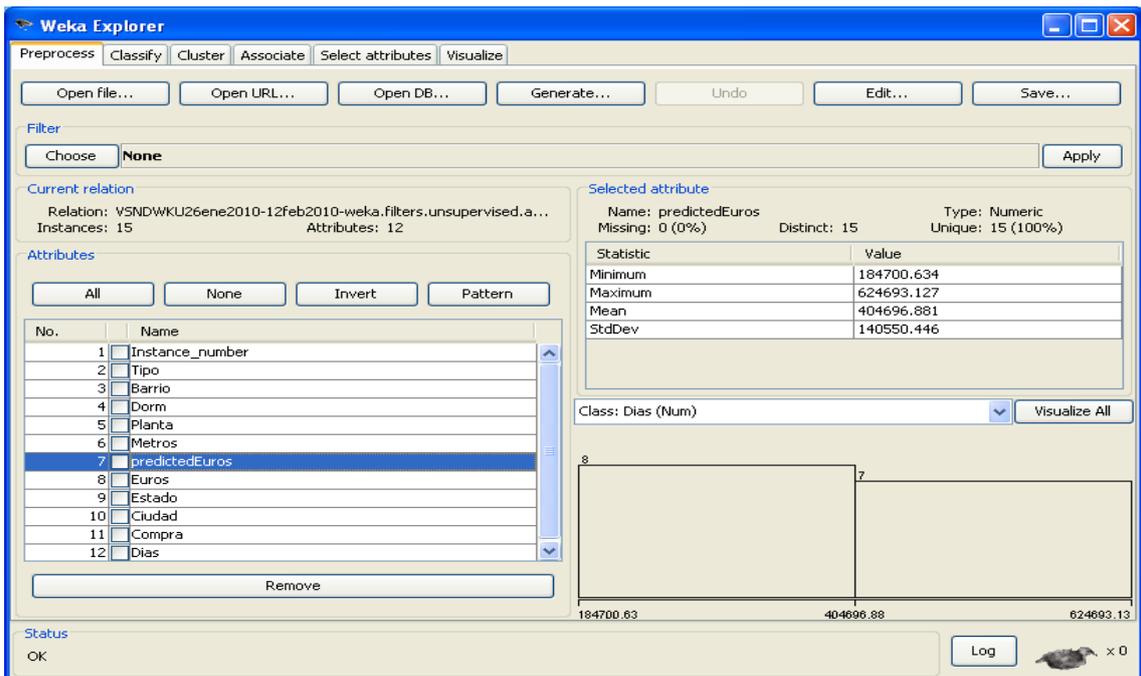


Figura 3.41: Se ha cargado el fichero de la predicción del atributo “euros” de las réplicas en “WEKA”.

Sin embargo, la información estadística que nos muestra “WEKA” no es lo que deseamos ver. Lo queremos apreciar son los valores de la predicción “predictedEuros” de cada una de las réplicas. Con lo que, observaremos el fichero de la predicción de las réplicas, por ejemplo, con cualquier editor que muestre el contenido de dicho fichero. La tabla 3.19, muestra tales réplicas. Este fichero con el valor predicho de los euros tiene de diferente que posee un nuevo atributo, que es la predicción del atributo “euros”.

Tipo	Barrio	Dorm	Planta	Metros	PrediccionEuros	...	Días
pisos	Extramurs	d5	p4	158	184700,634185	...	1
pisos	Extramurs	d5	p4	158	216128,669401	...	14
pisos	Extramurs	d5	p4	158	247556,704617	...	27
pisos	Extramurs	d5	p4	158	278984,739833	...	40
pisos	Extramurs	d5	p4	158	310412,775049	...	53
pisos	Extramurs	d5	p4	158	341840,810265	...	66
pisos	Extramurs	d5	p4	158	373268,845481	...	79
pisos	Extramurs	d5	p4	158	404696,880696	...	92
pisos	Extramurs	d5	p4	158	436124,915912	...	105
pisos	Extramurs	d5	p4	158	467552,951128	...	118
pisos	Extramurs	d5	p4	158	498980,986344	...	131
pisos	Extramurs	d5	p4	158	530409,02156	...	144
pisos	Extramurs	d5	p4	158	561837,056776	...	157
pisos	Extramurs	d5	p4	158	593265,091992	...	170
pisos	Extramurs	d5	p4	158	624693,127208	...	183

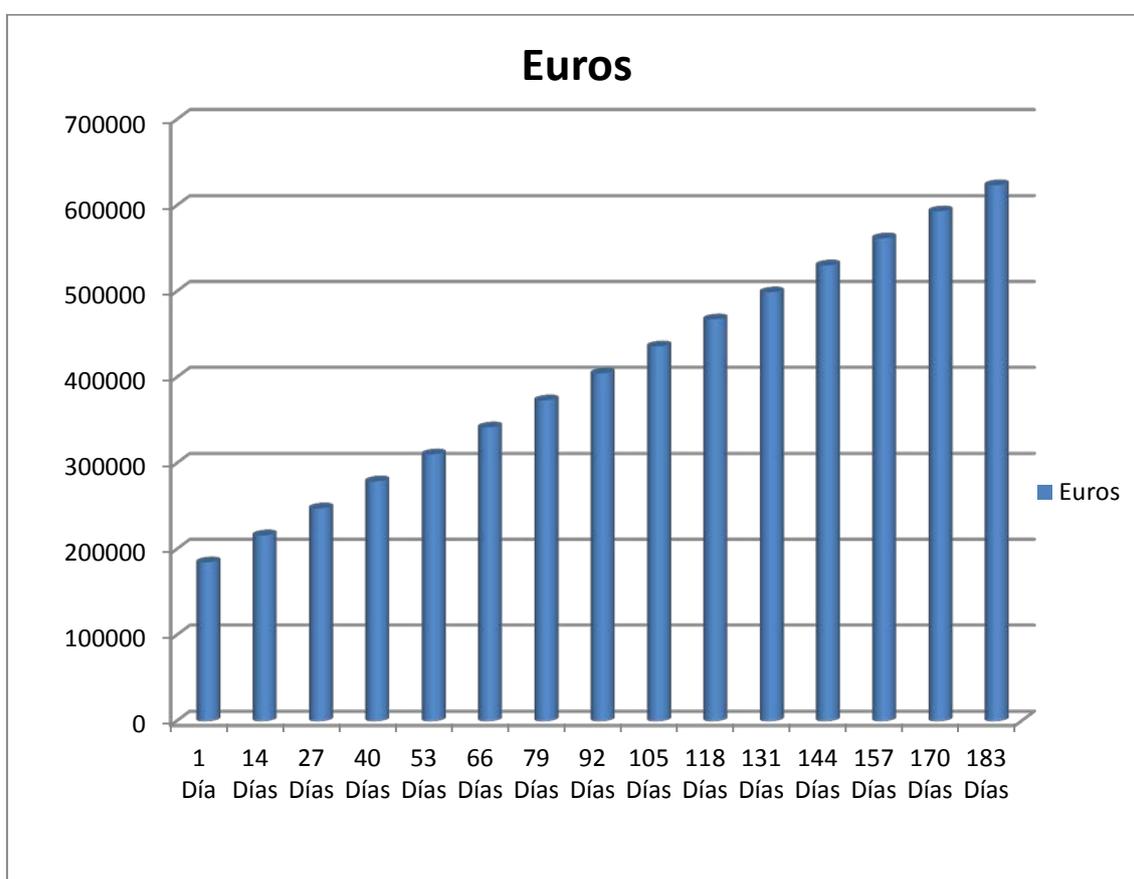
Tabla 3.19: “PrediccionReplicasViviendaModelo3VariandoDias26ene2010-12feb2010.arff”, se aprecian los valores del de la predicción de los euros.

A continuación, mostramos la tabla 3.20, con la variable independiente “x” que serán los días y la variable dependiente “y” que serán los euros predichos de las réplicas. Los valores de la tabla 3.20 se han extraído de la tabla 3.19, en concreto los valores de los días y los valores de la predicción de los euros. Los valores de la predicción de los euros de la tabla 3.20 han sido redondeados: exactamente ha habido redondeo en las unidades esto se ha hecho para que los valores queden más claros y más fáciles de rellenar de cara a la tabla 3.20 y a la gráfica 3.1.

X: (Días)	Y: (PredicciónEuros)
1	184701 €
14	216129 €
27	247557 €
40	278985 €
53	310413 €
66	341841 €
79	373269 €
92	404697 €
105	436125 €
118	467553 €
131	498981 €
144	530409 €
157	561837 €
170	593265 €
183	623693 €

Tabla 3.20: La predicción de los euros en función de los días de las réplicas.

Ahora mostraremos la gráfica 3.1, creada a partir de los datos de la tabla 3.20.



Gráfica 3.1: Euros en función de los días, se observa el aumento de la predicción de los euros de la vivienda con los días.

En la gráfica 3.1 se observa un aumento proporcional de la predicción del atributo “euros” en función del atributo “días”. Podemos colaborar la predicción de los euros en función de los días mostrando el patrón del atributo “días” en el ‘LM2’, que es en donde han ingresado las réplicas para predecirles el atributo “euros”.

2417.5412 * Días

Recuadro 3.24: Patrón de los días del ‘LM2’

Tal patrón que ya lo hemos explicado varias veces lo vamos a utilizar para verificar los resultados de la predicción de las réplicas en relación a la aportación del atributo “días”. Como se puede observar, en la gráfica 3.1, cada una de las réplicas tiene una diferencia de ‘13’ días respecto a la réplica siguiente, es decir, entre la primera y la segunda réplica hay ‘13’ días de diferencia, entre la segunda y la tercera también ‘13’ días y así sucesivamente. Por lo que, la diferencia entre una réplica y otra posterior es siempre constante en las ‘15’ réplicas. Siguiendo tal idea y en modo de verificación vamos a calcular la predicción de una vivienda posterior a otra, a partir de esa otra que es la anterior. El cálculo a realizar consistirá en:

1. Seleccionar el valor de los euros predichos de la vivienda anterior a una posterior.
2. Sustituir el atributo “días” por el valor “13”, en su miembro en el ‘LM2’ de la construcción seleccionada, que se observa entre otros recuadros en el recuadro 3.24 y multiplicarlo por su peso adyacente.
3. Sumar el valor de los euros del punto 1 con el del punto 2.

Al realizar dicha serie de operaciones quedará verificada la concordancia en cuanto a la predicción, en relación al atributo “días”, si obtenemos con estos cálculos la predicción de la réplica posterior a partir de la anterior. Como hay que escoger una de las réplicas para realizar la comparación a excepción de la última, seleccionamos la primera réplica. Por consiguiente, a partir de la predicción de los euros de la primera réplica tendremos que llegar a la predicción de los euros de la segunda réplica. Previamente, pondremos el atributo explicativo “días” y la predicción de los euros de la réplica primera y segunda.

Réplica	Días	Predicción euros
1	1	184700,634185
2	14	216128,669401

Tabla 3.21: El atributo “días” y la predicción de los euros de la réplica primera y segunda.

Ahora vamos a realizar el cálculo de llegar a la predicción de la réplica segunda a partir de la réplica primera con el patrón referente al atributo “días” utilizado en la predicción. Para lo cual vamos a seguir los pasos mostrados anteriormente:

1. Seleccionar el valor de la predicción de los euros en la réplica primera. Réplica primera predicción euros igual ‘184700,634185’ €
2. Sustituir el atributo “días” por el valor ‘13’ en su miembro perteneciente en el ‘LM2’ y multiplicarlo por el peso:
 - $f(\text{Días})=2417,5412 \times \text{Días}$; $f(13) = 2417,5412 \times 13 = 31428,0356$ €
3. Sumar el valor del punto 1 y el del punto 2.
 - Predicción euros réplica segunda = Punto 1 + Punto 2 =
 $= 184700,634185 + 31428,0356 = 216128,669785$ €

Por lo que se verifica, ya que, al observar el cálculo realizado vemos que llegamos a la predicción del atributo “euros” de la réplica segunda. Se puede apreciar que los tres últimos decimales del cálculo no coinciden, esto se debe a que “WEKA” nos muestra los resultados con algún tipo de redondeo o truncamiento en la parte decimal, ya que, es muy extensa. El valor de la predicción primera que nos ha ofrecido “WEKA” está aproximado como el resto de resultados de la predicción a seis decimales y entonces, al realizar el cálculo de hallar la predicción del atributo “euros” de la segunda réplica a partir de la predicción de la primera réplica aparecen algunas discrepancias que se han visto reflejadas en los tres últimos decimales que se ofrecen del cálculo.

Con la explicación de este último estudio relacionado con el atributo “días” y la predicción del atributo “euros” concluimos este apartado.

4. Evaluación

En este apartado llegamos a la fase de evaluación del proceso “KDD”. Concretamente, en este proyecto en la fase de evaluación nos centraremos en la parte del estudio de algoritmos de cada modelo referida a la prueba. En el apartado anterior, ya se dijo la construcción seleccionada para cada modelo, porque se querían explicar los patrones de

dichas construcciones seleccionadas. Y por tanto, ya sabemos que el algoritmo “RotationForest” es el que realiza la construcción seleccionada del modelo 1, y que el algoritmo “M5P” es el que realiza la construcción seleccionada del modelo 2, al igual que el algoritmo “M5P” es el que realiza la construcción seleccionada del modelo 3. No obstante, en este apartado vamos a dar resultados de porque se escogieron tales construcciones al valorarlas al lado de las otras construcciones no seleccionadas en cada modelo.

Antes de comentar la evaluación de cada una de las construcciones de cada modelo vamos a volver a citar, al igual que en los apartados anteriores, el enunciado de los tres modelos:

1. Modelo 1: Hallar la compra o la no compra de una vivienda.
2. Modelo 2: Precio por el que vender una vivienda, independientemente de los días.
3. Modelo 3: Precio por el que vender una vivienda, teniendo en cuenta los días.

Una vez se tengan los resultados de las pruebas de las construcciones realizadas en cada uno de los tres modelos, hay que considerar un criterio de evaluación, para poder comparar las construcciones de un mismo modelo, y decidir cual es la mejor, para así seleccionarla. Ese criterio de evaluación que utilizaremos para discernir será la precisión. Otros criterios que no utilizaremos podrían ser, por ejemplo, la claridad de los patrones o la utilización de un algoritmo de una familia determinada.

Ahora que sabemos que criterio se tiene en cuenta para la evaluación vamos a decir que: en el modelo 1 en que se realiza una tarea de clasificación, la forma habitual de medir la precisión es mediante los aciertos en la clasificación obtenidos en la prueba del modelo. Precisamente, por ello, el “porcentaje de aciertos” es la forma de medir dicho criterio de evaluación en el modelo 1 y en cualquier tarea de clasificación.

El “porcentaje de aciertos” se define como “el número de ejemplares correctamente clasificados partido el número total de ejemplares clasificados y multiplicado por cien.

$$\frac{\textit{Aciertos en la clasificación}}{\textit{Total clasificados}} \times 100$$

Por el contrario, en el modelo 2 y en el modelo 3 que se realiza una tarea de regresión, la forma habitual de medirla es mediante lo aproximado que esté el valor predicho de la clase respecto al valor real de ella en la prueba. Esto se consigue con el cálculo de “la raíz del error cuadrático medio” y del “error absoluto medio”. El “error absoluto

medio” y la "raíz del error cuadrático medio” son muy conocidos por el nombre que adoptan en el idioma inglés el cual es “mean absolute error” y “root mean squared error” respectivamente.

El “error absoluto medio” se puede definir como “la media calculada a los ejemplares de la diferencia en valor absoluto entre el valor que se tiene real y el valor predicho en cada ejemplar.

$$\frac{1}{N} \times \sum_{t=1}^N |x_t - \hat{x}_t|$$

En donde x_t sin los dos puntos superiores es el valor real y \hat{x}_t el valor predicho.

Y la “raíz del error cuadrático medio”, se puede definir como “la raíz de la media calculada a los ejemplares de la diferencia elevada al cuadrado entre el valor que se tiene real y el valor predicho en cada ejemplar.

$$\sqrt{\frac{1}{N} \times \sum_{t=1}^N (x_t - \hat{x}_t)^2}$$

En donde x_t sin los dos puntos superiores es el valor real y \hat{x}_t el valor predicho.

Después de citar el criterio a tener en cuenta, hay explicar el método de validación utilizado para probar las construcciones realizadas en el estudio de algoritmos de cada modelo. El método de validación que se utiliza en el estudio de algoritmos, como se dijo en el apartado anterior, es el método de “validación cruzada 10 X 10”. En este apartado centrado en la evaluación vamos a explicar dicho método, de “validación cruzada 10 X 10”:

- El método de “validación cruzada 10 X 10” consiste en dividir los datos disponibles en diez grupos y realizar la prueba de cada uno de ellos a partir de la construcción del modelo con los datos del resto de los grupos. Una vez se hayan probado todos, como conjunto de prueba, exactamente, una vez, se promediarán los resultados de las evaluaciones obtenidas. En cuanto al número de grupos, decir, que en este caso son un total de diez, debido a la expresión “10 X 10”. El tamaño de estos grupos de datos debe ser similar. Respecto a la proporción de los diferentes valores de la clase en cada uno de ellos: tiene que ser similar al conjunto completo de los datos sin segmentar si la validación cruzada es estratificada, “stratified”. La estratificación se utiliza en el método de validación cruzada en los modelos de clasificación.

En el método de “validación cruzada” que utilizamos en los modelos, como hemos dicho, se utilizan diez grupos o pliegues (10 X 10) pero se podrían utilizar cualquier otro número de ellos que no sea inferior lógicamente a dos. Uno se podría preguntar: ¿por qué escoger diez y por qué la aplicación ‘Explorer’ de “WEKA” adopta por defecto la “validación cruzada?” La cuestión es: porque parece ser que numerosos estudios han comprobado que la forma más eficaz de utilizar el “método de validación cruzada”, debido a que, se obtienen mejores resultados, es cuando se dividen los datos disponibles en diez partes.

La otra cuestión se puede explicar en que el método de “validación cruzada” es el más usual de utilizar en la aplicación “WEKA”, ya que, los datos disponibles se utilizan para el conjunto de entrenamiento y también para el conjunto de prueba sin sobrestimar la prueba. Ya que, la prueba no se realiza sobre el conjunto total de datos de entrada con los que se ha creado el modelo, sino que como hemos dicho para realizar la prueba divide los datos y prueba cada pliegue o grupo de datos con la construcción del modelo con el resto de pliegues o grupos, con lo que no se mezcla la construcción y la prueba del modelo, y por tanto, no hay sobrestimación en la prueba del modelo.

Después de haber explicado el método utilizado de validación cruzada “cross validation”, hay que comentar que la aplicación ‘Explorer’ de “WEKA”, en la ventana ‘Classify’ nos ofrece otros tres métodos de evaluación. Como se puede observar en la sección ‘test options’ de la figura 4.1, que muestra la ventana ‘Classify’, o en cualquier otra figura de la memoria que muestre la ventana ‘Classify’.

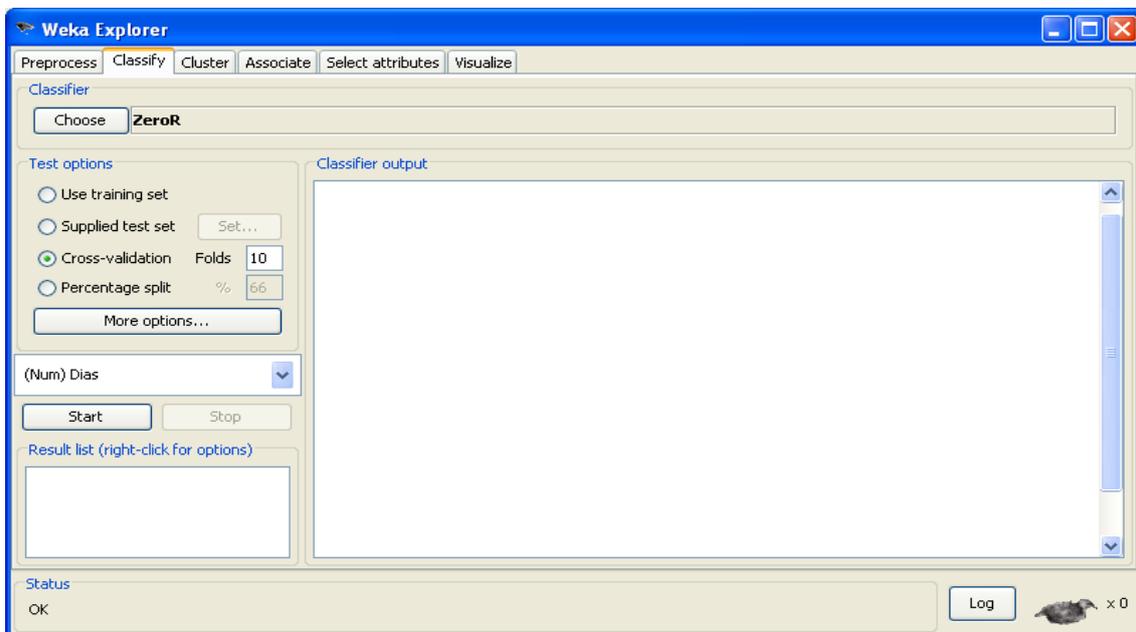


Figura 4.1: Se observa en la sección ‘Test options’ además del método de evaluación validación cruzada “Cross-validation”, los métodos de validación, “Use training set”, “Supplied test set” y “Percentage Split”.

Antes de comentar los tres métodos de evaluación restantes, decir que, “WEKA” para el método de validación cruzada posee un ‘textBox’, “Folds” (pliegues) que es en donde se indica el número de grupos en que dividir los datos de entrada, que como decíamos al referirnos a la expresión “10 X 10” debe tener y tiene un valor de ‘10’.

Respecto a los tres restantes métodos de evaluación, decir que, el “Supply test set” como ya se sabe lo utilizamos en el estudio exclusivo del modelo 3 referente: a la predicción del atributo “euros” de las réplicas de una vivienda. Dicho método “Supply test set”, se utiliza para poseer de un conjunto de prueba distinto a las viviendas utilizadas para construir el modelo.

En cuanto a “Use training set” (uso del conjunto de entrenamiento) decir que, se sirve para probar el modelo de los datos utilizados para construir el modelo. Por consiguiente, al utilizar los mismos datos que con los que se ha construido el modelo se sobrestima la evaluación, y por lo tanto, es un método de evaluación no conveniente de utilizar. Este método “Use training set” sólo tiene en común con el método de validación cruzada que se reutilizan los datos destinados a la construcción del modelo. Ya que, como comentamos en una de las cuestiones que nos hicimos con anterioridad, el método de “validación cruzada” no sobrestima los resultados de la prueba del modelo. Además, el método de “validación cruzada” tarda bastante más tiempo en ejecutarse.

Por último, queda el método de evaluación “Percentage Split” (partición en porcentaje) este método separa los datos de entrada disponibles en dos partes: una para construir el modelo y otra para probar el modelo. El porcentaje que se introduce en el ‘textBox’ adyacente a tal método se utiliza como el porcentaje de los datos disponibles de entrada que se utilizarán para construir el modelo y el resto como el porcentaje de los datos que se reservarán para la prueba del modelo. Por defecto, dicho método de evaluación tiene asignado un valor de ‘66 %’ en su ‘textBox’ adyacente. Por lo que, el ‘66 %’ de los datos se utilizarían para construir el modelo y el resto que sería el ‘34 %’ para probarlo.

Una cuestión que hay que tener en cuenta es que cada vez que se utiliza este método de evaluación, “WEKA” permuta los datos de entrada, con lo que los resultados de dos ejecuciones, de un modelo determinado con los mismos datos de entrada, con el mismo algoritmo y parámetros de configuración del algoritmo, serían diferentes. No obstante, si tenemos marcada la opción “preserve order for percentage Split” dicha variación no ocurriría, ya que, si se tiene marcada tal opción no permutarían los datos de entrada. Esta peculiaridad de la aleatoriedad de los datos de entrada sólo se da si tenemos asignado como método de evaluación “Percentage Split” y no tenemos marcada la opción comentada anteriormente “preserve order for percentage Split” que encontraremos al presionar el botón “More options” presente debajo de los métodos de evaluación. En este proyecto, como utilizamos el método de “validación cruzada” para probar los modelos no nos aparece dicha permutación en los datos de entrada, por lo que, como explicamos en el apartado “3. Aprendizaje”, una misma ejecución de un modelo, con los mismos datos de entrada, con el mismo algoritmo y con los mismos

parámetros de configuración del algoritmo nos proporcionará resultados en la ejecución del modelo similares.

Después de explicar los detalles respecto a los métodos de evaluación mostraremos la evaluación de cada una de las construcciones del estudio de algoritmos de los modelos. Primero trataremos la evaluación del estudio de algoritmos del modelo 1, después la del estudio de algoritmos del modelo 2 y después la del estudio de algoritmos del modelo 3. Además como el modelo 2 y el modelo 3 predicen la misma clase, que es el atributo “euros”, compararemos la evaluación del modelo 2 y la del modelo 3.

4.1. La evaluación del modelo 1

Seguidamente vamos a tratar la evaluación del estudio del “modelo 1: Hallar la compra o la no compra de una vivienda”. Para lo cual, mostraremos en primer lugar la figura 4.2, que muestra la parte de la evaluación del primer algoritmo del estudio del modelo 1, “BayesNet”.

La figura 4.2 muestra los resultados de la evaluación de la construcción del modelo 1 realizada con el algoritmo “BayesNet”, que es el primer algoritmo del estudio del modelo 1.

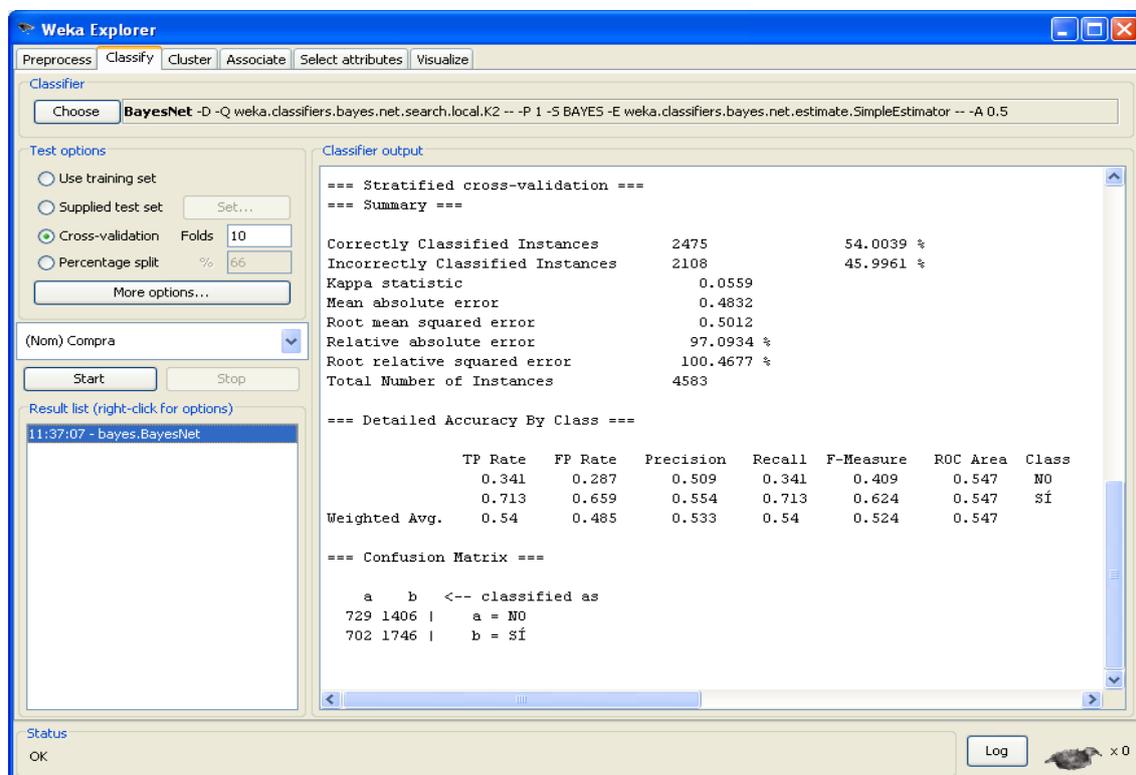


Figura 4.2: Resultados de la evaluación de la construcción del modelo 1 con el algoritmo “BayesNet”, que es el algoritmo primero del estudio del modelo 1.

En la figura 4.2 se observan las instancias de las viviendas correctamente clasificadas, “Correctly Classified Instances” y las instancias incorrectamente clasificadas, “Incorrectly Classified Instances” de la prueba del modelo 1 con el algoritmo “BayesNet”. Las instancias de las viviendas clasificadas correctamente son ‘2475’ y el porcentaje de acierto es ‘54,0039 %’ como pondremos en la tabla de la evaluación del modelo 1 en la fila correspondiente al algoritmo “BayesNet”. Las viviendas incorrectamente clasificadas, por tanto, serán el resto de las viviendas, es decir, ‘2108’, que hacen que resulte un porcentaje de ‘45,9961 %’ de instancias viviendas incorrectamente clasificadas. Los errores en la clasificación se pueden observar en la matriz de confusión “Confusion Matrix”. En tal matriz que se observa en la última parte de la salida, se aprecia que ‘1406’ instancia de viviendas de la clase ‘NO’ han sido clasificadas como ‘SÍ’ y ‘702’ instancia de viviendas de la clase ‘SÍ’ han sido clasificadas como ‘NO’. Otra medida que se observa después de las instancias incorrectamente clasificadas es el estadístico Kappa, “Kappa Statistic”, que mide la concordancia que aparece entre la evaluación de los ‘10’ segmentos de datos. El valor que nos indica el estadístico kappa ‘0,0559’ viene a decir que hay bastante poca concordancia entre la evaluación de los segmentos, ya que, un valor por debajo de ‘0,20’ en el estadístico Kappa indica poca concordancia. En el resto de medidas no nos vamos a fijar en la clasificación, ya que, las medidas como: el error absoluto medio “mean absolute error” son más clarificadoras en tareas numéricas.

Después viene la evaluación del resto de las construcciones realizadas del modelo 1 en el estudio de algoritmos. La última evaluación del estudio será la realizada a la construcción del modelo 2 hecha por el último algoritmo del estudio, “SimpleCLI”.

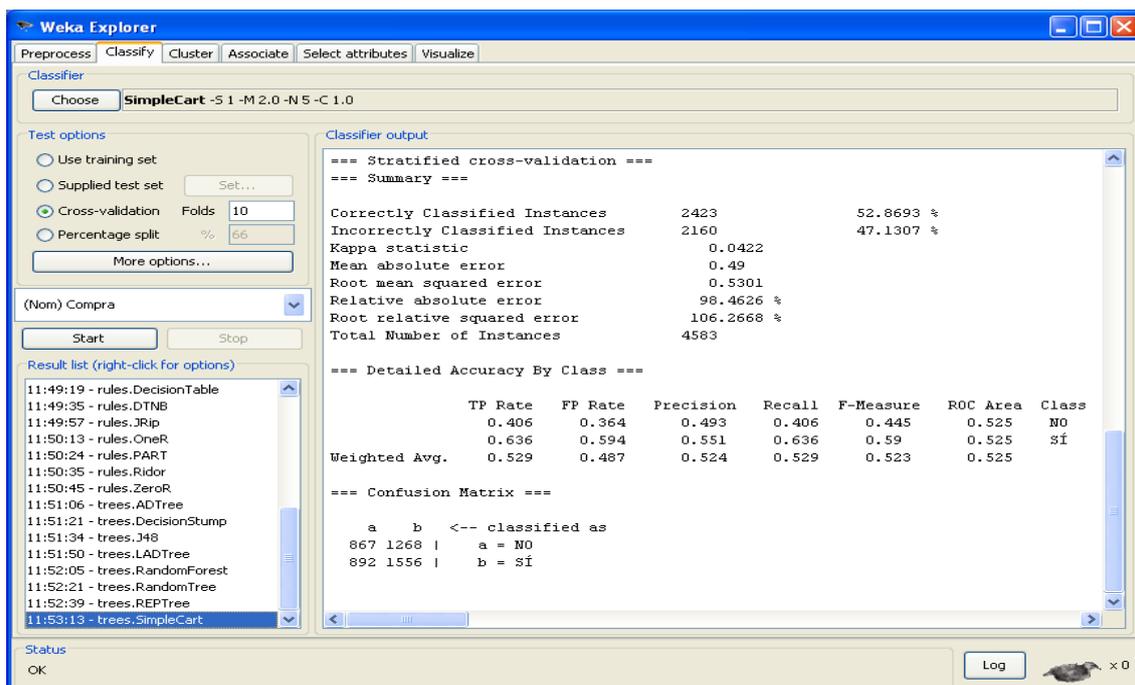


Figura 4.3: Resultados de la evaluación de la construcción del modelo 1 con el algoritmo “SimpleCLI”, que es el algoritmo último del estudio del modelo 1.

De los resultados que se observan de la evaluación de las construcciones del modelo 1 tendremos en cuenta como dijimos el “porcentaje de aciertos” del conjunto de prueba y el número de instancias correctamente clasificadas “Correctly Classified Instances” ya que, se trata de evaluar una clasificación.

Con lo que se seleccionará como algoritmo del modelo 1 el que mejor “porcentaje de aciertos” resulte tener entre todos los algoritmos del estudio del modelo 1 que, lógicamente, será también el que mayor número de instancias tenga correctamente clasificadas, ya que los dos conceptos están relacionados entre sí. Debido a que el número de instancias correctamente clasificadas forma parte del cálculo del “porcentaje de aciertos”.

A continuación, vamos a introducir la tabla 4.1, referida a la evaluación de las construcciones del modelo 1, realizadas en el estudio de algoritmos de dicho modelo. Para lo cual, en dicha tabla se informará para cada construcción realizada del estudio de algoritmos del modelo 1 de lo siguiente: del número que tiene asignada la construcción del modelo en relación al conjunto de construcciones realizadas del modelo, del nombre del algoritmo que ha realizado el modelo y de los resultados de la evaluación que se tienen en cuenta en la prueba de cada construcción del modelo. Dichos resultados de la evaluación que se tienen en cuenta como se sabe son el porcentaje de aciertos y las instancias correctamente clasificadas. Al respecto de la organización de la tabla 4.1, decir que como es bastante extensa se ha partido en trozos por familias de algoritmos.

Número	Algoritmo	Porcentaje de aciertos (%)	Instancias correctas
1	BayesNet	54,0039 %	2475
2	NaiveBayes	52,6075 %	2411
3	NaiveBayesSimple	52,6075 %	2411
4	NaiveBayesUpdateable	52,6075 %	2411

Tabla 4.1 primera parte: Se muestra de la prueba de cada modelo 1 realizado por cada algoritmo de la familia “Bayes”: el “porcentaje de aciertos” y el “número de instancias correctamente clasificadas”.

Número	Algoritmo	Porcentaje de aciertos (%)	Instancias correctas
5	Logistic	53,5239 %	2453
6	RBFNetwork	54,6803 %	2506
7	SimpleLogistic	56,6005 %	2594
8	SMO	55,3568 %	2537
9	VotedPerceptron	53,4148 %	2448

Tabla 4.1 continuación: Se muestra de la prueba de cada modelo 1 realizado por cada algoritmo de la familia “functions”: el “porcentaje de aciertos” y el “número de instancias correctamente clasificadas”.

Número	Algoritmo	Porcentaje de Aciertos (%)	Instancias correctas
10	IB1	53,2839 %	2442
11	IBK	53,2621 %	2441
12	Kstar	52,2583 %	2395
13	LWL	56,6005 %	2537

Tabla 4.1 continuación: Se muestra de la prueba de cada modelo 1 realizado por cada algoritmo de la familia “lazy”: el “porcentaje de aciertos” y el “número de instancias correctamente clasificadas”.

Número	Algoritmo	Porcentaje de Aciertos (%)	Instancias correctas
14	AdaBoostM1	56,6659 %	2597
15	AttributeSelectedClassifier	55,9677 %	2565
16	Bagging	52,1929 %	2392
17	ClassificationViaClustering	51,1892 %	2346
18	ClassificationViaRegression	53,5239 %	2453
19	CVParameterSelection	53,4148 %	2448
20	Dagging	53,6548 %	2459
21	Decorate	52,782 %	2419
22	END	55,9022 %	2562
23	EnsembleSelection	52,0183 %	2384
24	FilteredClassifier	55,8804 %	2561
25	Grading	53,4148 %	2448
26	LogitBoost	56,0332 %	2568
27	MultiBoostAB	56,6005 %	2594
28	MultiClassClassifier	53,5239 %	2453
29	MultiScheme	53,4148 %	2448
30	OrdinalClassClassifier	55,9022 %	2562
31	RacedIncrementalLogitBoost	56,6005 %	2594
32	RandomSubSpace	53,7203 %	2462
33	RotationForest	57,386 %	2630
34	Stacking	53,4148 %	2448
35	StackingC	53,4148 %	2448
36	ThresholdSelector	46,6288 %	2137
37	Vote	53,4148 %	2448

Tabla 4.1 continuación: Se muestra de la prueba de cada modelo 1 realizado por cada algoritmo de la familia “metaLearned”: el “porcentaje de aciertos” y el “número de instancias correctamente clasificadas”.

Número	Algoritmo	Aciertos	Instancias correctas
38	ClassBalancedND	55,9022 %	2562
39	DataNearBalancedND	55,9022 %	2562
40	ND	55,9022 %	2562

Tabla 4.1 continuación: Se muestra de la prueba de cada modelo 1 realizado por cada algoritmo de la familia “nestedDichotomies” de la familia “metaLearned”: el “porcentaje de aciertos” y el “número de instancias correctamente clasificadas”.

Número	Algoritmo	Aciertos	Instancias correctas
41	HyperPipes	46,5852 %	2135
42	VFI	53,2621 %	2441

Tabla 4.1 continuación: Se muestra de la prueba de cada modelo 1 realizado por cada algoritmo de la familia “miscellaneous”: el “porcentaje de aciertos” y el “número de instancias correctamente clasificadas”.

Número	Algoritmo	Aciertos	Instancias correctas
43	ConjunctiveRule	56,6005 %	2594
44	DecisionTable	55,4659 %	2542
45	DTNB	54,484 %	2497
46	JRip	56,295 %	2580
47	OneR	52,1274 %	2389
48	PART	52,6511 %	2413
49	Ridor	55,7495 %	2555
50	ZeroR	53,4148 %	2448

Tabla 4.1 continuación: Se muestra de la prueba de cada modelo 1 realizado por cada algoritmo de la familia “rules”: el “porcentaje de aciertos” y el “número de instancias correctamente clasificadas”.

Número	Algoritmo	Aciertos	Instancias correctas
51	ADTree	56,7314 %	2600
52	DecisionStump	56,6005 %	2594
53	J48	55,9022 %	2562
54	J48graft	55,8804 %	2561
55	LADTree	57,1023 %	2617
56	RandomForest	52,6511 %	2399
57	RandomTree	52,9784 %	2283
58	REPTree	53,4148 %	2448
59	SimpleCart	52,8693 %	2423

Tabla 4.1 finalización: Se muestra de la prueba de cada modelo 1 realizado por cada algoritmo de la familia “tree”: el “porcentaje de aciertos” y el “número de instancias correctamente clasificadas”.

Como se aprecia en la tabla 4.1 y se dijo en el apartado “3. Aprendizaje”, efectivamente, el algoritmo “RotationForest” de la familia “metaLearned” es el que presenta una mejor evaluación en la prueba de la construcción del modelo 1 que realiza. Ya que, la prueba de la construcción del modelo 1 que ha realizado dispone del mayor porcentaje de aciertos con ‘57,386 %’, y por consiguiente, del mayor número de instancias correctamente clasificadas ‘2630’ de un total de ‘4583’.

La figura 4.3 siguiente muestra la salida que nos ofrece “WEKA” de la prueba del modelo 1 que mejor evaluación ha conseguido realizada con el algoritmo “RotationForest”.

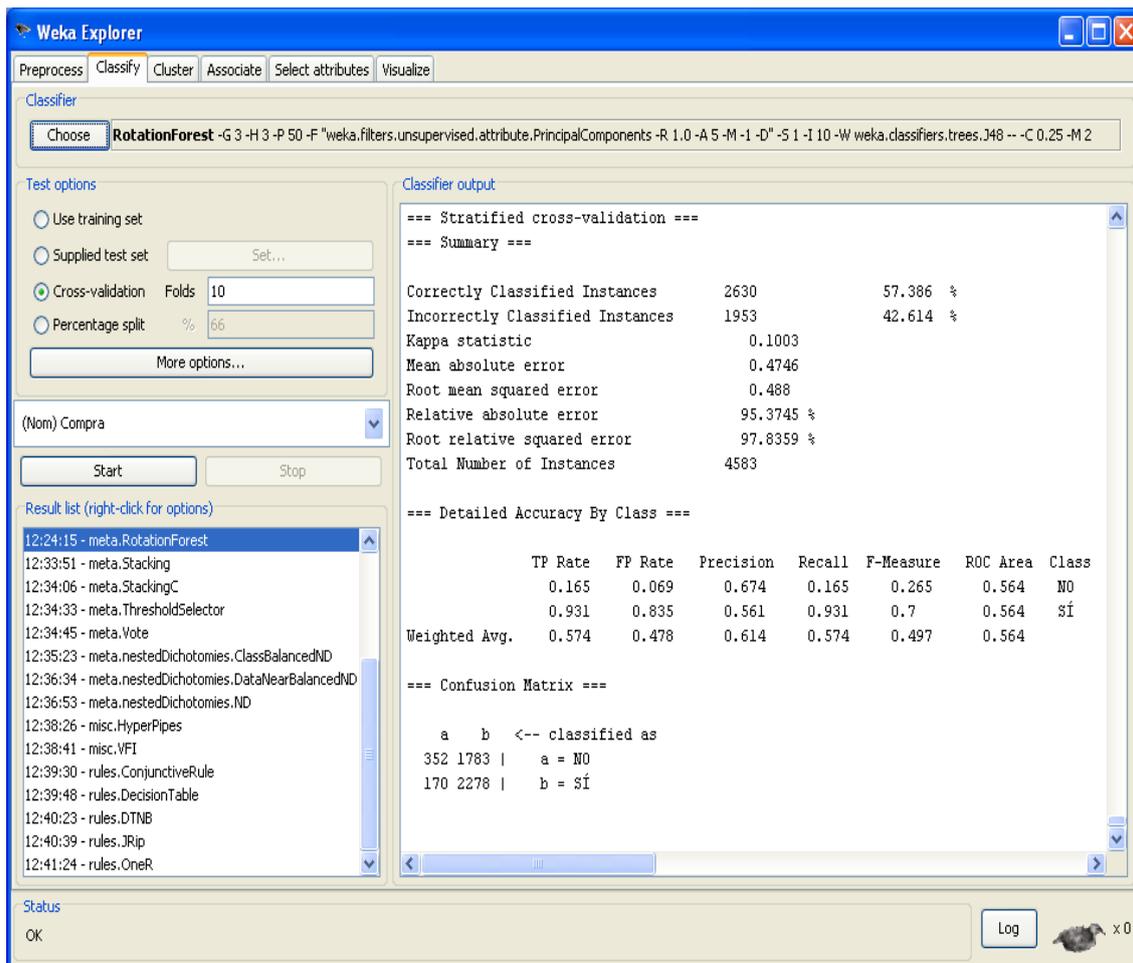


Figura 4.3: Se muestran los resultados de la evaluación del modelo 1 realizado con el algoritmo “RotationForest” que es el que mejor precisión presenta de entre todas las construcciones realizadas del modelo 1 en el estudio de algoritmos del modelo.

4.2. La evaluación del modelo 2

En referencia a la evaluación del “modelo 2: Precio por el que vender una vivienda, independientemente de los días”. Vamos a hacer referencia también a la parte del estudio de algoritmos del modelo 2 relativa a la evaluación. Para lo cual, mostraremos en primer lugar en la figura 4.4, los resultados de la evaluación de la construcción realizada del modelo 2 con el algoritmo “LeastMedSq”, que es el primer algoritmo del estudio de algoritmos del modelo 2.

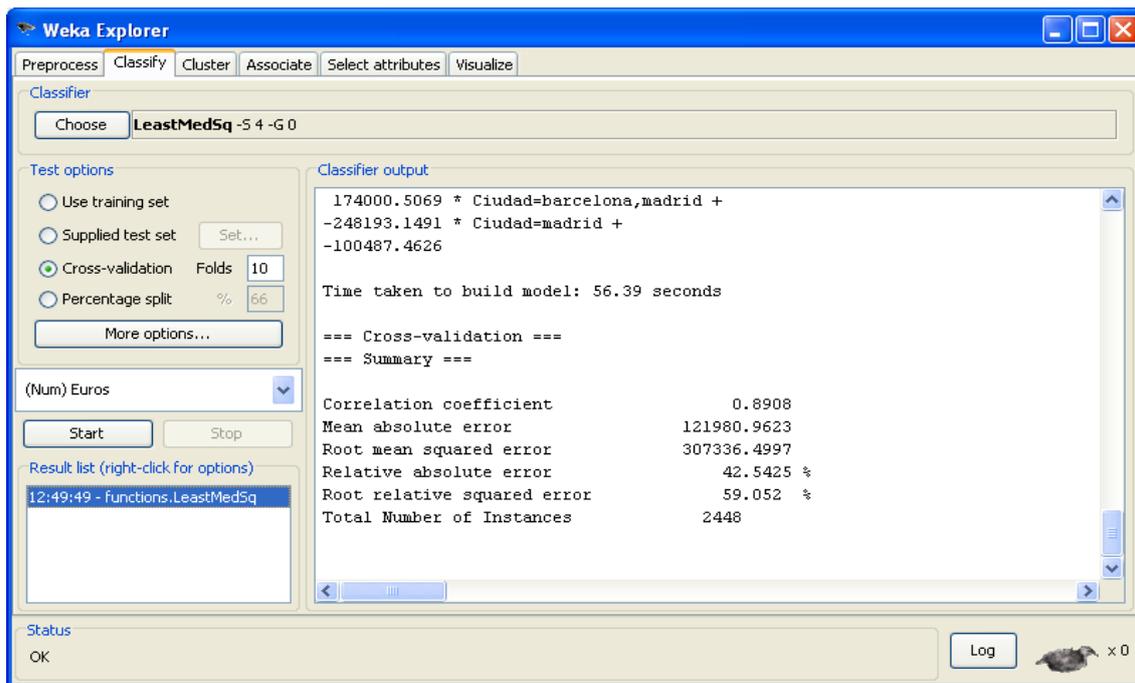


Figura 4.4: Resultados de la evaluación de la construcción del modelo 2 realizada con el algoritmo “LeastMedSq”, que es el algoritmo primero del estudio del modelo 2.

Luego seguiría el resto de las evaluaciones de las construcciones del modelo 2. En la figura 4.5 se muestra la evaluación de la construcción del modelo 2 realizada con el algoritmo “REPTree”, que es el último del estudio del modelo 2.

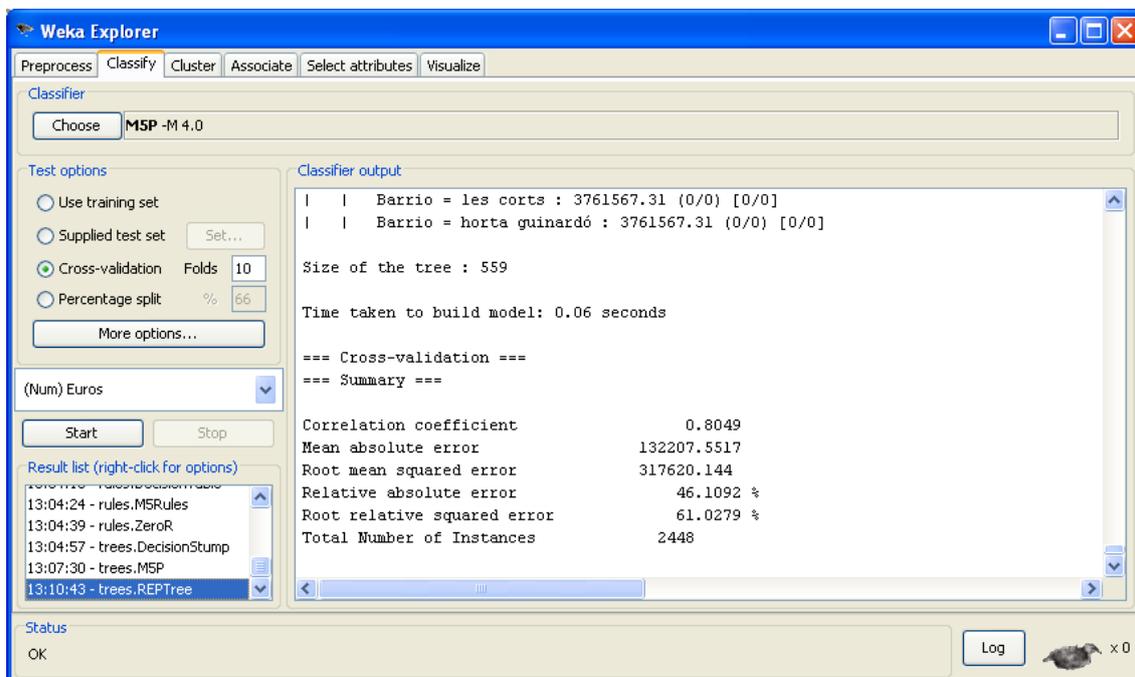


Figura 4.5: Resultados de la evaluación de la construcción del modelo 2 realizada con el “REPTree”, que es el algoritmo último del estudio del modelo 2.

Después de insertar la figura de la evaluación del primer algoritmo del estudio “LeastMedSq” y la del último algoritmo “REPTree”, en la tabla 4.2 se observa los resultados de la evaluación de las construcciones realizadas del modelo 2 en el estudio de algoritmos de dicho modelo 2. Para evaluar las construcciones realizadas tendremos en cuenta de los resultados de la evaluación, el “error absoluto medio” y la “raíz del error cuadrático medio”, porque como se ha mencionado anteriormente en este mismo apartado el criterio de evaluación utilizado es la precisión. Y la forma más común de medir la precisión de este modelo es mediante el cálculo del “error absoluto medio”, y de la “raíz del error cuadrático medio”, ya que, la tarea que se realiza en este modelo es una tarea predictiva de regresión. La prueba del modelo que presente más minimizados el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio” será el que mejor evaluación presente y por tanto, será el que seleccionemos. Asimismo el algoritmo que haya realizado dicha construcción será el más óptimo para construir el modelo 2. La tabla 4.2 que trata de la evaluación de los algoritmos utilizados en el estudio del modelo 2 se muestra a continuación dividida por familias de algoritmos.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
1	LeastMedSq	121980,9623 €	307336,4997 €
2	LiniarRegression	119076,1394 €	234759,8629 €
3	RBFNetwork	277272,9306 €	499484,5784 €

Tabla 4.2 primera parte: Se muestra de la prueba de cada modelo 2 realizado por cada algoritmo de la familia “functions”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
4	IBK	156848,3437 €	376427,9228 €
5	Kstar	116007,4403 €	301287,8088 €
6	LWL	184528,6733 €	351847,3863 €

Tabla 4.2 continuación: Se muestra de la prueba de cada modelo 2 realizado por cada algoritmo de la familia “lazy”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
7	AdditiveRegression	141304,0613 €	283620,1216 €
8	Bagging	114885,711 €	268972,4454 €
9	CVParameterSelection	286727,0572 €	520450,9607 €
10	MultiScheme	286727,0572 €	520450,9607 €
11	RandomSubSpace	139798,7466 €	309133,2464 €
12	RegressionByDiscretization	175497,6222 €	346073,6191 €
13	Stacking	286727,0572 €	520450,9607 €
14	Vote	286727,0572 €	520450,9607 €

Tabla 4.2 continuación: Se muestra de la prueba de cada modelo 2 realizado por cada algoritmo de la familia “metaLearned”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
15	ConjunctiveRule	189559,1058 €	364616,2461 €
16	DecisionTable	150441,4981 €	293923,6795 €
17	M5Rules	106421,5669 €	229286,2666 €
18	ZeroR	286727,0572 €	520450,9607 €

Tabla 4.2 continuación: Se muestra de la prueba de cada modelo 2 realizado por cada algoritmo de la familia “rules”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
19	DecisionStump	189706,1242 €	363083,7603 €
20	M5P	105980,2765 €	227315,5084 €
21	REPTree	132207,5517 €	317620,144 €

Tabla 4.2 finalización: Se muestra de la prueba de cada modelo 2 realizado por cada algoritmo de la familia “tree”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

El algoritmo “M5P” como se dijo en el apartado “3. Aprendizaje”, efectivamente, es el algoritmo que aporta la construcción del modelo 2 con una mayor precisión y por tanto una mejor evaluación. Ya que, los resultados de la prueba de la construcción realizada del modelo 2 con el algoritmo “M5P” es la presenta valores más minimizados en referencia al “error absoluto medio” con un total de ‘105980,2765 € y a la “raíz del error cuadrático medio” con un total de ‘227315,5084 € de entre las pruebas de las construcciones realizadas del modelo 2 en el estudio de algoritmos de dicho modelo.

La figurar 4.4 muestra los resultados de la salida de dicha mejor evaluación de la construcción del modelo 2 en el estudio de algoritmos del modelo 2 realizada por el algoritmo “M5P”.

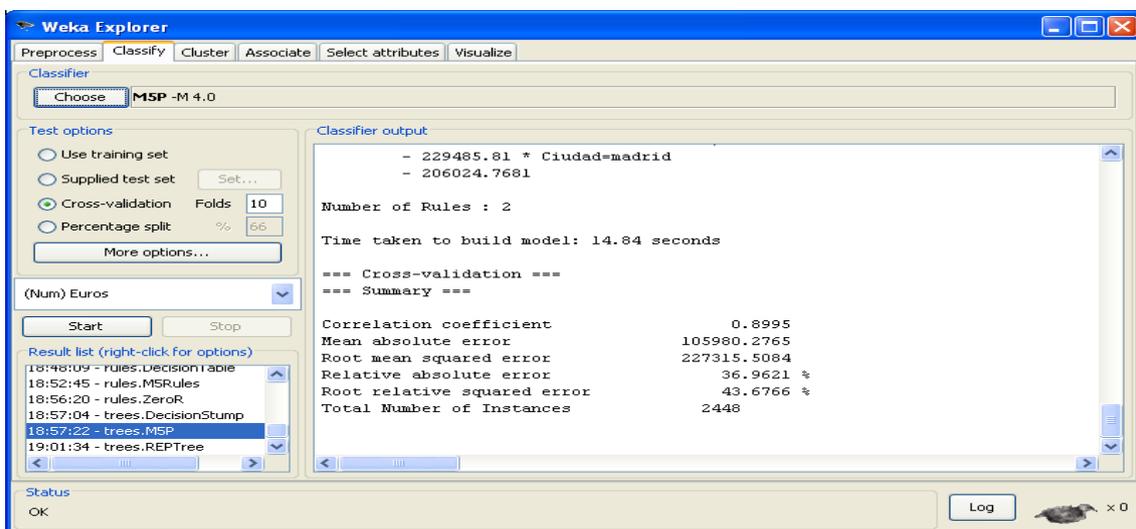


Figura 4.4: Resultados de la evaluación de la construcción elegida del modelo 2 realizada con el algoritmo “M5P”.

4.3. La evaluación del modelo 3

Ahora respecto a la evaluación del “modelo 3: Precio por el que vender una vivienda, teniendo en cuenta los días.” Vamos a evaluar las construcciones realizadas del modelo 3 en el estudio de algoritmos. Seguidamente introducimos la figura 4.5 referente a los resultados de la evaluación de la construcción realizada del modelo 3 con el algoritmo, “LeastMedSq”, que es el primer algoritmo del estudio de algoritmos del modelo 3.

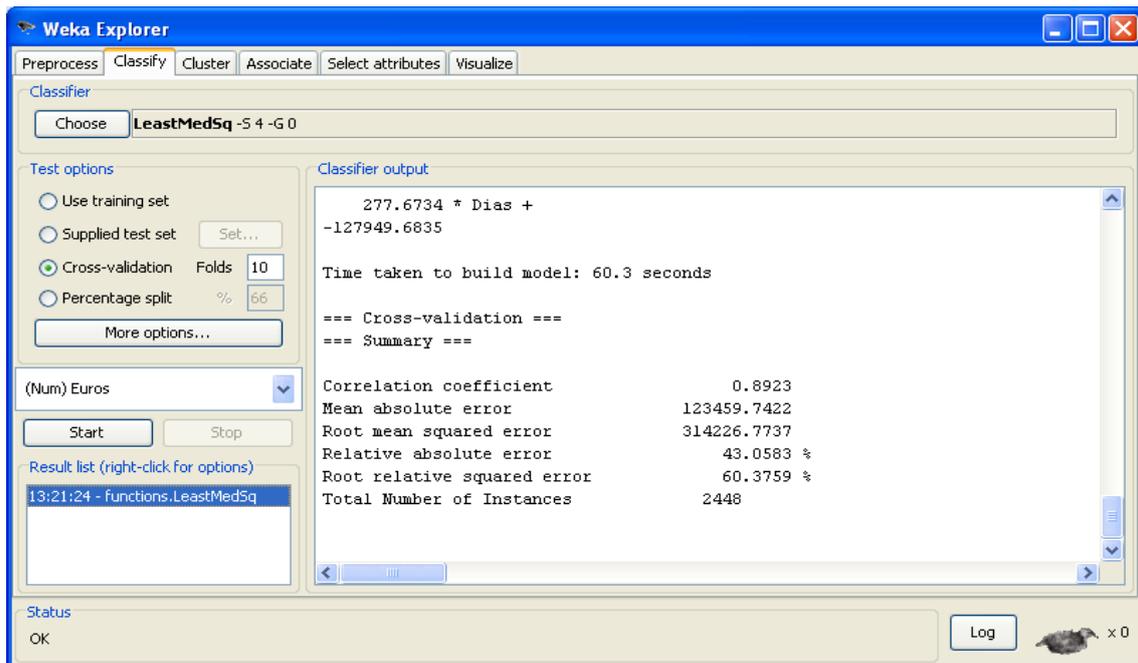


Figura 4.5: Resultados de la evaluación de la construcción del modelo 2 con el algoritmo “LeastMeSq”, que es el primer algoritmo del estudio del modelo 3.

Después seguiría el resto de las evaluaciones de las construcciones del modelo 3.

La figura 4.6 visualiza la evaluación de la construcción del modelo 3 con el último algoritmo del estudio.

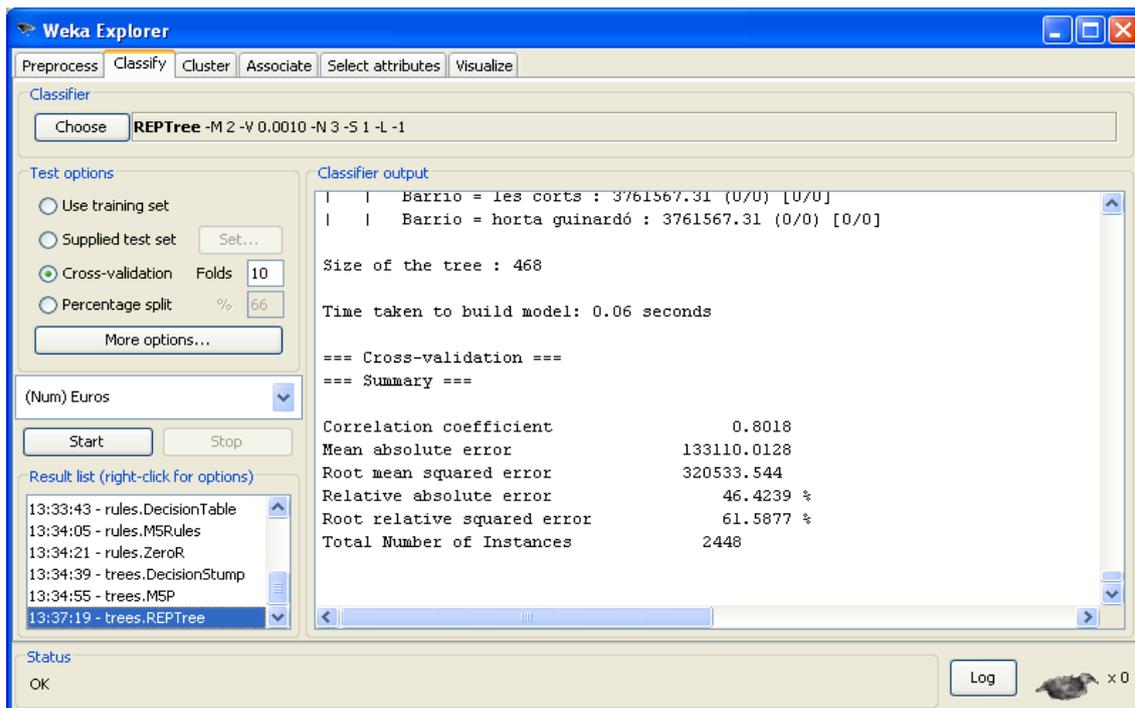


Figura 4.6: Resultados de la evaluación de la construcción del modelo 3 realizada con el algoritmo “REPTree”, que es el último algoritmo del estudio del modelo 3.

Seguidamente, vamos a introducir la tabla 4.3 que tendrá los resultados referidos al “error absoluto medio” y a la “raíz del error cuadrático medio” de la evaluación de cada una de las construcciones realizadas en el modelo 3.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
1	LeastMedSq	123459,7422 €	314226,7737 €
2	LiniarRegression	120243,4073 €	233693,6921 €
3	RBFNetwork	283988,813 €	509170,7174 €

Tabla 4.3 primera parte: Se muestra de la prueba de cada modelo 3 realizado por cada algoritmo de la familia “functions”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
4	IBK	187305,0233 €	411879,4116 €
5	Kstar	121723,9391 €	306388,2039 €
6	LWL	121723,9391 €	306388,2039 €

Tabla 4.3 continuación: : Se muestra de la prueba de cada modelo 3 realizado por cada algoritmo de la familia “lazy”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
7	AdditiveRegression	144263,4875 €	279063,7623 €
8	Bagging	114426,6501 €	267140,1866 €
9	CVParameterSelection	286727,0572 €	520450,9607 €
10	MultiScheme	286727,0572 €	520450,9607 €
11	RandomSubSpace	127952,4061 €	300063,4635 €
12	RegressionByDiscretization	178257,5922 €	369188,7921 €
13	Stacking	286727,0572 €	520450,9607 €
14	Vote	286727,0572 €	520450,9607 €

Tabla 4.3 continuación: : Se muestra de la prueba de cada modelo 3 realizado por cada algoritmo de la familia “metaLearned”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
15	ConjunctiveRule	189559,1058 €	364616,2461 €
16	DecisionTable	150441,4981 €	293923,6795 €
17	M5Rules	105549,8037 €	230655,1345 €
18	ZeroR	286727,0572 €	520450,9607 €

Tabla 4.3 continuación: Se muestra de la prueba de cada modelo 3 realizado por cada algoritmo de la familia “rules”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

Número	Algoritmo	Mean Absolute Error	Root mean squared error
19	DecisionStump	189706,1242 €	363083,7603 €
20	M5P	104808,0026 €	227188,9084 €
21	REPTree	133110,0128 €	320533,544 €

Tabla 4.3 finalización: Se muestra de la prueba de cada modelo 3 realizado por cada algoritmo de la familia “tree”: el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio”.

El algoritmo “M5P”, como ya se dijo en el apartado “3. Aprendizaje”, efectivamente, es el que mayor precisión presenta en el estudio del modelo 3, ya que, no hay ninguna construcción realiza en el modelo 3 aparte de la realiza con el algoritmo “M5P” que presenta un menor error absoluto medio con un valor de ‘104808,0026 € y una menor raíz del error cuadrático con un valor de ‘227188,9084 €.

En la figura 4.7 se observa, precisamente, la evaluación del algoritmo “M5P”.

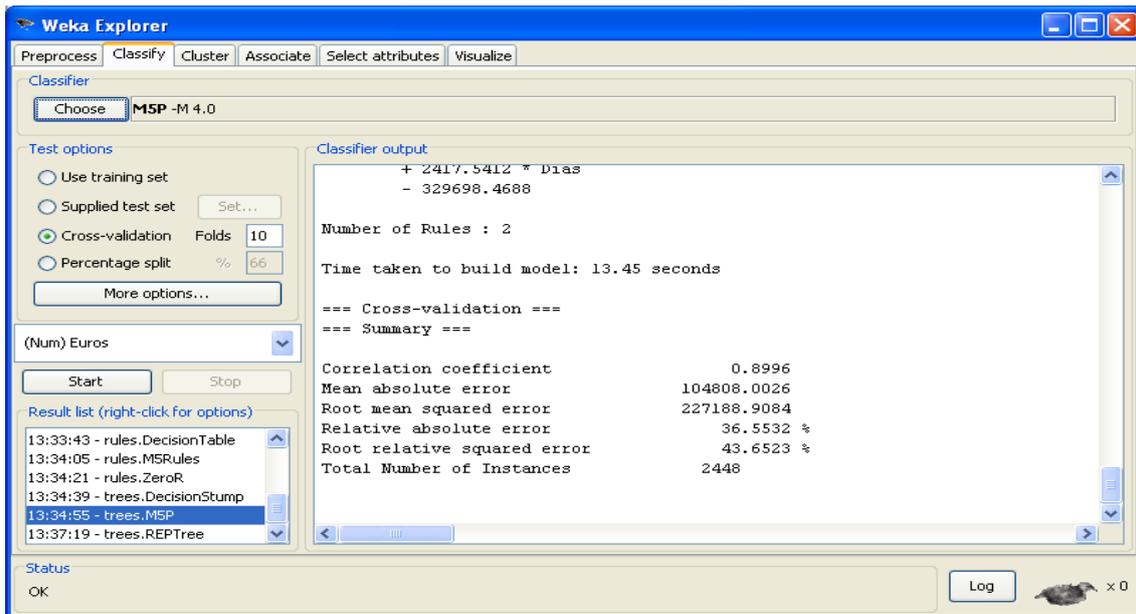


Figura 4.7: Resultados de la evaluación de la construcción del modelo 3 con el algoritmo “M5P”.

Como se ha dicho varias veces, el algoritmo “M5P” es el que ha realizado la construcción evaluada con mayor precisión del estudio de algoritmos del modelo 3. Igualmente, en el modelo 2 el algoritmo “M5P” es el que ha realizado la mejor construcción del modelo 2 según nuestros criterios de evaluación. El modelo 2 y el modelo 3 como se ha podido apreciar en las construcciones de ambos modelos en el apartado “3.Aprendizaje de datos” se diferencian en que en el modelo 3 se tiene en cuenta el atributo “días” como una variable explicativa o predictiva más para predecir el atributo “euros”, mientras, que en el modelo 2 no se tiene en cuenta dicha variable explicativa para predecir el atributo “euros”. Ambos modelos predicen el atributo “euros”, sin embargo, el modelo 3 ofrece mejor precisión en la predicción del atributo “euros” que la que ofrece el modelo 2 en la predicción del atributo “euros”. Ya que el modelo 3 seleccionado construido con el algoritmo “M5P”, aporta una mejor precisión, que el modelo 2 seleccionado construido, precisamente, también con el algoritmo “M5P”. Esto se puede observar comparando el “error absoluto medio” y la “raíz del error cuadrático medio” que aporta la construcción elegida del modelo 3 y la elegida del modelo 2. Como se muestra a continuación en la tabla 4.4.

Modelo	Algoritmo seleccionado	Mean Absolute Error	Root mean squared error
Modelo 2	M5P	105980,2765 €	227315,5084 €
Modelo 3	M5P	104808,0026 €	227188,9084 €

Tabla 4.4: Se muestra el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio” de la prueba de la construcción elegida del modelo 2 contra el valor del “error absoluto medio” y el de la “raíz del error cuadrático medio” de la prueba de la construcción elegida del modelo 3.

Con los datos reflejados en la tabla 4.4 se puede llegar a decir que el atributo “días” ayuda positivamente a predecir el atributo “euros” de las viviendas. Debido a que la predicción del atributo “euros” elegida es más precisa cuando se tiene en cuenta el atributo “días”, modelo 3, que cuando no se tiene en cuenta el atributo “días”, modelo 2. Una cuestión importante que hay que resaltar, respecto a la conclusión a la que se ha llegado de que el modelo 3 tiene más precisión que el modelo 2 es que: se ha alcanzado dicha conclusión a partir de los datos que se disponen para crear los modelos. No obstante, lo que no podríamos hacer de ninguna de las maneras sería concluir que el modelo 3 sería más preciso que el modelo 2, fueran cuales fueran, los datos de entrada (instancias de viviendas disponibles) con los que se construyeran los modelos, debido a que, esto no se puede asegurar. De hecho, esta es una de las peculiaridades de los modelos de datos, que como se construyen a partir de los datos, no se puede generalizar que resultaría lo mismo con otros datos distintos, en cuanto a la construcción de los modelos y a la evaluación de los modelos.

Con esto concluimos la fase de evaluación referente al proceso “KDD” en este proyecto.

5. Conclusiones

En este apartado, “5. Conclusiones”, vamos a relatar en primer lugar las ideas más relevantes. En segundo lugar discutiremos sobre los resultados de la evaluación de las construcciones seleccionadas, y también hablaremos de la comprensibilidad que presentan esas construcciones elegidas. En tercer lugar argumentaremos sobre el tiempo aproximado utilizado en este proyecto para cada fase del proceso “KDD”, y por último discutiremos sobre otras tareas que se podrían realizar como trabajo futuro.

Seguidamente introduciremos las ideas más relevantes del proyecto:

- En todo momento se utiliza el proceso “KDD”.
- En la preparación de los datos se pretende que los datos sean óptimos, reales y sin ningún tipo de ambigüedad.
- En este proyecto, se accede en dos fechas distintas al portal web inmobiliario seleccionado, para recopilar datos de las viviendas. Las instancias de viviendas recopiladas en la “fecha segunda” de extracción tienen como propósito intentar estimar el atributo “compra”, a las instancias de las viviendas recopiladas en la “fecha primera” de extracción.

- Las instancias de las viviendas obtenidas en la “fecha primera” de extracción a las que se les ha podido estimar el atributo “compra” se seleccionan, mientras que el resto de instancias se desechan.
- A las instancias de las viviendas seleccionadas se les calcula el atributo “días”. Con esto llegamos a tener todos los datos que deseamos de los disponibles.
- Se tratan las instancias de las viviendas seleccionadas: primero de forma común, hasta llegar a la vista de datos general, y después de forma separada para cada uno de los tres modelos de datos a tratar, hasta alcanzar la vista minable de cada uno de ellos. Estos tres modelos de datos se llaman:
 - Modelo1: Hallar la compra o la no compra de una vivienda.
 - Modelo2: Precio por el que vender una vivienda independientemente del número de días.
 - Modelo3: Precio por el que vender una vivienda teniendo en cuenta el número de días.
- En el modelo 1, la clase a predecir es el atributo “compra”. El atributo “compra” es de tipo nominal, lo que conlleva que en el modelo 1 se realice una tarea de clasificación.
- En el modelo 2 y en el modelo 3, la clase a predecir es el atributo “euros”, que es un atributo de tipo numérico, lo que conlleva que tanto en el modelo 2 como en el modelo 3 se realice una tarea de regresión.
- Para el modelo 1/modelo 2/modelo 3 se realiza un estudio con los algoritmos disponibles en “WEKA” que puedan construir el modelo 1/modelo 2/ modelo 3. Se selecciona en cada uno de los estudios de algoritmos, la construcción con mejor evaluación, es decir, que mejor precisión haya obtenido, como la construcción del modelo 1/modelo 2/modelo3. El método de evaluación que se utiliza en todo momento en los modelos es validación cruzada 10 X 10.
- Se extraen patrones de la construcción elegida del modelo 1/modelo 2/ modelo 3. En el caso del modelo 1, para explicar el atributo “compra” y en el caso del modelo 2 y del modelo 3 para explicar el atributo “euros”, para lo cual: en modelo 2 no se tiene en cuenta el atributo explicativo “días” y en el modelo 3 sí.
- De forma adicional, desearemos utilizar el modelo 3 seleccionado para: predecir el atributo “euros” a las réplicas de una misma instancia de una vivienda vendida, en función del atributo “días”, que tendrá un valor diferente en cada una de las réplicas.

Ahora vamos a tratar sobre lo buenos que son los resultados que hemos visto, en el apartado “4. Evaluación”, en las construcciones elegidas. En el modelo 1 la construcción elegida como se relató es la realizada con el algoritmo “RotationForest”. Dicha construcción ha resultado tener en la prueba del modelo un porcentaje de acierto del ‘57,368 %’. Este porcentaje de acierto se acerca aproximadamente al ‘60 %’ de probabilidad. Con lo que la incorrección se aproxima a un ‘40 %’. Si el acierto fuera del ‘66 %’, vendría a decir, que ‘2/3’ de las viviendas predichas del conjunto de prueba estarían bien clasificadas. Con lo que el modelo, aunque estaría bien que tuviera más precisión, es para tener en cuenta la precisión que adopta, sobretudo en datos de viviendas. No obstante, las buenas clasificaciones suelen estar por encima del ‘90%’ de aciertos.

En cuanto al modelo 2, la construcción seleccionada con el algoritmo “M5P” ha presentado en la prueba del modelo, un valor en el error absoluto medio de ‘105980,2765 €’ y un valor en la raíz del error cuadrático medio igual a ‘227315,5084 €’ con lo que, con estos errores hay bastantes euros de error, sin embargo, como las cantidades del atributo “euros” son elevadas, hace que el error no sea tan abultado, sobre todo tratándose del precio en euros por los que vender una vivienda. Respecto al modelo 3, se podría decir prácticamente lo mismo que para el modelo 2, ya que, también trata de predecir el atributo “euros” y la construcción seleccionada que es también la realizada con el algoritmo “M5P” ha presentado en la prueba del modelo un valor en el error absoluto medio de ‘104808,0026 €’ y uno en la raíz del error cuadrático medio de ‘227188,9084 €’, que son mejores que los del modelo 2, como decíamos en el apartado “4. Evaluación”. Por lo tanto, con los datos que se dispone de las viviendas, el modelo 3 es más preciso que el modelo 2.

Respecto a la comprensibilidad del modelo 1 seleccionado, decir que, como se muestran los patrones mediante árboles de decisión (debido a que el algoritmo “RotationForest” aunque es de la familia “metaLearned” utiliza el algoritmo “J48” para su implementación) es bastante comprensible el modelo. La mayor dificultad que presenta es que la construcción posee una gran cantidad de información, ya que, está formada por ‘10’ árboles de clasificación y abultan bastante cada uno de ellos, es decir, poseen muchas ramas y a su vez muchos nodos. Además cada nodo suele presentar abundante información.

Por lo que hace a la comprensibilidad de la construcción elegida del modelo 2 y la del modelo 3, decir que, el algoritmo “M5P”, que realiza la construcción elegida para ambos modelos, muestra los patrones en forma de regresión lineal directamente, aunque el algoritmo “M5P” es de la familia “tree”. Al reflejarse en forma de regresión lineal la comprensibilidad de estos dos modelos también es bastante buena.

Ahora vamos a hablar del tercer punto a tratar en este apartado: el coste tiempo de cada fase del proceso “KDD” en este proyecto. Lo habitual en cualquier proyecto que utilice el proceso “KDD” es que la fase de preparación sea la que mayor duración presente entre un ‘55 %’ y ‘65 %’ del tiempo estimado del proceso “KDD”. Después le seguiría

la fase de minería de datos del proceso “KDD” con una duración entre el ‘20 %’ y el ‘25 %’ del computo total. La fase de evaluación suele poseer entre el ‘10 %’ y el ‘15 %’ del tiempo utilizado en el proceso, y el resto del tiempo sería para la fase de interpretación y visualización que como se sabe es la última. En este proyecto, concretamente, la fase de preparación, efectivamente, es la que más tiempo ha presentado, aproximadamente el ‘60%’ del tiempo del proceso “KDD”. Por otro lado, un ‘40 %’ del tiempo utilizado recalaría entre la fase de minería de datos, la de evaluación y la de interpretación y visualización que han estado especialmente relacionadas en este proyecto.

5.1. Trabajo futuro

Por último, en cuanto a otras tareas que se podrían realizar, decir que, en el futuro se podría implantar la construcción elegida del modelo 1, del modelo 2 y del modelo 3, en una aplicación específica para cada uno ellos, para explotarlos de tal modo que al pasar a dichos modelos instancias de viviendas, se predijera la clase de dichas viviendas, clase atributo “compra” si estamos en la aplicación del modelo 1 y atributo “euros” si estamos en la del modelo 2 o modelo 3.

Otra posible tarea consistiría en reutilizar el fichero de las instancias de viviendas que no se les ha podido estimar el atributo “compra”, que en este proyecto desechamos, llamado “VRestoFecha1-Fecha2.txt”. Y utilizarlo como “fichero base” de otra extracción posterior que resultaría esta última en un “fichero comparar”. La extracción posterior tendría que distar entre “15” días y “20” días” de la “Fecha2” del fichero “VRestoFecha1-Fecha2.txt”. Las instancias de viviendas del fichero “VRestoFecha1-Fecha2.txt” que se les pudiera estimar el atributo “compra” se unirían a las instancias de viviendas que ya se tuvieran con el atributo “compra” hallado. Y así se podrían tener también instancias de viviendas del fichero “VRestoFecha1-Fecha2.txt” disponibles para los modelos. Habría que cambiar el valor del atributo “identificador” a las instancias de viviendas del fichero “VRestoFecha1-Fecha2.txt” que son ofertas distintas, de instancias que se les ha podido estimar el atributo “compra” en este proyecto, ya que dispondrían del mismo identificador.

6. Referencias

Libros:

Libro 1:

Título: *Introducción a la Minería de Datos.*

Editorial: PEARSON Prentice Hall.

Autores: José Hernández Orallo, M^a José Ramirez Quintana, Cèsar Ferri Ramírez.

Recursos de la web:

Recurso 1:

Soporte aplicación "WEKA": www.cs.waikato.ac.nz

Apéndice 1: La construcción del modelo 1

En este apéndice vamos a poner íntegramente la salida ofrecida por “WEKA” de la ejecución del modelo 1 con el algoritmo seleccionado “RotationForest” de la familia “metaLearned”. Dicha construcción consta de ‘10’ árboles de clasificación como relatamos en el apartado “3. Aprendizaje de datos”. También se apreciará: la información de ejecución del modelo y la de la prueba del modelo. Entre paréntesis pondremos algunas aclaraciones.

Modelo1: Hallar la compra o no compra de una vivienda. Con el algoritmo seleccionado “RotationForest”:

=== Run information === (Información de ejecución.)

```
Schema: "WEKA".classifiers.meta.RotationForest -G 3 -H 3 -P 50 -F ""WEKA".filters.unsupervised.attribute.PrincipalComponents -R 1.0 -A 5 -M -1 -D" -S 1 -I 10 -W
"WEKA".classifiers.trees.J48 - -C 0.25 -M 2
Relation: VSNDWK26ene2010-12feb2010.txt-"WEKA".filters.unsupervised.attribute.Remove-R1,11
Instances: 4583
Attributes: 9
  Tipo
  Barrio
  Dorm
  Planta
  Metros
  Euros
  Estado
  Ciudad
  Compra
Test mode: 10-fold cross-validation
```

=== Classifier model (full training set) === (Aprendizaje o Entrenamiento o construcción del modelo.)

All the base classifiers:

J48 pruned tree (Árbol 1º de los 10 que forman el modelo.)

```
-----
-0.771Estado-0.444Planta=p1-0.23Ciudad=valencia+0.169Planta=p5+0.163Ciudad=barcelona... <= -0.439575
| 0.991Estado-0.124Euros-0.052Metros <= -0.006365: NO (26.0/3.0)
| 0.991Estado-0.124Euros-0.052Metros > -0.006365
| | -0.501Barrio=ciutat vella-0.434Barrio=L'eixample-0.294Barrio=villa de vallecas-0.273Dorm=d8+0.216Barrio=sant andreu... <= 0.152513
| | | 0.638Planta=p13-0.624Planta=p12-0.326Planta=p15+0.312Planta=p11+0Planta=p14... <= 0
| | | -0.437Dorm=d7+0.305Tipo=áticos+0.282Barrio=carabanchel+0.275Barrio=sarrià-sant gervasi-0.25Dorm=d8... <= -0.211321: SÍ (787.0/304.0)
| | | -0.437Dorm=d7+0.305Tipo=áticos+0.282Barrio=carabanchel+0.275Barrio=sarrià-sant gervasi-0.25Dorm=d8... > -0.211321
| | | | 0.655Barrio=pueblo de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... <= -0.477962
| | | | 0.587Ciudad=valencia-0.517Ciudad=barcelona-0.387Planta=pen+0.22Planta=p6+0.205Planta=p7... <= 0.061925
| | | | | 0.867Planta=p9-0.41Planta=p8+0.227Planta=p10+0.097Planta=p11-0.073Planta=p7... <= -0.056796
| | | | | 0.704Metros+0.699Euros+0.124Estado <= 0.321119
| | | | | -0.473Tipo=chalets+0.398Tipo=estudios+0.391Dorm=d0-0.304Dorm=d5+0.24Tipo=pisos... <= 0.143378
| | | | | -0.708Metros+0.704Euros+0.051Estado <= 0.039496: SÍ (3.0/1.0)
| | | | | -0.708Metros+0.704Euros+0.051Estado > 0.039496: NO (2.0)
| | | | | -0.473Tipo=chalets+0.398Tipo=estudios+0.391Dorm=d0-0.304Dorm=d5+0.24Tipo=pisos... > 0.143378
| | | | | -0.708Metros+0.704Euros+0.051Estado <= 0.028996: NO (8.0/2.0)
| | | | | -0.708Metros+0.704Euros+0.051Estado > 0.028996: SÍ (15.0/2.0)
| | | | | 0.704Metros+0.699Euros+0.124Estado > 0.321119: SÍ (26.0/2.0)
| | | | | 0.867Planta=p9-0.41Planta=p8+0.227Planta=p10+0.097Planta=p11-0.073Planta=p7... > -0.056796: SÍ (8.0)
| | | | | 0.587Ciudad=valencia-0.517Ciudad=barcelona-0.387Planta=pen+0.22Planta=p6+0.205Planta=p7... > 0.061925
| | | | | -0.708Metros+0.704Euros+0.051Estado <= 0.028603: SÍ (9.0)
| | | | | -0.708Metros+0.704Euros+0.051Estado > 0.028603
| | | | | 0.991Estado-0.124Euros-0.052Metros <= 0.98326: NO (12.0/1.0)
| | | | | 0.991Estado-0.124Euros-0.052Metros > 0.98326: SÍ (2.0)
| | | | | 0.655Barrio=pueblo de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... > -0.477962
| | | | | 0.655Barrio=pueblo de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... <= -0.000658: NO
(1602.0/793.0)
| | | | | 0.655Barrio=pueblo de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... > -0.000658: SÍ
(1575.0/693.0)
| | | | | 0.638Planta=p13-0.624Planta=p12-0.326Planta=p15+0.312Planta=p11+0Planta=p14... > 0
| | | | | 0.655Barrio=pueblo de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... <= -0.039467: NO (5.0)
| | | | | 0.655Barrio=pueblo de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... > -0.039467
```

```

| | | | | -0.889Barrio=La saïdia+0.226Barrio=Rascanya+0.188Barrio=barajas-0.169Barrio=campanar - beni...+0.146Barrio=Quatre carreres... <= 0.007342: SÍ (8.0)
| | | | | -0.889Barrio=La saïdia+0.226Barrio=Rascanya+0.188Barrio=barajas-0.169Barrio=campanar - beni...+0.146Barrio=Quatre carreres... > 0.007342
| | | | | -0.695Barrio=horta guinardó+0.26 Barrio=gràcia-0.25Barrio=sants-montjuïc-0.227Barrio=sant martí+0.21 Barrio=nou barris... <= -0.010895: NO (2.0)
| | | | | -0.695Barrio=horta guinardó+0.26 Barrio=gràcia-0.25Barrio=sants-montjuïc-0.227Barrio=sant martí+0.21 Barrio=nou barris... > -0.010895
| | | | | -0.656Ciudad=madrid+0.485Ciudad=barcelona+0.357Ciudad=valencia-0.273Planta=p1+0.186Estado... <= -0.460829
| | | | | 0.991Estado-0.124Euros-0.052Metros <= 0.977753: NO (2.0)
| | | | | 0.991Estado-0.124Euros-0.052Metros > 0.977753: SÍ (6.0/1.0)
| | | | | -0.656Ciudad=madrid+0.485Ciudad=barcelona+0.357Ciudad=valencia-0.273Planta=p1+0.186Estado... > -0.460829: SÍ (3.0)
| | | | | -0.438Barrio=L'olivereta-0.424Barrio=El pla del real+0.348Barrio=Extramurs-0.261Barrio=Algirós+0.255Barrio=les corts... <= -0.155687
| | | | | -0.645Barrio=sant martí+0.307Barrio=Camins al grau+0.301Barrio=Extramurs+0.262Barrio=mortalaz+0.197Barrio=sants-montjuïc... <= 0.143901: SÍ (6.0)
| | | | | -0.645Barrio=sant martí+0.307Barrio=Camins al grau+0.301Barrio=Extramurs+0.262Barrio=mortalaz+0.197Barrio=sants-montjuïc... > 0.143901: NO (5.0/1.0)
| | | | | -0.438Barrio=L'olivereta-0.424Barrio=El pla del real+0.348Barrio=Extramurs-0.261Barrio=Algirós+0.255Barrio=les corts... > -0.155687
| | | | | -0.354Dorm=d4+0.322Barrio=les corts+0.269Dorm=d6+0.239Barrio=tetuán-0.23Barrio=El pla del real... <= 0.262415
| | | | | -0.675Barrio=san blas+0.614Barrio=arganzuela+0.247Barrio=usera-0.138Barrio=centro+0.121Barrio=retiro... <= -0.012801
| | | | | 0.991Estado-0.124Euros-0.052Metros <= -0.010076: NO (30.0)
| | | | | 0.991Estado-0.124Euros-0.052Metros > -0.010076
| | | | | 0.451Dorm=d8+0.423Barrio=carabanchel-0.404Dorm=d7-0.281Barrio=eixample-0.252Barrio=ciudad lineal... <= 0.069205
| | | | | -0.708Metros+0.704Euros+0.051Estado <= -0.005579
| | | | | 0.547Barrio=Ciutat vella-0.386Barrio=Quatre carreres-0.383Barrio=villa de vallecas+0.33 Barrio=mortalaz-0.288Barrio=L'eixample... <= -0.003712: SÍ (7.0)
| | | | | 0.547Barrio=Ciutat vella-0.386Barrio=Quatre carreres-0.383Barrio=villa de vallecas+0.33 Barrio=mortalaz-0.288Barrio=L'eixample... > -0.003712
| | | | | -0.721Planta=p5+0.653Planta=p6+0.171Planta=pen+0.1 Planta=pbj+0.066Planta=p7... <= -0.051503: SÍ (3.0)
| | | | | -0.721Planta=p5+0.653Planta=p6+0.171Planta=pen+0.1 Planta=pbj+0.066Planta=p7... > -0.051503
| | | | | -0.498Barrio=villa de vallecas-0.445Barrio=fuencarral-0.315Barrio=sants-montjuïc+0.293Barrio=L'eixample+0.241Barrio=El pla del real... <= -0.120446: SÍ (2.0)
| | | | | -0.498Barrio=villa de vallecas-0.445Barrio=fuencarral-0.315Barrio=sants-montjuïc+0.293Barrio=L'eixample+0.241Barrio=El pla del real... > -0.120446
| | | | | 0.991Estado-0.124Euros-0.052Metros <= -0.009922: SÍ (2.0)
| | | | | 0.991Estado-0.124Euros-0.052Metros > -0.009922: NO (11.0/1.0)
| | | | | -0.708Metros+0.704Euros+0.051Estado > -0.005579: NO (10.0)
| | | | | 0.451Dorm=d8+0.423Barrio=carabanchel-0.404Dorm=d7-0.281Barrio=eixample-0.252Barrio=ciudad lineal... > 0.069205: NO (12.0)
| | | | | -0.675Barrio=san blas+0.614Barrio=arganzuela+0.247Barrio=usera-0.138Barrio=centro+0.121Barrio=retiro... > -0.012801: NO (113.0/4.0)
| | | | | -0.354Dorm=d4+0.322Barrio=les corts+0.269Dorm=d6+0.239Barrio=tetuán-0.23Barrio=El pla del real... > 0.262415
| | | | | -0.708Metros+0.704Euros+0.051Estado <= -0.010459: SÍ (5.0)
| | | | | -0.708Metros+0.704Euros+0.051Estado > -0.010459
| | | | | -0.708Metros+0.704Euros+0.051Estado <= 0.001157: NO (5.0)
| | | | | -0.708Metros+0.704Euros+0.051Estado > 0.001157: SÍ (4.0/1.0)
| | | | | -0.514Dorm=d6+0.44 Barrio=sarrià-sant gervasi+0.348Dorm=d5-0.259Barrio=salamanca-0.259Barrio=fuencarral... > 0.0523
| | | | | 0.655Barrio=pueño de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... <= 0.037055
| | | | | 0.587Ciudad=valencia-0.517Ciudad=barcelona-0.387Planta=pen+0.22 Planta=p6+0.205Planta=p7... <= -0.500504: NO (7.0)
| | | | | 0.587Ciudad=valencia-0.517Ciudad=barcelona-0.387Planta=pen+0.22 Planta=p6+0.205Planta=p7... > -0.500504
| | | | | -0.437Dorm=d7+0.305Tipo=áticos+0.282Barrio=carabanchel+0.275Barrio=sarrià-sant gervasi-0.25Dorm=d8... <= 0.115324
| | | | | -0.675Barrio=san blas+0.614Barrio=arganzuela+0.247Barrio=usera-0.138Barrio=centro+0.121Barrio=retiro... <= -0.000691
| | | | | -0.708Metros+0.704Euros+0.051Estado <= -0.03245: NO (2.0)
| | | | | -0.708Metros+0.704Euros+0.051Estado > -0.03245: SÍ (35.0/4.0)
| | | | | -0.675Barrio=san blas+0.614Barrio=arganzuela+0.247Barrio=usera-0.138Barrio=centro+0.121Barrio=retiro... > -0.000691
| | | | | -0.699Planta=p3+0.638Planta=p4-0.219Planta=p2+0.134Planta=pen+0.108Planta=p1... <= 0.109594
| | | | | -0.708Metros+0.704Euros+0.051Estado <= -0.004465
| | | | | 0.573Planta=p7-0.559Planta=p6+0.384Planta=pbj-0.368Planta=p5+0.163Planta=p11... <= -0.070073: NO (23.0/4.0)
| | | | | 0.573Planta=p7-0.559Planta=p6+0.384Planta=pbj-0.368Planta=p5+0.163Planta=p11... > -0.070073
| | | | | 0.704Metros+0.699Euros+0.124Estado <= 0.055288: NO (2.0)
| | | | | 0.704Metros+0.699Euros+0.124Estado > 0.055288: SÍ (2.0)
| | | | | -0.708Metros+0.704Euros+0.051Estado > -0.004465
| | | | | 0.991Estado-0.124Euros-0.052Metros <= -0.008035: SÍ (8.0)
| | | | | 0.991Estado-0.124Euros-0.052Metros > -0.008035: NO (2.0)
| | | | | -0.699Planta=p3+0.638Planta=p4-0.219Planta=p2+0.134Planta=pen+0.108Planta=p1... > 0.109594: SÍ (2.0)
| | | | | -0.437Dorm=d7+0.305Tipo=áticos+0.282Barrio=carabanchel+0.275Barrio=sarrià-sant gervasi-0.25Dorm=d8... > 0.115324: NO (6.0)
| | | | | 0.655Barrio=pueño de vallecas-0.479Barrio=chamartín-0.318Barrio=arganzuela+0.269Barrio=chamberí+0.208Barrio=eixample... > 0.037055: SÍ (12.0)

```

Number of Leaves : 46
Size of the tree : 91

J48 pruned tree (Árbol 2º de los 10 que forman el modelo.)

```

0.782Estado+0.49 Tipo=dúplex-0.237Tipo=áticos-0.196Euros+0.186Tipo=estudios... <= 0.495042
0.54 Barrio=san blas-0.432Barrio=arganzuela-0.405Barrio=latina+0.244Barrio=ciudad lineal-0.216Barrio=hortaleza... <= 0.185645
| | | | | -0.633Barrio=El pla del real-0.599Barrio=Algirós+0.252Barrio=Poblats marítims+0.245Barrio=L'olivereta+0.174Barrio=Patraix... <= -0.00376
| | | | | 0.552Dorm=d4-0.344Dorm=d1+0.246Dorm=d5+0.206Barrio=sarrià-sant gervasi-0.196Barrio=centro... <= 0.265158: NO (101.0/4.0)
| | | | | 0.552Dorm=d4-0.344Dorm=d1+0.246Dorm=d5+0.206Barrio=sarrià-sant gervasi-0.196Barrio=centro... > 0.265158
| | | | | 0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... <= 0.188458: NO (3.0)
| | | | | 0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... > 0.188458: SÍ (3.0)
-0.633Barrio=El pla del real-0.599Barrio=Algirós+0.252Barrio=Poblats marítims+0.245Barrio=L'olivereta-0.174Barrio=Patraix... > -0.00376
| | | | | 0.656Barrio=Extramurs-0.466Barrio=campanar - beni...-0.366Barrio=Camins al grau+0.312Barrio=mortalaz+0.18 Barrio=L'eixample... <= -0.047062: SÍ (15.0)
| | | | | 0.656Barrio=Extramurs-0.466Barrio=campanar - beni...-0.366Barrio=Camins al grau+0.312Barrio=mortalaz+0.18 Barrio=L'eixample... > -0.047062
| | | | | 0.609Barrio=moncloa-0.353Barrio=usera-0.343Barrio=fuencarral-0.279Barrio=villaverde+0.215Barrio=hortaleza... <= -0.158818: NO (73.0/5.0)
| | | | | 0.609Barrio=moncloa-0.353Barrio=usera-0.343Barrio=fuencarral-0.279Barrio=villaverde+0.215Barrio=hortaleza... > -0.158818
| | | | | 0.609Barrio=moncloa-0.353Barrio=usera-0.343Barrio=fuencarral-0.279Barrio=villaverde+0.215Barrio=hortaleza... <= -0.099424
| | | | | 0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... <= 0.166063: SÍ (10.0)
| | | | | 0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... > 0.166063
| | | | | -0.687Planta=p15-0.68Planta=p14+0.174Planta=pss+0.12 Planta=pst-0.108Planta=p22... <= 0.000978
| | | | | -0.578Ciudad=madrid+0.44 Ciudad=barcelona+0.3 Ciudad=valencia+0.177Barrio=eixample+0.163Barrio=sarrià-sant gervasi... <= -0.508599
| | | | | -0.806Tipo=estudios+0.503Tipo=dúplex+0.26 Tipo=áticos-0.166Tipo=chalets-0.039Estado... <= 0.020394: SÍ (3.0)
| | | | | -0.806Tipo=estudios+0.503Tipo=dúplex+0.26 Tipo=áticos-0.166Tipo=chalets-0.039Estado... > 0.020394: NO (3.0/1.0)
| | | | | -0.578Ciudad=madrid+0.44 Ciudad=barcelona+0.3 Ciudad=valencia+0.177Barrio=eixample+0.163Barrio=sarrià-sant gervasi... > -0.508599: NO (3.0)
| | | | | -0.687Planta=p15-0.68Planta=p14+0.174Planta=pss+0.12 Planta=pst-0.108Planta=p22... > 0.000978: SÍ (2.0)
| | | | | 0.609Barrio=moncloa-0.353Barrio=usera-0.343Barrio=fuencarral-0.279Barrio=villaverde+0.215Barrio=hortaleza... > -0.099424
| | | | | 0.375Dorm=d4-0.316Barrio=L'eixample-0.289Dorm=d3+0.246Barrio=nou barris-0.239Barrio=El pla del real... <= 0.177967
| | | | | 0.562Dorm=d1-0.547Dorm=d2+0.242Barrio=centro+0.181Barrio=campanar - beni...+0.18 Barrio=salamanca... <= -0.021825
| | | | | -0.806Tipo=estudios+0.503Tipo=dúplex+0.26 Tipo=áticos-0.166Tipo=chalets-0.039Estado... <= 0.019855
| | | | | 0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... <= 0.175623: SÍ (12.0/1.0)

```

0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... > 0.175623: NO (8.0)
-0.806Tipo=estudios+0.503Tipo=dúplex+0.26 Tipo=áticos-0.166Tipo=chalets-0.039Estado... > 0.019855: SÍ (9.0)
0.562Dorm=d1-0.547Dorm=d2+0.242Barrio=centro+0.181Barrio=campanar - beni...+0.18 Barrio=salamanca... > -0.021825
-0.8377Planta=pen-0.372Planta=p7+0.299Planta=pbj+0.131Planta=p5+0.105Planta=p4... <= -0.004171: SÍ (4.0)
-0.8377Planta=pen-0.372Planta=p7+0.299Planta=pbj+0.131Planta=p5+0.105Planta=p4... > -0.004171
-0.658Ciudad=madrid-0.637Ciudad=barcelona-0.29Ciudad=valencia+0.076Barrio=eixample+0.068Barrio=sarrià-sant gervasi... <= -0.646078
-0.578Ciudad=madrid+0.44 Ciudad=barcelona+0.3 Ciudad=valencia+0.177Barrio=eixample+0.163Barrio=sarrià-sant gervasi... <= -0.722982
-0.806Tipo=estudios+0.503Tipo=dúplex+0.26 Tipo=áticos-0.166Tipo=chalets-0.039Estado... <= 0.019612: NO (6.0/1.0)
-0.806Tipo=estudios+0.503Tipo=dúplex+0.26 Tipo=áticos-0.166Tipo=chalets-0.039Estado... > 0.019612: SÍ (4.0)
-0.578Ciudad=madrid+0.44 Ciudad=barcelona+0.3 Ciudad=valencia+0.177Barrio=eixample+0.163Barrio=sarrià-sant gervasi... > -0.722982: SÍ (2.0)
-0.658Ciudad=madrid-0.637Ciudad=barcelona-0.29Ciudad=valencia+0.076Barrio=eixample+0.068Barrio=sarrià-sant gervasi... > -0.646078
0.722Euros-0.619Tipo=chalets+0.237Tipo=estudios+0.103Tipo=pisos+0.099Tipo=dúplex... <= 0.133677
-0.578Ciudad=madrid+0.44 Ciudad=barcelona+0.3 Ciudad=valencia+0.177Barrio=eixample+0.163Barrio=sarrià-sant gervasi... <= -0.778456: SÍ (5.0)
-0.578Ciudad=madrid+0.44 Ciudad=barcelona+0.3 Ciudad=valencia+0.177Barrio=eixample+0.163Barrio=sarrià-sant gervasi... > -0.778456
-0.726Planta=p5+0.461Planta=p6+0.428Planta=pbj+0.194Planta=p7+0.103Planta=p8... <= -0.048951
-0.56Planta=p2+0.559Metros+0.287Estado+0.275Planta=p6-0.225Planta=pbj... <= 0.107386
0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... <= 0.16457: NO (2.0)
0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... > 0.16457: SÍ (2.0)
-0.56Planta=p2+0.559Metros+0.287Estado+0.275Planta=p6-0.225Planta=pbj... > 0.107386: SÍ (2.0)
-0.726Planta=p5+0.461Planta=p6+0.428Planta=pbj+0.194Planta=p7+0.103Planta=p8... > -0.048951: NO (2.0/2.0)
0.722Euros-0.619Tipo=chalets+0.237Tipo=estudios+0.103Tipo=pisos+0.099Tipo=dúplex... > 0.133677: NO (2.0)
0.375Dorm=d4-0.316Barrio=L'eixample-0.289Dorm=d3+0.246Barrio=nou barris-0.239Barrio=El pla del real... > 0.177967: NO (11.0)
0.54 Barrio=san blas-0.432Barrio=arganzuela-0.405Barrio=latina+0.244Barrio=ciudad lineal-0.216Barrio=hortaleza... > 0.185645
-0.589Dorm=d8-0.464Barrio=les corts-0.382Barrio=carabanchel-0.265Dorm=d0-0.187Dorm=d5... <= 0.136009: SÍ (29.0/5.0)
-0.589Dorm=d8-0.464Barrio=les corts-0.382Barrio=carabanchel-0.265Dorm=d0-0.187Dorm=d5... > 0.136009: NO (3.0)
0.782Estado+0.49 Tipo=dúplex-0.237Tipo=áticos-0.196Euros+0.186Tipo=estudios... > 0.495042
0.974Barrio=Benimaclet-0.131Barrio=Patraix-0.081Barrio=Algirós+0.08 Barrio=El pla del real-0.056Barrio=moratalaz... <= 0.018588
-0.687Planta=p15-0.68Planta=p14+0.174Planta=pss+0.12 Planta=pst-0.108Planta=p22... <= 0.020235: SÍ (3947.0/1801.0)
-0.687Planta=p15-0.68Planta=p14+0.174Planta=pss+0.12 Planta=pst-0.108Planta=p22... > 0.020235
-0.478Barrio=moratalaz+0.476Barrio=gràcia-0.403Barrio=sant andreu+0.342Barrio=Extramurs+0.248Barrio=vicálvaro... <= 0.155464
-0.66Barrio=tetuán+0.552Barrio=puente de vallecas+0.374Barrio=salamanca-0.155Barrio=ciudad lineal-0.152Barrio=centro... <= -0.017875
-0.625Planta=p22+0.469Planta=p15-0.423Planta=p14-0.38Planta=pss-0.157Planta=p12... <= 0.11797
0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... <= 0.550216
0.682Barrio=Rascanya-0.514Barrio=Poblets marítims-0.436Barrio=La saïdia+0.127Barrio=campanar - beni...-0.121Barrio=Jesús... <= 0.000942: NO (7.0)
0.682Barrio=Rascanya-0.514Barrio=Poblets marítims-0.436Barrio=La saïdia+0.127Barrio=campanar - beni...-0.121Barrio=Jesús... > 0.000942: SÍ (3.0/1.0)
0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... > 0.550216: SÍ (2.0)
-0.625Planta=p22+0.469Planta=p15-0.423Planta=p14-0.38Planta=pss-0.157Planta=p12... > 0.11797: SÍ (5.0)
-0.66Barrio=tetuán+0.552Barrio=puente de vallecas+0.374Barrio=salamanca-0.155Barrio=ciudad lineal-0.152Barrio=centro... > -0.017875
0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... <= 0.5573: SÍ (22.0)
0.564Euros+0.533Tipo=chalets+0.366Estado-0.348Tipo=estudios-0.274Tipo=dúplex... > 0.5573
-0.697Dorm=d3+0.403Dorm=d2-0.22Barrio=latina+0.188Dorm=d1-0.157Barrio=moratalaz... <= -0.774683: NO (2.0)
-0.697Dorm=d3+0.403Dorm=d2-0.22Barrio=latina+0.188Dorm=d1-0.157Barrio=moratalaz... > -0.774683: SÍ (4.0)
-0.478Barrio=moratalaz+0.476Barrio=gràcia-0.403Barrio=sant andreu+0.342Barrio=Extramurs+0.248Barrio=vicálvaro... > 0.155464: NO (3.0)
0.974Barrio=Benimaclet-0.131Barrio=Patraix-0.081Barrio=Algirós+0.08 Barrio=El pla del real-0.056Barrio=moratalaz... > 0.018588
0.562Dorm=d1-0.547Dorm=d2+0.242Barrio=centro+0.181Barrio=campanar - beni...+0.18 Barrio=salamanca... <= 0.102598
0.558Dorm=d7-0.333Barrio=fuencarral+0.317Dorm=d6-0.228Barrio=hortaleza-0.215Dorm=d4... <= -0.055285: SÍ (84.0/13.0)
0.558Dorm=d7-0.333Barrio=fuencarral+0.317Dorm=d6-0.228Barrio=hortaleza-0.215Dorm=d4... > -0.055285
-0.58Barrio=san blas-0.5Barrio=san blas+0.5 Barrio=puente de vallecas+0.5 Barrio=puente de vallecas+0Barrio=arganzuela... <= -1
-0.726Planta=p5+0.461Planta=p6+0.428Planta=pbj+0.194Planta=p7+0.103Planta=p8... <= 0.453171: SÍ (85.0/24.0)
-0.726Planta=p5+0.461Planta=p6+0.428Planta=pbj+0.194Planta=p7+0.103Planta=p8... > 0.453171: NO (7.0/1.0)
0.562Dorm=d1-0.547Dorm=d2+0.242Barrio=centro+0.181Barrio=campanar - beni...+0.18 Barrio=salamanca... > 0.102598
-0.669Barrio=retiro+0.309Barrio=villaverde+0.304Barrio=eixample-0.301Barrio=ciutat vella-0.219Barrio=barajas... <= -0.058315: NO (8.0)
-0.669Barrio=retiro+0.309Barrio=villaverde+0.304Barrio=eixample-0.301Barrio=ciutat vella-0.219Barrio=barajas... > -0.058315: SÍ (42.0/17.0)

Number of Leaves : 40

Size of the tree : 79

J48 pruned tree (Árbol 3º de los 10 que forman el modelo.)

0.807Estado+0.526Tipo=dúplex-0.159Tipo=pisos-0.139Euros-0.1Tipo=chalets... <= 0.523768
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza+0Barrio=fuencarral... <= 0
-0.4Barrio=ciutat vella-0.4Barrio=ciutat vella-0.4Barrio=gràcia-0.4Barrio=gràcia+0.345Barrio=sant andreu... <= 0.002434
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza+0Barrio=fuencarral... <= -1
0.855Tipo=estudios-0.351Tipo=áticos-0.328Tipo=dúplex+0.167Tipo=chalets+0.081Estado... <= -0.043794: SÍ (10.0)
0.855Tipo=estudios-0.351Tipo=áticos-0.328Tipo=dúplex+0.167Tipo=chalets+0.081Estado... > -0.043794
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... <= 0.672579: NO (14.0/2.0)
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... > 0.672579: SÍ (3.0)
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza+0Barrio=fuencarral... > -1
-0.632Barrio=retiro-0.632Barrio=retiro+0.291Barrio=villaverde+0.291Barrio=villaverde+0.062Barrio=usera... <= 0.000801: NO (52.0/1.0)
-0.632Barrio=retiro-0.632Barrio=retiro+0.291Barrio=villaverde+0.291Barrio=villaverde+0.062Barrio=usera... > 0.000801
-0.5Barrio=san blas-0.5Barrio=san blas+0.5 Barrio=puente de vallecas+0.5 Barrio=puente de vallecas+0Barrio=arganzuela... <= -1
-0.621Planta=p7+0.427Planta=p6-0.363Planta=p8-0.241Planta=pen-0.218Planta=p10... <= -0.032651: NO (4.0)
-0.621Planta=p7+0.427Planta=p6-0.363Planta=p8-0.241Planta=pen-0.218Planta=p10... > -0.032651: SÍ (8.0)
-0.5Barrio=san blas-0.5Barrio=san blas+0.5 Barrio=puente de vallecas+0.5 Barrio=puente de vallecas+0Barrio=arganzuela... > -1
-0.66Barrio=centro-0.66Barrio=centro+0.168Barrio=chamartín+0.168Barrio=chamartín+0.138Barrio=salamanca... <= 0.109548
0.359Planta=p4-0.309Dorm=d7+0.301Dorm=d0-0.299Planta=pbj+0.278Planta=pen... <= 0.198743
0.685Barrio=les corts+0.685Barrio=les corts-0.12Barrio=nou barris-0.12Barrio=nou barris-0.064Barrio=ciutat vella... <= -0.001233
0.523Planta=p9-0.485Dorm=d6+0.318Planta=p10-0.297Planta=pbj-0.214Planta=pss... <= 0.150789: NO (80.0/7.0)
0.523Planta=p9-0.485Dorm=d6+0.318Planta=p10-0.297Planta=pbj-0.214Planta=pss... > 0.150789: SÍ (6.0/1.0)
0.685Barrio=les corts+0.685Barrio=les corts-0.12Barrio=nou barris-0.12Barrio=nou barris-0.064Barrio=ciutat vella... > -0.001233
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... <= 0.671973
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... <= -0.540496
0.807Estado+0.526Tipo=dúplex-0.159Tipo=pisos-0.139Euros-0.1Tipo=chalets... <= -0.16511
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... <= -0.149468: SÍ (5.0/1.0)
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... > -0.149468: NO (17.0)
0.807Estado+0.526Tipo=dúplex-0.159Tipo=pisos-0.139Euros-0.1Tipo=chalets... > -0.16511: SÍ (7.0)
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... > -0.540496
-0.471Planta=p6+0.313Dorm=d0-0.294Planta=pst+0.288Planta=p5+0.261Dorm=d9... <= -0.134046: NO (8.0)
-0.471Planta=p6+0.313Dorm=d0-0.294Planta=pst+0.288Planta=p5+0.261Dorm=d9... > -0.134046
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... <= 0.667788: NO (9.0)
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... > 0.667788: SÍ (2.0)
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... > 0.671973: SÍ (6.0)
0.359Planta=p4-0.309Dorm=d7+0.301Dorm=d0-0.299Planta=pbj+0.278Planta=pen... > 0.198743
-0.565Planta=pen+0.518Planta=pss-0.509Planta=p11+0.272Planta=p10+0.202Planta=p7... <= 0.021463: SÍ (16.0/2.0)
-0.565Planta=pen+0.518Planta=pss-0.509Planta=p11+0.272Planta=p10+0.202Planta=p7... > 0.021463: NO (3.0)
-0.66Barrio=centro-0.66Barrio=centro+0.168Barrio=chamartín+0.168Barrio=chamartín+0.138Barrio=salamanca... > 0.109548: NO (61.0/3.0)
-0.4Barrio=ciutat vella-0.4Barrio=ciutat vella-0.4Barrio=gràcia-0.4Barrio=gràcia+0.345Barrio=sant andreu... > 0.002434

-0.639Barrio=barajas-0.639Barrio=barajas-0.223Barrio=villa de vallecas-0.223Barrio=villa de vallecas+0.162Barrio=mortalaz... <= 0.001575
0.511Barrio=chamberí+0.511Barrio=chamberí-0.473Barrio=latina-0.473Barrio=latina+0.068Barrio=moncloa... <= -0.001576: NO (3.0/1.0)
0.511Barrio=chamberí+0.511Barrio=chamberí-0.473Barrio=latina-0.473Barrio=latina+0.068Barrio=moncloa... > -0.001576: SÍ (17.0)
-0.639Barrio=barajas-0.639Barrio=barajas-0.223Barrio=villa de vallecas-0.223Barrio=villa de vallecas+0.162Barrio=mortalaz... > 0.001575: NO (6.0)
0.5Barrio=ciudad lineal+0.5Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza+0Barrio=fuencarral... > 0
-0.409Planta=p1+0.379Dorm=d0+0.361Planta=pss+0.359Planta=p8+0.325Dorm=d7... <= -0.448833: NO (4.0/1.0)
-0.409Planta=p1+0.379Dorm=d0+0.361Planta=pss+0.359Planta=p8+0.325Dorm=d7... > -0.448833: SÍ (16.0/1.0)
0.807Eestado+0.526Tipo=dúplex-0.159Tipo=pisos-0.139Euros-0.1Tipo=chalets... > 0.523768
-0.533Barrio=sant andreu-0.533Barrio=sant andreu+0.377Barrio=horta guinardó+0.377Barrio=horta guinardó+0.21Barrio=sants-montjuïc... <= -1.067222
0.632Dorm=d8+0.321Planta=p11+0.289Planta=pss+0.257Planta=p13+0.226Planta=p10... <= -0.051038: NO (5.0)
0.632Dorm=d8+0.321Planta=p11+0.289Planta=pss+0.257Planta=p13+0.226Planta=p10... > -0.051038
-0.408Planta=p8-0.378Dorm=d9+0.372Dorm=d0+0.369Dorm=d7+0.33Planta=pbj... <= -0.043094: SÍ (17.0)
-0.408Planta=p8-0.378Dorm=d9+0.372Dorm=d0+0.369Dorm=d7+0.33Planta=pbj... > -0.043094
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... <= 0.789638
0.855Tipo=estudios-0.351Tipo=áticos-0.328Tipo=dúplex+0.167Tipo=chalets+0.081Estado... <= 0.038717: NO (16.0/6.0)
0.855Tipo=estudios-0.351Tipo=áticos-0.328Tipo=dúplex+0.167Tipo=chalets+0.081Estado... > 0.038717: SÍ (15.0/1.0)
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... > 0.789638: SÍ (9.0)
-0.533Barrio=sant andreu-0.533Barrio=sant andreu+0.377Barrio=horta guinardó+0.377Barrio=horta guinardó+0.21Barrio=sants-montjuïc... > -1.067222
-0.492Barrio=salamanca-0.492Barrio=salamanca-0.333Barrio=chamartín-0.333Barrio=chamartín+0.244Barrio=carabanchel... <= -0.00194
0.4Barrio=sant martí+0.4Barrio=sant martí+0.36Barrio=eixample+0.36Barrio=eixample-0.334Barrio=sarrià-sant gervasi... <= 0.005064
0.684Tipo=pisos-0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... <= 0.49319
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... <= 0.391075
0.511Barrio=chamberí+0.511Barrio=chamberí-0.473Barrio=latina-0.473Barrio=latina+0.068Barrio=moncloa... <= -0.023925
-0.546Dorm=d9-0.45Planta=pst+0.323Planta=pbj-0.226Planta=pss-0.22Planta=p12... <= 0.179552
0.405Planta=p4-0.4Planta=p3+0.337Planta=p2-0.327Planta=p1+0.314Planta=pbj... <= -0.205726: NO (6.0/1.0)
0.405Planta=p4-0.4Planta=p3+0.337Planta=p2-0.327Planta=p1+0.314Planta=pbj... > -0.205726
0.877Planta=p22-0.399Planta=p14-0.209Planta=p13-0.131Dorm=d8+0.069Planta=p12... <= -0.010844
-0.546Dorm=d9-0.45Planta=pst+0.323Planta=pbj-0.226Planta=pss-0.22Planta=p12... <= 0.128656: NO (2.0)
-0.546Dorm=d9-0.45Planta=pst+0.323Planta=pbj-0.226Planta=pss-0.22Planta=p12... > 0.128656: SÍ (4.0/1.0)
0.877Planta=p22-0.399Planta=p14-0.209Planta=p13-0.131Dorm=d8+0.069Planta=p12... > -0.010844: SÍ (18.0)
-0.546Dorm=d9-0.45Planta=pst+0.323Planta=pbj-0.226Planta=pss-0.22Planta=p12... > 0.179552: NO (4.0)
0.511Barrio=chamberí+0.511Barrio=chamberí-0.473Barrio=latina-0.473Barrio=latina+0.068Barrio=moncloa... > -0.023925
-0.556Barrio=chamartín-0.556Barrio=chamartín+0.431Barrio=salamanca+0.431Barrio=salamanca-0.039Barrio=centro... <= -0.000522
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... <= -0.064313
0.855Tipo=estudios-0.351Tipo=áticos-0.328Tipo=dúplex+0.167Tipo=chalets+0.081Estado... <= 0.238375: NO (2.0)
0.855Tipo=estudios-0.351Tipo=áticos-0.328Tipo=dúplex+0.167Tipo=chalets+0.081Estado... > 0.238375: SÍ (2.0)
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... > -0.064313: SÍ (10.0)
-0.556Barrio=chamartín-0.556Barrio=chamartín+0.431Barrio=salamanca+0.431Barrio=salamanca-0.039Barrio=centro... > -0.000522
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... <= 0.635403
0.64Metros+0.362Dorm=d5+0.288Planta=p1-0.263Dorm=d2+0.255Dorm=d4... <= 0.638826: SÍ (17.0/2.0)
0.64Metros+0.362Dorm=d5+0.288Planta=p1-0.263Dorm=d2+0.255Dorm=d4... > 0.638826: NO (2.0)
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... > 0.635403
0.542Planta=p7+0.519Planta=p6-0.375Planta=p8-0.26Planta=p4+0.222Planta=pen... <= -0.14921: NO (3.0)
0.542Planta=p7+0.519Planta=p6-0.375Planta=p8-0.26Planta=p4+0.222Planta=pen... > -0.14921
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... <= -0.488966: NO (3.0/1.0)
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... > -0.488966: SÍ (7.0)
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... > 0.391075
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... <= 0.401585: NO (9.0)
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... > 0.401585: SÍ (2.0)
0.684Tipo=pisos-0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... > 0.49319
-0.597Barrio=Camins al grau-0.597Barrio=Camins al grau+0.159Barrio=Algirós-0.159Barrio=Algirós+0.159Barrio=La saïdia... <= -0.00033
0.64Metros+0.362Dorm=d5+0.288Planta=p1-0.263Dorm=d2+0.255Dorm=d4... <= 0.261511
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... <= 0.635403: SÍ (57.0/20.0)
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... > 0.635403
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... <= -0.587874: NO (10.0/2.0)
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... > -0.587874
0.64Metros+0.362Dorm=d5+0.288Planta=p1-0.263Dorm=d2+0.255Dorm=d4... <= -0.329541: NO (4.0)
0.64Metros+0.362Dorm=d5+0.288Planta=p1-0.263Dorm=d2+0.255Dorm=d4... > -0.329541: SÍ (52.0/14.0)
0.64Metros+0.362Dorm=d5+0.288Planta=p1-0.263Dorm=d2+0.255Dorm=d4... > 0.261511: NO (10.0/1.0)
-0.597Barrio=Camins al grau-0.597Barrio=Camins al grau+0.159Barrio=Algirós-0.159Barrio=Algirós+0.159Barrio=La saïdia... > -0.00033
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... <= -0.540496
-0.408Planta=p8-0.378Dorm=d9+0.372Dorm=d0+0.369Dorm=d7+0.33Planta=pbj... <= -0.025098: NO (117.0/50.0)
-0.408Planta=p8-0.378Dorm=d9+0.372Dorm=d0+0.369Dorm=d7+0.33Planta=pbj... > -0.025098: SÍ (66.0/21.0)
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... > -0.540496
-0.482Dorm=d1-0.481Planta=p1+0.31Dorm=d4+0.299Dorm=d2+0.262Planta=p2... <= -0.561428: SÍ (8.0)
-0.482Dorm=d1-0.481Planta=p1+0.31Dorm=d4+0.299Dorm=d2+0.262Planta=p2... > -0.561428
0.523Planta=p9-0.485Dorm=d6+0.318Planta=p10-0.297Planta=pbj-0.214Planta=pss... <= 0.220609: NO (83.0/38.0)
0.523Planta=p9-0.485Dorm=d6+0.318Planta=p10-0.297Planta=pbj-0.214Planta=pss... > 0.220609: SÍ (6.0)
0.4Barrio=sant martí+0.4Barrio=sant martí+0.36Barrio=eixample+0.36Barrio=eixample-0.334Barrio=sarrià-sant gervasi... > 0.005064
-0.607Barrio=Quatre carreres-0.607Barrio=Quatre carreres+0.341Barrio=Ciutat vella+0.341Barrio=Ciutat vella+0.076Barrio=L'eixample... <= -1.213405
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... <= -0.036696
-0.565Planta=pen+0.518Planta=pss-0.509Planta=p11+0.272Planta=p10+0.202Planta=p7... <= -0.02141: SÍ (4.0)
-0.565Planta=pen+0.518Planta=pss-0.509Planta=p11+0.272Planta=p10+0.202Planta=p7... > -0.02141
0.625Planta=p14-0.624Planta=p13-0.329Planta=p12+0.197Planta=p22+0.144Planta=pst... <= 0.007639: NO (5.0)
0.625Planta=p14-0.624Planta=p13-0.329Planta=p12+0.197Planta=p22+0.144Planta=pst... > 0.007639
0.684Tipo=pisos+0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... <= 0.49319: NO (3.0/1.0)
0.684Tipo=pisos+0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... > 0.49319
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... <= 0.096322
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... > 0.096322
-0.565Planta=pen+0.518Planta=pss-0.509Planta=p11+0.272Planta=p10+0.202Planta=p7... <= 0.006085: SÍ (2.0)
-0.565Planta=pen+0.518Planta=pss-0.509Planta=p11+0.272Planta=p10+0.202Planta=p7... > 0.006085: NO (8.0/2.0)
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... > 0.096322: SÍ (3.0)
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... > -0.036696
0.684Tipo=pisos+0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... <= 0.49319
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... <= 0.187498: SÍ (3.0)
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... > 0.187498: NO (2.0)
0.684Tipo=pisos+0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... > 0.49319: SÍ (22.0/1.0)
-0.607Barrio=Quatre carreres-0.607Barrio=Quatre carreres+0.341Barrio=Ciutat vella+0.341Barrio=Ciutat vella+0.076Barrio=L'eixample... > -1.213405
-0.552Barrio=Ciutat vella-0.552Barrio=Ciutat vella+0.276Barrio=L'eixample+0.276Barrio=L'eixample-0.235Barrio=Quatre carreres... <= -0.000591
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... <= -0.662985
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... <= 0.245715: SÍ (183.0/66.0)
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... > 0.245715: NO (10.0/1.0)
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample... > -0.662985
0.684Tipo=pisos+0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... <= 0.49319
0.359Planta=p4-0.309Dorm=d7+0.301Dorm=d0-0.299Planta=pbj+0.278Planta=pen... <= 0.231191
0.542Planta=p7+0.519Planta=p6-0.375Planta=p8-0.26Planta=p4+0.222Planta=pen... <= 0.048641: SÍ (11.0)
0.542Planta=p7+0.519Planta=p6-0.375Planta=p8-0.26Planta=p4+0.222Planta=pen... > 0.048641
0.68Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... <= -0.251786: SÍ (2.0)
0.68Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... > -0.251786: NO (2.0)
0.359Planta=p4-0.309Dorm=d7+0.301Dorm=d0-0.299Planta=pbj+0.278Planta=pen... > 0.231191: NO (5.0)
0.684Tipo=pisos+0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... > 0.49319
0.877Planta=p22-0.399Planta=p14-0.209Planta=p13-0.131Dorm=d8+0.069Planta=p12... <= 0.000941

-0.536Dorm=d6-0.395Planta=p3+0.368Planta=p4-0.305Planta=p9+0.238Dorm=d4... <= -0.4253
-0.409Planta=p1+0.379Dorm=d0+0.361Planta=pss+0.359Planta=p8+0.325Dorm=d7... <= 0.123847: SÍ (3.0/1.0)
-0.409Planta=p1+0.379Dorm=d0+0.361Planta=pss+0.359Planta=p8+0.325Dorm=d7... > 0.123847: NO (5.0)
-0.536Dorm=d6-0.395Planta=p3+0.368Planta=p4-0.305Planta=p9+0.238Dorm=d4... > -0.4253: SÍ (89.0/20.0)
0.877Planta=p22-0.399Planta=p14-0.209Planta=p13-0.131Dorm=d8+0.069Planta=p12... > 0.000941
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... <= -0.249051
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... <= -0.509671: NO (7.0/2.0)
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... > -0.509671: SÍ (18.0/1.0)
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... > -0.249051: NO (67.0/27.0)
-0.552Barrio=Ciutat vella-0.552Barrio=Ciutat vella+0.276Barrio=L'eixample+0.276Barrio=L'eixample-0.235Barrio=Quatre carreres... > -0.000591: SÍ (270.0/100.0)
-0.492Barrio=salamanca-0.492Barrio=salamanca-0.333Barrio=chamartín-0.333Barrio=chamartín+0.244Barrio=carabanchel... > -0.00194
-0.5Barrio=san blas-0.5Barrio=san blas+0.5 Barrio=puente de vallecas+0.5 Barrio=puente de vallecas+0Barrio=arganzuela... <= 0
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... <= 0.851962
-0.533Barrio=sant andreu-0.533Barrio=sant andreu+0.377Barrio=horta guinardó+0.377Barrio=horta guinardó+0.21 Barrio=sants-montjuïc... <= 0.001265
-0.5Barrio=san blas-0.5Barrio=san blas+0.5 Barrio=puente de vallecas+0.5 Barrio=puente de vallecas+0Barrio=arganzuela... <= -1
0.684Tipo=pisos-0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... <= 0.328874
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... <= -0.098144: NO (3.0/1.0)
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... > -0.098144: SÍ (6.0/1.0)
0.684Tipo=pisos-0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... > 0.328874: NO (118.0/52.0)
-0.5Barrio=san blas-0.5Barrio=san blas+0.5 Barrio=puente de vallecas+0.5 Barrio=puente de vallecas+0Barrio=arganzuela... > -1
0.5 Barrio=gràcia+0.5 Barrio=gràcia-0.5Barrio=ciutat vella-0.5Barrio=ciutat vella+0Barrio=usera... <= -1
0.735Euros-0.621Tipo=chalets+0.219Tipo=estudios+0.119Tipo=pisos+0.075Tipo=dúplex... <= 0.173222
0.684Tipo=pisos-0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... <= 0.49319
-0.471Planta=p6+0.313Dorm=d0-0.294Planta=pst+0.288Planta=p5+0.261Dorm=d9... <= -0.270969: NO (3.0)
-0.471Planta=p6+0.313Dorm=d0-0.294Planta=pst+0.288Planta=p5+0.261Dorm=d9... > -0.270969: SÍ (7.0/1.0)
0.684Tipo=pisos-0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... > 0.49319: SÍ (35.0/15.0)
0.735Euros-0.621Tipo=chalets+0.219Tipo=estudios+0.119Tipo=pisos+0.075Tipo=dúplex... > 0.173222: NO (19.0/1.0)
0.5 Barrio=gràcia+0.5 Barrio=gràcia-0.5Barrio=ciutat vella-0.5Barrio=ciutat vella+0Barrio=usera... <= -1
-0.5Barrio=Extramurs-0.5Barrio=Extramurs+0.5 Barrio=Jesús+0.5 Barrio=Jesús+0Barrio=L'eixample... <= -1
0.684Tipo=pisos-0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... <= 0.49319: SÍ (4.0/1.0)
0.684Tipo=pisos-0.493Tipo=áticos+0.329Tipo=chalets+0.311Tipo=dúplex+0.291Tipo=estudios... > 0.49319
0.422Dorm=d6+0.398Dorm=d0-0.384Planta=pen-0.376Planta=pss-0.238Planta=pbj... <= -0.078657: NO (4.0)
0.422Dorm=d6+0.398Dorm=d0-0.384Planta=pen-0.376Planta=pss-0.238Planta=pbj... > -0.078657
0.484Planta=p2-0.436Planta=p5-0.385Planta=pbj+0.378Planta=p3-0.287Dorm=d9... <= 0.553982
-0.724Planta=p5+0.287Planta=p8+0.26 Dorm=d9+0.237Dorm=d4+0.221Dorm=d0... <= -0.689203: NO (2.0)
-0.724Planta=p5+0.287Planta=p8+0.26 Dorm=d9+0.237Dorm=d4+0.221Dorm=d0... > -0.689203
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... <= 0.850102: SÍ (11.0)
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... > 0.850102
-0.621Planta=p7+0.427Planta=p6-0.363Planta=p8-0.241Planta=pen-0.218Planta=p10... <= 0.06539: SÍ (3.0)
-0.621Planta=p7+0.427Planta=p6-0.363Planta=p8-0.241Planta=pen-0.218Planta=p10... > 0.06539: NO (3.0)
0.484Planta=p2-0.436Planta=p5-0.385Planta=pbj+0.378Planta=p3-0.287Dorm=d9... > 0.553982: NO (3.0)
-0.5Barrio=Extramurs-0.5Barrio=Extramurs+0.5 Barrio=Jesús+0.5 Barrio=Jesús+0Barrio=L'eixample... > -1
0.735Euros-0.621Tipo=chalets+0.219Tipo=estudios+0.119Tipo=pisos+0.075Tipo=dúplex... <= 0.147138
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza0Barrio=fuencarral... <= -1: SÍ (63.0/26.0)
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza0Barrio=fuencarral... > -1
0.529Barrio=benicalap+0.529Barrio=benicalap-0.317Barrio=Poblats marítims-0.317Barrio=Poblats marítims-0.222Barrio=Algirós... <= -0.365145
-0.471Planta=p6+0.313Dorm=d0-0.294Planta=pst+0.288Planta=p5+0.261Dorm=d9... <= 0.198858
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... <= -0.059673: NO (10.0)
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... > -0.059673
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... <= 0.283294
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... <= 0.850526: SÍ (4.0)
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... > 0.850526: NO (3.0/1.0)
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... > 0.283294: NO (3.0)
-0.471Planta=p6+0.313Dorm=d0-0.294Planta=pst+0.288Planta=p5+0.261Dorm=d9... > 0.198858: SÍ (5.0)
0.529Barrio=benicalap+0.529Barrio=benicalap-0.317Barrio=Poblats marítims-0.317Barrio=Poblats marítims-0.222Barrio=Algirós... > -0.365145: NO (642.0/292.0)
0.735Euros-0.621Tipo=chalets+0.219Tipo=estudios+0.119Tipo=pisos+0.075Tipo=dúplex... > 0.147138
0.514Barrio=Poblats marítims+0.514Barrio=Poblats marítims-0.435Barrio=Algirós-0.435Barrio=Algirós+0.144Barrio=benicalap... <= -0.172235
0.422Dorm=d6+0.398Dorm=d0-0.384Planta=pen-0.376Planta=pss-0.238Planta=pbj... <= -0.078749
-0.565Planta=pen+0.518Planta=pss-0.509Planta=p11+0.272Planta=p10+0.202Planta=p7... <= -0.012422
0.484Planta=p2-0.436Planta=p5-0.385Planta=pbj+0.378Planta=p3-0.287Dorm=d9... <= 0.01219: NO (2.0)
0.484Planta=p2-0.436Planta=p5-0.385Planta=pbj+0.378Planta=p3-0.287Dorm=d9... > 0.01219: SÍ (2.0)
-0.565Planta=pen+0.518Planta=pss-0.509Planta=p11+0.272Planta=p10+0.202Planta=p7... > -0.012422: NO (7.0)
0.422Dorm=d6+0.398Dorm=d0-0.384Planta=pen-0.376Planta=pss-0.238Planta=pbj... > -0.078749
0.632Dorm=d8+0.321Planta=p11+0.289Planta=pss+0.257Planta=p13+0.226Planta=p10... <= 0.104215
-0.727Metros+0.374Dorm=d5+0.283Dorm=d7+0.279Dorm=d8+0.223Dorm=d6... <= 0.026494: SÍ (11.0)
-0.727Metros+0.374Dorm=d5+0.283Dorm=d7+0.279Dorm=d8+0.223Dorm=d6... > 0.026494: NO (2.0)
0.632Dorm=d8+0.321Planta=p11+0.289Planta=pss+0.257Planta=p13+0.226Planta=p10... > 0.104215: NO (2.0)
0.514Barrio=Poblats marítims+0.514Barrio=Poblats marítims-0.435Barrio=Algirós-0.435Barrio=Algirós+0.144Barrio=benicalap... > -0.172235
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza0Barrio=fuencarral... <= -1
-0.724Planta=p5+0.287Planta=p8+0.26 Dorm=d9+0.237Dorm=d4+0.221Dorm=d0... <= 0.025695: NO (35.0/8.0)
-0.724Planta=p5+0.287Planta=p8+0.26 Dorm=d9+0.237Dorm=d4+0.221Dorm=d0... > 0.025695: SÍ (50.0/17.0)
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza0Barrio=fuencarral... > -1
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza0Barrio=fuencarral... <= 0
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... <= 0.108364
0.5 Barrio=gràcia+0.5 Barrio=gràcia-0.5Barrio=ciutat vella-0.5Barrio=ciutat vella+0Barrio=usera... <= 0: SÍ (777.0/356.0)
0.5 Barrio=gràcia+0.5 Barrio=gràcia-0.5Barrio=ciutat vella-0.5Barrio=ciutat vella+0Barrio=usera... > 0
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... <= -0.561594: NO (5.0)
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... > -0.561594
0.575Euros+0.549Tipo=chalets+0.378Estado-0.378Estado-0.269Tipo=dúplex... <= 0.55493: SÍ (25.0/8.0)
0.575Euros+0.549Tipo=chalets+0.378Estado-0.378Estado-0.269Tipo=dúplex... > 0.55493: NO (3.0)
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... > 0.108364
0.632Dorm=d8+0.321Planta=p11+0.289Planta=pss+0.257Planta=p13+0.226Planta=p10... <= -0.00666
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... <= -0.337086: NO (9.0/1.0)
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... > -0.337086
-0.597Barrio=Camins al grau-0.597Barrio=Camins al grau+0.159Barrio=Algirós+0.159Barrio=Algirós+0.159Barrio=La saïdia... <= -0.000402
-0.546Dorm=d9-0.45Planta=pst+0.323Planta=pbj-0.226Planta=pss-0.22Planta=p12... <= -0.135686: SÍ (6.0)
-0.546Dorm=d9-0.45Planta=pst+0.323Planta=pbj-0.226Planta=pss-0.22Planta=p12... > -0.135686
0.685Barrio=les corts+0.685Barrio=les corts-0.12Barrio=nou barris-0.12Barrio=nou barris-0.064Barrio=ciutat vella... <= -0.001238: NO (8.0)
0.685Barrio=les corts+0.685Barrio=les corts-0.12Barrio=nou barris-0.12Barrio=nou barris-0.064Barrio=ciutat vella... > -0.001238
-0.647Planta=p10+0.427Planta=p9+0.341Planta=p12+0.231Dorm=d8-0.228Planta=pst... <= 0.005444
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... <= 0.843349: SÍ (3.0)
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... > 0.843349: NO (3.0/1.0)
-0.647Planta=p10+0.427Planta=p9+0.341Planta=p12+0.231Dorm=d8-0.228Planta=pst... > 0.005444: NO (3.0)
-0.597Barrio=Camins al grau-0.597Barrio=Camins al grau+0.159Barrio=Algirós+0.159Barrio=Algirós+0.159Barrio=La saïdia... > -0.000402
0.306Barrio=Extramurs-0.303Barrio=Extramurs-0.294Barrio=usera+0.292Barrio=usera-0.197Barrio=villa de vallecas... <= 0.001427: SÍ (8.0)

0.306Barrio=Extramurs-0.303Barrio=Extramurs-0.294Barrio=usera+0.292Barrio=usera-0.197Barrio=usera-villa de vallecas... > 0.001427
0.5 Barrio=gràcia+0.5 Barrio=gràcia-0.5Barrio=ciutat vella-0.5Barrio=ciutat vella+0Barrio=usera... <= 0
-0.562Barrio=sarrià-sant gervasi-0.562Barrio=sarrià-sant gervasi+0.281Barrio=sants-montjuïc+0.281Barrio=sants-
montjuïc+0.218Barrio=horta guinardó... <= -0.003471
-0.727Metros+0.374Dorm=d5+0.283Dorm=d7+0.279Dorm=d8+0.223Dorm=d6... <= -0.138651
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... <= -0.284544: SÍ (10.0)
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... > -0.284544
-0.533Barrio=sant andreu-0.533Barrio=sant andreu+0.377Barrio=horta guinardó+0.377Barrio=horta guinardó+0.21
Barrio=sants-montjuïc... <= 0.000952
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample...
<= -0.635816
0.575Euros+0.549Tipo=chalets+0.378Estado-0.3Tipo=estudios-0.269Tipo=dúplex... <= 0.53108: NO (2.0)
0.575Euros+0.549Tipo=chalets+0.378Estado-0.3Tipo=estudios-0.269Tipo=dúplex... > 0.53108: SÍ (2.0)
-0.504Ciudad=madrid+0.412Ciudad=barcelona+0.221Ciudad=valencia+0.167Barrio=eixample+0.167Barrio=eixample...
> -0.635816
-0.409Planta=p1+0.379Dorm=d0+0.361Planta=pss+0.359Planta=p8+0.325Dorm=d7... <= -0.120298: SÍ (6.0/1.0)
-0.409Planta=p1+0.379Dorm=d0+0.361Planta=pss+0.359Planta=p8+0.325Dorm=d7... > -0.120298
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... <= 0.017561: SÍ (4.0/1.0)
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... > 0.017561: NO (3.0)
-0.533Barrio=sant andreu-0.533Barrio=sant andreu+0.377Barrio=horta guinardó+0.377Barrio=horta guinardó+0.21
Barrio=sants-montjuïc... > 0.000952
0.64 Metros+0.362Dorm=d5+0.288Planta=p1-0.263Dorm=d2+0.255Dorm=d4... <= -0.114451: NO (4.0)
0.64 Metros+0.362Dorm=d5+0.288Planta=p1-0.263Dorm=d2+0.255Dorm=d4... > -0.114451: SÍ (3.0/1.0)
-0.727Metros+0.374Dorm=d5+0.283Dorm=d7+0.279Dorm=d8+0.223Dorm=d6... > -0.138651: SÍ (19.0)
-0.562Barrio=sarrià-sant gervasi-0.562Barrio=sarrià-sant gervasi+0.281Barrio=sants-montjuïc+0.281Barrio=sants-
montjuïc+0.218Barrio=horta guinardó... > -0.003471
0.527Barrio=sants-montjuïc+0.527Barrio=sants-montjuïc-0.456Barrio=horta guinardó-0.456Barrio=horta guinardó-
0.091Barrio=sant andreu... <= 0.000821
<= 0.17075
0.539Barrio=fuencarral+0.539Barrio=fuencarral-0.313Barrio=hortaleza-0.313Barrio=hortaleza-0.313Barrio=ciudad lineal...
<= 0.17075
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... <= 0.181959: NO (18.0/8.0)
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... > 0.181959: SÍ (5.0)
0.539Barrio=fuencarral+0.539Barrio=fuencarral-0.313Barrio=hortaleza-0.313Barrio=hortaleza-0.313Barrio=ciudad lineal... >
0.17075
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... <= 0.147539
0.405Planta=p4-0.4Planta=p3+0.337Planta=p2-0.327Planta=p1+0.314Planta=pbj... <= 0.323935: NO (8.0/1.0)
0.405Planta=p4-0.4Planta=p3+0.337Planta=p2-0.327Planta=p1+0.314Planta=pbj... > 0.323935: SÍ (2.0)
-0.496Planta=pst+0.491Planta=p11+0.281Planta=pss-0.264Planta=p13-0.25Dorm=d0... > 0.147539: SÍ (3.0)
0.527Barrio=sants-montjuïc+0.527Barrio=sants-montjuïc-0.456Barrio=horta guinardó-0.456Barrio=horta guinardó-
0.091Barrio=sant andreu... > 0.000821
-0.409Planta=p1+0.379Dorm=d0+0.361Planta=pss+0.359Planta=p8+0.325Dorm=d7... <= -0.120736: NO (4.0)
-0.409Planta=p1+0.379Dorm=d0+0.361Planta=pss+0.359Planta=p8+0.325Dorm=d7... > -0.120736
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... <= -0.488656: NO (4.0/1.0)
-0.507Dorm=d4-0.432Dorm=d1+0.416Dorm=d2-0.261Planta=pst+0.212Planta=p6... > -0.488656: SÍ (6.0)
0.5 Barrio=gràcia+0.5 Barrio=gràcia-0.5Barrio=ciutat vella-0.5Barrio=ciutat vella+0Barrio=usera... > 0: SÍ (7.0/1.0)
0.632Dorm=d8+0.321Planta=p11+0.289Planta=pss+0.257Planta=p13+0.226Planta=p10... > -0.00666
0.539Barrio=fuencarral+0.539Barrio=fuencarral-0.313Barrio=hortaleza-0.313Barrio=hortaleza-0.313Barrio=ciudad lineal... <= -0.056681:
NO (2.0)
0.539Barrio=fuencarral+0.539Barrio=fuencarral-0.313Barrio=hortaleza-0.313Barrio=hortaleza-0.313Barrio=ciudad lineal... > -0.056681
-0.552Barrio=Ciutat vella-0.552Barrio=Ciutat vella+0.276Barrio=L'eixample+0.276Barrio=L'eixample-0.235Barrio=Quatre carreres...
<= -0.00067
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... <= -0.089347: NO (2.0)
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... > -0.089347: SÍ (4.0)
-0.552Barrio=Ciutat vella-0.552Barrio=Ciutat vella+0.276Barrio=L'eixample+0.276Barrio=L'eixample-0.235Barrio=Quatre carreres... >
-0.00067: SÍ (37.0/2.0)
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza+0Barrio=fuencarral... > 0
0.807Estado+0.526Tipo=dúplex-0.159Tipo=pisos-0.139Euros-0.1Tipo=chalets... <= 0.640018
0.422Dorm=d6+0.398Dorm=d0-0.384Planta=pen-0.376Planta=pss-0.238Planta=pbj... <= -0.0695
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... <= -0.171198: NO (4.0)
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... > -0.171198
-0.536Dorm=d6-0.395Planta=p3+0.368Planta=p4-0.305Planta=p9+0.238Dorm=d4... <= 0.173228: SÍ (13.0/2.0)
-0.536Dorm=d6-0.395Planta=p3+0.368Planta=p4-0.305Planta=p9+0.238Dorm=d4... > 0.173228
0.405Planta=p4-0.4Planta=p3+0.337Planta=p2-0.327Planta=p1+0.314Planta=pbj... <= -0.175161: SÍ (6.0/1.0)
0.405Planta=p4-0.4Planta=p3+0.337Planta=p2-0.327Planta=p1+0.314Planta=pbj... > -0.175161: NO (5.0)
0.422Dorm=d6+0.398Dorm=d0-0.384Planta=pen-0.376Planta=pss-0.238Planta=pbj... > -0.0695: SÍ (19.0/1.0)
0.807Estado+0.526Tipo=dúplex-0.159Tipo=pisos-0.139Euros-0.1Tipo=chalets... > 0.640018
0.484Planta=p2-0.436Planta=p5-0.385Planta=pbj+0.378Planta=p3-0.287Dorm=d9... <= -0.004359
0.484Planta=p2-0.436Planta=p5-0.385Planta=pbj+0.378Planta=p3-0.287Dorm=d9... <= -0.366
0.542Planta=p7+0.519Planta=p6-0.375Planta=p8-0.26Planta=p4+0.222Planta=pen... <= 0.015449: SÍ (4.0/1.0)
0.542Planta=p7+0.519Planta=p6-0.375Planta=p8-0.26Planta=p4+0.222Planta=pen... > 0.015449: NO (5.0)
0.484Planta=p2-0.436Planta=p5-0.385Planta=pbj+0.378Planta=p3-0.287Dorm=d9... > -0.366: NO (16.0)
0.484Planta=p2-0.436Planta=p5-0.385Planta=pbj+0.378Planta=p3-0.287Dorm=d9... > -0.004359
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... <= -0.248292: NO (24.0/6.0)
-0.524Planta=p2+0.476Planta=p3+0.434Planta=p4+0.275Dorm=d5-0.275Dorm=d0... > -0.248292
0.515Dorm=d8-0.455Planta=p13-0.423Planta=p14-0.312Planta=p11-0.24Planta=p22... <= 0.055687: SÍ (13.0/1.0)
0.515Dorm=d8-0.455Planta=p13-0.423Planta=p14-0.312Planta=p11-0.24Planta=p22... > 0.055687
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... <= 0.400321: NO (7.0/2.0)
-0.482Dorm=d1-0.481Planta=p1+0.31 Dorm=d4+0.299Dorm=d2+0.262Planta=p2... > 0.400321: SÍ (2.0)
-0.533Barrio=sant andreu-0.533Barrio=sant andreu+0.377Barrio=horta guinardó+0.377Barrio=horta guinardó+0.21 Barrio=sants-montjuïc... > 0.001265
-0.727Metros+0.374Dorm=d5+0.283Dorm=d7+0.279Dorm=d8+0.223Dorm=d6... <= -0.242291: SÍ (4.0)
-0.727Metros+0.374Dorm=d5+0.283Dorm=d7+0.279Dorm=d8+0.223Dorm=d6... > -0.242291: NO (55.0/14.0)
0.68 Tipo=pisos-0.417Tipo=áticos-0.371Tipo=chalets-0.299Euros-0.259Tipo=dúplex... > 0.851962
0.663Barrio=nou barris+0.663Barrio=nou barris-0.138Barrio=ciutat vella-0.138Barrio=ciutat vella-0.138Barrio=gràcia... <= -0.001216: SÍ (84.0/22.0)
0.663Barrio=nou barris+0.663Barrio=nou barris-0.138Barrio=ciutat vella-0.138Barrio=ciutat vella-0.138Barrio=gràcia... > -0.001216
-0.575Barrio=Patraix-0.575Barrio=Patraix+0.315Barrio=benicalap+0.315Barrio=benicalap+0.264Barrio=Algirós... <= 0
0.855Tipo=estudios-0.351Tipo=áticos-0.328Tipo=dúplex+0.167Tipo=chalets+0.081Estado... <= 0.039296
0.577Barrio=L'eixample+0.577Barrio=L'eixample-0.266Barrio=Jesús-0.266Barrio=Jesús-0.266Barrio=Extramurs... <= 0.000345
-0.647Planta=p10+0.427Planta=p9+0.341Planta=p12+0.231Dorm=d8-0.228Planta=pst... <= 0.024454
-0.599Barrio-La saïdia-0.599Barrio-La saïdia+0.302Barrio=Algirós+0.302Barrio=Algirós+0.189Barrio=Poblat marítims... <= 0
-0.621Planta=p7+0.427Planta=p6-0.363Planta=p8-0.241Planta=pen-0.218Planta=p10... <= -0.002419: SÍ (14.0)
-0.621Planta=p7+0.427Planta=p6-0.363Planta=p8-0.241Planta=pen-0.218Planta=p10... > -0.002419
-0.5Barrio-san blas-0.5Barrio-san blas+0.5 Barrio=puento de vallecas+0.5 Barrio=puento de vallecas+0Barrio=arganzuela... <= -1
-0.727Metros+0.374Dorm=d5+0.283Dorm=d7+0.279Dorm=d8+0.223Dorm=d6... <= -0.116754: SÍ (4.0)
-0.727Metros+0.374Dorm=d5+0.283Dorm=d7+0.279Dorm=d8+0.223Dorm=d6... > -0.116754: NO (4.0)
-0.5Barrio-san blas-0.5Barrio-san blas+0.5 Barrio=puento de vallecas+0.5 Barrio=puento de vallecas+0Barrio=arganzuela... > -1
0.5 Barrio=ciudad lineal+0.5 Barrio=ciudad lineal-0.5Barrio=hortaleza-0.5Barrio=hortaleza+0Barrio=fuencarral... <= 0
0.725Dorm=d3-0.501Dorm=d2-0.201Planta=pbj-0.173Dorm=d1-0.162Planta=p1... <= 0.796517
0.514Barrio=eixample+0.514Barrio=eixample-0.485Barrio=sant martí-0.485Barrio=sant martí-0.012Barrio=sarrià-sant gervasi... <= 0.000441
0.539Barrio=fuencarral+0.539Barrio=fuencarral-0.313Barrio=hortaleza-0.313Barrio=hortaleza-0.313Barrio=ciudad lineal... <= -
0.023725: NO (3.0)

-0.713Dorm=d3+0.469Dorm=d2+0.301Planta=p1+0.23 Dorm=d1-0.21Planta=p4... > -0.147226: SÍ (13.0)
-0.484Planta=p3+0.407Planta=p5-0.351Dorm=d5-0.316Planta=p2-0.291Dorm=d1... > 0.286483: NO (8.0)
0.689Barrio=L'olivereta-0.596Barrio=Patraix+0.298Barrio=El pla del real-0.242Barrio=Algirós-0.149Barrio=benicalap... > -0.148613
-0.674Ciudad=madrid-0.581Ciudad=barcelona-0.456Ciudad=valencia+0Estado+0Euros <= -0.58066
-0.674Ciudad=madrid-0.581Ciudad=barcelona-0.456Ciudad=valencia+0Estado+0Euros <= -0.674163
0.58 Tipo=dúplex-0.358Tipo=áticos-0.272Barrio=Poblats marítims+0.231Barrio=horta guinardó-0.219Barrio=Ciutat vella... <= 0.154273: NO (997.0/490.0)
0.58 Tipo=dúplex-0.358Tipo=áticos-0.272Barrio=Poblats marítims+0.231Barrio=horta guinardó-0.219Barrio=Ciutat vella... > 0.154273: SÍ (96.0/28.0)
-0.674Ciudad=madrid-0.581Ciudad=barcelona-0.456Ciudad=valencia+0Estado+0Euros > -0.674163: SÍ (300.0/133.0)
-0.674Ciudad=madrid-0.581Ciudad=barcelona-0.456Ciudad=valencia+0Estado+0Euros > -0.58066
-0.638Barrio=ciutat vella-0.459Barrio=gràcia+0.35 Barrio=sants-montjuïc+0.316Barrio=horta guinardó+0.247Barrio=sant andreu... <= 0.107649
-0.484Planta=p3+0.407Planta=p5-0.351Dorm=d5-0.316Planta=p2-0.291Dorm=d1... <= -0.142518
0.689Barrio=L'olivereta-0.596Barrio=Patraix+0.298Barrio=El pla del real-0.242Barrio=Algirós-0.149Barrio=benicalap... <= 0
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado <= -0.441987: SÍ (2.0)
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado > -0.441987
0.825Barrio=Benimaclet-0.381Barrio=vicálvaro-0.34Barrio=El pla del real+0.15 Barrio=campanar - beni...-0.145Barrio=Patraix... <= -0.037375
0.652Metros+0.334Dorm=d5+0.31 Dorm=d4+0.299Dorm=d6-0.297Dorm=d2... <= -0.330943: NO (2.0)
0.652Metros+0.334Dorm=d5+0.31 Dorm=d4+0.299Dorm=d6-0.297Dorm=d2... > -0.330943: SÍ (3.0)
0.825Barrio=Benimaclet-0.381Barrio=vicálvaro-0.34Barrio=El pla del real+0.15 Barrio=campanar - beni...-0.145Barrio=Patraix... > -0.037375: NO (21.0)
0.689Barrio=L'olivereta-0.596Barrio=Patraix+0.298Barrio=El pla del real-0.242Barrio=Algirós-0.149Barrio=benicalap... > 0: NO (4.0)
-0.484Planta=p3+0.407Planta=p5-0.351Dorm=d5-0.316Planta=p2-0.291Dorm=d1... > -0.142518
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado <= -0.436414
-0.759Barrio=benicalap+0.642Barrio=Algirós+0.1 Barrio=L'olivereta+0.035Barrio=Patraix-0.018Barrio=El pla del real... <= -0.017577
0.502Tipo=áticos-0.39Tipo=pisos+0.322Tipo=estudios-0.322Metros-0.293Tipo=chalets... <= -0.356637: NO (2.0)
0.502Tipo=áticos-0.39Tipo=pisos+0.322Tipo=estudios-0.322Metros-0.293Tipo=chalets... > -0.356637: SÍ (3.0)
-0.759Barrio=benicalap+0.642Barrio=Algirós+0.1 Barrio=L'olivereta+0.035Barrio=Patraix-0.018Barrio=El pla del real... > -0.017577
-0.407Planta=pen+0.345Dorm=d6+0.324Dorm=d0+0.314Dorm=d1+0.272Planta=p7... <= 0.014898
-0.759Barrio=benicalap+0.642Barrio=Algirós+0.1 Barrio=L'olivereta+0.035Barrio=Patraix-0.018Barrio=El pla del real... <= 0.035155: SÍ (51.0/16.0)
-0.759Barrio=benicalap+0.642Barrio=Algirós+0.1 Barrio=L'olivereta+0.035Barrio=Patraix-0.018Barrio=El pla del real... > 0.035155
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado <= -0.441786: SÍ (2.0)
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado > -0.441786: NO (2.0)
-0.407Planta=pen+0.345Dorm=d6+0.324Dorm=d0+0.314Dorm=d1+0.272Planta=p7... > 0.014898
0.838Barrio=La saïdia+0.252Barrio=L'olivereta+0.208Barrio=benicalap+0.198Barrio=Algirós-0.198Barrio=Extramurs... <= -0.150234: SÍ (6.0/1.0)
0.838Barrio=La saïdia+0.252Barrio=L'olivereta+0.208Barrio=benicalap+0.198Barrio=Algirós-0.198Barrio=Extramurs... > -0.150234
0.495Barrio=fuencarral-0.486Barrio=chamberí+0.46 Barrio=sant martí-0.32Barrio=latina+0.289Barrio=sarrià-sant gervasi... <= 0.026327: NO (13.0)
0.495Barrio=fuencarral-0.486Barrio=chamberí+0.46 Barrio=sant martí-0.32Barrio=latina+0.289Barrio=sarrià-sant gervasi... > 0.026327
0.748Barrio=moratatalaz-0.606Barrio=usera-0.233Barrio=nou barris+0.065Barrio=horta guinardó+0.061Barrio=Camins al grau... <= -0.018693: NO (3.0)
0.748Barrio=moratatalaz-0.606Barrio=usera-0.233Barrio=nou barris+0.065Barrio=horta guinardó+0.061Barrio=Camins al grau... > -0.018693: SÍ (4.0/1.0)
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado > -0.436414
0.43 Planta=p5-0.407Planta=p9-0.349Planta=p6-0.324Planta=pen-0.275Planta=p7... <= 0.038141: NO (10.0)
0.43 Planta=p5-0.407Planta=p9-0.349Planta=p6-0.324Planta=pen-0.275Planta=p7... > 0.038141
-0.623Barrio=puede de vallecas+0.394Barrio=tetuán-0.301Barrio=carabanchel-0.278Barrio=ciudad lineal+0.27 Barrio=san blas... <= 0.011106: NO (2.0)
-0.623Barrio=puede de vallecas+0.394Barrio=tetuán-0.301Barrio=carabanchel-0.278Barrio=ciudad lineal+0.27 Barrio=san blas... > 0.011106: SÍ (2.0)
-0.638Barrio=ciutat vella-0.459Barrio=gràcia+0.35 Barrio=sants-montjuïc+0.316Barrio=horta guinardó+0.247Barrio=sant andreu... > 0.107649: SÍ (7.0)
0.653Planta=p14-0.515Planta=pst-0.303Planta=p13+0.295Planta=p12-0.263Planta=p10... > -0.005539
-0.759Barrio=benicalap+0.642Barrio=Algirós+0.1 Barrio=L'olivereta+0.035Barrio=Patraix-0.018Barrio=El pla del real... <= 0
-0.674Ciudad=madrid-0.581Ciudad=barcelona-0.456Ciudad=valencia+0Estado+0Euros <= -0.674163: SÍ (1783.0/778.0)
-0.674Ciudad=madrid-0.581Ciudad=barcelona-0.456Ciudad=valencia+0Estado+0Euros > -0.674163
-0.674Ciudad=madrid-0.581Ciudad=barcelona-0.456Ciudad=valencia+0Estado+0Euros <= -0.58066
-0.638Barrio=ciutat vella-0.459Barrio=gràcia+0.35 Barrio=sants-montjuïc+0.316Barrio=horta guinardó+0.247Barrio=sant andreu... <= -0.638457
-0.698Barrio=gràcia+0.637Barrio=ciutat vella+0.162Barrio=villaverde-0.131Barrio=retiro+0.091Barrio=moncloa... <= 0.635363: SÍ (3.0)
-0.698Barrio=gràcia+0.637Barrio=ciutat vella+0.162Barrio=villaverde-0.131Barrio=retiro+0.091Barrio=moncloa... > 0.635363
0.653Planta=p14-0.515Planta=pst-0.303Planta=p13+0.295Planta=p12-0.263Planta=p10... <= 0.012026: NO (14.0)
0.653Planta=p14-0.515Planta=pst-0.303Planta=p13+0.295Planta=p12-0.263Planta=p10... > 0.012026
-0.738Planta=p4+0.422Planta=p2-0.292Dorm=d9-0.227Dorm=d5-0.166Planta=pbj... <= 0.267493: SÍ (10.0/3.0)
-0.738Planta=p4+0.422Planta=p2-0.292Dorm=d9-0.227Dorm=d5-0.166Planta=pbj... > 0.267493: NO (4.0)
-0.638Barrio=ciutat vella-0.459Barrio=gràcia+0.35 Barrio=sants-montjuïc+0.316Barrio=horta guinardó+0.247Barrio=sant andreu... > -0.638457: SÍ (522.0/199.0)
-0.674Ciudad=madrid-0.581Ciudad=barcelona-0.456Ciudad=valencia+0Estado+0Euros > -0.58066
0.689Barrio=L'olivereta-0.596Barrio=Patraix+0.298Barrio=El pla del real-0.242Barrio=Algirós-0.149Barrio=benicalap... <= -0.148613
-0.743Barrio=centro+0.438Barrio=salamanca+0.248Barrio=Camins al grau+0.222Barrio=ciudad lineal+0.217Barrio=tetuán... <= -0.024911: NO (4.0)
-0.743Barrio=centro+0.438Barrio=salamanca+0.248Barrio=Camins al grau+0.222Barrio=ciudad lineal+0.217Barrio=tetuán... > -0.024911: SÍ (9.0/1.0)
0.689Barrio=L'olivereta-0.596Barrio=Patraix+0.298Barrio=El pla del real-0.242Barrio=Algirós-0.149Barrio=benicalap... > -0.148613
-0.448Dorm=d6+0.407Planta=p9+0.384Dorm=d5+0.284Planta=pen+0.25 Dorm=d9... <= -0.281998: NO (12.0/2.0)
-0.448Dorm=d6+0.407Planta=p9+0.384Dorm=d5+0.284Planta=pen+0.25 Dorm=d9... > -0.281998: SÍ (210.0/79.0)
-0.759Barrio=benicalap+0.642Barrio=Algirós+0.1 Barrio=L'olivereta+0.035Barrio=Patraix-0.018Barrio=El pla del real... > 0
-0.583Planta=p12-0.499Planta=p10-0.33Planta=p11-0.295Planta=p15+0.269Planta=p9... <= -0.010913: NO (6.0)
-0.583Planta=p12-0.499Planta=p10-0.33Planta=p11-0.295Planta=p15+0.269Planta=p9... > -0.010913
-0.759Barrio=benicalap+0.642Barrio=Algirós+0.1 Barrio=L'olivereta+0.035Barrio=Patraix-0.018Barrio=El pla del real... <= 0.035155
0.652Metros+0.334Dorm=d5+0.31 Dorm=d4+0.299Dorm=d6-0.297Dorm=d2... <= 0.030284
-0.669Planta=p3+0.617Planta=p2+0.2 Dorm=d0+0.182Dorm=d5+0.171Planta=p4... <= 0.017701: NO (3.0)
-0.669Planta=p3+0.617Planta=p2+0.2 Dorm=d0+0.182Dorm=d5+0.171Planta=p4... > 0.017701: SÍ (7.0/1.0)
0.652Metros+0.334Dorm=d5+0.31 Dorm=d4+0.299Dorm=d6-0.297Dorm=d2... > 0.030284: NO (5.0)
-0.759Barrio=benicalap+0.642Barrio=Algirós+0.1 Barrio=L'olivereta+0.035Barrio=Patraix-0.018Barrio=El pla del real... > 0.035155
-0.713Dorm=d3+0.469Dorm=d2+0.301Planta=p1+0.23 Dorm=d1-0.21Planta=p4... <= -0.798477: NO (3.0)
-0.713Dorm=d3+0.469Dorm=d2+0.301Planta=p1+0.23 Dorm=d1-0.21Planta=p4... > -0.798477
0.725Metros-0.349Dorm=d5-0.331Dorm=d6-0.214Dorm=d4+0.212Dorm=d1... <= 0.035058: NO (3.0)
0.725Metros-0.349Dorm=d5-0.331Dorm=d6-0.214Dorm=d4+0.212Dorm=d1... > 0.035058
-0.669Planta=p8-0.468Planta=p11+0.246Planta=p13+0.226Dorm=d7-0.208Dorm=d5... <= 0.080005: SÍ (13.0)
-0.669Planta=p8-0.468Planta=p11+0.246Planta=p13+0.226Dorm=d7-0.208Dorm=d5... > 0.080005: NO (5.0/2.0)
-0.78Estado-0.402Euros-0.385Ciudad=valencia+0.287Ciudad=barcelona+0.013Ciudad=madrid > -0.493104
-0.595Barrio=sant martí+0.451Barrio=fuencarral-0.325Barrio=sarrià-sant gervasi+0.287Barrio=latina+0.27 Barrio=Camins al grau... <= 0.150727
0.872Barrio=chamartín-0.331Barrio=eixample-0.286Barrio=hortaleza-0.087Barrio=moncloa-0.079Barrio=horta guinardó... <= -0.047605
0.249Barrio=centro+0.222Barrio=salamanca+0.217Barrio=ciudad lineal+0.214Barrio=carabanchel+0.21 Barrio=chamartín... <= 0.230455: NO (6.0)
0.249Barrio=centro+0.222Barrio=salamanca+0.217Barrio=ciudad lineal+0.214Barrio=carabanchel+0.21 Barrio=chamartín... > 0.230455
-0.698Barrio=gràcia+0.637Barrio=ciutat vella+0.162Barrio=villaverde-0.131Barrio=retiro+0.091Barrio=moncloa... <= -0.031583: NO (5.0)
-0.698Barrio=gràcia+0.637Barrio=ciutat vella+0.162Barrio=villaverde-0.131Barrio=retiro+0.091Barrio=moncloa... > -0.031583
0.249Barrio=centro+0.222Barrio=salamanca+0.217Barrio=ciudad lineal+0.214Barrio=carabanchel+0.21 Barrio=chamartín... <= 0.278774
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado <= 0.732693
0.803Barrio=Jesús+0.313Barrio=Rascanya-0.234Barrio=Quatre carreres-0.225Barrio=L'eixample+0.161Barrio=Ciutat vella... <= 0.008711
-0.481Planta=pbj+0.408Planta=p1-0.355Planta=pen-0.337Dorm=d6+0.323Dorm=d4... <= 0.413698: SÍ (19.0/2.0)
-0.481Planta=pbj+0.408Planta=p1-0.355Planta=pen-0.337Dorm=d6+0.323Dorm=d4... > 0.413698: NO (5.0/1.0)
0.803Barrio=Jesús+0.313Barrio=Rascanya-0.234Barrio=Quatre carreres-0.225Barrio=L'eixample+0.161Barrio=Ciutat vella... > 0.008711
0.43 Planta=p5-0.407Planta=p9-0.349Planta=p6-0.324Planta=pen-0.275Planta=p7... <= 0.05693: NO (11.0/2.0)
0.43 Planta=p5-0.407Planta=p9-0.349Planta=p6-0.324Planta=pen-0.275Planta=p7... > 0.05693: SÍ (3.0)
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado > 0.732693: SÍ (15.0)
0.249Barrio=centro+0.222Barrio=salamanca+0.217Barrio=ciudad lineal+0.214Barrio=carabanchel+0.21 Barrio=chamartín... > 0.278774
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado <= 0.731565: SÍ (4.0)
0.726Ciudad=madrid-0.571Ciudad=barcelona-0.346Ciudad=valencia+0.135Euros-0.097Estado > 0.731565
-0.484Planta=p3+0.407Planta=p5-0.351Dorm=d5-0.316Planta=p2-0.291Dorm=d1... <= 0.311298: NO (17.0)
-0.484Planta=p3+0.407Planta=p5-0.351Dorm=d5-0.316Planta=p2-0.291Dorm=d1... > 0.311298: SÍ (5.0/1.0)

0.872Barrio=chamartín-0.331Barrio=eixample-0.286Barrio=hortaleza-0.087Barrio=moncloa-0.079Barrio=horta guinardó... > -0.047605
-0.623Barrio=puede de vallecas+0.394Barrio=tetuán-0.301Barrio=carabanchel-0.278Barrio=ciudad lineal+0.27 Barrio=san blas... <= 0.047369
-0.623Barrio=puede de vallecas+0.394Barrio=tetuán-0.301Barrio=carabanchel-0.278Barrio=ciudad lineal+0.27 Barrio=san blas... <= -0.619077
-0.713Dorm=d3+0.469Dorm=d2+0.301Planta=p1+0.23 Dorm=d1-0.21Planta=p4... <= 0.244766: SÍ (3.0)
-0.713Dorm=d3+0.469Dorm=d2+0.301Planta=p1+0.23 Dorm=d1-0.21Planta=p4... > 0.244766
0.864Barrio=ciudad lineal-0.346Barrio=salamanca-0.232Barrio=carabanchel-0.205Barrio=puede de vallecas+0.1 Barrio=chamartín... <= -0.211524: SÍ (3.0/1.0)
0.864Barrio=ciudad lineal-0.346Barrio=salamanca-0.232Barrio=carabanchel-0.205Barrio=puede de vallecas+0.1 Barrio=chamartín... > -0.211524: NO (4.0)
-0.623Barrio=puede de vallecas+0.394Barrio=tetuán-0.301Barrio=carabanchel-0.278Barrio=ciudad lineal+0.27 Barrio=san blas... > -0.619077: NO (169.0/8.0)
-0.623Barrio=puede de vallecas+0.394Barrio=tetuán-0.301Barrio=carabanchel-0.278Barrio=ciudad lineal+0.27 Barrio=san blas... > 0.047369
-0.698Barrio=gràcia+0.637Barrio=ciutat vella+0.162Barrio=villaverde-0.131Barrio=retiro+0.091Barrio=moncloa... <= 0.055009
-0.482Planta=p11+0.393Planta=p12+0.342Planta=p13-0.293Dorm=d9-0.273Dorm=d8... <= -0.066056: SÍ (7.0)
-0.482Planta=p11+0.393Planta=p12+0.342Planta=p13-0.293Dorm=d9-0.273Dorm=d8... > -0.066056
-0.698Ciudad=valencia+0.505Ciudad=barcelona+0.38 Euros+0.335Estado+0.038Ciudad=madrid <= -0.687985: SÍ (7.0/1.0)
-0.698Ciudad=valencia+0.505Ciudad=barcelona+0.38 Euros+0.335Estado+0.038Ciudad=madrid > -0.687985
0.653Planta=p14-0.515Planta=pst-0.303Planta=p13+0.295Planta=p12-0.263Planta=p10... <= 0.005284
0.591Barrio=chamberí-0.417Barrio=arganzuela-0.304Barrio=tetuán+0.274Barrio=san blas-0.249Barrio=eixample... <= -0.304732: SÍ (3.0/1.0)
0.591Barrio=chamberí-0.417Barrio=arganzuela-0.304Barrio=tetuán+0.274Barrio=san blas-0.249Barrio=eixample... > -0.304732: NO (37.0/1.0)
0.653Planta=p14-0.515Planta=pst-0.303Planta=p13+0.295Planta=p12-0.263Planta=p10... > 0.005284: SÍ (3.0/1.0)
-0.698Barrio=gràcia+0.637Barrio=ciutat vella+0.162Barrio=villaverde-0.131Barrio=retiro+0.091Barrio=moncloa... > 0.055009: SÍ (9.0)
-0.595Barrio=sant martí+0.451Barrio=fuencarral+0.325Barrio=sarrià-sant gervasi+0.287Barrio=latina+0.27 Barrio=Camins al grau... > 0.150727
-0.7Barrio=retiro+0.365Barrio=sant martí+0.273Barrio=sant andreu+0.202Barrio=latina-0.184Barrio=gràcia... <= 0.116351: SÍ (19.0)
-0.7Barrio=retiro+0.365Barrio=sant martí+0.273Barrio=sant andreu+0.202Barrio=latina-0.184Barrio=gràcia... > 0.116351: NO (3.0)

Number of Leaves : 68

Size of the tree : 135

J48 pruned tree (Árbol 5° de los 10 que forman el modelo.)

0.517Estado-0.388Tipo=dúplex+0.345Tipo=chalets-0.342Barrio=campanar - beni...+0.214Tipo=áticos... <= -0.18277
0.442Tipo=chalets+0.381Tipo=dúplex-0.361Barrio=campanar - beni...-0.288Tipo=áticos-0.274Barrio=fuencarral... <= 0.407313
0.287Barrio=ciutat vella+0.268Barrio=Ciutat vella+0.248Barrio=Camins al grau+0.24 Barrio=sant andreu+0.233Tipo=pisos... <= 0.406894: NO (130.0/8.0)
0.287Barrio=ciutat vella+0.268Barrio=Ciutat vella+0.248Barrio=Camins al grau+0.24 Barrio=sant andreu+0.233Tipo=pisos... > 0.406894
-0.649Planta=p11+0.602Planta=p10-0.398Planta=p13-0.205Planta=pst-0.073Planta=p12... <= 0.007162: NO (11.0/1.0)
-0.649Planta=p11+0.602Planta=p10-0.398Planta=p13-0.205Planta=pst-0.073Planta=p12... > 0.007162: SÍ (7.0)
0.442Tipo=chalets+0.381Tipo=dúplex-0.361Barrio=campanar - beni...-0.288Tipo=áticos-0.274Barrio=fuencarral... > 0.407313
-0.9Barrio=gràcia+0.326Barrio=villaverde+0.155Barrio=retiro+0.086Barrio=Quatre carreres+0.085Barrio=usera... <= 0.012081: NO (3.0)
-0.9Barrio=gràcia+0.326Barrio=villaverde+0.155Barrio=retiro+0.086Barrio=Quatre carreres+0.085Barrio=usera... > 0.012081: SÍ (6.0)
0.517Estado-0.388Tipo=dúplex+0.345Tipo=chalets-0.342Barrio=campanar - beni...+0.214Tipo=áticos... > -0.18277
0.661Tipo=pisos-0.436Tipo=áticos-0.317Tipo=dúplex-0.239Tipo=chalets-0.208Tipo=estudios... <= -0.380486
0.658Barrio=sarrià-sant gervasi-0.319Barrio=sant martí+0.265Barrio=eixample-0.26Barrio=retiro-0.249Barrio=moratalaz... <= 0.13383: NO (35.0/3.0)
0.658Barrio=sarrià-sant gervasi-0.319Barrio=sant martí+0.265Barrio=eixample-0.26Barrio=retiro-0.249Barrio=moratalaz... > 0.13383: SÍ (2.0)
0.661Tipo=pisos-0.436Tipo=áticos-0.317Tipo=dúplex-0.239Tipo=chalets-0.208Tipo=estudios... > -0.380486
-0.504Tipo=estudios+0.369Tipo=dúplex-0.247Barrio=ciutat vella+0.238Estado+0.236Barrio=campanar - beni... <= -0.462329
-0.649Planta=p11+0.602Planta=p10-0.398Planta=p13-0.205Planta=pst-0.073Planta=p12... <= 0.007202
0.542Barrio=nou barris+0.542Barrio=les corts-0.343Barrio=villa de vallecas+0.323Barrio=Ciutat vella-0.212Barrio=barajas... <= 0.003798: NO (4.0)
0.542Barrio=nou barris+0.542Barrio=les corts-0.343Barrio=villa de vallecas+0.323Barrio=Ciutat vella-0.212Barrio=barajas... > 0.003798: SÍ (6.0/1.0)
-0.649Planta=p11+0.602Planta=p10-0.398Planta=p13-0.205Planta=pst-0.073Planta=p12... > 0.007202: NO (16.0)
-0.504Tipo=estudios+0.369Tipo=dúplex-0.247Barrio=ciutat vella+0.238Estado+0.236Barrio=campanar - beni... > -0.462329
0.621Barrio=arganzuela-0.375Barrio=chamberí-0.335Barrio=latina-0.322Barrio=san blas-0.184Barrio=usera... <= 0.018035
-0.604Barrio=Extramurs-0.428Barrio=sant andreu+0.319Barrio=Camins al grau+0.307Barrio=Quatre carreres+0.281Barrio=La saïdia... <= -0.040504
-0.707Barrio=Algirós+0.707Barrio=El pla del real+0Barrio=L'olivereta+0Barrio=Jesús+0Barrio=Patraix... <= 0
-0.826Planta=p13+0.405Planta=p11+0.392Planta=pst+0.022Planta=p10-0.011Planta=p14... <= 0.000354
0.738Barrio=puede de vallecas-0.486Barrio=carabanchel-0.309Barrio=ciudad lineal-0.179Barrio=salamanca-0.132Barrio=centro... <= 0.020025: NO (11.0/1.0)
0.738Barrio=puede de vallecas-0.486Barrio=carabanchel-0.309Barrio=ciudad lineal-0.179Barrio=salamanca-0.132Barrio=centro... > 0.020025
-0.517Dorm=d3-0.517Dorm=d3+0.46 Dorm=d2+0.46 Dorm=d2+0.131Ciudad=madrid... <= 0.045499: SÍ (10.0/2.0)
-0.517Dorm=d3-0.517Dorm=d3+0.46 Dorm=d2+0.46 Dorm=d2+0.131Ciudad=madrid... > 0.045499: NO (3.0/1.0)
-0.826Planta=p13+0.405Planta=p11+0.392Planta=pst+0.022Planta=p10-0.011Planta=p14... > 0.000354: NO (19.0/2.0)
-0.707Barrio=Algirós+0.707Barrio=El pla del real+0Barrio=L'olivereta+0Barrio=Jesús+0Barrio=Patraix... > 0: SÍ (3.0/1.0)
-0.604Barrio=Extramurs-0.428Barrio=sant andreu+0.319Barrio=Camins al grau+0.307Barrio=Quatre carreres+0.281Barrio=La saïdia... > -0.040504: SÍ (1939.0/919.0)
0.621Barrio=arganzuela-0.375Barrio=chamberí-0.335Barrio=latina-0.322Barrio=san blas-0.184Barrio=usera... > 0.018035
0.484Barrio=sant andreu+0.475Barrio=Rascanya-0.412Barrio=benicalap-0.348Barrio=La saïdia-0.348Barrio=Jesús... <= 0.128114
0.606Planta=pss-0.597Planta=pst+0.432Planta=p12-0.214Planta=p13+0.162Planta=p11... <= -0.069396
-0.619Barrio=san blas+0.405Barrio=chamartín-0.349Barrio=hortaleza+0.323Barrio=moncloa+0.227Barrio=usera... <= 0.001361: SÍ (18.0)
-0.619Barrio=san blas+0.405Barrio=chamartín-0.349Barrio=hortaleza+0.323Barrio=moncloa+0.227Barrio=usera... > 0.001361
0.518Dorm=d6-0.518Dorm=d6+0.344Dorm=d0-0.344Dorm=d0+0.171Dorm=d3... <= -0.026797: SÍ (2.0)
0.518Dorm=d6-0.518Dorm=d6+0.344Dorm=d0-0.344Dorm=d0+0.171Dorm=d3... > -0.026797
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... <= -0.03636: SÍ (2.0)
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... > -0.03636: NO (3.0)
0.606Planta=pss-0.597Planta=pst+0.432Planta=p12-0.214Planta=p13+0.162Planta=p11... > -0.069396
0.428Dorm=d6+0.428Dorm=d6+0.42 Dorm=d0+0.42 Dorm=d0-0.257Dorm=d5... <= -0.341542: SÍ (204.0/68.0)
0.428Dorm=d6+0.428Dorm=d6+0.42 Dorm=d0+0.42 Dorm=d0-0.257Dorm=d5... > -0.341542
-0.647Dorm=d8-0.647Dorm=d8-0.258Dorm=d7-0.258Dorm=d7+0.081Dorm=d6... <= 0.054504
-0.532Dorm=d7-0.532Dorm=d7+0.297Dorm=d6+0.297Dorm=d6+0.268Dorm=d8... <= -0.115096
0.428Dorm=d6+0.428Dorm=d6+0.42 Dorm=d0+0.42 Dorm=d0-0.257Dorm=d5... <= 0.092592: SÍ (368.0/145.0)
0.428Dorm=d6+0.428Dorm=d6+0.42 Dorm=d0+0.42 Dorm=d0-0.257Dorm=d5... > 0.092592
0.518Dorm=d6-0.518Dorm=d6+0.344Dorm=d0-0.344Dorm=d0+0.171Dorm=d3... <= -0.026928
-0.991Planta=p15+0.105Planta=p14+0.046Planta=pss+0.045Planta=p12+0.031Planta=p10... <= 0.002288: NO (3.0/1.0)
-0.991Planta=p15+0.105Planta=p14+0.046Planta=pss+0.045Planta=p12+0.031Planta=p10... > 0.002288: SÍ (7.0)
0.518Dorm=d6-0.518Dorm=d6+0.344Dorm=d0-0.344Dorm=d0+0.171Dorm=d3... > -0.026928
0.636Barrio=arganzuela-0.366Barrio=eixample-0.328Barrio=sarrià-sant gervasi+0.327Barrio=san blas+0.253Barrio=chamberí... <= -0.032192
0.458Planta=p1+0.398Planta=p2+0.383Planta=p3+0.352Planta=p4+0.32 Planta=p5... <= 0.39764: SÍ (98.0/39.0)
0.458Planta=p1+0.398Planta=p2+0.383Planta=p3+0.352Planta=p4+0.32 Planta=p5... > 0.39764: NO (31.0/12.0)
0.636Barrio=arganzuela-0.366Barrio=eixample-0.328Barrio=sarrià-sant gervasi+0.327Barrio=san blas+0.253Barrio=chamberí... > -0.032192
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... <= -0.029298: SÍ (27.0)
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... > -0.029298
-0.649Planta=p11+0.602Planta=p10-0.398Planta=p13-0.205Planta=pst-0.073Planta=p12... <= 0.006349: SÍ (9.0)
-0.649Planta=p11+0.602Planta=p10-0.398Planta=p13-0.205Planta=pst-0.073Planta=p12... > 0.006349
0.804Planta=p1-0.389Planta=p3-0.213Planta=p5-0.177Planta=p4-0.168Euros... <= -0.201493: NO (5.0)
0.804Planta=p1-0.389Planta=p3-0.213Planta=p5-0.177Planta=p4-0.168Euros... > -0.201493
-0.504Tipo=estudios+0.369Tipo=dúplex-0.247Barrio=ciutat vella+0.238Estado+0.236Barrio=campanar - beni... <= 0.238528
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... <= -0.015427: NO (3.0)
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... > -0.015427
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... <= 0.003398: SÍ (4.0)
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... > 0.003398: NO (2.0)

-0.504Tipo=estudios+0.369Tipo=dúplex-0.247Barrio=ciutat vella+0.238Estado+0.236Barrio=campanar - beni..... > 0.238528: Sí (6.0/1.0)
-0.532Dorm=d7-0.532Dorm=d7+0.297Dorm=d6+0.297Dorm=d6+0.268Dorm=d8... > -0.115096
0.658Barrio=latina+0.53 Barrio=hortaleza-0.328Barrio=moncloa-0.177Barrio=chamberí-0.158Barrio=san blas... <= -0.00507
-0.707Barrio=les corts+0.707Barrio=nou barris+0Barrio=Quatre carreres+0Barrio=villa de vallecas+0Barrio=horta guinardó... <= 0
-0.707Barrio=les corts+0.707Barrio=nou barris+0Barrio=Quatre carreres+0Barrio=villa de vallecas+0Barrio=horta guinardó... <= -0.707107
0.661Tipo=pisos-0.436Tipo=áticos-0.317Tipo=dúplex-0.239Tipo=chalets-0.208Tipo=estudios... <= 0.098727: NO (3.0)
0.661Tipo=pisos-0.436Tipo=áticos-0.317Tipo=dúplex-0.239Tipo=chalets-0.208Tipo=estudios... > 0.098727
0.804Planta=p1-0.389Planta=p3-0.213Planta=p5-0.177Planta=p4-0.168Euros... <= -0.198738: NO (12.0/3.0)
0.804Planta=p1-0.389Planta=p3-0.213Planta=p5-0.177Planta=p4-0.168Euros... > -0.198738: Sí (22.0/7.0)
-0.707Barrio=les corts+0.707Barrio=nou barris+0Barrio=Quatre carreres+0Barrio=villa de vallecas+0Barrio=horta guinardó... > -0.707107: Sí (936.0/388.0)
-0.707Barrio=les corts+0.707Barrio=nou barris+0Barrio=Quatre carreres+0Barrio=villa de vallecas+0Barrio=horta guinardó... > 0
-0.87Barrio=sant martí+0.326Barrio=retiro+0.198Barrio=moratalaz-0.154Barrio=Ciutat vella+0.136Barrio=gràcia... <= 0.031219
-0.826Planta=p13+0.405Planta=p11+0.392Planta=pst+0.022Planta=p10-0.011Planta=p14... <= 0.000319: Sí (2.0)
-0.826Planta=p13+0.405Planta=p11+0.392Planta=pst+0.022Planta=p10-0.011Planta=p14... > 0.000319
-0.744Planta=pbj+0.496Planta=p5-0.213Planta=p8-0.204Planta=p7-0.17Planta=p6... <= 0.152552: NO (2.0)
-0.744Planta=pbj+0.496Planta=p5-0.213Planta=p8-0.204Planta=p7-0.17Planta=p6... > 0.152552: Sí (2.0)
-0.87Barrio=sant martí+0.326Barrio=retiro+0.198Barrio=moratalaz-0.154Barrio=Ciutat vella+0.136Barrio=gràcia... > 0.031219: NO (46.0/21.0)
0.658Barrio=latina+0.53 Barrio=hortaleza-0.328Barrio=moncloa-0.177Barrio=chamberí-0.158Barrio=san blas... > -0.00507: NO (276.0/125.0)
-0.647Dorm=d8-0.647Dorm=d8-0.258Dorm=d7-0.258Dorm=d7+0.081Dorm=d6... > 0.054504
-0.707Barrio=Algirós+0.707Barrio=El pla del real+0Barrio=L'olivereta+0Barrio=Jesús+0Barrio=Patraix... <= 0: NO (170.0/80.0)
-0.707Barrio=Algirós+0.707Barrio=El pla del real+0Barrio=L'olivereta+0Barrio=Jesús+0Barrio=Patraix... > 0
-0.547Planta=pst+0.424Planta=pen-0.417Planta=pss-0.324Planta=p9-0.264Planta=p12... <= -0.010646: Sí (3.0)
-0.547Planta=pst+0.424Planta=pen-0.417Planta=pss-0.324Planta=p9-0.264Planta=p12... > -0.010646: NO (6.0)
0.484Barrio=sant andreu+0.475Barrio=Rascanya-0.412Barrio=benicalap-0.348Barrio=La saüdia-0.348Barrio=Jesús... > 0.128114
-0.517Dorm=d3-0.517Dorm=d3+0.46 Dorm=d2+0.46 Dorm=d2+0.131Ciudad=madrid... <= -0.062647
0.428Dorm=d6+0.428Dorm=d6+0.42 Dorm=d0+0.42 Dorm=d0-0.257Dorm=d5... <= -0.123396
-0.547Planta=pst+0.424Planta=pen-0.417Planta=pss-0.324Planta=p9-0.264Planta=p12... <= 0.054722: Sí (20.0/1.0)
-0.547Planta=pst+0.424Planta=pen-0.417Planta=pss-0.324Planta=p9-0.264Planta=p12... > 0.054722
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... <= -0.017094
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... <= -0.04402: NO (2.0)
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... > -0.04402: Sí (8.0/1.0)
-0.709Metros+0.703Euros-0.035Planta=p8+0.025Planta=p7-0.023Planta=pbj... > -0.017094: NO (3.0)
0.428Dorm=d6+0.428Dorm=d6+0.42 Dorm=d0+0.42 Dorm=d0-0.257Dorm=d5... > -0.123396: Sí (30.0/2.0)
-0.517Dorm=d3-0.517Dorm=d3+0.46 Dorm=d2+0.46 Dorm=d2+0.131Ciudad=madrid... > -0.062647
0.518Dorm=d6-0.518Dorm=d6+0.344Dorm=d0-0.344Dorm=d0+0.171Dorm=d3... <= -0.022942: NO (4.0)
0.518Dorm=d6-0.518Dorm=d6+0.344Dorm=d0-0.344Dorm=d0+0.171Dorm=d3... > -0.022942
0.805Planta=p2-0.54Planta=p3-0.138Planta=p1-0.119Planta=pbj-0.09Planta=p4... <= 0.811595
-0.705Planta=p4+0.577Planta=p3+0.31 Planta=p2-0.201Planta=p5-0.128Planta=pbj... <= -0.70497: NO (3.0)
-0.705Planta=p4+0.577Planta=p3+0.31 Planta=p2-0.201Planta=p5-0.128Planta=pbj... > -0.70497: Sí (32.0/8.0)
0.805Planta=p2-0.54Planta=p3-0.138Planta=p1-0.119Planta=pbj-0.09Planta=p4... > 0.811595: NO (4.0)

Number of Leaves : 54

Size of the tree : 107

J48 pruned tree (Árbol 6° de los 10 que forman el modelo.)

-0.702Estado-0.293Barrio=hortaleza-0.271Barrio=Ciutat vella-0.268Barrio=barajas-0.236Barrio=Rascanya... <= -0.013621
-0.761Barrio=villa de vallecas+0.509Barrio=Extramurs+0.381Barrio=Camins al grau-0.129Barrio=les corts+0Barrio=Ciutat vella... <= -0.76101
0.692Euros+0.342Barrio=moncloa+0.253Barrio=hortaleza+0.23 Barrio=chamartín+0.227Barrio=salamanca... <= -0.077595: Sí (13.0/1.0)
0.692Euros+0.342Barrio=moncloa+0.253Barrio=hortaleza+0.23 Barrio=chamartín+0.227Barrio=salamanca... > -0.077595
0.757Barrio=ciudad lineal-0.599Barrio=puede de vallecas-0.166Barrio=hortaleza-0.153Barrio=tetuán+0.081Barrio=san blas... <= 0.004597
-0.673Ciudad=madrid-0.582Ciudad=barcelona-0.456Ciudad=valencia-0.002Tipo=pisos-0.001Tipo=áticos... <= -0.674589
-0.813Barrio=centro+0.539Barrio=carabanchel+0.183Barrio=hortaleza+0.078Barrio=chamartín-0.046Barrio=salamanca... <= -0.004131
-0.762Barrio=carabanchel-0.49Barrio=centro+0.217Barrio=puede de vallecas+0.202Barrio=tetuán+0.138Barrio=ciudad lineal... <= 0.022034: NO (3.0)
-0.762Barrio=carabanchel-0.49Barrio=centro+0.217Barrio=puede de vallecas+0.202Barrio=tetuán+0.138Barrio=ciudad lineal... > 0.022034
-0.684Dorm=d3-0.335Dorm=d2+0.299Planta=p1+0.283Dorm=d4+0.241Ciudad=valencia... <= -0.727893: NO (6.0)
-0.684Dorm=d3-0.335Dorm=d2+0.299Planta=p1+0.283Dorm=d4+0.241Ciudad=valencia... > -0.727893
-0.642Tipo=pisos+0.459Tipo=chalets+0.424Metros+0.357Tipo=áticos+0.228Tipo=dúplex... <= -0.625043: NO (4.0)
-0.642Tipo=pisos+0.459Tipo=chalets+0.424Metros+0.357Tipo=áticos+0.228Tipo=dúplex... > -0.625043: Sí (8.0/2.0)
-0.813Barrio=centro+0.539Barrio=carabanchel+0.183Barrio=hortaleza+0.078Barrio=chamartín-0.046Barrio=salamanca... > -0.004131
-0.626Ciudad=madrid+0.455Ciudad=barcelona+0.35 Ciudad=valencia-0.234Planta=p1-0.189Dorm=d1... <= -0.714401: NO (3.0)
-0.626Ciudad=madrid+0.455Ciudad=barcelona+0.35 Ciudad=valencia-0.234Planta=p1-0.189Dorm=d1... > -0.714401: Sí (6.0)
-0.673Ciudad=madrid-0.582Ciudad=barcelona-0.456Ciudad=valencia-0.002Tipo=pisos-0.001Tipo=áticos... > -0.674589: NO (2.0)
0.757Barrio=ciudad lineal-0.599Barrio=puede de vallecas-0.166Barrio=hortaleza-0.153Barrio=tetuán+0.081Barrio=san blas... > 0.004597: NO (15.0/1.0)
-0.761Barrio=villa de vallecas+0.509Barrio=Extramurs+0.381Barrio=Camins al grau-0.129Barrio=les corts+0Barrio=Ciutat vella... > -0.76101
-0.807Barrio=Patraix+0.487Barrio=L'olivereta+0.333Barrio=vicálvaro-0.013Barrio=campanar - beni...+0Barrio=El pla del real... <= 0
-0.807Barrio=Patraix+0.487Barrio=L'olivereta+0.333Barrio=vicálvaro-0.013Barrio=campanar - beni...+0Barrio=El pla del real... <= -0.013395
-0.875Barrio=villaverde+0.351Barrio=Quatre carreres+0.257Barrio=ciutat vella+0.139Barrio=moratalaz-0.101Barrio=nou barris... <= -0.008747
0.563Planta=p15+0.375Dorm=d6-0.303Planta=pen+0.288Dorm=d0-0.247Planta=p8... <= -0.019948: Sí (10.0/2.0)
0.563Planta=p15+0.375Dorm=d6-0.303Planta=pen+0.288Dorm=d0-0.247Planta=p8... > -0.019948
-0.807Barrio=Patraix+0.487Barrio=L'olivereta+0.333Barrio=vicálvaro-0.013Barrio=campanar - beni...+0Barrio=El pla del real... <= -0.807103
0.467Planta=p2-0.402Planta=p3-0.347Planta=p5+0.328Dorm=d4+0.249Planta=p4... <= -0.147111: NO (6.0)
0.467Planta=p2-0.402Planta=p3-0.347Planta=p5+0.328Dorm=d4+0.249Planta=p4... > -0.147111
-0.528Dorm=d3-0.516Dorm=d2-0.408Dorm=d4-0.387Dorm=d1-0.198Dorm=d5... <= -0.337986
-0.684Dorm=d3-0.335Dorm=d2+0.299Planta=p1+0.283Dorm=d4+0.241Ciudad=valencia... <= -0.011927
-0.784Ciudad=valencia+0.534Ciudad=barcelona-0.195Tipo=dúplex+0.149Tipo=estudios+0.139Tipo=áticos... <= -0.87088: NO (2.0)
-0.784Ciudad=valencia+0.534Ciudad=barcelona-0.195Tipo=dúplex+0.149Tipo=estudios+0.139Tipo=áticos... > -0.87088: Sí (4.0/1.0)
-0.684Dorm=d3-0.335Dorm=d2+0.299Planta=p1+0.283Dorm=d4+0.241Ciudad=valencia... > -0.011927: NO (3.0)
-0.528Dorm=d3-0.516Dorm=d2-0.408Dorm=d4-0.387Dorm=d1-0.198Dorm=d5... > -0.337986: Sí (4.0)
-0.807Barrio=Patraix+0.487Barrio=L'olivereta+0.333Barrio=vicálvaro-0.013Barrio=campanar - beni...+0Barrio=El pla del real... > -0.807103: NO (7.0)
-0.875Barrio=villaverde+0.351Barrio=Quatre carreres+0.257Barrio=ciutat vella+0.139Barrio=moratalaz-0.101Barrio=nou barris... > -0.008747: Sí (7.0/1.0)
-0.807Barrio=Patraix+0.487Barrio=L'olivereta+0.333Barrio=vicálvaro-0.013Barrio=campanar - beni...+0Barrio=El pla del real... > -0.013395
-0.673Ciudad=madrid-0.582Ciudad=barcelona-0.456Ciudad=valencia-0.002Tipo=pisos-0.001Tipo=áticos... <= -0.457781: Sí (4197.0/1856.0)
-0.673Ciudad=madrid-0.582Ciudad=barcelona-0.456Ciudad=valencia-0.002Tipo=pisos-0.001Tipo=áticos... > -0.457781
-0.626Ciudad=madrid+0.455Ciudad=barcelona+0.35 Ciudad=valencia-0.234Planta=p1-0.189Dorm=d1... <= 0.141815: Sí (6.0)
-0.626Ciudad=madrid+0.455Ciudad=barcelona+0.35 Ciudad=valencia-0.234Planta=p1-0.189Dorm=d1... > 0.141815
-0.761Barrio=villa de vallecas+0.509Barrio=Extramurs+0.381Barrio=Camins al grau-0.129Barrio=les corts+0Barrio=Ciutat vella... <= 0.380505: NO (60.0/20.0)
-0.761Barrio=villa de vallecas+0.509Barrio=Extramurs+0.381Barrio=Camins al grau-0.129Barrio=les corts+0Barrio=Ciutat vella... > 0.380505: Sí (4.0/1.0)
-0.807Barrio=Patraix+0.487Barrio=L'olivereta+0.333Barrio=vicálvaro-0.013Barrio=campanar - beni...+0Barrio=El pla del real... > 0
-0.53Planta=p11+0.474Planta=p12-0.358Planta=p14+0.341Planta=p9-0.249Planta=pss... <= 0.02578
0.786Barrio=san blas-0.481Barrio=sant martí-0.242Barrio=hortaleza+0.228Barrio=moncloa-0.109Barrio=chamberí... <= -0.011965
-0.813Barrio=centro+0.539Barrio=carabanchel+0.183Barrio=hortaleza+0.078Barrio=chamartín-0.046Barrio=salamanca... <= -0.009764
-0.642Tipo=pisos+0.459Tipo=chalets+0.424Metros+0.357Tipo=áticos+0.228Tipo=dúplex... <= -0.584514: NO (7.0/1.0)
-0.642Tipo=pisos+0.459Tipo=chalets+0.424Metros+0.357Tipo=áticos+0.228Tipo=dúplex... > -0.584514
-0.626Ciudad=madrid-0.455Ciudad=barcelona+0.35 Ciudad=valencia-0.234Planta=p1-0.189Dorm=d1... <= 0.620626: Sí (7.0/1.0)

|-0.626Ciudad=madrid+0.455Ciudad=barcelona+0.35 Ciudad=valencia-0.234Planta=p1-0.189Dorm=d1... > 0.620626: NO (2.0)
|-0.813Barrio=centro+0.539Barrio=carabanchel+0.183Barrio=hortaleza+0.078Barrio=chamartín-0.046Barrio=salamanca... > -0.009764: NO (2.0)
|0.786Barrio=san blas-0.481Barrio=sant martí-0.242Barrio=hortaleza+0.228Barrio=moncloa-0.109Barrio=chamberí... > -0.011965
|-0.495Dorm=d5+0.347Planta=p6-0.339Planta=p8-0.311Planta=p10+0.309Dorm=d1... <= 0.022119
|-0.642Tipo=pisos+0.459Tipo=chalets+0.424Metros+0.357Tipo=áticos+0.228Tipo=dúplex... <= -0.617974: NO (11.0/4.0)
|-0.642Tipo=pisos+0.459Tipo=chalets+0.424Metros+0.357Tipo=áticos+0.228Tipo=dúplex... > -0.617974: SÍ (6.0/1.0)
|-0.495Dorm=d5+0.347Planta=p6-0.339Planta=p8-0.311Planta=p10+0.309Dorm=d1... > 0.022119: NO (5.0)
|-0.53Planta=p11+0.474Planta=p12-0.358Planta=p14+0.341Planta=p9-0.249Planta=pss... > 0.02578: SÍ (4.0)
|-0.702Estado-0.293Barrio=hortaleza-0.271Barrio=Ciutat vella-0.268Barrio=barajas-0.236Barrio=Rascanya... > -0.013621
|-0.364Planta=pbj-0.358Planta=pen+0.356Dorm=d0-0.312Dorm=d4+0.296Ciudad=barcelona... <= 0.183231: NO (137.0/5.0)
|-0.364Planta=pbj-0.358Planta=pen+0.356Dorm=d0-0.312Dorm=d4+0.296Ciudad=barcelona... > 0.183231
|0.678Barrio=chamberí-0.491Barrio=fuencarral+0.277Barrio=sant martí-0.204Barrio=hortaleza-0.187Barrio=latina... <= 0.158663
|0.678Barrio=arganzuela-0.448Barrio=usera-0.308Barrio=sarrià-sant gervasi+0.193Barrio=retiro-0.192Barrio=hortaleza... <= 0.018621: NO (22.0)
|0.678Barrio=arganzuela-0.448Barrio=usera-0.308Barrio=sarrià-sant gervasi+0.193Barrio=retiro-0.192Barrio=hortaleza... > 0.018621
|-0.813Barrio=centro+0.539Barrio=carabanchel+0.183Barrio=hortaleza+0.078Barrio=chamartín-0.046Barrio=salamanca... <= 0.008395: SÍ (2.0)
|-0.813Barrio=centro+0.539Barrio=carabanchel+0.183Barrio=hortaleza+0.078Barrio=chamartín-0.046Barrio=salamanca... > 0.008395: NO (3.0)
|0.678Barrio=chamberí-0.491Barrio=fuencarral+0.277Barrio=sant martí-0.204Barrio=hortaleza-0.187Barrio=latina... > 0.158663: SÍ (5.0)

Number of Leaves : 34
Size of the tree : 67

J48 pruned tree (Árbol 7º de los 10 que forman el modelo.)

|-0.397Estado+0.248Barrio=campanar - beni...+0.248Barrio=campanar - beni...+0.222Barrio=Ciutat vella+0.222Barrio=Ciutat vella... <= 0.098591
|-0.436Barrio=Quatre carreres-0.436Barrio=Quatre carreres-0.436Barrio=sant andreu-0.436Barrio=sant andreu+0.265Barrio=Camins al grau... <= -0.87453
|-0.563Dorm=d3-0.518Dorm=d2-0.41Dorm=d4-0.365Dorm=d1-0.262Dorm=d5... <= -0.587224: SÍ (36.0/4.0)
|-0.563Dorm=d3-0.518Dorm=d2-0.41Dorm=d4-0.365Dorm=d1-0.262Dorm=d5... > -0.587224
|-0.849Tipo=estudios+0.344Tipo=áticos+0.31 Ciudad=valencia-0.223Ciudad=barcelona+0.101Euros... <= -0.184776: SÍ (12.0)
|-0.849Tipo=estudios+0.344Tipo=áticos+0.31 Ciudad=valencia-0.223Ciudad=barcelona+0.101Euros... > -0.184776
|0.558Dorm=d0+0.422Planta=pss+0.414Dorm=d8+0.214Planta=p5-0.202Dorm=d5... <= -0.152173: NO (6.0)
|0.558Dorm=d0+0.422Planta=pss+0.414Dorm=d8+0.214Planta=p5-0.202Dorm=d5... > -0.152173
|0.677Dorm=d6-0.458Planta=p9-0.338Planta=pen+0.189Planta=p7+0.184Planta=p6... <= -0.124403: SÍ (9.0)
|0.677Dorm=d6-0.458Planta=p9-0.338Planta=pen+0.189Planta=p7+0.184Planta=p6... > -0.124403
|-0.686Tipo=pisos-0.512Tipo=áticos-0.363Tipo=chalets-0.279Tipo=dúplex-0.214Tipo=estudios... <= -0.574411
|0.5 Barrio=sant andreu+0.5 Barrio=sant andreu-0.5Barrio=Quatre carreres-0.5Barrio=Quatre carreres+0Barrio=les corts... <= 0
|0.641Planta=p3-0.427Planta=p1+0.282Dorm=d8+0.276Planta=p2-0.203Dorm=d4... <= -0.090047: SÍ (13.0/1.0)
|0.641Planta=p3-0.427Planta=p1+0.282Dorm=d8+0.276Planta=p2-0.203Dorm=d4... > -0.090047
|-0.682Planta=p15-0.465Planta=p12+0.386Planta=p13-0.251Planta=p14+0.249Planta=pst... <= -0.001002
|-0.593Dorm=d1+0.442Dorm=d2-0.393Dorm=d4+0.297Planta=p4-0.236Planta=p3... <= 0.354161: SÍ (3.0)
|-0.593Dorm=d1+0.442Dorm=d2-0.393Dorm=d4+0.297Planta=p4-0.236Planta=p3... > 0.354161: NO (8.0/3.0)
|-0.682Planta=p15-0.465Planta=p12+0.386Planta=p13-0.251Planta=p14+0.249Planta=pst... > -0.001002: NO (4.0)
|0.5 Barrio=sant andreu+0.5 Barrio=sant andreu-0.5Barrio=Quatre carreres-0.5Barrio=Quatre carreres+0Barrio=les corts... > 0
|0.47 Planta=p14+0.417Planta=p8-0.374Planta=p15-0.3Planta=pss+0.291Planta=pen... <= 0.036557
|-0.38Dorm=d4+0.378Planta=p9-0.332Planta=pbj+0.325Dorm=d1+0.27 Dorm=d0... <= 0.065971: NO (10.0/2.0)
|-0.38Dorm=d4+0.378Planta=p9-0.332Planta=pbj+0.325Dorm=d1+0.27 Dorm=d0... > 0.065971: SÍ (2.0)
|0.47 Planta=p14+0.417Planta=p8-0.374Planta=p15-0.3Planta=pss+0.291Planta=pen... > 0.036557: SÍ (6.0)
|-0.686Tipo=pisos-0.512Tipo=áticos-0.363Tipo=chalets-0.279Tipo=dúplex-0.214Tipo=estudios... > -0.574411: NO (5.0/1.0)
|-0.436Barrio=Quatre carreres-0.436Barrio=Quatre carreres-0.436Barrio=sant andreu-0.436Barrio=sant andreu+0.265Barrio=Camins al grau... <= -0.87453
|0.916Estado+0.126Barrio=campanar - beni...+0.126Barrio=campanar - beni...+0.112Barrio=Ciutat vella+0.112Barrio=Ciutat vella... <= -0.045977
|-0.58Barrio=carabanchel-0.58Barrio=carabanchel+0.384Barrio=salamanca+0.384Barrio=salamanca+0.068Barrio=tetuán... <= -0.003544: NO (3.0)
|-0.58Barrio=carabanchel-0.58Barrio=carabanchel+0.384Barrio=salamanca+0.384Barrio=salamanca+0.068Barrio=tetuán... > -0.003544
|-0.354Barrio=villaverde-0.354Barrio=villaverde+0.29 Barrio=gràcia+0.231Barrio=sant martí... <= -0.162778: NO (4.0/1.0)
|-0.354Barrio=villaverde-0.354Barrio=villaverde+0.29 Barrio=gràcia+0.231Barrio=sant martí... > -0.162778
|-0.661Barrio=centro-0.661Barrio=centro+0.171Barrio=carabanchel+0.171Barrio=carabanchel+0.131Barrio=salamanca... <= 0.03277: SÍ (20.0)
|-0.661Barrio=centro-0.661Barrio=centro+0.171Barrio=carabanchel+0.171Barrio=carabanchel+0.131Barrio=salamanca... > 0.03277
|-0.397Estado+0.248Barrio=campanar - beni...+0.248Barrio=campanar - beni...+0.222Barrio=Ciutat vella+0.222Barrio=Ciutat vella... <= -0.231342
|0.47 Planta=p14+0.417Planta=p8-0.374Planta=p15-0.3Planta=pss+0.291Planta=pen... <= -0.072585: NO (2.0)
|0.47 Planta=p14+0.417Planta=p8-0.374Planta=p15-0.3Planta=pss+0.291Planta=pen... > -0.072585: SÍ (18.0/2.0)
|-0.397Estado+0.248Barrio=campanar - beni...+0.248Barrio=campanar - beni...+0.222Barrio=Ciutat vella+0.222Barrio=Ciutat vella... > -0.231342
|-0.471Planta=pss-0.348Planta=p8+0.294Dorm=d0+0.279Planta=p7-0.265Dorm=d7... <= -0.035969: NO (4.0)
|-0.471Planta=pss-0.348Planta=p8+0.294Dorm=d0+0.279Planta=p7-0.265Dorm=d7... > -0.035969: SÍ (8.0)
|0.916Estado+0.126Barrio=campanar - beni...+0.126Barrio=campanar - beni...+0.112Barrio=Ciutat vella+0.112Barrio=Ciutat vella... > -0.045977
|0.916Estado+0.126Barrio=campanar - beni...+0.126Barrio=campanar - beni...+0.112Barrio=Ciutat vella+0.112Barrio=Ciutat vella... <= 0.027121
|-0.498Planta=p6-0.423Planta=pen+0.362Dorm=d7+0.26 Planta=pbj+0.257Planta=p5... <= -0.094035: SÍ (5.0)
|-0.498Planta=p6-0.423Planta=pen+0.362Dorm=d7+0.26 Planta=pbj+0.257Planta=p5... > -0.094035
|0.727Metros-0.374Dorm=d5-0.343Dorm=d8-0.263Dorm=d7+0.189Dorm=d1... <= 0.056657: SÍ (3.0)
|0.727Metros-0.374Dorm=d5-0.343Dorm=d8-0.263Dorm=d7+0.189Dorm=d1... > 0.056657
|0.417Dorm=d8-0.363Dorm=d0-0.345Dorm=d7-0.286Planta=pen-0.266Planta=pss... <= 0.315649: NO (45.0)
|0.417Dorm=d8-0.363Dorm=d0-0.345Dorm=d7-0.286Planta=pen-0.266Planta=pss... > 0.315649
|-0.711Dorm=d3+0.551Dorm=d2+0.192Planta=p1+0.185Planta=pbj+0.149Dorm=d1... <= 0.33993: SÍ (3.0)
|-0.711Dorm=d3+0.551Dorm=d2+0.192Planta=p1+0.185Planta=pbj+0.149Dorm=d1... > 0.33993: NO (2.0)
|0.916Estado+0.126Barrio=campanar - beni...+0.126Barrio=campanar - beni...+0.112Barrio=Ciutat vella+0.112Barrio=Ciutat vella... > 0.027121
|0.5 Barrio=sant martí+0.5 Barrio=sant martí-0.5Barrio=ciutat vella-0.5Barrio=ciutat vella+0Barrio=gràcia... <= -1
|-0.701Ciudad=madrid+0.557Ciudad=barcelona+0.326Ciudad=valencia-0.236Euros-0.173Tipo=chalets... <= 0.57314
|-0.563Dorm=d3-0.518Dorm=d2-0.41Dorm=d4-0.365Dorm=d1-0.262Dorm=d5... <= -0.520448
|0.535Dorm=d4-0.453Planta=p1-0.362Dorm=d1+0.234Planta=p5+0.219Planta=p7... <= 0.044453
|-0.593Dorm=d1+0.442Dorm=d2-0.393Dorm=d4+0.297Planta=p4-0.236Planta=p3... <= 0.193324
|0.678Planta=p4-0.342Planta=p1-0.321Planta=p3+0.294Dorm=d5+0.213Dorm=d1... <= -0.343481: SÍ (2.0)
|0.678Planta=p4-0.342Planta=p1-0.321Planta=p3+0.294Dorm=d5+0.213Dorm=d1... > -0.343481: NO (2.0)
|-0.593Dorm=d1+0.442Dorm=d2-0.393Dorm=d4+0.297Planta=p4-0.236Planta=p3... > 0.193324: NO (6.0)
|0.535Dorm=d4-0.453Planta=p1-0.362Dorm=d1+0.234Planta=p5+0.219Planta=p7... > 0.044453: SÍ (2.0)
|-0.563Dorm=d3-0.518Dorm=d2-0.41Dorm=d4-0.365Dorm=d1-0.262Dorm=d5... > -0.520448: NO (14.0)
|-0.701Ciudad=madrid+0.557Ciudad=barcelona+0.326Ciudad=valencia-0.236Euros-0.173Tipo=chalets... > 0.57314
|-0.563Dorm=d3-0.518Dorm=d2-0.41Dorm=d4-0.365Dorm=d1-0.262Dorm=d5... <= -0.395671
|-0.686Tipo=pisos-0.512Tipo=áticos-0.363Tipo=chalets-0.279Tipo=dúplex-0.214Tipo=estudios... <= -0.731819
|-0.593Dorm=d1+0.442Dorm=d2-0.393Dorm=d4+0.297Planta=p4-0.236Planta=p3... <= 0.431208: NO (26.0/12.0)
|-0.593Dorm=d1+0.442Dorm=d2-0.393Dorm=d4+0.297Planta=p4-0.236Planta=p3... > 0.431208: SÍ (7.0)
|-0.686Tipo=pisos-0.512Tipo=áticos-0.363Tipo=chalets-0.279Tipo=dúplex-0.214Tipo=estudios... > -0.731819
|0.639Planta=p5+0.432Dorm=d6+0.298Planta=pen-0.241Dorm=d0-0.197Planta=p6... <= -0.05921: NO (3.0)
|0.639Planta=p5+0.432Dorm=d6+0.298Planta=pen-0.241Dorm=d0-0.197Planta=p6... > -0.05921: SÍ (5.0/1.0)
|-0.563Dorm=d3-0.518Dorm=d2-0.41Dorm=d4-0.365Dorm=d1-0.262Dorm=d5... > -0.395671: SÍ (9.0)
|0.5 Barrio=sant martí+0.5 Barrio=sant martí-0.5Barrio=ciutat vella-0.5Barrio=ciutat vella+0Barrio=gràcia... > -1
|-0.682Planta=p15-0.465Planta=p12+0.386Planta=p13-0.251Planta=p14+0.249Planta=pst... <= 0.02533
|-0.5Barrio=Poblats marítims-0.5Barrio=Poblats marítims+0.5 Barrio=Algirós+0.5 Barrio=Algirós+0Barrio=El pla del real... <= -1

-0.643Metros-0.376Dorm=d5-0.281Planta=p1-0.267Dorm=d8-0.249Dorm=d4... <= -0.387565: NO (4.0)
-0.643Metros-0.376Dorm=d5-0.281Planta=p1-0.267Dorm=d8-0.249Dorm=d4... > -0.387565
-0.711Dorm=d3+0.551Dorm=d2+0.192Planta=p1+0.185Planta=pbj+0.149Dorm=d1... <= 0.440272
-0.777Planta=pst+0.491Planta=p11+0.368Planta=p13-0.127Planta=p12-0.046Dorm=d8... <= -0.006026: NO (2.0)
-0.777Planta=pst+0.491Planta=p11+0.368Planta=p13-0.127Planta=p12-0.046Dorm=d8... > -0.006026
0.501Ciudad=valencia-0.469Ciudad=barcelona+0.376Tipo=estudios-0.369Euros-0.354Tipo=chalets... <= 0.317707
-0.498Planta=p6-0.423Planta=pen+0.362Dorm=d7+0.26 Planta=pbj+0.257Planta=p5... <= 0.069408: NO (4.0/1.0)
-0.498Planta=p6-0.423Planta=pen+0.362Dorm=d7+0.26 Planta=pbj+0.257Planta=p5... > 0.069408: SÍ (3.0)
0.501Ciudad=valencia-0.469Ciudad=barcelona+0.376Tipo=estudios-0.369Euros-0.354Tipo=chalets... > 0.317707: SÍ (11.0)
-0.711Dorm=d3+0.551Dorm=d2+0.192Planta=p1+0.185Planta=pbj+0.149Dorm=d1... > 0.440272: NO (3.0)
-0.5Barrio=Poblats marítims-0.5Barrio=Poblats marítims+0.5 Barrio=Algirós+0.5 Barrio=Algirós+0Barrio=El pla del real... > -1
0.5 Barrio=Jesús+0.5 Barrio=Jesús-0.5Barrio=Patraix-0.5Barrio=Patraix+0Barrio=L'olivereta... <= 0: SÍ (3692.0/1694.0)
0.5 Barrio=Jesús+0.5 Barrio=Jesús-0.5Barrio=Patraix-0.5Barrio=Patraix+0Barrio=L'olivereta... > 0
-0.701Ciudad=madrid+0.557Ciudad=barcelona+0.326Ciudad=valencia-0.236Euros-0.173Tipo=chalets... <= 0.349982: NO (9.0/1.0)
-0.701Ciudad=madrid+0.557Ciudad=barcelona+0.326Ciudad=valencia-0.236Euros-0.173Tipo=chalets... > 0.349982
0.679Planta=p13-0.53Planta=p10+0.264Planta=p14+0.24 Planta=pst+0.235Planta=p15... <= -0.011667: NO (2.0)
0.679Planta=p13-0.53Planta=p10+0.264Planta=p14+0.24 Planta=pst+0.235Planta=p15... > -0.011667
-0.762Planta=p2+0.45 Planta=p3+0.289Planta=p4-0.241Dorm=d8+0.181Dorm=d0... <= 0.273297: SÍ (12.0)
-0.762Planta=p2+0.45 Planta=p3+0.289Planta=p4-0.241Dorm=d8+0.181Dorm=d0... > 0.273297
-0.619Planta=p12+0.486Planta=p11+0.394Planta=p14-0.339Planta=p13+0.249Planta=pst... <= 0.010429: SÍ (4.0/1.0)
-0.619Planta=p12+0.486Planta=p11+0.394Planta=p14-0.339Planta=p13+0.249Planta=pst... > 0.010429: NO (2.0)
-0.682Planta=p15-0.465Planta=p12+0.386Planta=p13-0.251Planta=p14+0.249Planta=pst... > 0.025533
0.583Barrio=barajas+0.583Barrio=barajas-0.214Barrio=gràcia-0.214Barrio=gràcia-0.21Barrio=ciutat vella... <= 0.04234: SÍ (301.0/96.0)
0.583Barrio=barajas+0.583Barrio=barajas-0.214Barrio=gràcia-0.214Barrio=gràcia-0.21Barrio=ciutat vella... > 0.04234
-0.777Planta=pst+0.491Planta=p11+0.368Planta=p13-0.127Planta=p12-0.046Dorm=d8... <= 0.008062: SÍ (6.0/1.0)
-0.777Planta=pst+0.491Planta=p11+0.368Planta=p13-0.127Planta=p12-0.046Dorm=d8... > 0.008062: NO (16.0/2.0)
-0.397Estado+0.248Barrio=campanar - beni...+0.248Barrio=campanar - beni...+0.222Barrio=Ciutat vella+0.222Barrio=Ciutat vella... > 0.098591
-0.465Barrio=mortalaz-0.465Barrio=mortalaz+0.452Barrio=vicálvaro+0.452Barrio=vicálvaro+0.207Barrio=les corts... <= -0.387106: NO (13.0)
-0.465Barrio=mortalaz-0.465Barrio=mortalaz+0.452Barrio=vicálvaro+0.452Barrio=vicálvaro+0.207Barrio=les corts... > -0.387106
-0.534Barrio=chamartín-0.534Barrio=chamartín+0.45 Barrio=chamberí+0.45 Barrio=chamberí-0.054Barrio=ciudad lineal... <= -0.007706
-0.701Ciudad=madrid+0.557Ciudad=barcelona+0.326Ciudad=valencia-0.236Euros-0.173Tipo=chalets... <= 0.34912
-0.397Estado+0.248Barrio=campanar - beni...+0.248Barrio=campanar - beni...+0.222Barrio=Ciutat vella+0.222Barrio=Ciutat vella... <= 0.219429
-0.849Tipo=estudios+0.344Tipo=áticos+0.31 Ciudad=valencia-0.223Ciudad=barcelona+0.101Euros... <= 0.037813: SÍ (8.0)
-0.849Tipo=estudios+0.344Tipo=áticos+0.31 Ciudad=valencia-0.223Ciudad=barcelona+0.101Euros... > 0.037813
0.678Planta=p4-0.342Planta=p1-0.321Planta=p3+0.294Dorm=d5+0.213Dorm=d1... <= -0.512841: SÍ (5.0/1.0)
0.678Planta=p4-0.342Planta=p1-0.321Planta=p3+0.294Dorm=d5+0.213Dorm=d1... > -0.512841: NO (17.0)
-0.397Estado+0.248Barrio=campanar - beni...+0.248Barrio=campanar - beni...+0.222Barrio=Ciutat vella+0.222Barrio=Ciutat vella... > 0.219429
0.778Planta=p7-0.453Planta=p6+0.255Planta=pen-0.21Dorm=d0+0.151Planta=pss... <= -0.007169: NO (20.0)
0.778Planta=p7-0.453Planta=p6+0.255Planta=pen-0.21Dorm=d0+0.151Planta=pss... > -0.007169
0.639Planta=p5+0.432Dorm=d6+0.298Planta=pen-0.241Dorm=d0-0.197Planta=p6... <= -0.055847
-0.91Tipo=dúplex+0.231Tipo=áticos+0.215Ciudad=valencia+0.198Tipo=chalets+0.183Tipo=estudios... <= 0.249656: NO (13.0/1.0)
-0.91Tipo=dúplex+0.231Tipo=áticos+0.215Ciudad=valencia+0.198Tipo=chalets+0.183Tipo=estudios... > 0.249656
-0.675Tipo=pisos+0.458Tipo=áticos+0.378Tipo=chalets+0.319Euros+0.222Tipo=dúplex... <= -0.673205: SÍ (3.0)
-0.675Tipo=pisos+0.458Tipo=áticos+0.378Tipo=chalets+0.319Euros+0.222Tipo=dúplex... > -0.673205: NO (2.0)
0.639Planta=p5+0.432Dorm=d6+0.298Planta=pen-0.241Dorm=d0-0.197Planta=p6... > -0.055847: SÍ (4.0/1.0)
-0.701Ciudad=madrid+0.557Ciudad=barcelona+0.326Ciudad=valencia-0.236Euros-0.173Tipo=chalets... > 0.34912: SÍ (5.0)
-0.534Barrio=chamartín-0.534Barrio=chamartín+0.45 Barrio=chamberí+0.45 Barrio=chamberí-0.054Barrio=ciudad lineal... > -0.007706: NO (115.0/9.0)

Number of Leaves : 60

Size of the tree : 119

J48 pruned tree (Árbol 8º de los 6 que forman el modelo.)

0.902Barrio=centro-0.324Barrio=salamanca-0.247Barrio=carabanchel-0.093Barrio=ciudad lineal-0.077Barrio=tetuán... <= -0.322958: NO (32.0/1.0)
0.902Barrio=centro-0.324Barrio=salamanca-0.247Barrio=carabanchel-0.093Barrio=ciudad lineal-0.077Barrio=tetuán... > -0.322958
0.687Euros+0.284Barrio=hortaleza+0.27 Barrio=chamartín+0.266Barrio=salamanca+0.231Barrio=moncloa... <= -0.05634
0.668Barrio=salamanca-0.479Barrio=carabanchel-0.47Barrio=hortaleza-0.183Barrio=chamartín+0.128Barrio=centro... <= 0.095113
-0.672Barrio=tetuán+0.515Barrio=puede de vallecas-0.33Barrio=ciudad lineal-0.254Barrio=carabanchel+0.169Barrio=san blas... <= 0.514366: NO (93.0/4.0)
-0.672Barrio=tetuán+0.515Barrio=puede de vallecas-0.33Barrio=ciudad lineal-0.254Barrio=carabanchel+0.169Barrio=san blas... > 0.514366
0.61 Planta=p12-0.5Planta=p8-0.478Planta=p11+0.206Planta=pen+0.194Planta=p9... <= -0.011539: SÍ (3.0)
0.61 Planta=p12-0.5Planta=p8-0.478Planta=p11+0.206Planta=pen+0.194Planta=p9... > -0.011539
0.46 Planta=p3-0.401Dorm=d0-0.345Dorm=d4-0.329Planta=pbj-0.26Planta=p6... <= -0.237991
0.687Euros+0.284Barrio=hortaleza+0.27 Barrio=chamartín+0.266Barrio=salamanca+0.231Barrio=moncloa... <= -0.163457: NO (2.0)
0.687Euros+0.284Barrio=hortaleza+0.27 Barrio=chamartín+0.266Barrio=salamanca+0.231Barrio=moncloa... > -0.163457: SÍ (2.0)
0.46 Planta=p3-0.401Dorm=d0-0.345Dorm=d4-0.329Planta=pbj-0.26Planta=p6... > -0.237991: NO (8.0)
0.668Barrio=salamanca-0.479Barrio=carabanchel-0.47Barrio=hortaleza-0.183Barrio=chamartín+0.128Barrio=centro... > 0.095113
0.608Dorm=d3-0.445Planta=p1-0.27Dorm=d2+0.25 Ciudad=madrid-0.236Dorm=d4... <= 0.228067
-0.637Ciudad=madrid+0.495Ciudad=barcelona+0.319Ciudad=valencia+0.231Dorm=d3-0.221Dorm=d1... <= -1.065676
0.687Euros+0.284Barrio=hortaleza+0.27 Barrio=chamartín+0.266Barrio=salamanca+0.231Barrio=moncloa... <= -0.076355: NO (2.0)
0.687Euros+0.284Barrio=hortaleza+0.27 Barrio=chamartín+0.266Barrio=salamanca+0.231Barrio=moncloa... > -0.076355: SÍ (2.0)
-0.637Ciudad=madrid+0.495Ciudad=barcelona+0.319Ciudad=valencia+0.231Dorm=d3-0.221Dorm=d1... > -1.065676: SÍ (7.0)
0.608Dorm=d3-0.445Planta=p1-0.27Dorm=d2+0.25 Ciudad=madrid-0.236Dorm=d4... > 0.228067: NO (2.0)
0.687Euros+0.284Barrio=hortaleza+0.27 Barrio=chamartín+0.266Barrio=salamanca+0.231Barrio=moncloa... > -0.05634
-0.812Barrio=Rascanya+0.344Barrio=Ciutat vella-0.255Barrio=Camins al grau-0.223Barrio=campanar - beni...+0.203Barrio=villa de vallecas... <= -0.003634: NO (394.0/182.0)
-0.812Barrio=Rascanya+0.344Barrio=Ciutat vella-0.255Barrio=Camins al grau-0.223Barrio=campanar - beni...+0.203Barrio=villa de vallecas... > -0.003634
0.966Barrio=El pla del real-0.126Barrio=vicálvaro+0.117Barrio=La saïdia-0.108Barrio=Patraix-0.089Barrio=Jesús... <= -0.003652
0.966Barrio=El pla del real-0.126Barrio=vicálvaro+0.117Barrio=La saïdia-0.108Barrio=Patraix-0.089Barrio=Jesús... > -0.003652
-0.506Planta=p13+0.363Dorm=d7+0.289Planta=p5-0.287Dorm=d5+0.283Dorm=d8... <= -0.025915
0.76 Barrio=sarrià-sant gervasi-0.335Barrio=moncloa+0.326Barrio=arganzuela+0.29 Barrio=fuencarral-0.224Barrio=retiro... <= -0.037104: SÍ (545.0/181.0)
0.76 Barrio=sarrià-sant gervasi-0.335Barrio=moncloa+0.326Barrio=arganzuela+0.29 Barrio=fuencarral-0.224Barrio=retiro... > -0.037104
-0.689Barrio=gràcia+0.636Barrio=ciutat vella-0.197Barrio=usera+0.164Barrio=Ciutat vella+0.162Barrio=villa de vallecas... <= 0.032702: SÍ (367.0/160.0)
-0.689Barrio=gràcia+0.636Barrio=ciutat vella-0.197Barrio=usera+0.164Barrio=Ciutat vella+0.162Barrio=villa de vallecas... > 0.032702
0.72 Planta=p4+0.295Dorm=d1-0.22Planta=p9-0.213Planta=p3-0.18Dorm=d3... <= -0.092762
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... <= 0.628116: SÍ (3.0)
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... > 0.628116: NO (5.0)
0.72 Planta=p4+0.295Dorm=d1-0.22Planta=p9-0.213Planta=p3-0.18Dorm=d3... > -0.092762: NO (12.0)
-0.506Planta=p13+0.363Dorm=d7+0.289Planta=p5-0.287Dorm=d5+0.283Dorm=d8... > -0.025915
0.5 Planta=p1+0.399Planta=p3+0.382Planta=p2+0.347Planta=p4+0.28 Planta=p5... <= 0.057426
0.46 Planta=p3-0.401Dorm=d0-0.345Dorm=d4-0.329Planta=pbj-0.26Planta=p6... <= -0.329119: SÍ (12.0)
0.46 Planta=p3-0.401Dorm=d0-0.345Dorm=d4-0.329Planta=pbj-0.26Planta=p6... > -0.329119
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... <= 0.612298: SÍ (38.0/6.0)
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... > 0.612298
-0.618Barrio=eixample+0.342Barrio=arganzuela-0.336Barrio=latina+0.241Barrio=moncloa+0.227Barrio=retiro... <= -0.033774: SÍ (5.0)
-0.618Barrio=eixample+0.342Barrio=arganzuela-0.336Barrio=latina+0.241Barrio=moncloa+0.227Barrio=retiro... > -0.033774
-0.577Planta=p6-0.428Dorm=d6+0.388Planta=p2+0.301Dorm=d5-0.228Planta=pen... <= -0.022192

-0.721Euros+0.288Barrio=hortaleza+0.263Barrio=salamanca+0.244Barrio=chamartín+0.219Barrio=moncloa... <= 0.255567
-0.684Estado+0.354Barrio=Ciutat vella+0.235Barrio=villa de vallecas+0.218Barrio=Rascanya+0.207Barrio=salamanca... <= -0.772139: SÍ
(3.0/1.0)
-0.684Estado+0.354Barrio=Ciutat vella+0.235Barrio=villa de vallecas+0.218Barrio=Rascanya+0.207Barrio=salamanca... > -0.772139: NO (5.0)
-0.721Euros+0.288Barrio=hortaleza+0.263Barrio=salamanca+0.244Barrio=chamartín+0.219Barrio=moncloa... > 0.255567: SÍ (3.0)
-0.577Planta=p6-0.428Dorm=d6+0.388Planta=p2+0.301Dorm=d5-0.228Planta=pen... > -0.022192: NO (5.0)
0.5 Planta=p1+0.399Planta=p3+0.382Planta=p2+0.347Planta=p4+0.28 Planta=p5... > 0.057426
0.412Dorm=d0-0.319Dorm=d5-0.263Planta=p6-0.26Dorm=d1-0.257Dorm=d7... <= -0.203066
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... <= 0.610661
0.753Barrio=sant andreu-0.57Barrio=moratalaz-0.301Barrio=villa de vallecas-0.07Barrio=Quatre carreres+0.044Barrio=ciutat vella... <= -0.003759: NO
(29.0/9.0)
0.753Barrio=sant andreu-0.57Barrio=moratalaz-0.301Barrio=villa de vallecas-0.07Barrio=Quatre carreres+0.044Barrio=ciutat vella... > -0.003759: SÍ
(96.0/32.0)
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... > 0.610661: SÍ (30.0/4.0)
0.412Dorm=d0-0.319Dorm=d5-0.263Planta=p6-0.26Dorm=d1-0.257Dorm=d7... > -0.203066: SÍ (125.4/0.596.0)
0.966Barrio=El pla del real-0.126Barrio=vicálvaro+0.117Barrio=La saída-0.108Barrio=Patraix-0.089Barrio=Jesús... > -0.004221
0.412Dorm=d0-0.319Dorm=d5-0.263Planta=p6-0.26Dorm=d1-0.257Dorm=d7... <= -0.054607
0.902Barrio=centro-0.324Barrio=salamanca-0.247Barrio=carabanchel-0.093Barrio=ciudad lineal-0.077Barrio=tetuán... <= -0.006538: SÍ (4.0)
0.902Barrio=centro-0.324Barrio=salamanca-0.247Barrio=carabanchel-0.093Barrio=ciudad lineal-0.077Barrio=tetuán... > -0.006538: NO (15.0/5.0)
0.412Dorm=d0-0.319Dorm=d5-0.263Planta=p6-0.26Dorm=d1-0.257Dorm=d7... > -0.054607: SÍ (67.0/10.0)
0.966Barrio=El pla del real-0.126Barrio=vicálvaro+0.117Barrio=La saída-0.108Barrio=Patraix-0.089Barrio=Jesús... > -0.003652
0.729Barrio=Jesús-0.555Barrio=Extramurs-0.394Barrio=Patraix+0.051Barrio=campanar - beni...-0.029Barrio=benicalap... <= 0.00083
0.713Barrio=eixample-0.261Barrio=chamberí-0.251Barrio=latina-0.246Barrio=san blas-0.227Barrio=pueblo de vallecas... <= -0.175224
-0.675Barrio=latina+0.498Barrio=san blas+0.348Barrio=pueblo de vallecas-0.194Barrio=chamartín+0.153Barrio=moncloa... <= 0.36321
-0.682Planta=p5-0.3Dorm=d5-0.258Planta=pen+0.238Planta=p10+0.197Ciudad=valencia... <= 0.169385
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... <= 0.599752
0.46 Planta=p3-0.401Dorm=d0-0.345Dorm=d4-0.329Planta=pbj-0.26Planta=p6... <= 0.523953
-0.672Barrio=tetuán+0.515Barrio=pueblo de vallecas-0.33Barrio=ciudad lineal-0.254Barrio=carabanchel+0.169Barrio=san blas... <= 0.513143: SÍ
(4.0/1.0)
-0.672Barrio=tetuán+0.515Barrio=pueblo de vallecas-0.33Barrio=ciudad lineal-0.254Barrio=carabanchel+0.169Barrio=san blas... > 0.513143
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... <= 0.170242: SÍ (7.0)
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... > 0.170242
-0.675Barrio=latina+0.498Barrio=san blas+0.348Barrio=pueblo de vallecas-0.194Barrio=chamartín+0.153Barrio=moncloa... <= 0.363036
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... <= -0.033505: NO (9.0/4.0)
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... > -0.033505: SÍ (14.0)
-0.675Barrio=latina+0.498Barrio=san blas+0.348Barrio=pueblo de vallecas-0.194Barrio=chamartín+0.153Barrio=moncloa... > 0.363036
-0.675Barrio=latina+0.498Barrio=san blas+0.348Barrio=pueblo de vallecas-0.194Barrio=chamartín+0.153Barrio=moncloa... <= 0.363101
-0.67Dorm=d6+0.416Planta=pen+0.405Planta=p6-0.2Planta=pbj-0.195Planta=p7... <= -0.110982: SÍ (4.0)
-0.67Dorm=d6+0.416Planta=pen+0.405Planta=p6-0.2Planta=pbj-0.195Planta=p7... > -0.110982: NO (37.0/15.0)
-0.675Barrio=latina+0.498Barrio=san blas+0.348Barrio=pueblo de vallecas-0.194Barrio=chamartín+0.153Barrio=moncloa... > 0.363101
0.46 Planta=p3-0.401Dorm=d0-0.345Dorm=d4-0.329Planta=pbj-0.26Planta=p6... <= 0.060025: SÍ (3.0)
0.46 Planta=p3-0.401Dorm=d0-0.345Dorm=d4-0.329Planta=pbj-0.26Planta=p6... > 0.060025: NO (3.0)
0.46 Planta=p3-0.401Dorm=d0-0.345Dorm=d4-0.329Planta=pbj-0.26Planta=p6... > 0.523953: NO (9.0)
0.568Tipo=pisos-0.547Tipo=áticos-0.345Tipo=estudios-0.308Tipo=dúplex+0.269Metros... > 0.599752
-0.559Dorm=d4-0.511Planta=p3+0.292Planta=p1+0.218Dorm=d3-0.192Dorm=d9... <= 0.000062: SÍ (15.0)
-0.559Dorm=d4-0.511Planta=p3+0.292Planta=p1+0.218Dorm=d3-0.192Dorm=d9... > 0.000062
0.668Barrio=salamanca-0.479Barrio=carabanchel-0.47Barrio=hortaleza-0.183Barrio=chamartín+0.128Barrio=centro... <= 0.038856
-0.367Planta=pbj-0.332Ciudad=valencia+0.325Ciudad=barcelona+0.324Planta=p7-0.301Dorm=d4... <= 0.125257
0.714Estado+0.367Barrio=Ciutat vella+0.256Barrio=villa de vallecas+0.243Barrio=Rascanya+0.148Barrio=campanar - beni... <= 0.839225: NO
(2.0)
0.714Estado+0.367Barrio=Ciutat vella+0.256Barrio=villa de vallecas+0.243Barrio=Rascanya+0.148Barrio=campanar - beni... > 0.839225
0.54 Metros+0.54 Metros+0.46 Tipo=chalets-0.412Tipo=pisos+0.151Tipo=áticos... <= -0.319132: SÍ (10.0)
0.54 Metros+0.54 Metros+0.46 Tipo=chalets-0.412Tipo=pisos+0.151Tipo=áticos... > -0.319132: NO (2.0)
-0.367Planta=pbj-0.332Ciudad=valencia+0.325Ciudad=barcelona+0.324Planta=p7-0.301Dorm=d4... > 0.125257: NO (9.0/3.0)
0.668Barrio=salamanca-0.479Barrio=carabanchel-0.47Barrio=hortaleza-0.183Barrio=chamartín+0.128Barrio=centro... > 0.038856
0.412Dorm=d0-0.319Dorm=d5-0.263Planta=p6-0.26Dorm=d1-0.257Dorm=d7... <= -0.110136: NO (3.0/1.0)
0.412Dorm=d0-0.319Dorm=d5-0.263Planta=p6-0.26Dorm=d1-0.257Dorm=d7... > -0.110136: SÍ (19.0/1.0)
-0.682Planta=p5-0.3Dorm=d5-0.258Planta=pen+0.238Planta=p10+0.197Ciudad=valencia... > 0.169385: NO (8.0)
-0.675Barrio=latina+0.498Barrio=san blas+0.348Barrio=pueblo de vallecas-0.194Barrio=chamartín+0.153Barrio=moncloa... > 0.36321
0.668Barrio=fuencarral-0.54Barrio=sant martí-0.364Barrio=sarrià-sant gervasi+0.188Barrio=arganzuela+0.187Barrio=retiro... <= 0.00784: NO (8.0/1.0)
0.668Barrio=fuencarral-0.54Barrio=sant martí-0.364Barrio=sarrià-sant gervasi+0.188Barrio=arganzuela+0.187Barrio=retiro... > 0.00784: SÍ (4.0)
0.713Barrio=eixample-0.261Barrio=chamberí-0.251Barrio=latina-0.246Barrio=san blas-0.227Barrio=pueblo de vallecas... <= -0.175224
-0.683Barrio=carabanchel+0.493Barrio=ciudad lineal-0.366Barrio=salamanca-0.238Barrio=centro+0.224Barrio=tetuán... <= -0.366242
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... <= 0.344611
-0.684Estado+0.354Barrio=Ciutat vella+0.235Barrio=villa de vallecas+0.218Barrio=Rascanya+0.207Barrio=salamanca... <= -0.735741
-0.367Planta=pbj-0.332Ciudad=valencia+0.325Ciudad=barcelona+0.324Planta=p7-0.301Dorm=d4... <= -0.036589
0.668Barrio=salamanca-0.479Barrio=carabanchel-0.47Barrio=hortaleza-0.183Barrio=chamartín+0.128Barrio=centro... <= -0.461062: NO (2.0)
0.668Barrio=salamanca-0.479Barrio=carabanchel-0.47Barrio=hortaleza-0.183Barrio=chamartín+0.128Barrio=centro... > -0.461062
-0.682Planta=p7-0.486Planta=pen+0.204Planta=p5+0.196Planta=p9+0.187Planta=p6... <= 0.03019: SÍ (9.0/2.0)
-0.682Planta=p7-0.486Planta=pen+0.204Planta=p5+0.196Planta=p9+0.187Planta=p6... > 0.03019: NO (2.0)
-0.367Planta=pbj-0.332Ciudad=valencia+0.325Ciudad=barcelona+0.324Planta=p7-0.301Dorm=d4... > -0.036589: SÍ (15.0)
-0.684Estado+0.354Barrio=Ciutat vella+0.235Barrio=villa de vallecas+0.218Barrio=Rascanya+0.207Barrio=salamanca... > -0.735741
0.403Dorm=d5-0.332Planta=p8-0.324Planta=p9+0.277Planta=p2+0.275Dorm=d9... <= 0.222106
0.54 Metros+0.54 Metros+0.46 Tipo=chalets-0.412Tipo=pisos+0.151Tipo=áticos... <= -0.351948
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... <= -0.045664
-0.675Barrio=latina+0.498Barrio=san blas+0.348Barrio=pueblo de vallecas-0.194Barrio=chamartín+0.153Barrio=moncloa... <= 0.094891
-0.682Planta=p5-0.3Dorm=d5-0.258Planta=pen+0.238Planta=p10+0.197Ciudad=valencia... <= -0.019398: NO (4.0)
-0.682Planta=p5-0.3Dorm=d5-0.258Planta=pen+0.238Planta=p10+0.197Ciudad=valencia... > -0.019398
0.54 Metros+0.54 Metros+0.46 Tipo=chalets-0.412Tipo=pisos+0.151Tipo=áticos... <= -0.370543: NO (5.0/1.0)
0.54 Metros+0.54 Metros+0.46 Tipo=chalets-0.412Tipo=pisos+0.151Tipo=áticos... > -0.370543: SÍ (8.0/1.0)
-0.675Barrio=latina+0.498Barrio=san blas+0.348Barrio=pueblo de vallecas-0.194Barrio=chamartín+0.153Barrio=moncloa... > 0.094891: SÍ
(4.0/1.0)
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... > -0.045664: NO (29.0/3.0)
0.54 Metros+0.54 Metros+0.46 Tipo=chalets-0.412Tipo=pisos+0.151Tipo=áticos... > -0.351948
0.752Barrio=chamartín-0.543Barrio=hortaleza+0.248Barrio=pueblo de vallecas-0.169Barrio=latina-0.131Barrio=san blas... <= 0.052695
0.412Dorm=d0-0.319Dorm=d5-0.263Planta=p6-0.26Dorm=d1-0.257Dorm=d7... <= 0.363366: SÍ (8.0/33.0)
0.412Dorm=d0-0.319Dorm=d5-0.263Planta=p6-0.26Dorm=d1-0.257Dorm=d7... > 0.363366: NO (5.0)
0.752Barrio=chamartín-0.543Barrio=hortaleza+0.248Barrio=pueblo de vallecas-0.169Barrio=latina-0.131Barrio=san blas... > 0.052695
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... <= 0.128252: SÍ (9.0/3.0)
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... > 0.128252: NO (4.0)
0.403Dorm=d5-0.332Planta=p8-0.324Planta=p9+0.277Planta=p2+0.275Dorm=d9... > 0.222106: SÍ (14.0/2.0)
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... > 0.344611: SÍ (13.0/1.0)
-0.683Barrio=carabanchel+0.493Barrio=ciudad lineal-0.366Barrio=salamanca-0.238Barrio=centro+0.224Barrio=tetuán... > -0.366242
-0.367Planta=pbj-0.332Ciudad=valencia+0.325Ciudad=barcelona+0.324Planta=p7-0.301Dorm=d4... <= -0.283909
-0.683Barrio=carabanchel+0.493Barrio=ciudad lineal-0.366Barrio=salamanca-0.238Barrio=centro+0.224Barrio=tetuán... <= -0.003004: NO (9.0/1.0)
-0.683Barrio=carabanchel+0.493Barrio=ciudad lineal-0.366Barrio=salamanca-0.238Barrio=centro+0.224Barrio=tetuán... > -0.003004: SÍ (2.0)
-0.367Planta=pbj-0.332Ciudad=valencia+0.325Ciudad=barcelona+0.324Planta=p7-0.301Dorm=d4... > -0.283909: SÍ (8.0)
0.729Barrio=Jesús-0.555Barrio=Extramurs-0.394Barrio=Patraix+0.051Barrio=campanar - beni...-0.029Barrio=benicalap... <= 0.00083
0.715Dorm=d8-0.565Dorm=d7+0.271Dorm=d9-0.167Planta=p8+0.166Planta=p4... <= 0.00693
-0.721Euros+0.288Barrio=hortaleza+0.263Barrio=salamanca+0.244Barrio=chamartín+0.219Barrio=moncloa... <= -0.02599
-0.683Barrio=carabanchel+0.493Barrio=ciudad lineal-0.366Barrio=salamanca-0.238Barrio=centro+0.224Barrio=tetuán... <= -0.23789

0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... <= -0.125899: NO (8.0)
0.529Ciudad=valencia-0.484Ciudad=barcelona+0.28 Planta=p7-0.271Planta=pen+0.238Planta=p5... > -0.125899: SÍ (12.0/1.0)
-0.683Barrio=carabanchel+0.493Barrio=ciudad lineal-0.366Barrio=salamanca-0.238Barrio=centro+0.224Barrio=tetuán... > -0.23789
-0.683Barrio=carabanchel+0.493Barrio=ciudad lineal-0.366Barrio=salamanca-0.238Barrio=centro+0.224Barrio=tetuán... <= 0.218123: NO (144.0/40.0)
-0.683Barrio=carabanchel+0.493Barrio=ciudad lineal-0.366Barrio=salamanca-0.238Barrio=centro+0.224Barrio=tetuán... > 0.218123
-0.682Planta=p5-0.3Dorm=d5-0.258Planta=pen+0.238Planta=p10+0.197Ciudad=valencia... <= 0.006484
0.777Tipo=estudios-0.603Tipo=dúplex+0.167Tipo=chalets-0.123Tipo=áticos-0.011Tipo=pisos... <= -0.013522: SÍ (7.0)
0.777Tipo=estudios-0.603Tipo=dúplex+0.167Tipo=chalets-0.123Tipo=áticos-0.011Tipo=pisos... > -0.013522
0.687Euros+0.284Barrio=hortaleza+0.27 Barrio=chamartín+0.266Barrio=salamanca+0.231Barrio=moncloa... <= 0.134011: SÍ (3.0)
0.687Euros+0.284Barrio=hortaleza+0.27 Barrio=chamartín+0.266Barrio=salamanca+0.231Barrio=moncloa... > 0.134011: NO (4.0/1.0)
-0.682Planta=p5-0.3Dorm=d5-0.258Planta=pen+0.238Planta=p10+0.197Ciudad=valencia... > 0.006484: NO (3.0)
-0.721Euros+0.288Barrio=hortaleza+0.263Barrio=salamanca+0.244Barrio=chamartín+0.219Barrio=moncloa... > -0.02599: SÍ (595.0/297.0)
0.715Dorm=d8-0.565Dorm=d7+0.271Dorm=d9-0.167Planta=p8+0.166Planta=p4... > 0.00693
0.54 Metros+0.54 Metros+0.46 Tipo=chalets-0.412Tipo=pisos+0.151Tipo=áticos... <= -0.36398
0.74 Barrio=ciudad lineal-0.601Barrio=tetuán+0.186Barrio=carabanchel-0.111Barrio=puede de vallecas+0.105Barrio=salamanca... <= 0.188506: NO (15.0/1.0)
0.74 Barrio=ciudad lineal-0.601Barrio=tetuán+0.186Barrio=carabanchel-0.111Barrio=puede de vallecas+0.105Barrio=salamanca... > 0.188506: SÍ (4.0/1.0)
0.54 Metros+0.54 Metros+0.46 Tipo=chalets-0.412Tipo=pisos+0.151Tipo=áticos... > -0.36398: SÍ (351.0/145.0)

Number of Leaves : 78
Size of the tree : 155

J48 pruned tree (Árbol 9º de los 10 que forman el modelo.)

0.564Estado+0.376Tipo=chalets-0.334Tipo=dúplex-0.301Barrio=Ciutat vella-0.248Barrio=campanar - beni..... <= -0.039514
0.554Barrio=moratalaz-0.469Barrio=horta guinardó+0.311Barrio=Camins al grau+0.299Barrio=sant andreu+0.255Barrio=L'exemple... <= -0.022577
0.603Planta=pbj-0.486Planta=p4+0.441Planta=p5-0.266Planta=p3+0.254Planta=p6... <= -0.271606: SÍ (5.0)
0.603Planta=pbj-0.486Planta=p4+0.441Planta=p5-0.266Planta=p3+0.254Planta=p6... > -0.271606
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... <= -0.004871
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... <= 0.690559: SÍ (9.0)
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... > 0.690559: NO (4.0/1.0)
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... > -0.004871
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... <= -0.168976: SÍ (4.0)
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... > -0.168976
0.406Barrio=hortaleza-0.372Barrio=latina-0.324Barrio=ciudad lineal+0.323Barrio=salamanca+0.285Barrio=ciutat vella... <= 0.050705
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... <= 0.026355
-0.773Barrio=sarrià-sant gervasi+0.361Barrio=arganzuela+0.344Barrio=usera+0.245Barrio=retiro+0.108Barrio=sant martí... <= -0.037348
-0.601Dorm=d5+0.588Dorm=d6+0.512Dorm=d0+0.119Dorm=d7+0.089Dorm=d8... <= -0.016439
-0.767Dorm=d3+0.621Dorm=d2+0.111Dorm=d4+0.062Dorm=d5+0.057Euros... <= 0.049549
-0.767Dorm=d3+0.621Dorm=d2+0.111Dorm=d4+0.062Dorm=d5+0.057Euros... <= 0.049441: NO (7.0)
-0.767Dorm=d3+0.621Dorm=d2+0.111Dorm=d4+0.062Dorm=d5+0.057Euros... > 0.049441: SÍ (3.0)
-0.767Dorm=d3+0.621Dorm=d2+0.111Dorm=d4+0.062Dorm=d5+0.057Euros... > 0.049549: NO (16.0)
-0.601Dorm=d5+0.588Dorm=d6+0.512Dorm=d0+0.119Dorm=d7+0.089Dorm=d8... > -0.016439: SÍ (2.0)
-0.773Barrio=sarrià-sant gervasi+0.361Barrio=arganzuela+0.344Barrio=usera+0.245Barrio=retiro+0.108Barrio=sant martí... > -0.037348: NO (25.0)
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... <= 0.026355: SÍ (4.0/1.0)
0.406Barrio=hortaleza-0.372Barrio=latina-0.324Barrio=ciudad lineal+0.323Barrio=salamanca+0.285Barrio=ciutat vella... > 0.050705
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... <= -0.05397
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... <= -0.475391: SÍ (2.0)
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... > -0.475391: NO (3.0)
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... > -0.05397: SÍ (4.0)
0.554Barrio=moratalaz-0.469Barrio=horta guinardó+0.311Barrio=Camins al grau+0.299Barrio=sant andreu+0.255Barrio=L'exemple... > -0.022577
0.552Barrio=tetuán-0.531Barrio=carabanchel+0.418Barrio=centro-0.316Barrio=chamberí-0.212Barrio=ciudad lineal... <= 0.159649
0.664Barrio=centro-0.549Barrio=tetuán-0.308Barrio=ciudad lineal+0.229Barrio=chamberí-0.208Barrio=puede de vallecas... <= -0.100194
-0.905Dorm=d9+0.29 Dorm=d7-0.279Dorm=d8+0.128Dorm=d6+0.042Dorm=d5... <= -0.004408
-0.517Ciudad=madrid-0.517Ciudad=madrid+0.382Ciudad=barcelona+0.382Ciudad=barcelona+0.276Ciudad=valencia... <= -1.115917
-0.662Barrio=chamberí+0.445Barrio=puede de vallecas+0.411Barrio=chamartín-0.205Barrio=tetuán+0.16 Barrio=centro... <= 0.434202: SÍ (3.0/1.0)
-0.662Barrio=chamberí+0.445Barrio=puede de vallecas+0.411Barrio=chamartín-0.205Barrio=tetuán+0.16 Barrio=centro... > 0.434202: NO (4.0)
-0.517Ciudad=madrid-0.517Ciudad=madrid+0.382Ciudad=barcelona+0.382Ciudad=barcelona+0.276Ciudad=valencia... > -1.115917: SÍ (4.0)
-0.905Dorm=d9+0.29 Dorm=d7-0.279Dorm=d8+0.128Dorm=d6+0.042Dorm=d5... > -0.004408: NO (31.0/1.0)
0.664Barrio=centro-0.549Barrio=tetuán-0.308Barrio=ciudad lineal+0.229Barrio=chamberí-0.208Barrio=puede de vallecas... > -0.100194: NO (103.0/2.0)
0.552Barrio=tetuán-0.531Barrio=carabanchel+0.418Barrio=centro-0.316Barrio=chamberí-0.212Barrio=ciudad lineal... > 0.159649
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... <= -0.00425: SÍ (5.0)
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... > -0.00425
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... <= -0.090111: SÍ (6.0)
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... > -0.090111: NO (18.0)
0.564Estado+0.376Tipo=chalets-0.334Tipo=dúplex-0.301Barrio=Ciutat vella-0.248Barrio=campanar - beni..... <= -0.039514
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... <= -0.426644
-0.846Planta=p1+0.319Planta=p2+0.296Planta=p3+0.194Planta=p4+0.14 Planta=pbj... <= 0.380724
0.603Planta=pbj-0.486Planta=p4+0.441Planta=p5-0.266Planta=p3+0.254Planta=p6... <= -0.089449: NO (24.0)
0.603Planta=pbj-0.486Planta=p4+0.441Planta=p5-0.266Planta=p3+0.254Planta=p6... > -0.089449
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... <= -0.318827: SÍ (3.0)
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... > -0.318827: NO (4.0)
-0.846Planta=p1+0.319Planta=p2+0.296Planta=p3+0.194Planta=p4+0.14 Planta=pbj... > 0.380724: SÍ (2.0)
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... <= -0.426644
0.675Barrio=vicálvaro+0.492Barrio=Benimaclet-0.421Barrio=Jesús-0.316Barrio=Quatre carreres-0.082Barrio=campanar - beni..... <= -0.017276
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... <= -0.542037: SÍ (234.0/72.0)
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... > -0.542037
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... <= 0.395992
-0.571Planta=p13-0.571Planta=p14-0.484Planta=pst+0.299Planta=pss-0.117Planta=p12... <= 0.03454
0.664Barrio=centro-0.549Barrio=tetuán-0.308Barrio=ciudad lineal+0.229Barrio=chamberí-0.208Barrio=puede de vallecas... <= 0.000712
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... <= 0.032104: NO (4.0)
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... > 0.032104
0.537Ciudad=valencia+0.537Ciudad=valencia-0.429Ciudad=barcelona-0.429Ciudad=barcelona+0.124Planta=p12... <= 1.117243
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... <= -0.417253
-0.846Planta=p1+0.319Planta=p2+0.296Planta=p3+0.194Planta=p4+0.14 Planta=pbj... <= -0.020503: SÍ (5.0)
-0.846Planta=p1+0.319Planta=p2+0.296Planta=p3+0.194Planta=p4+0.14 Planta=pbj... > -0.020503
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... <= -0.528432
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... <= 0.827432
0.752Dorm=d7+0.571Dorm=d8-0.318Dorm=d6-0.072Dorm=d5+0.023Metros... <= -0.004315
0.752Dorm=d7+0.571Dorm=d8-0.318Dorm=d6-0.072Dorm=d5+0.023Metros... <= -0.004569: SÍ (3.0/1.0)
0.752Dorm=d7+0.571Dorm=d8-0.318Dorm=d6-0.072Dorm=d5+0.023Metros... > -0.004569: NO (6.0)
0.752Dorm=d7+0.571Dorm=d8-0.318Dorm=d6-0.072Dorm=d5+0.023Metros... > -0.004315: SÍ (2.0)

0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... > 0.827432: NO (16.0/6.0)
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... > -0.528432
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... <= -0.004736: SÍ (3.0)
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... > -0.004736: NO (2.0)
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... > -0.417253
-0.804Planta=p6+0.523Planta=pbj+0.17 Planta=p12-0.133Planta=p5+0.104Planta=p11... <= 0.093703
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... <= -0.01166: NO (15.0/1.0)
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... > -0.01166
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... <= -0.011419: SÍ (3.0)
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... > -0.011419: NO (2.0)
-0.804Planta=p6+0.523Planta=pbj+0.17 Planta=p12-0.133Planta=p5+0.104Planta=p11... > 0.093703: SÍ (5.0/1.0)
0.537Ciudad=valencia+0.537Ciudad=valencia-0.429Ciudad=barcelona-0.429Ciudad=barcelona+0.124Planta=p12... > 1.117243
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... <= -0.417253
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... <= -0.474957: SÍ (5.0)
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... > -0.474957: NO (2.0)
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... > -0.417253: SÍ (7.0)
0.664Barrio=centro-0.549Barrio=tetuán-0.308Barrio=ciudad lineal+0.229Barrio=chamberí-0.208Barrio=pueblo de valdecarlos... > 0.000712
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... <= -0.417253
0.499Planta=p1+0.404Planta=p2+0.384Planta=p3+0.352Planta=p4+0.286Planta=pbj... <= 0.376851
0.894Planta=p9-0.402Planta=pen+0.103Planta=pst-0.094Planta=p11-0.094Planta=p12... <= -0.021618: SÍ (6.0)
0.894Planta=p9-0.402Planta=pen+0.103Planta=pst-0.094Planta=p11-0.094Planta=p12... > -0.021618
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... <= -0.528432
0.7 Barrio=carabanchel-0.528Barrio=ciudad lineal-0.36Barrio=salamanca+0.203Barrio=tetuán+0.109Barrio=latina... <= 0.017055: SÍ (51.0/16.0)
0.7 Barrio=carabanchel-0.528Barrio=ciudad lineal-0.36Barrio=salamanca+0.203Barrio=tetuán+0.109Barrio=latina... > 0.017055: NO (3.0)
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... > -0.528432
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... <= 0.803956: SÍ (4.0)
0.679Tipo=pisos-0.464Tipo=áticos-0.293Tipo=chalets-0.265Tipo=dúplex-0.194Tipo=estudios... > 0.803956
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... <= -0.005108: NO (4.0)
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... > -0.005108: SÍ (8.0/1.0)
0.499Planta=p1+0.404Planta=p2+0.384Planta=p3+0.352Planta=p4+0.286Planta=pbj... > 0.376851
0.587Barrio=moncloa-0.516Barrio=fuencarral+0.231Barrio=ciudad vella-0.22Barrio=salamanca-0.206Barrio=pueblo de valdecarlos... <= 0.014568
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... <= -0.474403
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... <= 0.008196: NO (4.0)
-0.757Dorm=d8+0.509Dorm=d7+0.374Dorm=d9-0.127Dorm=d6-0.079Dorm=d5... > 0.008196
0.752Dorm=d7+0.571Dorm=d8-0.318Dorm=d6-0.072Dorm=d5+0.023Metros... <= -0.005166: NO (2.0)
0.752Dorm=d7+0.571Dorm=d8-0.318Dorm=d6-0.072Dorm=d5+0.023Metros... > -0.005166: SÍ (2.0)
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... > -0.474403: SÍ (2.0)
0.587Barrio=moncloa-0.516Barrio=fuencarral+0.231Barrio=ciudad vella-0.22Barrio=salamanca-0.206Barrio=pueblo de valdecarlos... > 0.014568: NO (3.0)
-0.542Dorm=d3-0.528Dorm=d2-0.417Dorm=d1-0.393Dorm=d4-0.208Dorm=d5... > -0.417253
0.424Barrio=villa de valdecarlos+0.405Barrio=sant martí-0.362Barrio=horta guinardó-0.347Barrio=nou barris-0.282Barrio=moratalaz... <= -0.372465
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... <= 0.298774: SÍ (2.0)
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... > 0.298774: NO (5.0/1.0)
0.424Barrio=villa de valdecarlos+0.405Barrio=sant martí-0.362Barrio=horta guinardó-0.347Barrio=nou barris-0.282Barrio=moratalaz... > -0.372465
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... <= -0.020308: NO (2.0)
0.786Dorm=d1-0.472Dorm=d2-0.307Dorm=d3+0.214Dorm=d4+0.097Dorm=d0... > -0.020308: SÍ (29.0/2.0)
-0.571Planta=p13-0.571Planta=p14-0.484Planta=pst+0.299Planta=ps-0.117Planta=p12... > 0.03454: SÍ (14.0/1.0)
0.596Metros+0.566Euros+0.303Dorm=d6+0.266Dorm=d5+0.256Dorm=d4... > 0.395992: NO (12.0/1.0)
0.675Barrio=vicálvaro+0.492Barrio=Benimaclet-0.421Barrio=Jesús-0.316Barrio=Quatre carreres-0.082Barrio=campanar - beni... > -0.017276: SÍ (3821.0/1731.0)

Number of Leaves : 59
Size of the tree : 117

J48 pruned tree (Árbol 10º de los 10 que forman el modelo.)

-0.757Estado-0.341Planta=p1-0.307Dorm=d1-0.222Dorm=d0+0.166Dorm=d5... <= -0.436024
0.514Estado-0.512Planta=p1+0.443Dorm=d3-0.313Dorm=d1-0.228Dorm=d2... <= -0.640523
0.838Barrio=ciudad lineal-0.368Barrio=chamartín-0.241Barrio=centro+0.199Barrio=salamanca-0.16Barrio=carabanchel... <= 0.202706
-0.734Barrio=barajas+0.339Barrio=les corts+0.33 Barrio=Ciutat vella+0.286Barrio=nou barris-0.225Barrio=Camins al grau... <= -0.222231: SÍ (6.0)
-0.734Barrio=barajas+0.339Barrio=les corts+0.33 Barrio=Ciutat vella+0.286Barrio=nou barris-0.225Barrio=Camins al grau... > -0.222231
0.556Ciudad=valencia+0.556Ciudad=valencia-0.428Ciudad=barcelona-0.428Ciudad=barcelona-0.119Euros... <= 1.107899
-0.946Barrio=L'olivereta-0.151Barrio=campanar - beni...-0.15Barrio=La saïdia+0.118Barrio=Algirós+0.105Barrio=Jesús... <= 0.006696
0.597Barrio=Extramurs-0.472Barrio=L'eixample+0.394Barrio=les corts+0.278Barrio=Poblat marítim-0.205Barrio=Rascanya... <= -0.01827: NO (19.0)
0.597Barrio=Extramurs-0.472Barrio=L'eixample+0.394Barrio=les corts+0.278Barrio=Poblat marítim-0.205Barrio=Rascanya... > -0.01827
0.708Barrio=pueblo de valdecarlos-0.453Barrio=carabanchel-0.3Barrio=tetuán+0.278Barrio=chamartín-0.194Barrio=chamberí... <= -0.136458
-0.439Barrio=Extramurs-0.366Barrio=villa de valdecarlos+0.334Barrio=ciutat vella+0.288Barrio=Poblat marítim-0.273Barrio=Camins al grau... <= -0.009605
0.51 Tipo=dúplex-0.469Tipo=estudios+0.3 Barrio=Ciutat vella-0.284Barrio=tetuán+0.263Barrio=Quatre carreres... <= -0.294307: SÍ (2.0)
0.51 Tipo=dúplex-0.469Tipo=estudios+0.3 Barrio=Ciutat vella-0.284Barrio=tetuán+0.263Barrio=Quatre carreres... > -0.294307: NO (2.0)
-0.439Barrio=Extramurs-0.366Barrio=villa de valdecarlos+0.334Barrio=ciutat vella+0.288Barrio=Poblat marítim-0.273Barrio=Camins al grau... > -0.009605: SÍ (7.0)
0.708Barrio=pueblo de valdecarlos-0.453Barrio=carabanchel-0.3Barrio=tetuán+0.278Barrio=chamartín-0.194Barrio=chamberí... > -0.136458
-0.734Barrio=barajas+0.339Barrio=les corts+0.33 Barrio=Ciutat vella+0.286Barrio=nou barris-0.225Barrio=Camins al grau... <= 0.00228: NO (8.0)
-0.734Barrio=barajas+0.339Barrio=les corts+0.33 Barrio=Ciutat vella+0.286Barrio=nou barris-0.225Barrio=Camins al grau... > 0.00228
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.045074: NO (6.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.045074
-0.734Barrio=barajas+0.339Barrio=les corts+0.33 Barrio=Ciutat vella+0.286Barrio=nou barris-0.225Barrio=Camins al grau... <= 0.002323: SÍ (4.0)
-0.734Barrio=barajas+0.339Barrio=les corts+0.33 Barrio=Ciutat vella+0.286Barrio=nou barris-0.225Barrio=Camins al grau... > 0.002323: NO (8.0/2.0)
-0.946Barrio=L'olivereta-0.151Barrio=campanar - beni...-0.15Barrio=La saïdia+0.118Barrio=Algirós+0.105Barrio=Jesús... > 0.006696: NO (49.0)
0.556Ciudad=valencia+0.556Ciudad=valencia-0.428Ciudad=barcelona-0.428Ciudad=barcelona-0.119Euros... > 1.107899: SÍ (7.0/1.0)
0.838Barrio=ciudad lineal-0.368Barrio=chamartín-0.241Barrio=centro+0.199Barrio=salamanca-0.16Barrio=carabanchel... > 0.202706: SÍ (11.0/1.0)
0.514Estado-0.512Planta=p1+0.443Dorm=d3-0.313Dorm=d1-0.228Dorm=d2... > -0.640523
0.992Euros+0.077Ciudad=valencia+0.077Ciudad=valencia-0.036Ciudad=barcelona-0.036Ciudad=barcelona... <= -0.033829
-0.48Ciudad=madrid-0.47Ciudad=madrid-0.436Ciudad=barcelona-0.365Ciudad=barcelona-0.337Ciudad=valencia... <= -0.950285
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.043579
-0.79Barrio=Jesús+0.51 Barrio=Algirós+0.286Barrio=vicálvaro-0.097Barrio=Poblat marítim+0.082Barrio=Camins al grau... <= 0.003522
0.734Barrio=Algirós+0.369Barrio=Jesús-0.357Barrio=benicalap-0.286Barrio=vicálvaro-0.254Barrio=L'eixample... <= -0.005249: SÍ (43.0/7.0)
0.734Barrio=Algirós+0.369Barrio=Jesús-0.357Barrio=benicalap-0.286Barrio=vicálvaro-0.254Barrio=L'eixample... > -0.005249
-0.743Metros+0.529Tipo=chalets-0.168Tipo=pisos-0.146Tipo=estudios+0.128Barrio=chamberí... <= -0.259357
-0.625Planta=p5+0.372Dorm=d7+0.336Planta=ps+0.308Dorm=d0+0.287Planta=p6... <= -0.151102: SÍ (23.0/2.0)
-0.625Planta=p5+0.372Dorm=d7+0.336Planta=ps+0.308Dorm=d0+0.287Planta=p6... > -0.151102
0.6 Barrio=gràcia-0.488Barrio=usera-0.371Barrio=horta guinardó+0.267Barrio=villaverde-0.219Barrio=villa de valdecarlos... <= 0.019916
0.524Planta=p15-0.481Planta=p14+0.38 Dorm=d8-0.344Planta=p22+0.282Planta=p11... <= 0.007188: SÍ (4.0)
0.524Planta=p15-0.481Planta=p14+0.38 Dorm=d8-0.344Planta=p22+0.282Planta=p11... > 0.007188: NO (7.0/1.0)
0.6 Barrio=gràcia-0.488Barrio=usera-0.371Barrio=horta guinardó+0.267Barrio=villaverde-0.219Barrio=villa de valdecarlos... > 0.019916
0.503Dorm=d9+0.486Planta=pst+0.264Planta=p11-0.251Planta=p13-0.234Planta=p10... <= -0.009996: SÍ (9.0)
0.503Dorm=d9+0.486Planta=pst+0.264Planta=p11-0.251Planta=p13-0.234Planta=p10... > -0.009996

-0.755Barrio=villaverde+0.331Barrio=sant andreu-0.254Barrio=villa de vallecas+0.239Barrio=retiro-0.153Barrio=nou barris... <= 0.029516: SÍ (3.0)
-0.755Barrio=villaverde+0.331Barrio=sant andreu-0.254Barrio=villa de vallecas+0.239Barrio=retiro-0.153Barrio=nou barris... > 0.029516
-0.806Barrio=campanar - beni...-0.333Barrio=El pla del real-0.262Barrio=La saïdia+0.238Barrio=L'olivereta-0.179Barrio=vicálvaro... <= -
0.004603: NO (2.0)
-0.806Barrio=campanar - beni...-0.333Barrio=El pla del real-0.262Barrio=La saïdia+0.238Barrio=L'olivereta-0.179Barrio=vicálvaro... > -0.004603
-0.835Barrio=centro+0.383Barrio=tetuán+0.323Barrio=chamartín+0.171Barrio=carabanchel-0.113Barrio=hortaleza... <= 0.069632: SÍ (2.0)
-0.835Barrio=centro+0.383Barrio=tetuán+0.323Barrio=chamartín+0.171Barrio=carabanchel-0.113Barrio=hortaleza... > 0.069632: NO (4.0/1.0)
-0.743Metros+0.529Tipo=chalets-0.168Tipo=pisos-0.146Tipo=estudios+0.128Barrio=chamberí... > -0.259357
0.626Planta=p11-0.575Planta=pst+0.274Planta=p8-0.189Planta=p7-0.155Planta=p22... <= -0.164455: NO (5.0)
0.626Planta=p11-0.575Planta=pst+0.274Planta=p8-0.189Planta=p7-0.155Planta=p22... > -0.164455
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.043625
0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... <= -0.186029: SÍ (4.0)
0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... > -0.186029
0.514Estado-0.512Planta=p1+0.443Dorm=d3-0.313Dorm=d1-0.228Dorm=d2... <= 0.485603
-0.42Planta=pbj+0.409Planta=p8+0.34 Planta=p5-0.316Dorm=d9+0.295Dorm=d7... <= -0.017575: NO (42.0/15.0)
-0.42Planta=pbj+0.409Planta=p8+0.34 Planta=p5-0.316Dorm=d9+0.295Dorm=d7... > -0.017575
-0.743Metros+0.529Tipo=chalets-0.168Tipo=pisos-0.146Tipo=estudios+0.128Barrio=chamberí... <= -0.256174: NO (2.0)
-0.743Metros+0.529Tipo=chalets-0.168Tipo=pisos-0.146Tipo=estudios+0.128Barrio=chamberí... > -0.256174
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.043683: SÍ (10.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.043683
-0.557Tipo=pisos+0.48 Tipo=chalets+0.458Metros+0.271Tipo=àtics+0.231Barrio=hortaleza... <= -0.592986: SÍ (2.0)
-0.557Tipo=pisos+0.48 Tipo=chalets+0.458Metros+0.271Tipo=àtics+0.231Barrio=hortaleza... > -0.592986: NO (2.0)
0.514Estado-0.512Planta=p1+0.443Dorm=d3-0.313Dorm=d1-0.228Dorm=d2... > 0.485603: NO (6.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.043625: SÍ (6.0)
-0.79Barrio=Jesús+0.51 Barrio=Algirós+0.286Barrio=vicálvaro-0.097Barrio=Poblat marítims+0.082Barrio=Camins al grau... > 0.003522
0.489Planta=pen-0.441Dorm=d6-0.335Planta=p4+0.329Dorm=d0+0.247Planta=p6... <= 0.286365
-0.553Dorm=d7+0.429Planta=pss-0.338Dorm=d0+0.284Planta=p7+0.268Planta=p8... <= -0.048899
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.043616
-0.835Barrio=centro+0.383Barrio=tetuán+0.323Barrio=chamartín+0.171Barrio=carabanchel-0.113Barrio=hortaleza... <= 0.003761
0.666Planta=p2-0.575Planta=p3+0.228Dorm=d5-0.156Dorm=d2+0.138Dorm=d8... <= 0.077836: SÍ (4.0)
0.666Planta=p2-0.575Planta=p3+0.228Dorm=d5-0.156Dorm=d2+0.138Dorm=d8... > 0.077836: NO (5.0/1.0)
-0.835Barrio=centro+0.383Barrio=tetuán+0.323Barrio=chamartín+0.171Barrio=carabanchel-0.113Barrio=hortaleza... > 0.003761: NO (7.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.043616: SÍ (2.0)
0.417Planta=p5-0.329Dorm=d6+0.327Dorm=d5-0.316Planta=p9-0.3Planta=p4... > -0.048899: SÍ (6.0)
0.489Planta=pen-0.441Dorm=d6-0.335Planta=p4+0.329Dorm=d0+0.247Planta=p6... > 0.286365: NO (11.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.043579: SÍ (19.0/1.0)
-0.48Ciudad=madrid-0.47Ciudad=madrid-0.436Ciudad=barcelona-0.365Ciudad=barcelona-0.337Ciudad=valencia... > -0.950285
0.838Barrio=ciudad lineal-0.368Barrio=chamartín-0.241Barrio=centro+0.199Barrio=salamanca-0.16Barrio=carabanchel... <= 0.000027: SÍ (390.0/131.0)
0.838Barrio=ciudad lineal-0.368Barrio=chamartín-0.241Barrio=centro+0.199Barrio=salamanca-0.16Barrio=carabanchel... > 0.000027: NO (63.0/28.0)
0.992Euros+0.077Ciudad=valencia+0.077Ciudad=valencia-0.036Ciudad=barcelona-0.036Ciudad=barcelona... > -0.033829
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.049384
0.487Planta=p1+0.406Planta=p2+0.365Planta=p3+0.354Planta=p4+0.28 Planta=pbj... <= 0.196783: SÍ (12.0)
0.487Planta=p1+0.406Planta=p2+0.365Planta=p3+0.354Planta=p4+0.28 Planta=pbj... > 0.196783
-0.553Dorm=d7+0.429Planta=pss-0.338Dorm=d0+0.284Planta=p7+0.268Planta=p8... <= 0.02089: SÍ (146.0/42.0)
-0.553Dorm=d7+0.429Planta=pss-0.338Dorm=d0+0.284Planta=p7+0.268Planta=p8... > 0.02089
-0.619Planta=p4-0.429Dorm=d5+0.376Planta=p2+0.308Planta=p3-0.206Dorm=d9... <= 0.003856
-0.619Planta=p4-0.429Dorm=d5+0.376Planta=p2+0.308Planta=p3-0.206Dorm=d9... <= -0.060736
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.0561: NO (11.0/1.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.0561
0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... <= -0.090729: NO (2.0)
0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... > -0.090729: SÍ (8.0)
-0.619Planta=p4-0.429Dorm=d5+0.376Planta=p2+0.308Planta=p3-0.206Dorm=d9... > -0.060736: SÍ (4.0)
-0.619Planta=p4-0.429Dorm=d5+0.376Planta=p2+0.308Planta=p3-0.206Dorm=d9... > 0.003856: NO (6.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.049384
-0.553Dorm=d7+0.429Planta=pss-0.338Dorm=d0+0.284Planta=p7+0.268Planta=p8... <= -0.043858
0.703Barrio=Patraix-0.689Barrio=Benimaclet+0.176Barrio=El pla del real-0.008Barrio=L'olivereta-0.008Barrio=campanar - beni... <= -0.001383
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.044922
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.045422
0.417Planta=p5-0.329Dorm=d6+0.327Dorm=d5-0.316Planta=p9-0.3Planta=p4... <= 0.213736: SÍ (3.0)
0.417Planta=p5-0.329Dorm=d6+0.327Dorm=d5-0.316Planta=p9-0.3Planta=p4... > 0.213736: NO (4.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.045422: NO (13.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.044922: SÍ (281.0/134.0)
0.703Barrio=Patraix-0.689Barrio=Benimaclet+0.176Barrio=El pla del real-0.008Barrio=L'olivereta-0.008Barrio=campanar - beni... <= -0.001383: SÍ (874.0/353.0)
-0.553Dorm=d7+0.429Planta=pss-0.338Dorm=d0+0.284Planta=p7+0.268Planta=p8... > -0.043858: SÍ (2180.0/1069.0)
-0.757Estado-0.341Planta=p1-0.307Dorm=d1-0.222Dorm=d0+0.166Dorm=d5... > -0.436024
0.773Barrio=sarrià-sant gervasi-0.305Barrio=sants-montjuïc-0.222Barrio=sant andreu-0.187Barrio=chamberí-0.165Barrio=retiro... <= 0.143972
0.676Barrio=retiro-0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... <= -0.16556: SÍ (12.0/1.0)
0.676Barrio=retiro-0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... > -0.16556
-0.665Barrio=fuencarral-0.558Barrio=hortaleza-0.259Barrio=moncloa+0.22 Barrio=sant martí+0.208Barrio=latina... <= 0.565388
0.838Barrio=ciudad lineal-0.368Barrio=chamartín-0.241Barrio=centro+0.199Barrio=salamanca-0.16Barrio=carabanchel... <= 0.203723
0.777Barrio=sant andreu-0.484Barrio=sants-montjuïc-0.154Barrio=retiro-0.136Barrio=Quatre carreres-0.134Barrio=horta guinardó... <= -0.01926
0.703Barrio=Patraix-0.689Barrio=Benimaclet+0.176Barrio=El pla del real-0.008Barrio=L'olivereta-0.008Barrio=campanar - beni... <= -0.001229: SÍ (5.0)
0.703Barrio=Patraix-0.689Barrio=Benimaclet+0.176Barrio=El pla del real-0.008Barrio=L'olivereta-0.008Barrio=campanar - beni... > -0.001229
0.339Barrio=horta guinardó+0.284Barrio=Camins al grau+0.241Barrio=Patraix+0.238Barrio=Benimaclet+0.237Tipo=pisos... <= -0.078566: SÍ (4.0)
0.339Barrio=horta guinardó+0.284Barrio=Camins al grau+0.241Barrio=Patraix+0.238Barrio=Benimaclet+0.237Tipo=pisos... > -0.078566
-0.581Barrio=usera-0.451Barrio=Quatre carreres+0.431Barrio=villa de vallecas+0.25 Barrio=ciutat vella-0.223Barrio=Camins al grau... <= 0.014961
-0.734Barrio=barajas+0.339Barrio=les corts+0.33 Barrio=Ciutat vella+0.286Barrio=nou barris-0.225Barrio=Camins al grau... <= -0.011871: NO (6.0)
-0.734Barrio=barajas+0.339Barrio=les corts+0.33 Barrio=Ciutat vella+0.286Barrio=nou barris-0.225Barrio=Camins al grau... > -0.011871
-0.67Dorm=d2+0.523Dorm=d3+0.338Planta=p1-0.179Estado-0.177Planta=p2... <= -0.054574: SÍ (7.0/1.0)
-0.67Dorm=d2+0.523Dorm=d3+0.338Planta=p1-0.179Estado-0.177Planta=p2... > -0.054574: NO (4.0)
-0.581Barrio=usera-0.451Barrio=Quatre carreres+0.431Barrio=villa de vallecas+0.25 Barrio=ciutat vella-0.223Barrio=Camins al grau... > 0.014961
-0.553Dorm=d7+0.429Planta=pss-0.338Dorm=d0+0.284Planta=p7+0.268Planta=p8... <= 0.023937
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= 0.537684: NO (36.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > 0.537684
-0.806Barrio=campanar - beni...-0.333Barrio=El pla del real-0.262Barrio=La saïdia+0.238Barrio=L'olivereta-0.179Barrio=vicálvaro... <= 0.007815: SÍ (2.0)
-0.806Barrio=campanar - beni...-0.333Barrio=El pla del real-0.262Barrio=La saïdia+0.238Barrio=L'olivereta-0.179Barrio=vicálvaro... > 0.007815: NO (3.0)
-0.553Dorm=d7+0.429Planta=pss-0.338Dorm=d0+0.284Planta=p7+0.268Planta=p8... > 0.023937
-0.462Tipo=àtics+0.357Metros+0.357Tipo=pisos-0.337Tipo=estudios+0.31 Tipo=chalets... <= 0.24112: NO (2.0)
-0.462Tipo=àtics+0.357Metros+0.357Tipo=pisos-0.337Tipo=estudios+0.31 Tipo=chalets... > 0.24112: SÍ (2.0)
0.777Barrio=sant andreu-0.484Barrio=sants-montjuïc-0.154Barrio=retiro-0.136Barrio=Quatre carreres-0.134Barrio=horta guinardó... > -0.01926: NO (103.0/6.0)
0.838Barrio=ciudad lineal-0.368Barrio=chamartín-0.241Barrio=centro+0.199Barrio=salamanca-0.16Barrio=carabanchel... > 0.203723
0.514Estado-0.512Planta=p1+0.443Dorm=d3-0.313Dorm=d1-0.228Dorm=d2... <= -0.640523
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.044771: SÍ (6.0)
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.044771: NO (2.0)
0.514Estado-0.512Planta=p1+0.443Dorm=d3-0.313Dorm=d1-0.228Dorm=d2... > -0.640523: NO (2.0/1.0)
-0.665Barrio=fuencarral-0.558Barrio=hortaleza-0.259Barrio=moncloa+0.22 Barrio=sant martí+0.208Barrio=latina... > 0.565388
-0.705Dorm=d4+0.325Dorm=d3+0.271Planta=p2+0.263Planta=pbj+0.248Dorm=d2... <= 0.190229
-0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... <= -1.045468: SÍ (4.0)

```

| | | | -0.522Ciudad=madrid-0.522Ciudad=madrid+0.393Ciudad=barcelona+0.393Ciudad=barcelona+0.269Ciudad=valencia... > -1.045468: NO (5.0/1.0)
| | | | -0.705Dorm=d4+0.325Dorm=d3+0.271Planta=p2+0.263Planta=pbj+0.248Dorm=d2... > 0.190229: Sí (5.0)
| 0.773Barrio=sarrià-sant gervasi-0.305Barrio=sants-montjuïc-0.222Barrio=sant andreu-0.187Barrio=chamberí-0.165Barrio=retiro... > 0.143972
| | 0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... <= -0.104739: NO (2.0)
| | 0.676Barrio=retiro+0.435Barrio=sarrià-sant gervasi+0.21 Barrio=sants-montjuïc-0.19Barrio=san blas-0.189Barrio=sant martí... > -0.104739: Sí (15.0)

```

Number of Leaves : 72

Size of the tree : 143

Time taken to build model: 80.45 seconds

=== Stratified cross-validation === (Prueba del modelo con el método evaluación cruzada exigiendo la estratificación de los segmentos)
 === Summary ===

Correctly Classified Instances	2630	57.386 %
Incorrectly Classified Instances	1953	42.614 %
Kappa statistic	0.1003	
Mean absolute error	0.4746	
Root mean squared error	0.488	
Relative absolute error	95.3745 %	
Root relative squared error	97.8359 %	
Total Number of Instances	4583	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.165	0.069	0.674	0.165	0.265	0.564	NO
	0.931	0.835	0.561	0.931	0.7	0.564	Sí
Weighted Avg.	0.574	0.478	0.614	0.574	0.497	0.564	

=== Confusion Matrix ===

```

a b <- classified as
352 1783 | a = NO
170 2278 | b = Sí

```

Apéndice 2: La construcción del modelo 2

En este apéndice mostramos de forma completa la salida ofrecida por “WEKA” de la ejecución del modelo 2 con el algoritmo seleccionado “M5P” de la familia “tree”. Se apreciará en la construcción las dos regresiones lineales ‘LM1’ y ‘LM2’ y la parte de la construcción destinada, a la selección previa de uno de los dos modelos lineales, ‘LM1’ y ‘LM2’ teniendo en cuenta el atributo “metros”. También se observa la información de ejecución “Run information” y la de la prueba. Entre paréntesis pondremos algunas notas de aclaración.

Modelo2: Precio por el que vender una vivienda independientemente del número de días. Con el algoritmo seleccionado “M5P”:

=== Run information === (Información de ejecución.)

Scheme: "WEKA".classifiers.trees.M5P -M 4.0
Relation: VSNDWK26ene2010-12feb2010.txt-"WEKA".filters.unsupervised.attribute.Remove-R1-
"WEKA".filters.unsupervised.instance.RemoveWithValues-S0.0-C9-Lfirst-"WEKA".filters.unsupervised.attribute.Remove-R10
Instances: 2448
Attributes: 9
Tipo
Barrio
Dorm
Planta
Metros
Euros
Estado
Ciudad
Compra
Test mode: 10-fold cross-validation

=== Classifier model (full training set) === (Aprendizaje o Entrenamiento o construcción del modelo.)

M5 pruned model tree:
(using smoothed linear models) (Selección del modelo lineal 1 o del modelo lineal 2 dependiendo del valor del atributo "metros".)

Metros <= 122.5 : LM1 (1900/14.414%)
Metros > 122.5 : LM2 (548/82.127%)

LM num: 1 (Modelo lineal 1)

Euros =

23501.3676 * Tipo=áticos,dúplex,chalets
- 381.2652 * Tipo=dúplex,chalets
- 1573.2007 * Tipo=chalets
+ 19270.694 * Barrio=benicalap,latina,La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 450.6366 * Barrio=nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 37823.5041 * Barrio=villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 37797.9436 * Barrio=Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 192.3836 * Barrio=carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 22824.6611 * Barrio=moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 25079.2418 * Barrio=horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 17602.9658 * Barrio=sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 13421.874 * Barrio=Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 36543.1412 * Barrio=vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 61052.4143 * Barrio=ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 19072.5601 * Barrio=sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 694.2626 * Barrio=campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 1039.2879 * Barrio=tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 216.7501 * Barrio=Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 32115.2884 * Barrio=gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 63124.2232 * Barrio=san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 28981.4835 * Barrio=barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 452.6147 * Barrio=ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 43271.1579 * Barrio=Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
- 35427.188 * Barrio=L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
+ 51953.4764 * Barrio=centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
+ 11922.0161 * Barrio=El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
+ 705.1037 * Barrio=les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 19008.2381 * Barrio=eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 34677.7022 * Barrio=fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 63540.3739 * Barrio=salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 23096.4833 * Barrio=chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 62510.1766 * Barrio=sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 131663.0512 * Barrio=hortaleza,chamartín,moncloa
+ 52639.3656 * Barrio=chamartín,moncloa
- 60006.1702 * Barrio=moncloa
- 14509.5593 * Dorm=d2,d3,d4,d5,d9,d6,d7,d8
- 33871.7504 * Dorm=d3,d4,d5,d9,d6,d7,d8
- 21790.0035 * Dorm=d4,d5,d9,d6,d7,d8
- 2164.3337 * Dorm=d9,d6,d7,d8
+ 1052.6221 * Dorm=d6,d7,d8
+ 358.1139 * Planta=p14,pbj,p15,p2,p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
+ 12906.7895 * Planta=p15,p2,p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
+ 10635.365 * Planta=p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
+ 65524.7747 * Planta=p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
- 75613.7002 * Planta=p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
- 468.7697 * Planta=p10,p7,p1,p11,p9,p8,p22
+ 201.1062 * Planta=p7,p1,p11,p9,p8,p22
+ 242.4894 * Planta=p1,p11,p9,p8,p22
+ 31717.1932 * Planta=p11,p9,p8,p22
+ 4044.4422 * Metros
- 58917.4469 * Estado=a reformar
+ 100105.6718 * Ciudad=barcelona,madrid
- 23843.2991 * Ciudad=madrid
- 118415.6469

LM num: 2 (Modelo lineal 2)

Euros =

-1296.8435 * Tipo=dúplex,chalets
- 284873.0674 * Tipo=chalets
- 1532.8049 * Barrio=nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant
andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni....tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad
lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
+ 946.526 * Barrio=Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant
andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni....tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad
lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
+ 654.3775 * Barrio=carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat
vella,sant martí,campanar - beni....tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 626.9302 * Barrio=moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant
martí,campanar - beni....tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 238263.6247 * Barrio=Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar -
beni....tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 641.7915 * Barrio=horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar -
beni....tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 203169.1733 * Barrio=sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar -
beni....tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 2361.4793 * Barrio=campanar - beni....tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El
pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 3535.0556 * Barrio=tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 737.2582 * Barrio=Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 416.4649 * Barrio=san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 437.9036 * Barrio=barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 211718.7429 * Barrio=ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les
corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

```

+ 668.8305 * Barrio=Ciutat vella,L´eixample,centro,El pla del real,les cort,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
- 1092.866 * Barrio=L´eixample,centro,El pla del real,les cort,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
+ 2172.3687 * Barrio=centro,El pla del real,les cort,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
- 111471.3392 * Barrio=El pla del real,les cort,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant
gervasi,hortaleza,chamartín,moncloa
+ 2398.3546 * Barrio=les cort,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 104524.5781 * Barrio=retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 253595.6457 * Barrio=eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 1199.9819 * Barrio=fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 475923.7182 * Barrio=salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 525.3 * Barrio=chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 317343.8647 * Barrio=sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 2361.5535 * Barrio=hortaleza,chamartín,moncloa
+ 361729.5529 * Barrio=chamartín,moncloa
- 331071.4556 * Barrio=moncloa
- 1016.5507 * Dorm=d2,d3,d4,d5,d9,d6,d7,d8
- 1815.5223 * Dorm=d3,d4,d5,d9,d6,d7,d8
- 1058.9009 * Dorm=d4,d5,d9,d6,d7,d8
- 217687.6523 * Dorm=d9,d6,d7,d8
+ 3580.411 * Dorm=d6,d7,d8
+ 1218.0962 * Planta=p14,pbj,p15,p2,p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
+ 208680.8786 * Planta=p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
- 151405.4392 * Planta=p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22
- 158545.5569 * Planta=p10,p7,p1,p11,p9,p8,p22
+ 112799.1252 * Planta=p7,p1,p11,p9,p8,p22
+ 824.8085 * Planta=p1,p11,p9,p8,p22
+ 340369.2747 * Planta=p11,p9,p8,p22
+ 6466.5938 * Metros
- 493329.8192 * Estado=a reformar
+ 405783.282 * Ciudad=barcelona,madrid
- 229485.81 * Ciudad=madrid
- 206024.7681

```

Number of Rules : 2

Time taken to build model: 16.38 seconds

=== Cross-validation === (Prueba del modelo, con el método evaluación cruzada.)
=== Summary ===

Correlation coefficient	0.8995
Mean absolute error	105980.2765
Root mean squared error	227315.5084
Relative absolute error	36.9621 %
Root relative squared error	43.6766 %
Total Number of Instances	2448

Apéndice 3: La construcción del modelo 3

En este apéndice se observa íntegramente la salida ofrecida por “WEKA” de la ejecución del modelo 3 con el algoritmo seleccionado “M5P” de la familia “tree”. Se aprecia en la construcción del modelo las dos regresiones lineales ‘LM1’ y ‘LM2’ y la construcción destinada a la selección previa de uno de los dos modelos lineales ‘LM1’ y ‘LM2’ teniendo en cuenta el atributo “metros”. También se observa la información de ejecución “Run information” y la de la prueba del modelo. Entre paréntesis pondremos algunas aclaraciones.

Modelo3: Precio por el que vender una vivienda teniendo en cuenta el número de días. Con el algoritmo seleccionado “M5P”:

=== Run information === (Información de ejecución.)

Scheme: "WEKA".classifiers.trees.M5P -M 4.0
Relation: VSNDWK26ene2010-12feb2010.txt-"WEKA".filters.unsupervised.attribute.Remove-R1-
"WEKA".filters.unsupervised.instance.RemoveWithValues-S0.0-C9-Lfirst
Instances: 2448
Attributes: 10
 Tipo
 Barrio
 Dorm
 Planta
 Metros
 Euros
 Estado
 Ciudad
 Compra
 Dias
Test mode: 10-fold cross-validation

=== Classifier model (full training set) === (Aprendizaje o Entrenamiento o construcción del modelo.)

M5 pruned model tree:
(using smoothed linear models) (Selección del modelo lineal 1 o del modelo lineal 2 dependiendo del valor del atributo "metros".)

Metros <= 122.5 : LM1 (1900/14.331%)
Metros > 122.5 : LM2 (548/80.869%)

LM num: 1 (Modelo lineal 1)

Euros =
23433.8426 * Tipo=áticos,dúplex,chalets
- 381.3188 * Tipo=dúplex,chalets
- 1589.4741 * Tipo=chalets
- 934.2791 * Barrio=puerto de vallecas,usera,benicalap,latina,La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 22748.5688 * Barrio=benicalap,latina,La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 828.4423 * Barrio=latina,La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 720.0278 * Barrio=La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 1114.4428 * Barrio=nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 42529.9172 * Barrio=villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 56292.2351 * Barrio=Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
+ 15748.1914 * Barrio=Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa
- 54404.4751 * Barrio=villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 23486.3585 * Barrio=carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 57580.6503 * Barrio=Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 10466.6316 * Barrio=horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 15822.4329 * Barrio=sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 16338.8473 * Barrio=Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 691.9253 * Barrio=sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 41780.6615 * Barrio=vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 61046.2963 * Barrio=ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 17427.289 * Barrio=sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 179.9767 * Barrio=campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 324.94 * Barrio=tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 33166.4409 * Barrio=gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 65349.6445 * Barrio=san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 35089.1947 * Barrio=barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 341.1304 * Barrio=ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 1082.3627 * Barrio=Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 327.0475 * Barrio=L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 56722.3697 * Barrio=centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 11883.6986 * Barrio=El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 183.0555 * Barrio=retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 19327.9493 * Barrio=eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 33604.6816 * Barrio=fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 63576.7345 * Barrio=salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 24507.1815 * Barrio=chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 59893.3437 * Barrio=sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 131949.4249 * Barrio=hortaleza,chamartín,moncloa

+ 52583.8248 * Barrio=chamartín,moncloa

- 57794.245 * Barrio=moncloa

- 14509.0137 * Dorm=d2,d3,d4,d5,d9,d6,d7,d8

- 32969.2996 * Dorm=d3,d4,d5,d9,d6,d7,d8

- 21341.2294 * Dorm=d4,d5,d9,d6,d7,d8

- 1118.4153 * Dorm=d9,d6,d7,d8

+ 22158.1956 * Planta=p14,pbj,p15,p2,p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22

+ 13959.475 * Planta=p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22

+ 67061.3858 * Planta=p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22

- 77710.9638 * Planta=p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22

- 305.6359 * Planta=p10,p7,p1,p11,p9,p8,p22

+ 269.8974 * Planta=p1,p11,p9,p8,p22

+ 31843.4659 * Planta=p11,p9,p8,p22

+ 4050.2235 * Metros

- 58567.4409 * Estado=a reformar

+ 103042.4206 * Ciudad=barcelona,madrid

- 25256.7795 * Ciudad=madrid

+ 250.493 * Dias

- 145115.809

LM num: 2 (Modelo lineal 2)

Euros =

-1297.0258 * Tipo=dúplex,chalets

- 319551.7931 * Tipo=chalets

- 3177.8764 * Barrio=pueblo de vallecas,usara,benicalap,latina,La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 3851.1751 * Barrio=benicalap,latina,La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...,tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 2817.881 * Barrio=latina,La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 2449.1175 * Barrio=La saïdia,nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 3790.6891 * Barrio=nou barris,villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 996.1033 * Barrio=villaverde,Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 3149.3786 * Barrio=Algirós,Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 2941.255 * Barrio=Benimaclet,villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 5572.6734 * Barrio=villa de vallecas,carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 1220.5341 * Barrio=carabanchel,moratalaz,Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 4080.9599 * Barrio=Camins al grau,horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 3616.2077 * Barrio=horta guinardó,sant andreu,Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 2442.4039 * Barrio=Extramurs,sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 2353.5292 * Barrio=sants-montjuïc,vicálvaro,ciutat vella,sant martí,campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 612.1764 * Barrio=campanar - beni...tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 1105.2578 * Barrio=tetuán,Patraix,arganzuela,gràcia,san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 1146.8195 * Barrio=san blas,barajas,ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 204966.0497 * Barrio=ciudad lineal,Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 3681.571 * Barrio=Ciutat vella,L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 1112.4261 * Barrio=L'eixample,centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 551.4899 * Barrio=centro,El pla del real,les corts,retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 622.6487 * Barrio=retiro,eixample,fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 265166.3503 * Barrio=eixample, fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 1566.8862 * Barrio=fuencarral,salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

+ 484009.43 * Barrio=salamanca,chamberí,sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 301830.3483 * Barrio=sarrià-sant gervasi,hortaleza,chamartín,moncloa

- 2610.5112 * Barrio=hortaleza,chamartín,moncloa

+ 335145.6539 * Barrio=chamartín,moncloa

- 305018.757 * Barrio=moncloa

- 1027.1053 * Dorm=d2,d3,d4,d5,d9,d6,d7,d8

- 1769.7455 * Dorm=d3,d4,d5,d9,d6,d7,d8

- 1027.1157 * Dorm=d4,d5,d9,d6,d7,d8

- 201729.093 * Dorm=d9,d6,d7,d8

+ 1350.2445 * Planta=p14,pbj,p15,p2,p4,p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22

+ 235702.6474 * Planta=p13,p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22

- 196198.9842 * Planta=p3,p12,p5,p6,p10,p7,p1,p11,p9,p8,p22

- 1039.5964 * Planta=p10,p7,p1,p11,p9,p8,p22

+ 918.0346 * Planta=p1,p11,p9,p8,p22

+ 322562.5462 * Planta=p11,p9,p8,p22

+ 6434.3018 * Metros

- 502327.794 * Estado=a reformar

+ 398749.6773 * Ciudad=barcelona,madrid

- 214623.2172 * Ciudad=madrid

+ 2417.5412 * Dias

- 329698.4688

Number of Rules : 2

Time taken to build model: 21.55 seconds

=== Cross-validation === (Prueba del modelo, con el método evaluación cruzada.)
=== Summary ===

Correlation coefficient	0.8996
Mean absolute error	104808.0026
Root mean squared error	227188.9084
Relative absolute error	36.5532 %
Root relative squared error	43.6523 %
Total Number of Instances	2448

Apéndice 4: Programas propios

En este apéndice vamos a mencionar unos programas específicos que se han desarrollado para facilitar la fase de preparación. Dichos programas son los siguientes: “extraer”, “comparar”, “pasadosDias”, “retocarParaWEKA” y “eliminarRepeticiones”. Esta serie de programas han sido realizados en el lenguaje ‘Java’, concretamente, en ‘Java Standar Edition 6’ y en el entorno ‘BlueJ’.

Seguidamente vamos a comentar dichos programas específicos:

- El programa “extraer” se ha creado para extraer la información de las viviendas del portal web y almacenarla en un fichero del tipo “viviendasdd-mmm-yyyy.txt”. (Este tipo de ficheros “viviendasdd-mmm-yyyy.txt” los comentamos en el apartado “2. Obtención de los datos”.)

Como en el nombre de un fichero “viviendasdd-mmm-yyyy.txt” la parte “dd-mmm-yyyy” será la fecha en la que se realice la extracción de los datos de las viviendas, este programa sirve para realizar la tarea que trata de extraer los datos de las viviendas de la “fecha primera” o la tarea que trata de extraer los datos de las viviendas de la “fecha segunda” que se relata también en el apartado “2. Obtención de los datos”. El formato en el que se almacenan las instancias de las viviendas es el que se especifica en el apartado, “2. Obtención de los datos”, sobre los ficheros del tipo “viviendasdd-mmm-yyyy.txt”.

Respecto a las instancias de las viviendas que se almacenan, como también se dijo, son las instancias de las viviendas de las ciudades de Valencia, Madrid y Barcelona, ya que, las viviendas de otras ciudades no nos interesan en este proyecto, aunque

fácilmente se podría modificar si las necesidades respecto a las ciudades variaran y se necesitaran otras ciudades.

Primero se almacenan las instancias de las viviendas de Valencia, luego las de Madrid y por último las de Barcelona. El portal web inmobiliario tiene la información de las viviendas distribuidas en ciudades y cada ciudad tiene un número de páginas en las cuales aparecen sus instancias de viviendas. El programa “extraer” lo que realiza es acceder a las páginas de las instancias de las viviendas de la ciudad de Valencia, Madrid y Barcelona. Recorre las páginas de cada una de estas ciudades, y almacena la información de las instancias de las viviendas que nos interesa y que se encuentra. Para lo cual, en primer lugar se escoge la referencia de la primera ciudad, que es la ciudad de Valencia y se va a la primera página de las instancias de las viviendas de dicha “ciudad”. Se almacenan las instancias de las viviendas de la primera página en el fichero de salida y se obtiene la referencia a la página segunda. En la página segunda se realiza lo mismo que se ha hecho en la página primera, es decir, se almacenan las instancias de las viviendas de la página y se obtiene la referencia de la siguiente página que sería la tercera. Esto se realiza de la misma manera, para cada una de las páginas pertenecientes a las instancias de las viviendas de la ciudad de Valencia, hasta llegar a la última página. Cuando esto pasa se almacenan las instancias de las viviendas de dicha última página y luego repetimos el mismo proceso para las ciudades de Madrid y Barcelona.

Por último del programa “extraer”, decir, que como obtiene los datos del portal web está diseñado para recopilar las instancias de las viviendas según la estructura y organización del portal web. Si el portal web variara sería una complicación, como se dijo en la introducción, si ello afecta a la estructura y organización de las instancias de las viviendas, ya que, el programa podría dejar de obtener los datos. No obstante, tal programa tiene la cualidad de informar si no puede recoger los valores de las instancias de las viviendas e indicar que atributos fallan en la extracción. Además dispone de una opción de introducción de otros patrones para buscar los valores de los atributos que no puede extraer, aunque eso sí, habría que observarlos en el portal web para saber cuales serían dichos patrones de búsqueda. Otra opción alternativa sería ir al código fuente y poner los nuevos patrones directamente. Esta última es la opción más recomendable sobretodo si hay cambios muy drásticos que no se pudieran llevar a cabo sólo cambiando patrones de búsqueda.

- Por otro lado, el programa “comparar” sirve para realizar la tarea de la preparación que trata de comparar las instancias de las viviendas extraídas en la “fecha primera” y en la “fecha segunda”. Esto se realiza para estimar el atributo “compra” a las instancias de las viviendas extraídas en la “fecha primera”. Dicha estimación del atributo “compra” se explica también en el apartado “2. Obtención de los datos”.

El programa “comparar” envía las instancias de las viviendas extraídas en la “fecha primera” a las que ha podido estimar el atributo “compra” al fichero

“VSINOfecha1-fecha2.txt”, con las peculiaridades que se explican también en el apartado “2. Obtención de los datos”.

El programa “comparar” envía el resto de las instancias de viviendas a las que no les ha podido estimar el atributo “compra” a otro fichero de salida “VRestofecha1-fecha2.txt” con las peculiaridades que se indican también en el apartado “2.Obtención de los datos”.

Para que el programa “comparar” realice lo comentado: necesita dos ficheros de entrada del tipo “Viviendasdd-mmm-yyy.txt”. El de las viviendas extraídas en la “fecha primera” que se utilizará como “fichero base”. Y el de las viviendas extraídas en la “fecha segunda” que debe tener una fecha de entre ‘15’ y ‘20’ días mayor.

- El programa “pasadosDias” calcula el atributo “días” de cada instancia de vivienda a la que se le ha estimado el atributo “compra”. Recibe un fichero del tipo “VSINOFecha1-Fecha2.txt” y crea un fichero de salida del tipo “VSNDiasFecha1-Fecha2.txt”.
- El programa “retocarParaWEKA” realiza las transformaciones de formato y las eliminaciones de atributos sobre las instancias de las viviendas, de la forma que se relataba en la tarea “5ª” en el apartado “2. Obtención de los datos”. Recibe como entrada un fichero del programa anterior “pasadosDias” del tipo “VSNDiasFecha1-Fecha2.txt” y crea como salida un fichero con formato “csv” del tipo “VSNDWKFecha1-Fecha2.csv”.
- En cuanto al último programa específico denominado “eliminarRepeticiones”, elimina las posibles repeticiones de instancias de viviendas si es que encuentra alguna. Recibe como entrada un fichero del programa “retocarParaWEKA” del tipo “VSNDWKFecha1-Fecha2.csv” y crea como salida un fichero del tipo “VSNDWKUFecha1-Fecha2.csv”.

Los ficheros de salida “VSNDWKUfecha1-fecha2.csv” creados por el programa específico “eliminarRepeticiones” se importan a la herramienta o aplicación de minería de datos ‘Explorer’ de “WEKA”. Dicha aplicación “WEKA” creada por la universidad de Waikato es la que se utiliza para continuar con la preparación de los datos. Además la aplicación ‘Explorer’ de “WEKA” también es la que se utiliza como se ha dicho varias veces para el aprendizaje y la prueba de los modelos una vez preparados los datos.

El diagrama siguiente muestra de forma gráfica la forma en que se enlazan los programas específicos y la herramienta “WEKA”, para alcanzar la vista minable particular de cada uno de los tres modelos, que es lo que se persigue en la fase de preparación.

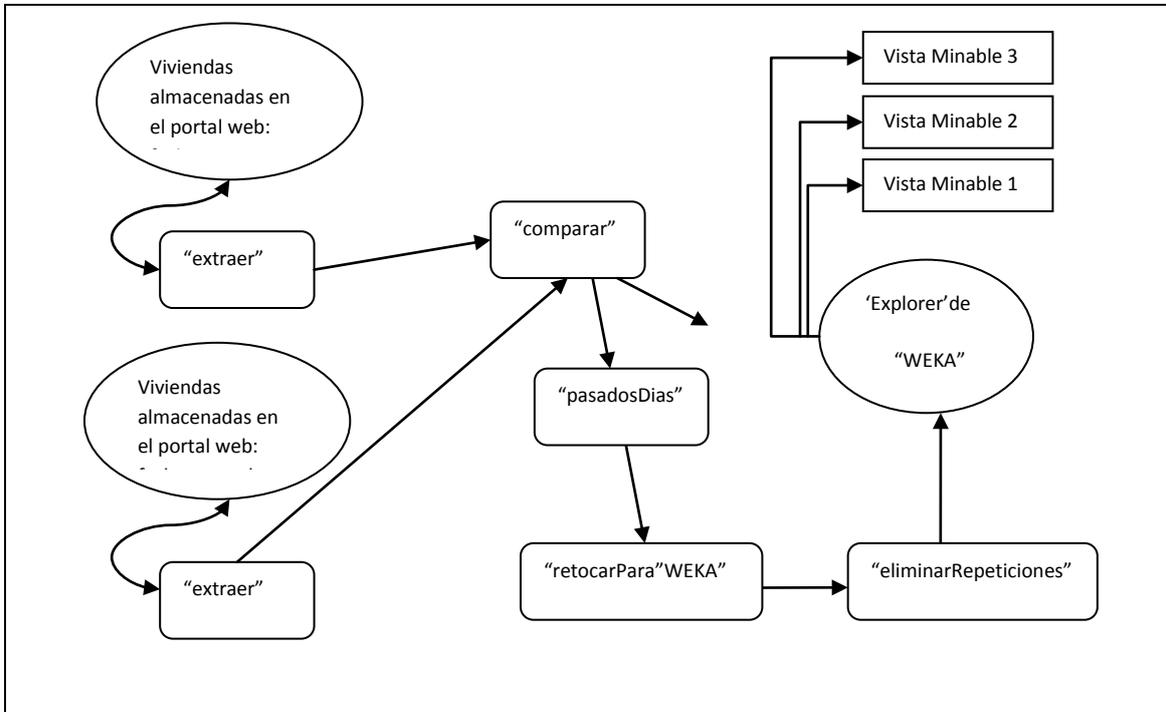


Diagrama Apéndice.1: Se observa cómo se utilizan y se enlazan los programas específicos en la fase de preparación. Y como el último de estos programas propios del proyecto se enlaza con la herramienta ‘Explorer’ de “WEKA” para terminar la preparación. Los enlaces entre programas representan que el fichero de salida o uno de los dos ficheros de salida en el caso del programa “comparar” se utilizan como entrada en el programa inmediatamente posterior. Ídem en referencia al enlace entre el programa propio “eliminarRepeticiones” y la herramienta ‘Explorer’ de “WEKA”. Por otra parte, las flechas bidireccionales que surgen del programa “extraer” y del portal web inmobiliario vienen a reflejar que el programa “extraer” solicita instancias de viviendas del portal web y éste se las envía. En este caso el enlace bidireccional viene a indicar un flujo de datos en la ejecución del programa “extraer”, entre dicho programa y el portal web.

Como se puede apreciar en el diagrama Apéndice.1, en las flechas unidireccionales se omite el nombre de los ficheros para que el diagrama presente un aspecto más simple y claro. En el diagrama se puede observar que aparecen dos instancias del programa “extraer”, esto se debe a que como se sabe se necesitan dos ejecuciones del programa “extraer” una en la “fecha primera” de extracción y otra en la “fecha segunda” de extracción que originarán dos ficheros “Viviendasdd-mmm-yyyy.txt” que se pasan al programa “comparar”.

El programa “comparar” crea dos ficheros de salida: uno con las instancias de las viviendas que se les ha podido estimar el atributo “compra” de las extraídas en la “fecha primera”. Y otro con el resto de las instancias de las viviendas. Este último fichero que

posee el resto de las instancias de salida se desecha, debido a eso en el diagrama Apéndice.1 aparece una flecha sin destino que surge del programa “comparar”.

El fichero seleccionado de las instancias de las viviendas con el atributo “compra” se pasa al programa “retocarParaWEKA”.

El fichero de salida de “retocarParaWEKA” se utiliza como entrada para el programa “eliminarRepeticiones”.

Por otro lado, el fichero de salida del programa “eliminarRepeticiones” se utiliza en la herramienta “WEKA” en la que se sigue la preparación como dijimos en el apartado “2. Preparación de los datos” hasta alcanzar la vista minable particular de cada uno de los tres modelos.

Después, aunque ya no forma parte del diagrama se utiliza el programa “WEKA” como hemos visto en el apartado “3. Aprendizaje de datos”, para la construcción de los modelos.

Apéndice 5: Organización de la carpeta del proyecto

Vamos a hablar de cómo está organizada la carpeta del proyecto en este apéndice. La carpeta de cualquier proyecto por norma se debe llamar con el nombre del alumno que realiza el proyecto. Además también por norma debe tener dentro de tal carpeta el documento de la memoria en formato “.pdf” en este caso, esta memoria que se está leyendo. Y opcionalmente una subcarpeta denominada “anexos” con el resto de información del proyecto. En este proyecto en la carpeta llamada “anexos” encontraremos dos subcarpetas llamadas: “preparación” y “modelos”.

- En la carpeta “preparación” tendremos una carpeta llamada “programas”. En dicha carpeta aparecerán los programas creados específicamente para la fase de preparación del proyecto, dichos programas los explicamos en el apéndice anterior: “Apéndice 3: Programas propios”. Y como dijimos en dicho apéndice se denominan: “extraer”, “comparar”, “pasadosDias”, “retocarParaWEKA” y “eliminarRepeticiones”. Además aparecerá la carpeta llamada “ficheros” en donde aparecerán todos los ficheros creados en la preparación a excepción del fichero que contiene la vista minable del modelo 1, del que contiene la vista minable del modelo 2 y del que contiene la vista minable del modelo 3.
- Respecto a la carpeta modelos aparecen tres carpetas contenidas para cada uno de los modelos llamadas: “modelo1”, “modelo2” y “modelo3”. Cada una de estas tres carpetas tendrá una carpeta llamada “vistaMinable” y otra denominada

“estudioDelModelo”. En la carpeta “modelo3” tendremos además de las dos carpetas comentadas otra carpeta denominada “prediccionReplicas”.

- La carpeta llamada “vistaMinable” de cada modelo contendrá la vista minable del modelo.
- Respecto a la carpeta “estudioDelModelo” de cada modelo contendrá dos carpetas, una denominada “seleccionado” y otra denominada “descartados”. En la carpeta “seleccionado” de cada modelo aparecerá una carpeta con el nombre de la familia del algoritmo seleccionado que contendrá un fichero de texto con la salida del modelo seleccionado. La carpeta “descartados” de cada modelo poseerá un conjunto de subcarpetas con el nombre de las familias de los algoritmos que han construido los modelos que han sido descartados. Cada una de estas carpetas de familias de los algoritmos que han sido descartadas contendrá, precisamente, las construcciones descartadas de los algoritmos de cada familia.
- Por último la carpeta exclusiva del modelo 3 “prediccionReplicas” contendrá el fichero de las réplicas de una vivienda variando únicamente el atributo “días” en ellas y con el atributo “euros” desconocido, la salida del modelo 3 seleccionado que se ejecutó esta vez para predecir las réplicas y el fichero de la predicción del atributo “euros” a cada réplica.