



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



ESCUELA TÉCNICA  
SUPERIOR INGENIEROS  
INDUSTRIALES VALENCIA

Curso Académico:

## **AGRADECIMIENTOS**

En primer lugar, me gustaría dar las gracias al Instituto Tecnológico de la Energía por ofrecerme la oportunidad de desarrollar mis prácticas de empresa, y en especial a Alfredo Quijano, tutor y persona responsable de que esté aquí.

Dentro de la empresa, me gustaría mencionar todos mis compañeros, que me han brindado un clima de comodidad, tranquilidad y buen compañerismo, y sobre todo a Julio Díaz por ayudarme a gestionar todo lo que haya podido necesitar en cualquier momento, a Javier Monreal Tolmo por todo el aprendizaje y la ayuda que me ha ofrecido, y a Julián Romero, por todas las dudas y problemas que me ha resuelto.

Por último, también me gustaría dar las gracias a mi familia y amigos por el apoyo moral y la ayuda que me han podido ofrecer todo este tiempo.

## **RESUMEN**

En la actualidad se está dedicando un gran esfuerzo para que los consumidores finales participen en la gestión de la red. Esto conlleva un beneficio conjunto, tanto para el sistema eléctrico, dado que contribuye a mantener sus requisitos de fiabilidad y estabilidad, como para el usuario, que, es capaz de realizar correcciones o ajustes en su demanda o contrato, al disponer de más información, para reducir su gasto económico.

Con el presente proyecto se pretende favorecer la integración de usuarios residenciales en dicha gestión a través de herramientas que automaticen el balance energético en una vivienda de forma óptima, ajustándose a una producción energética local (en caso de estar disponible), a la gestión de una red inteligente, o a diferentes tarifas eléctricas.

El desarrollo de dichas herramientas se suele basar en modelos capaces de predecir la evolución de las variables que participan en el balance. En el caso de los usuarios residenciales, la solución propuesta por la herramienta depende en gran medida de la capacidad que tenga de realizar una buena predicción de la demanda la, cual se caracteriza por ser muy cambiante y difícil de prever, debido a los diferentes factores que intervienen y que afectan al consumo.

Siguiendo esta línea de trabajo, en el presente proyecto se plantea el estudio, desarrollo y optimización de un algoritmo que, en base a unos datos históricos, sea capaz de predecir los estados de consumo de aquellos equipos de mayor demanda (electrodomésticos), cuyo gasto de electricidad corresponde al 55% del hogar, con el fin de detectar aquellas variables y patrones de que más influyen y determinan mejores resultados,

Para ello, se ha determinado unos escenarios y se han evaluado 3 modelos de predicción, analizando su comportamiento y efectividad en función de distintas variables. De esta manera se ha obtenido el óptimo, y se ha aplicado a distintos casos de consumo.

### **Palabras Clave:**

Datos históricos, Consumo energético, Algoritmo, Predicción.

## **RESUM**

En l'actualitat s'està dedicant un gran esforç a què els consumidors finals participen en la gestió de la xarxa. Açò comporta un benefici mutu, tant per al sistema elèctric, ja que contribueix a mantindre els seus requisits de fiabilitat i estabilitat, com per al usuari, que al disposar de més informació, és capaç de realitzar correccions o ajusts en la seua demanda o el seu contracte per a reduir el seu gasto econòmic.

Amb el present projecte es pretén afavorir la integració d'usuaris residencials en dita gestió a través de ferramentes que automatitzen el balanç energètic en un habitatge de forma òptima, ajustant-se a una producció energètica local, si la casa disposara, a la gestió d'una xarxa intel·ligent, o a diferents tarifes elèctriques.

El desenvolupament de dites ferramentes es sol basar en models capaços de predir l'evolució de les variables que participen en el balanç. En el cas dels usuaris residencial, la solució proposada per la ferramenta depèn en gran manera de la capacitat que tinga per a realitzar una bona predicció de la demanda, que es caracteritza per ser molt canviant i difícil de preveure, a causa dels diferents factors que intervenen i que afecten al consum.

Seguint esta línia de treball, en el present projecte es planteja l'estudi, desenvolupament i optimització d'un algoritme que, basant-se en unes dades històriques, siga capaç de predir els estats de consum d'aquells equips de major demanda (electrodomèstics) , el consum d'electricitat del qual correspon al 55% de la llar , a fi de detectar aquelles variables i patrons que més influeixen i determinen millors resultats.

Per a això, s'han determinat uns escenaris i s'han evaluat 3 models de predicció, analitzant el seu comportament i la seua efectivitat en funció de diferents variables. D'aquesta manera, s'ha obtingut el més òptim, i s'ha aplicar en distints casos de consum.

### **Paraules clau:**

Dades històriques, Consum energètic, Algoritme, Predicció.

## **ABSTRACT**

A major effort is currently being made to involve end consumers in the management of the network. This brings a joint benefit, both for the electrical system, as it contributes to maintaining its requirements for reliability and stability, as well as for the user, who, by having more information, is able to make corrections or adjustments to your claim or contract to reduce your financial expense.

The aim of this project is to promote the integration of residential users in such management through tools that automate the energy balance in a house in an optimal way, adjusting to a local energy production, if the dwelling had, to the management of a smart network, or at different electricity rates.

The development of such tools is usually based on models able to predict the evolution of the variables involved in the balance sheet. In the case of residential users, the solution proposed by the tool depends to a large extent on the ability to make a good prediction of demand, which is characterized by being very changeable and difficult to predict, because of the different factors involved and affecting consumption.

Following this line of work, this project proposes the study, development and optimization of an algorithm that, based on historical data, is able to predict the states of consumption of those most in demand (household appliances), for which 55% of the electricity is spent in them, in order to detect those variables and patterns that most influence and determine better results,

To this end, different scenarios have been determined and 3 models of predictions have been evaluated, analysing its behaviour and effectivity in function of different variables. By this way, the most optimum one has been obtained, and it has been applied in different demand cases.

### **Keywords:**

Historical data, Energy consumption, Algorithm, Prediction.

## ÍNDICE MEMORIA

CAPÍTULO 1. INTRODUCCIÓN.....	9
1.1. OBJETIVO DEL DOCUMENTO.....	9
1.2. ESTRUCTURA DEL DOCUMENTO .....	11
CAPÍTULO 2 . PREDICCIÓN DE CONSUMOS.....	13
2.1 APROXIMACIÓN AL PROBLEMA .....	13
2.2 HARDWARE NECESARIO .....	14
2.3 NORMATIVA .....	14
2.4 CONTEXTO.....	14
CAPÍTULO 3. ESTADO DEL ARTE .....	16
3.1 ALGORITMO DE HISTOGRAMA.....	16
3.2 ALGORITMO DE PREDICCIÓN BAYESIANO.....	17
3.3 ARMA (media móvil auto regresiva) .....	17
3.4 MODELO PROBABILÍSTICO .....	18
CAPÍTULO 4. SITUACIÓN INICIAL .....	19
4.1 DATOS DE PARTIDA .....	19
4.2 PREPROCESADO .....	19
4.3 CLASIFICACIÓN .....	20
4.3.1. K-medias.....	20
4.4 RESULTADOS .....	21
CAPÍTULO 5. MÉTODO PROPUESTO .....	24
5.1 GENERAR CONOCIMIENTO: ORACLE.....	24
5.2 PREDICTOR .....	26
5.2.1 Regresión Logística Multinomial .....	26
5.2.2 Cadenas ocultas de Márkov .....	27
5.2.3 Árbol de decisiones .....	28
5.3 MÉTODO DE EVALUACIÓN .....	29
CAPÍTULO 6. SIMULACIONES Y RESULTADOS.....	31
6.1 ESCENARIOS CÁLCULO Y DETERMINACIÓN VARIABLES .....	31

---

6.2 ESCENARIOS DETERMINACIÓN .....	33
6.2.1 Escenario 1 .....	33
6.2.2 Escenario 2. ....	33
6.2.3 Escenario 3. ....	34
6.3 RESULTADOS I .....	35
6.3.1 Regresión Logística.....	35
6.2.2 Márkov .....	36
6.2.3 Árbol de decisiones. ....	37
6.3 RESULTADOS II .....	38
6.3.1 Regresión Logística.....	38
6.3.2 Márkov .....	40
6.3.3 Árbol de decisiones. ....	41
7. CONCLUSIONES .....	42
7.1 CONCLUSIONES RESULTADOS .....	42
7.2 OTROS ELECTRODOMÉSTICOS .....	43
7.3 LÍNEAS FUTURAS .....	44
ANEJO 1 .....	49
BIBLIOGRAFÍA.....	71

## **ÍNDICE PRESUPUESTO**

1 NECESIDAD DEL PRESUPUESTO.....	46
2 PRESENTACIÓN DEL PRESUPUESTO.....	46
3 CONTENIDO.....	
3.1 Coste desarrollo.....	46
3.2 Coste adquisición datos.....	46
3.3 Coste mantenimiento.....	47

---

**Desarrollo, optimización e  
implementación de un algoritmo de  
predicción de consumo de energía de  
una vivienda**

---

**MEMORIA**

**GRADO EN INGENIERÍA DE LA ENERGÍA**

Razvan Valentin Stoica





# CAPÍTULO 1. INTRODUCCIÓN

## **1.1. Objetivo del documento**

La electricidad es un bien de consumo indispensable, cuyos primeros estudios datan de los siglos XVI y XVII, durante la Revolución Científica, con la aparición de prestigiosos investigadores, como William Gilbert, Benjamin Franklin o Charles Coloumb, y sus primeras aportaciones a la comunidad.

Posteriormente, diferentes personalidades, como Michael Faraday o Christian Oersted siguen las investigaciones y el uso de la electricidad se va aproximando al actual. También se ha de destacar en esta época el avance en iluminación, procesos industriales y telecomunicaciones, que potencia la necesidad de esta fuente de energía.

Esta es actualmente esencial en el día a día, dada la evolución tecnológica y medioambiental, dejando de lado fuentes precarias como el carbón o el petróleo y optando por la energía eléctrica, lo que provoca grandes demandas que han de ser satisfechas en el momento.

Debdo a que es imposible o excesivamente costoso almacenar tal cantidad, es muy importante conocer la demanda futura para poder suministrar la cantidad necesaria de energía, minimizando pérdidas por inutilización y evitando falta de abastecimiento, dado que la excesiva o insuficiente producción de electricidad genera un gran impacto medioambiental y económico.

Uno de los sucesos que han impulsado el estudio de la predicción ocurrió el 9 de noviembre de 1965, donde los habitantes de 8 estados del este de Estados Unidos sufrieron el apagón eléctrico más grande de la historia: 14 horas en las cuales semáforos, trenes, ascensores... dejaron de funcionar. Esto fue debido a un colapso en cadena de la red que vincula Canadá y la costa noroeste de EE. UU.

Esto fue una llamada de atención para el sistema eléctrico, y el Ministerio de Industria y Energías propuso a las empresas eléctricas a participar en el proyecto INDEL, basado en el análisis de la demanda eléctrica. Este buscaba conocer la demanda puntual de electricidad y cómo se demandará en el futuro, y estaba destinado a reducir la incertidumbre sobre su evolución y a facilitar las decisiones orientadas a aumentar su eficiencia, dadas las características principales de la demanda eléctrica:

- Esta debe coincidir estrictamente en el tiempo con su consumo
- El sistema eléctrico suministra en todo momento toda la energía que demandan sus clientes
- El perfil horario tiene gran importancia.

Es por ello que las herramientas de predicción tienen tanta relevancia hoy en día, dado que se deben tomar decisiones de futuro con cierta incertidumbre, la cual se verá ampliamente reducida si se dispone de un buen modelo de predicción que permita realizar estimaciones lo más exactas posibles.

Otro aliciente a los modelos de predicción es el desarrollo e implementación de sistemas de redes inteligentes, o SmartGrids.

Este concepto apareció como respuesta a los retos del sistema eléctrico. Una SmartGrid integra el uso de sensores, comunicaciones, habilidad computacional y control para mejorar la funcionalidad del sistema de distribución eléctrica. Con el objetivo de mejorar las operaciones, mantenimiento, y diseño usando tecnología moderna para gestionar mejor el uso de energía y su coste.

Las SmartGrids empezaron a aparecer en la década de 1980, en la cual la lectura automática de contadores se utilizó para el seguimiento de las cargas de los grandes clientes, y se convirtió en una infraestructura de medición avanzada en la década de 1990.

Las primeras formas de gestión se basaban en dispositivos, que, mediante el control en el cambio de la frecuencia de alimentación, eran capaces de evitar la activación en momentos pico de máquinas industriales y domésticas, como aires acondicionados, refrigeradores y calentadores.

Ya en la década de los 2000, aparecieron las primeras redes inteligentes masivas, con más de 27 millones de usuarios conectados mediante medidores inteligentes, como es el caso del proyecto Telegestore. Otras redes importantes fueron la red de Austin, Texas, o Boulder, Colorado.

Hoy en día, y con la progresiva descentralización de la generación energética, es importante contar con una red fiable, flexible, eficiente, y que sea capaz de ajustar la carga. Para esto último, el desarrollo de sistemas de almacenamiento y la generación renovable han permitido crear sistemas que proporcionen la energía necesaria sin afectar a su comportamiento, mientras que tradicionalmente se utilizaban técnicas basadas únicamente en el patrón de consumo de los usuarios. Aquí entra la importancia de contar con un buen modelo de predicción, para poder entre otras cosas ajustar la cantidad de energía a almacenar o a vender si se tiene la oportunidad, para anteponerse a posibles picos o valles de demanda, y en general para gestionar de manera óptima y eficiente la red.

Por último, la integración del usuario en la gestión de su propia red es un hecho cada vez más relevante, dado que esto tiene una repercusión económica.

El consumo eléctrico de los electrodomésticos supone un 55% del consumo total de la vivienda, por lo que, ante la imposibilidad teórica de reducirlo, es importante distribuirlo o al menos conocer sus características para poder adaptarse a la tarifa existente o adquirir una nueva si es necesario.

### 6.- % Consumo Eléctrico según Servicios

<b>Calefacción:</b>	<b>7,4%</b>
<b>ACS:</b>	<b>7,5%</b>
<b>Cocina:</b>	<b>9,3%</b>
<b>Refrigeración:</b>	<b>2,3%</b>
<b>Iluminación:</b>	<b>11,7%</b>
<b>Electrodomésticos:</b>	<b>55,2%</b>
<b>Standby:</b>	<b>6,6%</b>

Fig. 1. Distribución del consumo eléctrico en el sector residencial en España. Fuente: IDAE

Los modelos de predicción resultan óptimos en estos casos, debido a los datos que se han tenido que almacenar, a partir de los cuales se pueden obtener patrones, y detectar aquellas variables que más afectan o más condicionan la demanda.

Por lo tanto, el objetivo del TFG desarrollar un algoritmo en C# mediante Microsoft Visual que sea capaz de predecir con la mayor exactitud posible el estado de consumo de diferentes electrodomésticos durante un periodo futuro, que comprenderá de horas a días, a partir de unos datos históricos de consumo y otras variables. Se propondrán tres modelos de predicción y se evaluará el distinto comportamiento de cada uno para unos escenarios determinados.

Posteriormente se escogerá un óptimo y se presentarán sus resultados para otros electrodomésticos.

## 1.2. Estructura del documento

En el Capítulo 1 se hace una pequeña introducción de la introducción de la red eléctrica en la vida diaria. Se habla de la Red Eléctrica en España, del concepto SmartGrid, su evolución en el tiempo, y de la progresiva implicación del usuario en la gestión de su red. En todos estos puntos se hace hincapié indicando la importancia de tener un buen sistema de predicción, que evite pérdidas de energía no almacenable, o que pueda realizar un buen ajuste de carga/descarga de baterías para tener un mejor rendimiento económico, o que permita al usuario tener más información sobre sus hábitos, consumos, y factores que afecten a este. Además, se especifican los objetivos de este Trabajo de Fin de Grado y los pasos que se van a seguir para llevarlo a cabo.

En el capítulo 2, se explica los principales usos de los modelos de predicción, relacionándolo con el objetivo de este trabajo. A continuación, se da una breve exposición del hardware necesario para la obtención de los datos con los que se va a trabajar, y las normativas que se deberían tener en cuenta, y finalmente, se pone al lector en el contexto de la realización del trabajo, dentro del Instituto Tecnológico de la Energía.

En el capítulo 3 se explica el estado del arte de algunos algoritmos que están siendo utilizados para la predicción de uso de electrodomésticos.

En el capítulo 4 se plantea la situación de partida. Primero se expone de dónde se obtienen los datos con los que se va a probar y realizar el programa. A continuación, se aborda la necesidad de preprocesado de estos, y se explica el método de clasificación escogido y su teoría. Finalmente se presentan los resultados que éste otorga.

En el capítulo 5 se propone el método en el que se basará el algoritmo y la estructura que siguen las entradas del programa, así como los 3 predictores y las distintas variables que se van a analizar. A continuación, también se exponen los índices que evaluarán el comportamiento de éstos.

En el capítulo 6 se expone el escenario en el que se llevará a cabo la predicción y las simplificaciones realizadas para ello, y los distintos resultados obtenidos por cada predictor para la evaluación de su comportamiento.

En el capítulo 7 se explica las conclusiones obtenidas a partir de los resultados del capítulo anterior y se muestran resultados de otras simulaciones. Finalmente se proponen las diferentes líneas futuras que tiene este proyecto.

## **CAPÍTULO 2 . PREDICCIÓN DE CONSUMOS**

### **2.1 Aproximación al problema**

Los modelos de predicción son usados en diferentes dominios, por ejemplo, en el mundo de los mercados de valores, en negocios para inventariar o predecir demanda (Hyndman & Athanasopoulos 2014) [1], en medicina para monitorizar la extensión de enfermedades (Generous et al.2014) [2], o directamente para predecir el tiempo.

De la misma manera, la predicción juega un rol esencial en el control de plantas de generación de energía y en los intercambios de esta en sistemas interconectados. Predecir ayuda a los gestores de energía a entender la influencia de algunas variables en el consumo, y, por lo tanto, poder tomar decisiones.

En una escala temporal, la predicción puede ser a corto plazo, para equilibrar el suministro eléctrico, o a largo plazo, para analizar la capacidad de expansión, el retorno de una inversión o la renovación del sistema.

A través de los años, se han empleado diferentes modelos de predicción, como la regresión multivariable (Azadeh & Faiz 2011) [3], redes neuronales artificiales (Hernández et al.2014) [4], series temporales, media móvil autorregresiva...

No hay consenso sobre un sistema particular, y el uso de un método u otro queda finalmente a preferencia del experto.

Además. en el artículo "The M3-Competition: results, conclusions and implications" (Makridakis & Hibon 2000) [5], en el que se compara la precisión de diferentes modelos de predicción, se concluye que "Los modelos simples desarrollados por personas inexpertas predicen tan bien, o incluso mejor que aquellos sofisticados". Sin embargo, es interesante identificar qué modelos se comportan mejor en qué situación.

La necesidad de predicción varía de un caso a otro, por lo que la creación de un modelo está sujeto a muchas variaciones, desde los datos de los cuales se dispone como entradas, la ventana de tiempo deseada, la resolución temporal (minutal o anual), la escala (desde un electrodoméstico en cuestión a un país entero).

En este trabajo se han fijado algunas de esas variables, como son la extensión, que irá desde un electrodoméstico en cuestión a toda la línea de la casa. La resolución, que será cada minuto, y la ventana temporal, que irá desde 1 hora a un día.

## 2.2 Hardware necesario

Aun cuando el presente trabajo se centrará exclusivamente en las técnicas y algoritmos, puede resultar interesante una breve explicación sobre el hardware necesario para obtención de datos:

Existe una amplia diversidad de maneras de obtener los datos de consumos de una vivienda. La forma más sencilla es a través de adaptadores de conexión especiales que se conectan directamente a estos aparatos. Estos destacan por su gran facilidad de conexión y registro de información. No obstante, conlleva un coste económico importante, al necesitarse un adaptador por electrodoméstico.

Otra opción son los medidores inteligentes, los cuales pueden dar el valor de consumo por línea, y parece la solución óptima para obtener el consumo global de la vivienda.

## 2.3 Normativa

Al tratarse de un algoritmo, este proyecto no está limitado ni debe regirse por unas normativas en concreto.

No obstante, el caso de la obtención de datos, es necesario cumplir con El Reglamento Electrotécnico para Baja Tensión de España, en cuanto a la seguridad y dimensionamiento de la conexión del Smart-Meter, y la norma 802.11, la cual regula el modo de comunicación entre los medidores y el ordenador, y el posterior almacenamiento de la información en la nube.

## 2.4 Contexto

El presente Trabajo de fin de Grado forma parte de un proyecto del Instituto Tecnológico de la Energía, departamento de SmartGrids, que tiene como objetivo el desarrollo de técnicas analíticas de datos para la mejora del autoconsumo solar mediante la incorporación de algoritmos NIALM en las predicciones de demanda individualizadas. Igualmente, la predicción de demandas de cargas o demandas individualizadas es un objetivo en sí mismo debido a la importancia a la hora de desarrollar Sistemas de Gestión de Demanda para los cuáles una correcta previsión de este tipo de cargas a nivel de dispositivo es fundamental para realizar una planificación eficiente.

La predicción de demanda a corto plazo (STLF) individuales o desagregadas ya sea a nivel de vivienda individual, a nivel de fase/línea de alimentación dentro del hogar o a nivel de electrodoméstico individual, se puede considerar un enfoque de abajo hacia arriba: primero, la predicción de consumo de energía se realiza para cada aparato en una vivienda y después se realiza la predicción de la energía total consumida agregada en la vivienda.

Debido a las especiales características de estas demandas individuales (alta varianza y volatilidad), las técnicas habitualmente empleadas para la predicción de demandas agregadas pueden no ser adecuadas. La predicción de demanda para consumos individuales desagregados es más complicada debido a varios factores. La energía horaria consumida por una sola casa o electrodoméstico es pequeña y muy variable, ya que depende de la cantidad de personas que

permanecen en el hogar, el número y tipo de aparatos eléctricos conectados en ese momento en particular, etc. Debido a esto el consumo de energía en individual desagregada fluctúa mucho, incluso en horas consecutivas o en días consecutivos, respecto de cualquier patrón regular. La fluctuación es relativamente baja para consumidores a nivel de industrial a pequeña escala debido a una planificación fija de la producción. Por otra parte, en consumos agregados la volatilidad y la varianza de la demanda se ven reducidas debido a la propia agregación de los datos que compensa y cancela las fluctuaciones bruscas que puedan producirse de forma individual.

Los métodos más comúnmente utilizados en la actualidad para la predicción de la demanda y más concretamente para la predicción STLF son la regresión lineal múltiple, los métodos autorregresivos basados en series temporales y los basados en técnicas de machine-learning como Redes Neuronales, Regresión de Soporte Vectorial (SVR) y técnicas de clasificación y clustering.

Debido a la incertidumbre asociada con el uso de dispositivos, los modelos basados en regresión no han demostrado ser muy útiles en la predicción de uso de energía de dispositivos. El empleo de técnicas más cualitativas basadas clasificación o clustering con un enfoque probabilístico pueden proporcionar mejores resultados. Así mismo, el empleo técnicas dentro de un marco colaborativo o híbrido, donde se combinen métodos basados en regresión y series temporales con técnicas de clasificación y clustering puede permitir solventar o reducir las desventajas y debilidades de cada uno de los métodos por separado.

Otro factor importante a la hora de mejorar las prestaciones de precisión de la técnica de predicción empleada será el método de selección de características utilizado, donde se selecciona un subconjunto de características a partir de datos para la aplicación del algoritmo de predicción. El mejor subconjunto contiene la menor cantidad de características relevantes que contribuyen a la precisión; mientras descarta las características restantes menos relevantes. Una adecuada selección de los datos más relevantes para la predicción permite no solo utilizar la información más relevante que influye en la predicción y descartar aquella información que puede hacer menos fiable la predicción, sino reducir el tamaño de los patrones de entrenamiento de los modelos de predicción con el consiguiente ahorro computacional.



## **CAPÍTULO 3. ESTADO DEL ARTE**

En este capítulo se presentan y algunos de los algoritmos que están siendo aplicados en el ámbito de la predicción de consumos:

### **3.1 Algoritmo de Histograma**

Tal y como se explica en la publicación de Patrick Huber y otros, 2018 [6], Asumiendo que las actividades del hogar siguen un patrón semanal, uno puede construir un histograma de los tiempos de encendido de un electrodoméstico para cada día de la semana, basado en los datos de aprendizaje.

La aproximación usada condiciona los días relevantes mediante una ponderación Gaussiana alrededor de las horas de interés.

Basado en los anteriores  $N$  días, cada un dividido en  $T$  intervalos de tiempo, la probabilidad de que en el día  $n$  a la hora  $T$  el electrodoméstico  $l$  esté funcionando se calcula como:

$$p(x_{ntl}) \propto \sum_{m \in N} \sum_{\tau \leq T} w_{nm} e^{-\frac{(t-\tau)^2}{2\sigma^2}} x_{m\tau l} \quad (1)$$

Donde:

- $x_{m\tau l} = 1$  si el electrodoméstico  $l$  está funcionando en el intervalo  $\tau$  en el día  $m$ .
- $x_{m\tau l} = 0$  si es lo contrario
- $w_{nm} = 1$  si  $n = m$
- $w_{nm} = 0$  si es lo contrario
- La varianza  $\sigma$  es un parámetro que se establece experimentalmente

### 3.2 Algoritmo de predicción bayesiano

Este método (Patrick Huber y otros, 2018) [6] utiliza una aproximación de Cadenas de Márkov, basada en el método Montercarlo, para muestrear la posterior distribución de los parámetros del modelo. Los elementos claves del modelo son los “latent day-types  $k$ ”, que son usados para crear perfiles diarios para registrar correlaciones entre el uso de los electrodomésticos individuales.

En resumen, la probabilidad de un electrodoméstico  $l$  funcionando en la hora  $t$  en el día  $n$  se calcula como:

$$p(x_{nlt}) \propto \sum_K p(k | n) \mu_{kl}(t) \tag{2}$$

- Donde  $k$  recorre todos los días  $K$  tipo.
- $p(k | n)$  es la probabilidad de que el día  $n$  esté descrito por el día tipo  $k$
- Una de las ventajas de esta aproximación es que infiere los parámetros de cada electrodoméstico  $l$  desde los datos de todos estos resultando con un entrenamiento efectivo, dado por  $N * L$  puntos.
- $L$  es el número total de electrodomésticos.

### 3.3 ARMA (media móvil auto regresiva)

En este método (Nicoleta Arghira y otros, 2011) [7], el valor actual de una variable temporal es asumido como una función de sus valores anteriores, y es expresada como una suma ponderada (media móvil) (

$$A(q)y(t) = C(q)e(t) \tag{3}$$

Donde  $A$  y  $C$  son funciones polinómicas en el operador de cambio de tiempo  $q^{-1}$ ,  $y(t)$  es el output y  $e(t)$  es el sonido blanco. (componente usado para tener en cuenta la incertidumbre y se asume que está idénticamente distribuido con media cero y varianza finita.

Las expresiones de  $A(q)$  y  $C(q)$  son:

$$\begin{aligned} A(q) &= 1 + a_1q^{-1} + a_2q^{-2} + \dots + a_{N_A}q^{-N_A}, N_A = \deg(A(q)) \\ C(q) &= 1 + c_1q^{-1} + c_2q^{-2} + \dots + c_{N_C}q^{-N_C}, N_C = \deg(C(q)) \end{aligned} \tag{4}$$

El valor de la salida  $y(t)$  es expresado como una suma ponderada (media móvil) de todos los datos anteriores  $e(t)$  y sus valores iniciales  $y(0)$ .

### 3.4 Modelo probabilístico

Básicamente hay dos factores que necesitan tenerse en cuenta a la hora de calcular la probabilidad de que un electrodoméstico esté siendo usado: El tiempo que ha pasado desde que se ha utilizado por última vez, y la hora del día en la que normalmente se usa. ( Marc Wenninger y otros, 2017) [8].

Estos dos casos se modelan separadamente, como una función de distribución de probabilidades, donde  $E$  = tiempo transcurrido desde el evento, y  $D$ = hora del día del evento.

La probabilidad  $P(E)$  de que un electrodoméstico no se use para el periodo  $t$  (medido en minutos) es modelado mediante una distribución exponencial:

$$P(E) = P(E \leq t) = \begin{cases} 1 - e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{5}$$

La estimación de la probabilidad máxima del parámetro  $\lambda$  puede obtenerse a partir de un data set suficientemente largo calculando el valor medio de los tiempos entre encendidos consecutivos de los electrodomésticos. Una buena estimación será obtenida cuando el tiempo entre uso sea bastante regular, resultando en un valor pequeño de la varianza de los tiempos.

## CAPÍTULO 4. SITUACIÓN INICIAL

### 4.1 Datos de partida

Los datos con los que se va a trabajar corresponden a valores de potencia activa de diferentes electrodomésticos pertenecientes a la base de datos del ITE. Se tiene un valor cada minuto de aparatos como aire acondicionado, secadora, lavadora, horno, vitrocerámica, microondas, horno, nevera...entre otros.

Las mediciones de las que se disponen corresponden a un periodo de tiempo comprendido entre el 1 y 31 de enero de 2017.

El acceso a ellos se realiza mediante un fichero CSV del que se disponen de todos. Hay algunos casos en los cuales es posible encontrar un gap (dato vacío) entre ellos. Es por ello que el programa se ha configurado de tal manera que, si esto ocurre, a ese dato se le torga el valor de la medición inmediatamente anterior.

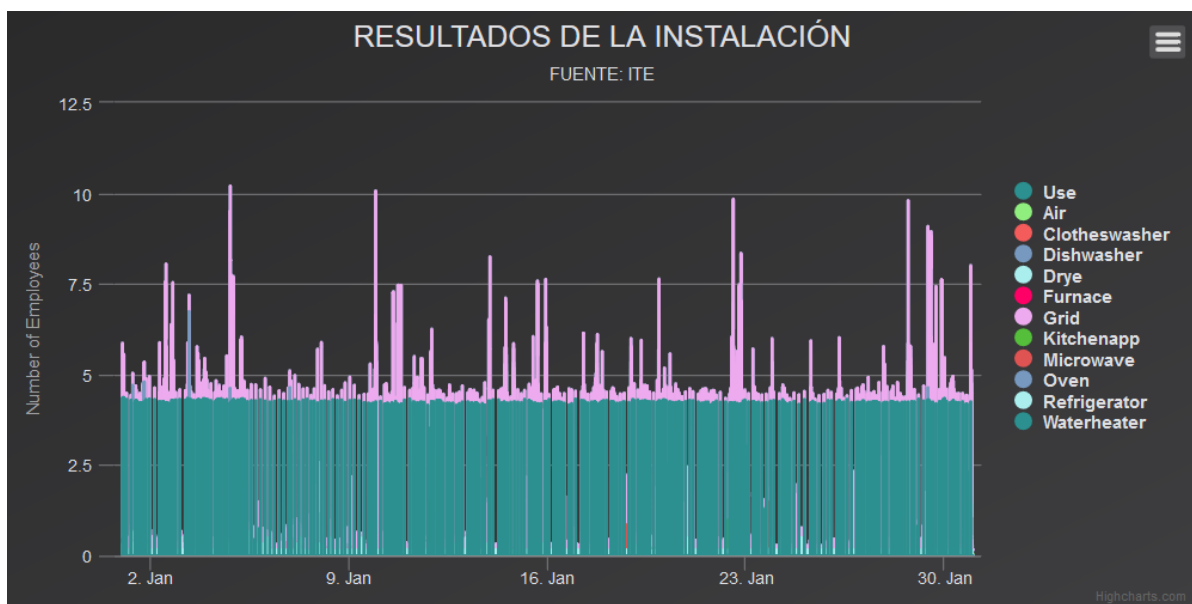


Fig. 2. Representación de los datos de la base del ITE. Fuente: ITE

### 4.2 Preprocesado

Se cuenta, por lo tanto, con datos de consumos de diferentes electrodomésticos. En la Figura 2 puede verse el consumo superpuesto de toda la base de datos, mientras que en la Figura 3 se puede ver únicamente el frigorífico.

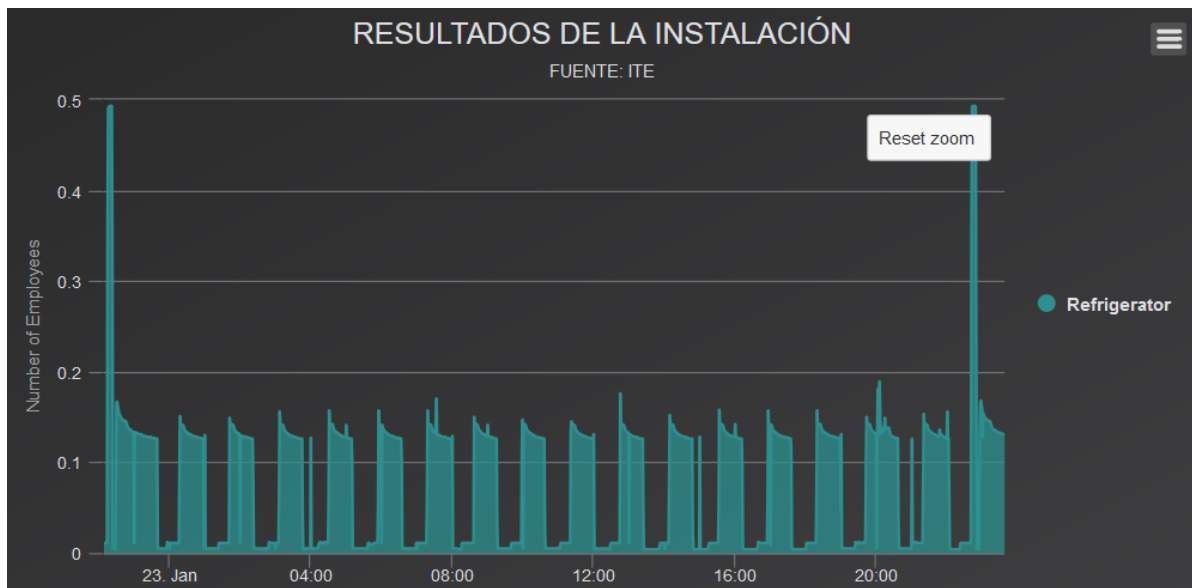


Fig. 3: Curva de carga de la nevera para el día 23 de enero. Fuente: ITE

Como se ha explicado antes, se obtienen estos datos y antes de trabajar con ellos, se analizan en busca de un hueco vacío. Si ocurre esto, se le da el valor del consumo interior.

### 4.3 Clasificación

Como ya se mencionó al inicio del trabajo, el objetivo del programa es predecir el estado en el que se encuentra el electrodoméstico  $X$  en el instante de tiempo  $T$ . Actualmente lo que se tiene son valores de potencia, por lo tanto, el primer paso será clasificar todos estos datos dándoles un valor que represente su estado.

El número de estados a clasificar/predecir se ha establecido en 3, de acuerdo a la curva típica de una nevera, que se caracteriza por tener momentos de conexión (0), momentos de funcionamiento continuo (1) y arranques, determinados por un pico de potencia (2), o el estado de consumo de una curva de carga diaria tipo, en la que se tiene el periodo valle, de muy baja demanda (0), el llano, donde la demanda sube y es variable (1), y el punta, los momentos de pico de potencia (2). Todo el programa se va a desarrollar para detectar 2 o 3 estados, aunque es posible y fácilmente implementable la detección de muchos más.

Para poder realizar esta clasificación, se ha escogido el método de k-medias.

**4.3.1. K-medias.** El algoritmo K-Medias, propuesto por MacQueen en el año 1967, es un método de agrupamiento, que tiene como objetivo la partición de un conjunto de  $n$  observaciones en  $k$  grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano. (Youguo Li, 2012) [9].

El algoritmo consta de tres pasos:

1. Inicialización: Una vez escogido el número de grupos  $k$ , se establecen  $k$  centroides en el espacio de datos, por ejemplo, escogiéndolos aleatoriamente.

2. Asignación objetos a los centroides: Cada objeto de los datos es asignado a su centroide más cercano.

3. Actualización de los centroides: Se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Se repiten los pasos 2 y 3 hasta que los centroides se mueven por debajo de una distancia umbral.

La función del K-medias es minimizar la suma de las distancias cuadráticas de cada objeto al centroide de su cluster. Para ello, los objetos se representan con vectores reales de  $d$  dimensiones  $((x_1, x_2, \dots, x_n))$  y el algoritmo k-means construye  $k$  grupos donde se minimiza la suma de distancias de los objetos, dentro de cada grupo  $S = \{S_1, S_2, \dots, S_k\}$  a su centroide. El problema se puede exponer de la siguiente manera:

$$\min_{\mathbf{S}} E(\boldsymbol{\mu}_i) = \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2 \quad (6)$$

Donde  $S$  es el conjunto de datos cuyos elementos son los objetos  $x_j$  representados por vectores, donde cada uno de sus elementos representa una característica o atributo. Tendremos  $k$  grupos o clústers con su correspondiente centroide  $\mu_i$

Las principales ventajas del método *k-means* son que es un método sencillo y rápido. Pero es necesario decidir el valor de  $k$  y el resultado final depende de la inicialización de los centroides. En principio no converge al mínimo global sino a un mínimo local.

#### 4.4 Resultados

En la figura XX se puede ver la clasificación de un periodo ciclo del frigorífico, mientras que en la figura XX se puede ver la clasificación de la curva agregada, con sus centroides y escogiendo  $k=3$ . (En esta última se ha hecho una gráfica debido a su extensión).

Tras realizar varias pruebas para escoger los datos de entrenamiento óptimo, se ha determinado que un periodo de 1 semana es perfectamente aceptable, dado que se tiene suficiente información como para determinar los centroides correctamente, evitándose utilizar una excesiva cantidad de datos, lo que ralentizaría el proceso de computación.

Desarrollo, optimización e implementación de un algoritmo de predicción de consumo de energía de una vivienda



Fig. 4 = Ciclo de 1 hora de frigorífico

Time	Power	Cluster	Time	Power	Cluster
07/07/2019 3:58:01	90	1	07/07/2019 3:28:01	0	0
07/07/2019 3:56:01	90	1	07/07/2019 3:26:01	0	0
07/07/2019 3:54:01	90	1	07/07/2019 3:24:01	0	0
07/07/2019 3:52:01	111	2	07/07/2019 3:22:01	0	0
07/07/2019 3:50:01	90	1	07/07/2019 3:20:01	0	0
07/07/2019 3:48:01	90	1	07/07/2019 3:18:01	0	0
07/07/2019 3:46:01	90	1	07/07/2019 3:16:01	0	0
07/07/2019 3:44:01	90	1	07/07/2019 3:14:01	0	0
07/07/2019 3:42:01	90	1	07/07/2019 3:12:01	111	2
07/07/2019 3:40:01	90	1	07/07/2019 3:10:01	90	1
07/07/2019 3:38:01	0	0	07/07/2019 3:08:01	90	1
07/07/2019 3:36:01	0	0	07/07/2019 3:06:01	90	1
07/07/2019 3:34:01	0	0	07/07/2019 3:04:01	90	1
07/07/2019 3:32:01	0	0	07/07/2019 3:02:01	90	1
07/07/2019 3:30:01	0	0	07/07/2019 3:00:01	90	1

Centroides: 0 = 0, 1 = 91.45340799, 2 = 111

Fig. 5 = Clasificación del ciclo de un frigorífico



Fig. 6 = 24 horas de consumo agregado

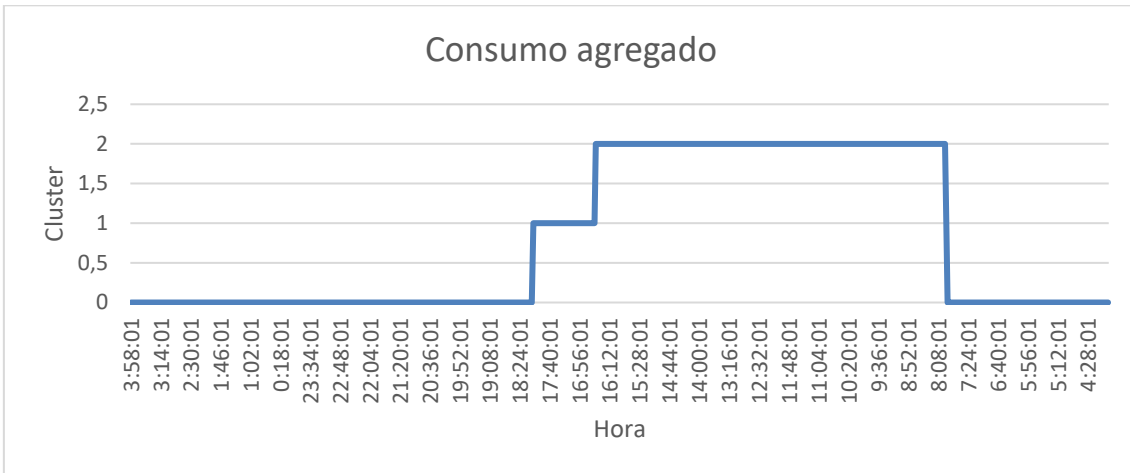


Fig. 7 = 24 horas de consumo agregado clasificado

Una vez los datos con los que se trabaja están clasificados, se puede trabajar con ellos mediante las propuestas mencionadas en el siguiente capítulo.



## CAPÍTULO 5. MÉTODO PROPUESTO

### 5.1 Generar conocimiento: Oracle

En el inicio únicamente se disponen de datos de consumo y de tiempo. Una opción es trabajar directamente con ellos, pero los resultados que esto da en otros proyectos son pobres.

Otra opción es intentar extraer más conocimiento de ellos, y es por ello por lo que se propone la utilización del modelo Oracle, tal y como afirman Kaustav Basu y otros, 2012. [10]

Este recibe los datos de partida de la base de datos. A continuación, los organiza para obtener más información de ellos y finalmente esa información es enviada al predictor, para su posterior aprendizaje y predicción. La información que Oracle incluye tiene que ser validada, y ese será uno de los objetivos del Trabajo de Fin de Grado, determinar qué datos influyen más o menos en función del predictor utilizado.

Los estados que se pueden representar en el modelo Oracle son:

- El consumo en el pasado
- La hora del día en la que se predice
- El minuto del día en el que se predice
- El día de la semana de predicción
- El consumo en el mismo momento de otros días

A continuación, se desarrolla cada una de las secciones:

- 1) Consumo en el pasado: Se tiene en cuenta el estado de consumo en el cual se encontraba el electrodoméstico en el pasado. Se representa matemáticamente como un vector  $\{0,1,2\}^n$ , donde  $n$  corresponde al número de minutos anteriores que el programa está considerando.
- 2) Hora del día: Se añade la hora del día que se está analizando. Para su representación, el día se discretiza en 24 horas. Se representa matemáticamente como un vector  $\{0 - 23\}^1$ . Por ejemplo, si se analiza la hora 5:00, su representación sería "5".
- 3) Minuto del día: Al igual que con la hora, se añade el minuto con el cual se trabaja. Para ello, la hora se discretiza en 60 espacios, representándose de 2 en 2, debido a que los datos de los que se disponen son cada 2 min. Se representa matemáticamente como un vector  $\{0 - 59\}^1$ . Por ejemplo, si se analiza el minuto 30, su representación sería "30".

- 4) Día de la semana: Similar a los otros, se representa el día con el que se está trabajando. La semana se discretiza en 7 espacios. Se representa matemáticamente como un vector  $\{0 - 6\}^1$ . Estas 3 últimas variables llevarán el nombre de "time".
- 5) Estado en el mismo momento de días pasados: Se analiza si ha habido o no consumo en el mismo momento de otros días pasados. Se toma como tiempo 1 semana, por lo cual se crean 7 espacios, representando cada uno el valor del estado en los días anteriores de la semana en la misma hora y minuto. Su representación matemática es un vector  $\{0,1,2\}^7$ . A esta variable se la designa como "estado".

Finalmente, la estructura de 1 dato tendrá la forma de un vector en el cual se concatenan todos los anteriores, donde M es el minuto para analizar.

Conocimiento	M
Estado(M-1)	0/1/2
Estado(M-2)	0/1/2
...	0/1/2
Estado(M-n)	0/1/2
Hora del día (H)	0/2/3
Minuto del día (M)	0/59
Día de la semana	0/6
Estado día-1	0/1/2
estado día-2	0/1/2
...	0/1/2
Estado día-7	0/1/2

Tabla 1: Esquema de las variables

Para poder realizar la predicción, son necesarios P datos, por lo que el input del programa será una matriz  $(n + 10) \times P$ , donde cada columna P representa un minuto anterior, que será un vector compuesto por n+10 valores, los mencionados anteriormente. Por lo tanto, el input que recibirá el programa, para el cual se querrá predecir el consumo en un minuto M tendrá la forma:

$$\begin{bmatrix} Estado M - 1 & \dots & Estado P - 1 \\ \vdots & \ddots & \vdots \\ Estado día(M) - 7 & \dots & Estado día(P) - 7 \end{bmatrix}$$

(7)

## 5.2 Predictor

Después de introducir el sistema de clasificación y de ordenación, se proponen 3 predictores, los cuales se van a analizar, probar, determinar variables y valorar.

**5.2.1 Regresión Logística Multinomial.** La regresión logística multinomial es utilizada en modelos de variables dependiente de tipo nominal con más de dos categorías (dado que se quiere predecir 3 o más estados), y es una extensión multivariante de la regresión logística binaria clásica. (Dr. Jon Starkweather, 2011) [11], (Abdalla M. EL-HABIL, 2012) [12]

Se considera una variable aleatoria dependiente  $Y$  con soporte  $(Y) = \{1,2,3\}$ , y con probabilidades  $p_1 = P(Y = 1)$ ,  $p_2 = P(Y = 2)$  y  $p_3 = 1 - p_1 - p_2$ .

Si se quiere analizar el efecto que ejercen dos variables  $X_1$  y  $X_2$  sobre las probabilidades  $p_1$  y  $p_2$ , se puede redefinir la variable  $Y$  mediante un vector  $(Y_1, Y_2)$  construido de la siguiente forma:

$$(Y_1, Y_2) = \begin{cases} (1,0) & \text{si } Y = 1 \\ (0,1) & \text{si } Y = 2 \\ (0,0) & \text{si } Y = 3 \end{cases} \quad (8)$$

Las variables  $Y_1$  e  $Y_2$  tienen una distribución de Bernoulli con  $E(Y_1) = p_1$  y  $E(Y_2) = p_2$ .

Se formula el modelo multivariante definido por las siguientes ecuaciones:

$$\left. \begin{aligned} p_1(X_1, X_2) = p_1 = E(Y_1) &= \frac{\exp(Z_1)}{1 + \exp(Z_1) + \exp(Z_2)} \\ p_2(X_1, X_2) = p_2 = E(Y_2) &= \frac{\exp(Z_2)}{1 + \exp(Z_1) + \exp(Z_2)} \end{aligned} \right\} \quad (9)$$

Y de donde se obtiene

$$p_3(X_1, X_2) = p_3 = 1 - p_1 - p_2 = \frac{1}{1 + \exp(Z_1) + \exp(Z_2)} \quad (10)$$

Donde  $Z_1 = \beta_{01} + \beta_{11} * X_1 + \beta_{21} * X_2$  y  $Z_2 = \beta_{02} + \beta_{12} * X_1 + \beta_{22} * X_2$ , siendo  $\beta_{01}, \beta_{11}, \beta_{21}, \beta_{02}, \beta_{12}, \beta_{22}$  parámetros que se desean estimar.

Tras realizar algunas transformaciones, se obtiene tomando logaritmos en ambas ecuaciones:

$$\left. \begin{aligned} \ln\left(\frac{p_1}{p_3}\right) &= Z_1 = \beta_{01} + \beta_{11} \cdot X_1 + \beta_{21} \cdot X_2 \\ \ln\left(\frac{p_2}{p_3}\right) &= Z_2 = \beta_{02} + \beta_{12} \cdot X_1 + \beta_{22} \cdot X_2 \end{aligned} \right\} \quad (11)$$

Ahora, dada una muestra de datos  $(Y_{1i}, Y_{2i}, X_{1i}, X_{2i})$  con  $i=1, 2..n$  se puede definir en función de los parámetros del modelo, las funciones  $Z_{1i}, Z_{2i}, p_{1i}, p_{2i}$  y abordar el problema de la estimación de los mismos mediante el método de máxima verosimilitud, como se muestra a continuación:

$$L = \prod_{i=1}^n \left( p_{1i}^{Y_{1i}} \cdot p_{2i}^{Y_{2i}} \cdot p_{3i}^{1-Y_{1i}-Y_{2i}} \right) = \prod_{i=1}^n \left( \left( \frac{p_{1i}}{p_{3i}} \right)^{Y_{1i}} \cdot \left( \frac{p_{2i}}{p_{3i}} \right)^{Y_{2i}} \cdot p_{3i} \right) \quad (12)$$

En lo que se refiere al programa, la función está dividida en dos segmentos: Aprendizaje y predicción.

Este primero ha de ser supervisado, es decir, el modelo se entrena con las matrices de la clasificación y con los estados correspondientes, para después pasar al modo predicción, en la que únicamente tendrá los datos clasificados. Lógicamente, se tiene que establecer un rango de datos para el entrenamiento que no incluyan a aquellos que se quieren predecir.

**5.2.2 Cadenas ocultas de Márkov.** Los modelos ocultos de Márkov han sido aplicados ampliamente en el estudio del NILM. Un modelo oculto de Márkov (HMM por sus siglas en inglés) es un modelo estadístico en el que se asume que el sistema a modelar es un proceso de Márkov de parámetros desconocidos. El objetivo es determinar los parámetros ocultos a partir de los parámetros observables. (Weicong Kong y otros 2018) [13].

En este modelo, el estado no es visible directamente para el observador, sino que sólo lo son las variables influidas por este.

Estos modelos son especialmente aplicados a reconocimiento de formas temporales, como el reconocimiento del habla, de escritura manual, de gestos....

En cuanto a su arquitectura, en la figura 8 se puede observar sus características. Cada óvalo representa una variable aleatoria que puede tomar determinados valores. La variable aleatoria  $x(t)$  es el valor de la variable oculta en el instante de tiempo t. La variable aleatoria  $y(t)$  es el valor de la variable observada en el mismo instante de tiempo t. Las flechas indican dependencias condicionales.

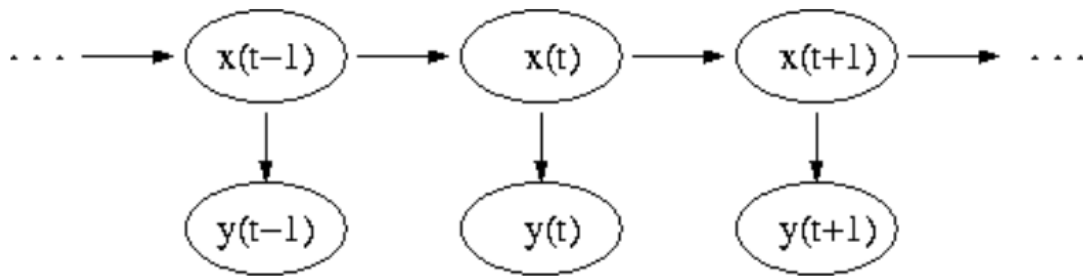


Fig. 8 Arquitectura cadena oculta de Márkov

La propiedad de Márkov establece que el valor de la variable oculta  $x(t)$  e el instante  $t$  solo depende del valor de esa en  $t-1$ . Del mismo modo, el valor de la variable observada  $y(t)$  solo depende de  $x(t)$ .

La probabilidad de observar una secuencia  $Y = y(0), y(1), \dots, y(L-1)$  de longitud  $L$  está dada por:

$$P(Y) = \sum_X P(Y | X)P(X) \quad (13)$$

Donde la sumatoria se extiende sobre todas las secuencias de nodos ocultos  $X = x(0), x(1), \dots, x(L-1)$ . Este cálculo se puede realizar por fuerza bruta, pero es impráctico por el número de secuencias de nodos, por lo que normalmente se utiliza un algoritmo conocido como procedimiento de avance-retroceso.

A igual que en la regresión logística, en el programa función está dividida en dos segmentos: entrenamiento y predicción.

Este primero será supervisado, para después pasar al modo predicción, en la que únicamente tendrá los datos clasificados para su predicción.

**5.2.3 Árbol de decisiones** Los árboles de decisión son representaciones gráficas de posibles soluciones a una decisión basadas en ciertas condiciones. Es uno de los algoritmos de aprendizaje supervisado más utilizado en machine learning y pueden realizar tareas de clasificación o regresión. Los árboles de decisión, donde la variable de destino puede tomar valores continuos (por lo general números reales) se llaman árboles de regresión. (Árboles de clasificación y regresión, UV) [14]

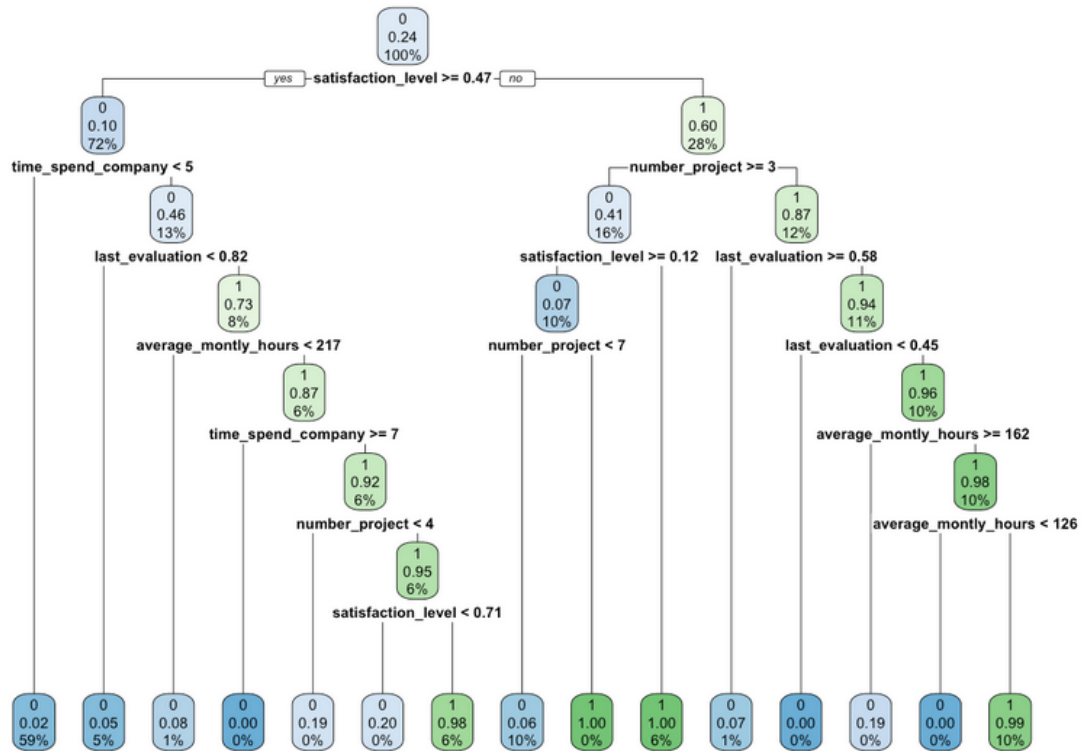


Figura 9. Representación gráfica de un árbol de decisiones

Los árboles de decisión tienen un primer nodo llamado raíz (root) y luego se descomponen el resto de los atributos de entrada en diferentes ramas. Cada nodo se bifurca y vuelven a subdividirse hasta a llegar a las hojas, que son los nodos finales y que equivalen a respuestas a la solución.

Este puede ser “aprendido” mediante el fraccionamiento del conjunto inicial en subconjuntos basados en una prueba de valor de atributo. Este proceso se repite en cada subconjunto derivado de una manera recursiva. La recursividad termina cuando el subconjunto en un nodo tiene el mismo valor que la variable objetivo, o cuando la partición ya no agrega valor a las predicciones.

En el proceso de aprendizaje, para medir y valorar las predicciones logradas se usan diferentes funciones, siendo las más conocidas y usadas el “Índice Gini”, y la “Ganancia de información”.

Finalmente, al igual que los casos anteriores, esta función necesitará unos datos para su entrenamiento supervisado, y posteriormente se centrará en la predicción.

### 5.3 Método de evaluación

Para evaluar el comportamiento de estos métodos de predicción se pueden disponer de distintos indicadores. No obstante, al tratarse de datos de estado, y no valor, se descartan aquellos como CV (Coeficiente de Varianza), MAPE (porcentaje absoluto medio de error), dado que para estos se necesita un dato de potencia predicho.

Es por ello que aquellos que la evaluación se llevará a cabo, tal y como lo realiza Kaustav Basu y otros, 2012. [10] mediante la precisión (AC), definida como:

$$AC = \frac{a + d}{a + b + c + d} \quad (14)$$

Y el siguiente indicador es TP, (Verdadero Positivo), definido como

$$TP = \frac{d}{c + d} \quad (15)$$

Donde

- $a$  es el número de predicciones correctas donde un caso es negativo
- $b$  es el número de predicciones incorrectas donde un caso es positivo
- $c$  es el número de predicciones incorrectas donde un caso es negativo
- $d$  es el número de predicciones correctas donde un caso es positivo

Se considera un caso como negativo cuando su valor es 0, y positivo cuando es de 1 o 2.

## **CAPÍTULO 6. SIMULACIONES Y RESULTADOS**

### **6.1 Escenarios cálculo y determinación de variables**

Dada la gran cantidad de datos y variables de las que se disponen, es necesario determinar los escenarios de cálculo y las combinaciones correctas de éstas para así poder correlacionar su influencia en el rendimiento del programa.

Una opción sería ejecutar todas las combinaciones posibles y así encontrar la óptima, no obstante, esto supone tiempos de cálculos de meses, por lo que queda descartado.

Las predicciones se realizarán para el frigorífico, dado que es el más adecuado para poder evaluar el programa, debido a que, aunque tiene cierta periodicidad, está presente una componente aleatoria en su comportamiento. Además, presenta consumos a lo largo de todo el día, siendo este tipo de comportamiento el de más difícil predicción.

También se presentarán algunas simulaciones para otros electrodomésticos, pero en menos detalle.

El horizonte de predicción del programa es de 1 día con actualización horaria. Esto quiere decir que cada hora éste se tiene que ejecutar para utilizar los datos reales de consumo anterior, y no sus anteriores predicciones. Esto aporta mucha fiabilidad al sistema, dado que un error en una predicción se verá magnificado en gran medida.

Por lo que respecta a la nevera, se ha establecido como escenario de cálculo el día 29 de enero.



## Desarrollo, optimización e implementación de un algoritmo de predicción de consumo de energía de una vivienda

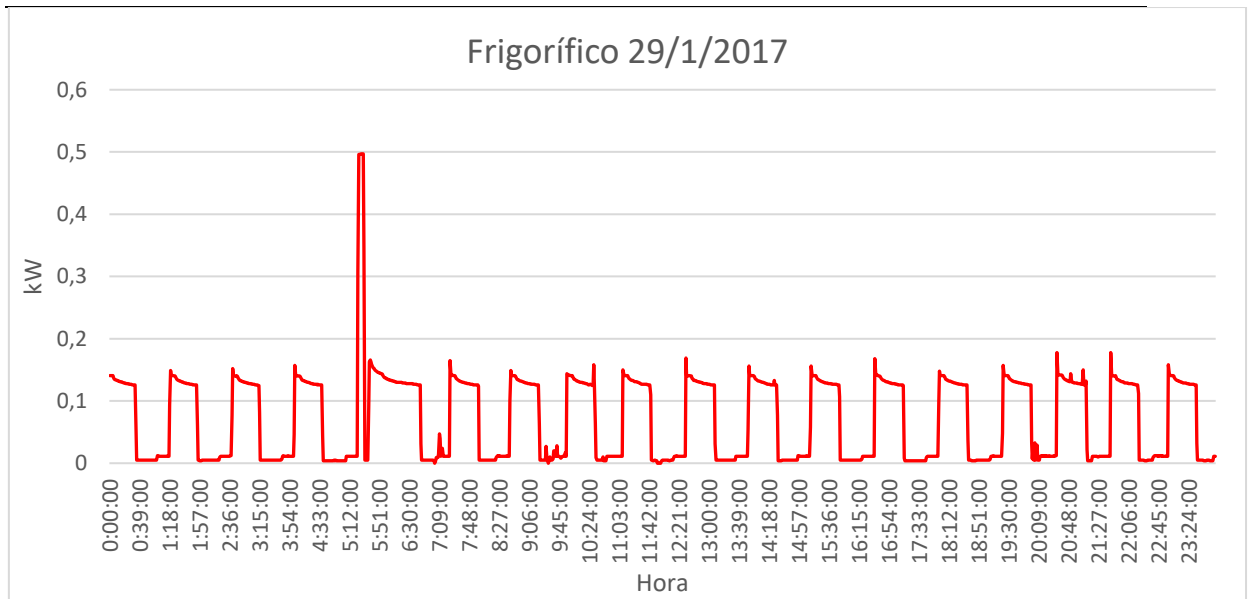


Fig. 10: Consumo frigorífico 29/1/2019

Tal y como se ha explicado en la distribución de datos, es necesario determinar las variables  $n$  y  $P$ , y posteriormente evaluar el comportamiento del programa según se utilicen o no los datos complementarios (“time” y “valores”).

Dado que  $n$  y  $P$  pueden abarcar un amplio abanico de posibilidades (según las horas que se deseen) se tienen miles de combinaciones posibles, el cálculo de las cuales no se puede realizar en un periodo razonable de tiempo. Es por ello por lo que el método de búsqueda de fuerza bruta es descartado, y en su defecto se propone un sistema simplificado de determinación de escenarios.

Por lo tanto, se han establecido 3 escenarios característicos de ese electrodoméstico a lo largo del día (sin utilizar ninguno del día utilizado para predicción, sino momentos anteriores), y se evaluará el comportamiento del programa prediciendo una hora con algunas combinaciones de  $n$ ,  $P$  y las variables.

A continuación, se determina el óptimo para cada método, y con ello se procederá a la predicción del día objetivo.

La variable  $P$ , es decir, el número de datos que se va a analizar está determinada en una semana en el programa del ITE. No obstante, en diferentes pruebas realizadas con el predictor se ha comprobado que esta información puede ser más escasa, dado que facilita e incluso aumenta la precisión final. Por lo tanto,  $P$  comprenderá el rango {24, 48, 72 ... 168} horas, debido a que es necesario conocer el comportamiento del usuario como mínimo durante un día.

Por lo que representa a la  $n$ , es decir, dentro de cada tiempo  $P$ , el número de minutos pasados que se valoran, comprende de {1 – 6} horas. La razón de esto es el elevado tiempo de cálculo si  $n$  es elevada, lo que conlleva al programa a necesitar más tiempo para calcular que el que necesita predecir, y porque durante la fase de desarrollo se obtuvo que generalmente valores de  $n$  reducidos otorgan mejores resultados.

Además, también se quiere evaluar la influencia de las variables “time” y “valores” en la predicción, por lo que cada combinación anterior se realiza con y sin estas variables.

En resumen, la cantidad de pruebas a realizar, para cada modo de predicción, son combinaciones de:

$$P=\{24 - 168\}$$

$$n =\{1 - 6\}$$

$$\text{Time}=\{Si, No\}$$

$$\text{Valores}=\{Si, No\}$$

## 6.2 Escenarios determinación

Los escenarios seleccionados para poder obtener las combinaciones óptimas son:

**6.2.1 Escenario 1.** Este empieza en un periodo de no consumo, y en un momento determinado se enciende el aparato.

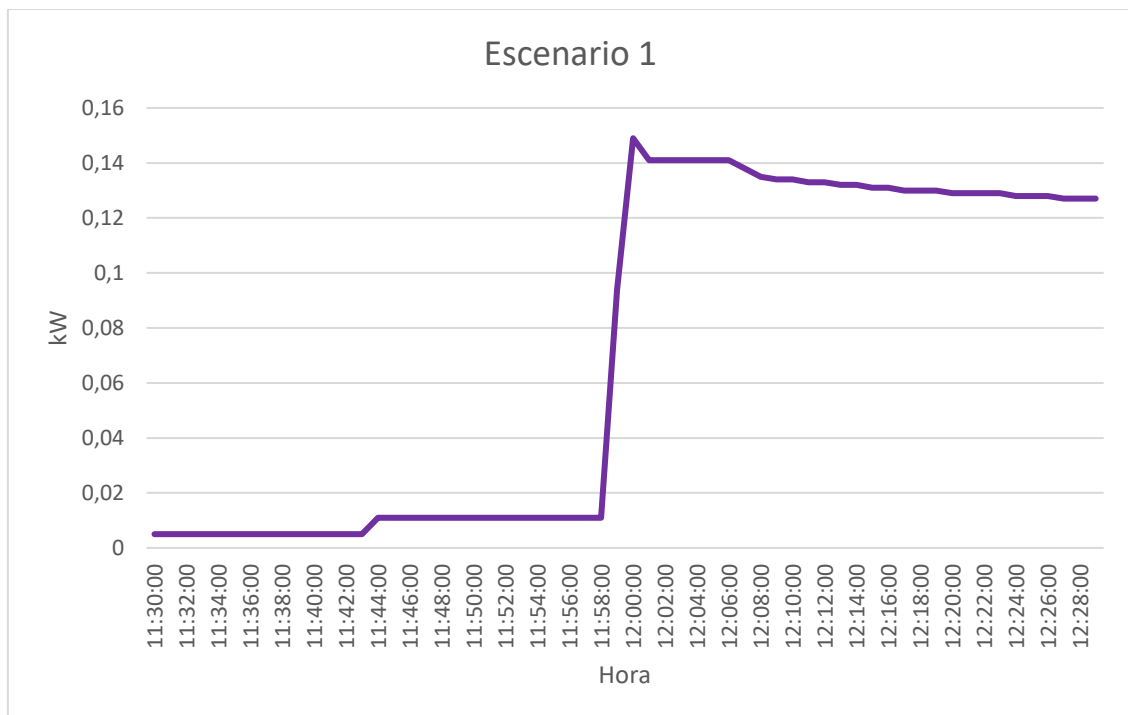


Fig. 11: Escenario de consumo 1 frigorífico

**6.2.2 Escenario 2.** Aquí se tiene consumo inicial y en un momento dado el aparato cambia a estado off.

Desarrollo, optimización e implementación de un algoritmo de predicción de consumo de energía de una vivienda

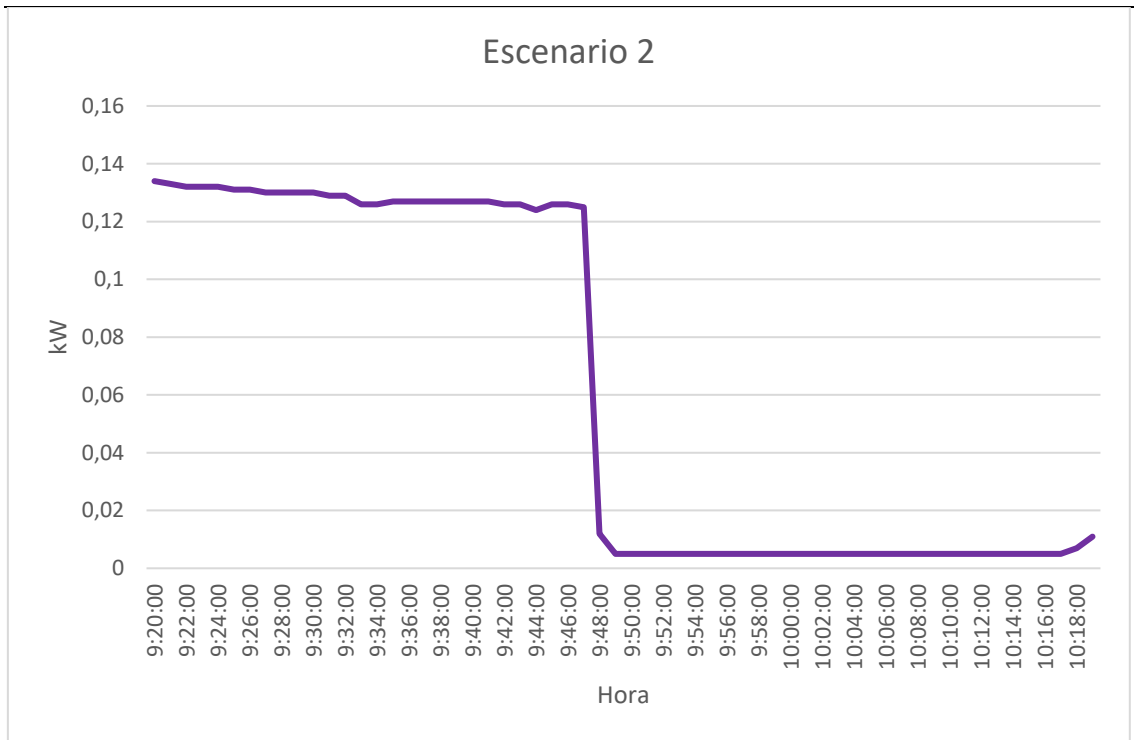


Fig. 11: Escenario de consumo 2 frigorífico

**6.2.3 Escenario 3.** Este representa picos periódicos de alto consumo del aparato, significando algún modo especial de funcionamiento de éste. Tiene un comportamiento ligeramente aleatorio, por lo que su influencia en las variables es interesante.

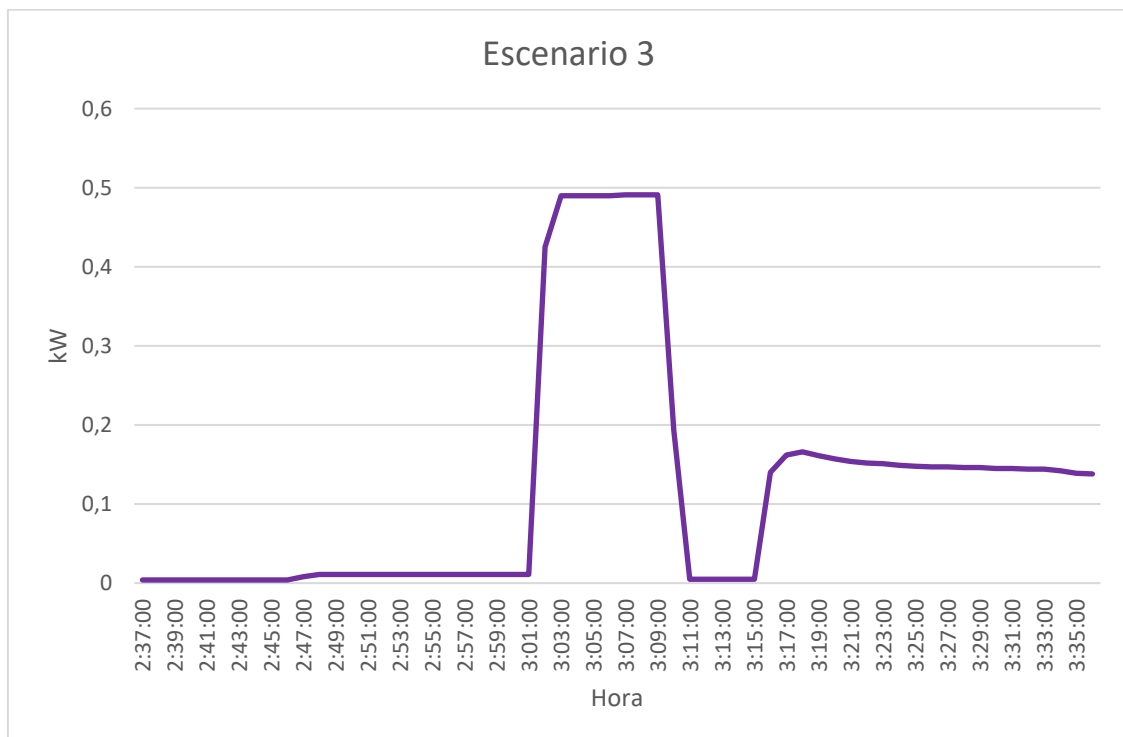


Fig. 12: Escenario de consumo 3 frigorífico

## 6.3 Resultados I

**6.3.1 Regresión Logística** Los resultados detallados del método de predicción Regresión Logística en función de las variables n, P, “time” y “valores” se pueden consultar en el Anejo 1 (al final del documento).

Con estos, se puede determinar n y P óptima, y permite realizar la predicción según las variables.

En este caso, prescindiendo de las variables “time” y “valores” la combinación que mejores resultados ha obtenido es una n de 2 horas y una P de 48 horas (2 días), presentando las siguientes tasas:

	ESCENARIO 1	ESCENARIO 2	ESCENARIO 3
<b>Verdadero Positivo</b>	100 %	96.55 %	81.48 %
<b>Precisión</b>	100 %	98.33 %	78.33 %

Tabla 2: Simulaciones Regresión Logística sin variables

Si se utiliza la variable “time”, la combinación más óptima se repite con una n de 2 horas y una P de 48 horas, y presenta las siguientes tasas.

	ESCENARIO 1	ESCENARIO 2	ESCENARIO 3
<b>Verdadero Positivo</b>	96.77 %	96.55 %	81.48 %
<b>Precisión</b>	96.66 %	98.33 %	78.33 %

Tabla 3: Simulaciones Regresión Logística con variable “time”

Al utilizar tanto las variables “time” como “valores”, obtenemos una n de 2 horas, pero esta vez P puede adquirir un valor de 48 o 144 horas, dando las mismas tasas.

En este caso, se escoge el valor de 48, debido a que se utilizan menos datos y el procesamiento es más rápido.

	ESCENARIO 1	ESCENARIO 2	ESCENARIO 3
<b>Verdadero Positivo</b>	100 %	96.55 %	81.48 %
<b>Precisión</b>	100 %	98.33 %	78.33 %

Tabla 4: Simulaciones Regresión Logística con variable “time” y “valores”

Finalmente, obtenemos para el método Regresión Logística

	n óptima (horas)	P óptima (horas)
<b>Sin variables</b>	2	48
<b>“time”</b>	2	48
<b>“time” + “estado”</b>	2	48

Tabla 5: n y P óptimas para cada variable en Regresión Logística

**6.2.2 Márkov.** Los resultados detallados del método de predicción Márkov en función de las variables n, P, “time” y “valores” se pueden consultar en el Anejo 1 (al final del documento).

Con estos, se puede determinar n y P óptima, y permite realizar la predicción según las variables.

En este caso, prescindiendo de las variables “time” y “valores” la combinación que mejores resultados ha obtenido es una n de 4 horas y una P de 24 horas (1 día), presentando las siguientes tasas:

	ESCENARIO 1	ESCENARIO 2	ESCENARIO 3
<b>Verdadero Positivo</b>	69.5 %	0 %	67.4 %
<b>Precisión</b>	75 %	53.33 %	66.66 %

Tabla 6: Simulaciones Márkov sin variables

Si se utiliza la variable “time”, la combinación más óptima se repite con una n de 4 horas y una P de 24 horas, y presenta las siguientes tasas.

	ESCENARIO 1	ESCENARIO 2	ESCENARIO 3
<b>Verdadero Positivo</b>	100 %	68.29 %	73.68 %
<b>Precisión</b>	98.33 %	78.33 %	65 %

Tabla 7: Simulaciones Márkov con variable “time”

Finalmente, al utilizar todas las variables, los mejores resultados se obtienen con los mismos valores de n y P, obteniendo las siguientes tasas.

	ESCENARIO 1	ESCENARIO 2	ESCENARIO 3
<b>Verdadero Positivo</b>	100 %	68.29 %	34.61 %
<b>Precisión</b>	98.33 %	78.33 %	36.66 %

Tabla 8: Simulaciones Márkov con variable “time” y “estado”

Así, se obtiene para Márkov:

	n óptima (horas)	P óptima (horas)
<b>Sin variables</b>	4	24
<b>“time”</b>	4	24
<b>“time” + “estado”</b>	4	24

Tabla 9: n y P óptimas para cada variable en Márkov

**6.2.3 Árbol de decisiones.** Los resultados detallados del método de predicción Árbol de decisiones en función de las variables n, P, “time” y “valores” se pueden consultar en el Anejo XX (al final del documento).

Desarrollo, optimización e implementación de un algoritmo de predicción de consumo de energía de una vivienda

Estos demuestran una influencia nula de las variables, dado que para cualquier combinación sus resultados son los mismos para cada escenario.

	ESCENARIO 1	ESCENARIO 2	ESCENARIO 3
<b>Verdadero Positivo</b>	0 %	46.66 %	0 %
<b>Precisión</b>	48.33 %	46.66 %	50 %

Tabla 10: Simulaciones árbol de decisiones

Por lo tanto, no se puede determinar una óptima relación de variables para la predicción.

### 6.3 Resultados II

Una vez determinadas estas variables, se puede proceder a la predicción del día objetivo para cada método.

#### 6.3.1 Regresión Logística

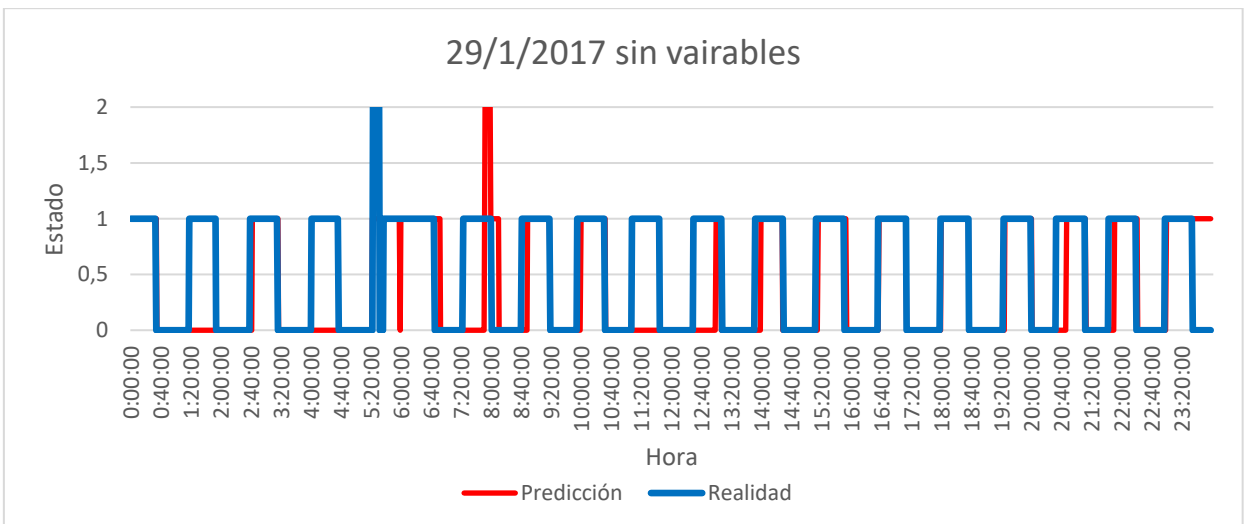


Fig 13: Predicción del día objetivo sin variables Regresión Logística

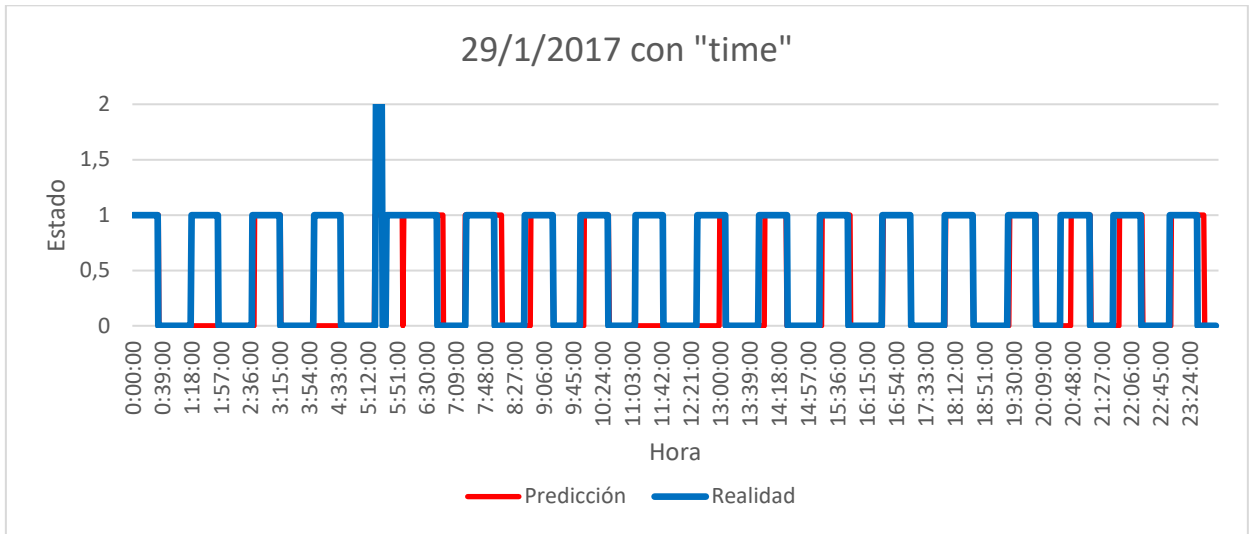


Fig 14: Predicción del día objetivo con variable "time" Regresión Logística

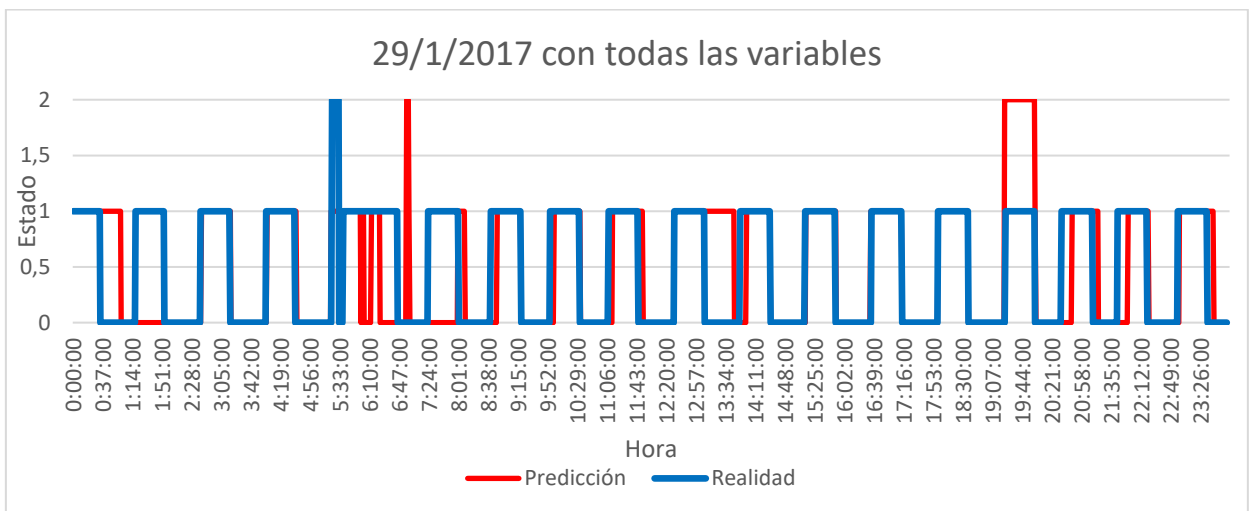


Fig. 15: Predicción del día objetivo con todas las variables Regresión Logística

Por lo que respecta a los indicadores, para cada simulación el resultado es:

	Sin variables	Con variable "time"	Con variables
<b>Precisión</b>	77.2 %	82.6 %	80.3 %
<b>Verdadero Positivo</b>	88 %	92 %	83 %

Tabla 11: Indicadores Regresión Logística para día de estudio



### 6.3.2 Márkov

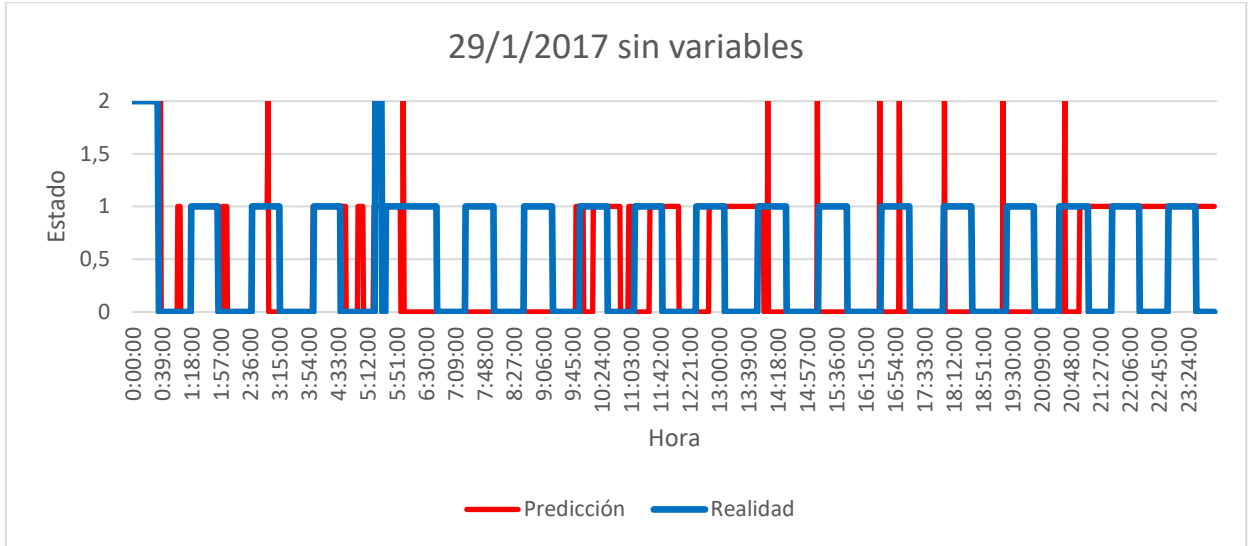


Fig 16: Predicción del día objetivo sin variables Márkov

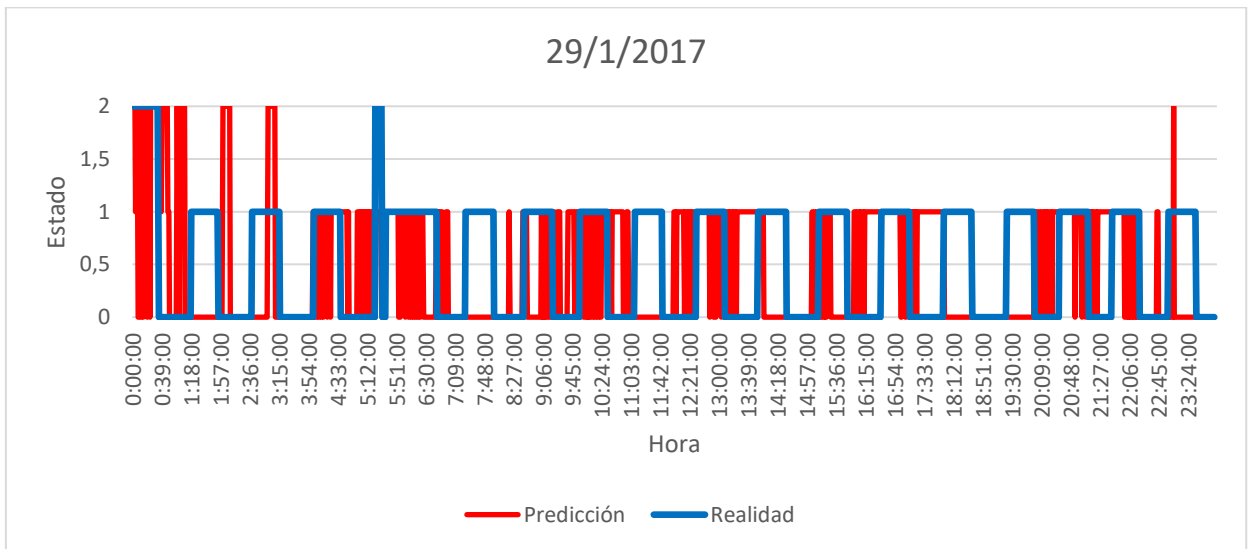


Fig. 17: Predicción del día objetivo con variable "time" Márkov

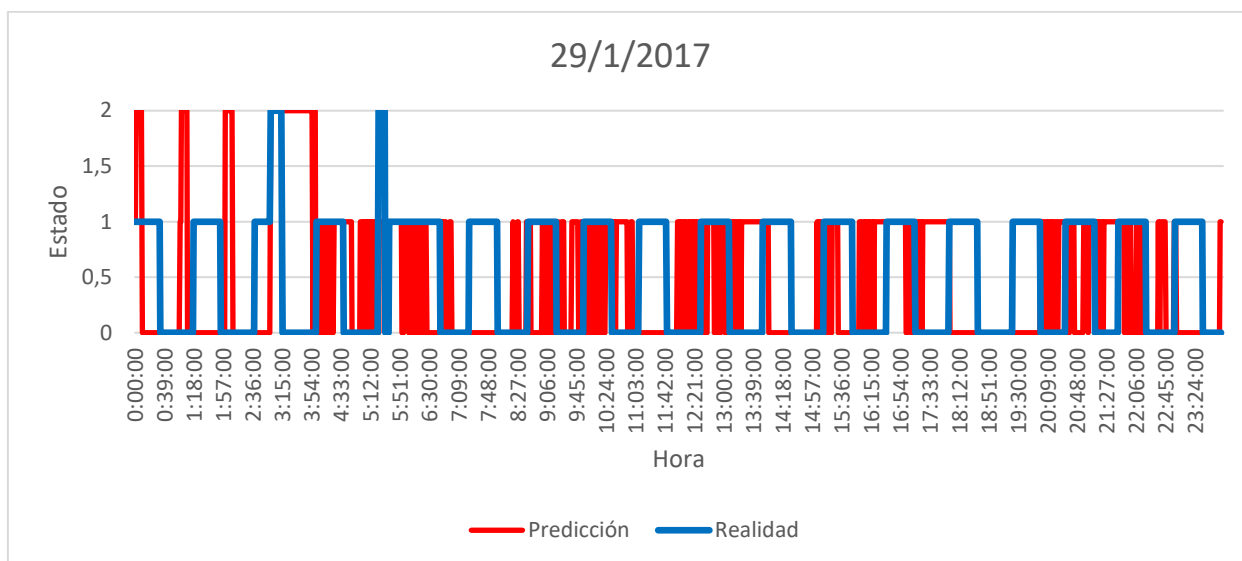


Fig. 18: Predicción del día objetivo con todas las variables Márkov

Por lo que respecta a los indicadores, para cada simulación el resultado es:

	Sin variables	Con variable "time"	Con variables
<b>Precisión</b>	55.9 %	46.6 %	41.8 %
<b>Verdadero Positivo</b>	57.1 %	45 %	39 %

Tabla 11: Indicadores Márkov para día de estudio

**6.3.3 Árbol de decisiones.** A partir de las soluciones obtenidas anteriormente, queda demostrado que Árbol de Decisiones no es un método óptimo para este caso, dado que no es capaz de determinar el cambio de estado del electrodoméstico en cuestión independientemente de las variables utilizadas.

## **7. CONCLUSIONES**

### **7.1 Conclusiones resultados**

A partir de los resultados explicados en los anteriores párrafos se pueden obtener las siguientes conclusiones:

Las elecciones y análisis de escenarios característicos del electrodoméstico es un buen sistema para poder determinar sus variables correctamente. Una de las cuestiones principales del programa desarrollado en el ITE era precisamente ese apartado, la diferente combinación entre variables, la cual no se puede llevar a cabo mediante prueba de fuerza bruta, dado que supone periodos de entrenamiento y predicción del orden de meses o años.

Sin lugar a duda el método de predicción que mejor se comporta de estos tres propuestos es la regresión logística, obteniendo precisiones de más de 80 % en una predicción diaria y un verdadero positivo de 90%. No obstante, este método está muy influenciado por la cantidad de datos P y n que se consideran, dado que el tiempo de cálculo para una P y n alta es considerable (del término de horas).

Cada electrodoméstico tiene unos valores P y n que hay que optimizar, y dado que realizar infinitas combinaciones entre ellas requiere mucho tiempo de cálculo, se deben realizar aproximaciones. En este caso, para el frigorífico, queda demostrado que con una P de 48 horas y una n de 2 es la mejor combinación.

Por lo que respecta a las variables, éstas ayudan a perfilar la predicción y hacerla más precisa, como es el añadir la variable "time". No obstante, añadir más información no supone siempre mejorar las predicciones, dado que, si se tienen en cuenta los "valores", aumenta el tiempo de cálculo y la efectividad del programa disminuye.

El método de cálculo Markov, pese a ser extremadamente rapido de calcular, presenta unos indicadores mejorables, con una precision de 56 % en el caso de no considerar variables.

En este caso, anadirlas tampoco aporta mucha informacion util al programa, dado que su rendimiento disminuye.

Finalmente, queda totalmente descartado el metodo de prediccion de arbol de decisiones, dado que no es capaz de predecir cambios de estado, indiferentemente de la cantidad de variables que se le anade.

## 7.2 Otros electrodomésticos

Se ha escogido el método de regresión logística para implementar, solamente con la variable “time”, y sus resultados con otros electrodomésticos son los siguientes:

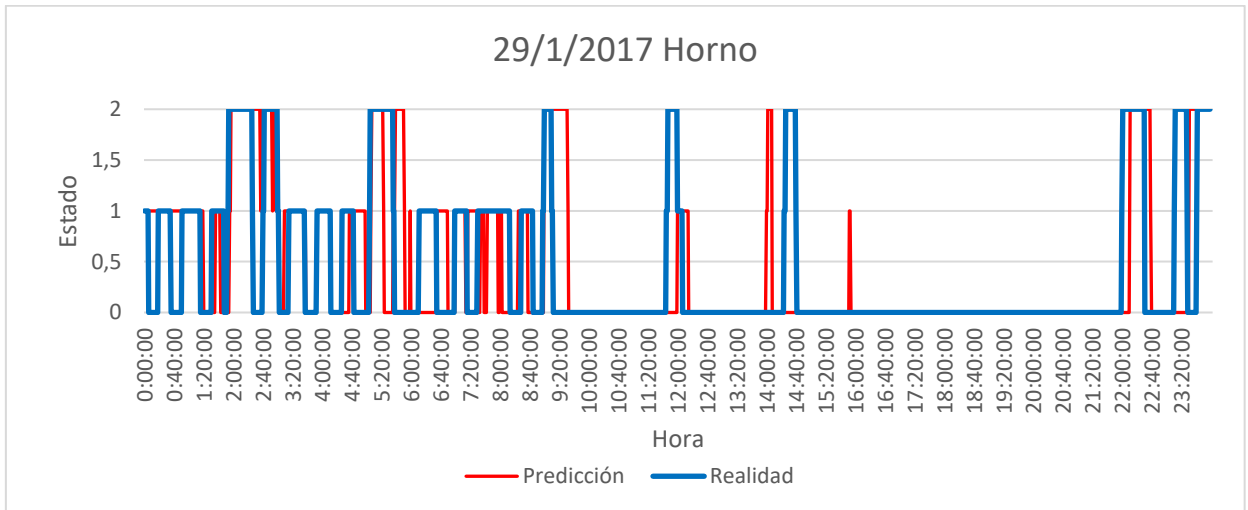


Fig. 19: Predicción del día objetivo para el horno con Regresión Logística

Con el horno el programa presenta una tasa de aciertos de 76.4% y una verdadero positivo de 61% para una P de 24 horas y una n de 2.

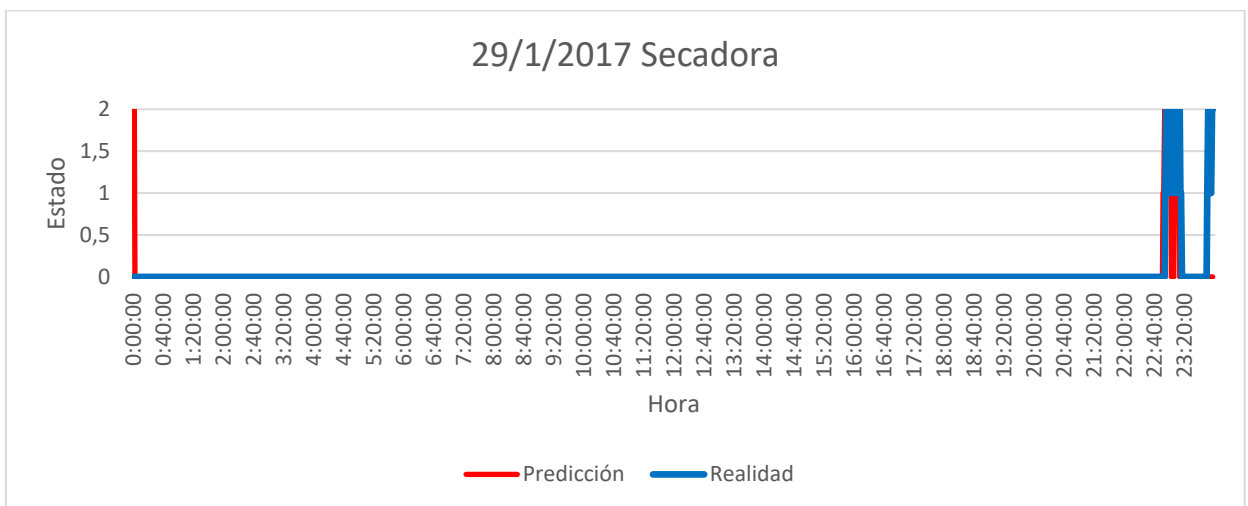


Fig. 19: Predicción del día objetivo para la secadora con Regresión Logística

Para la secadora, es destacable la predicción de no consumo durante el tiempo que esta no se lleva a cabo. Además, presenta una tasa de acierto de 98% con verdadero positivo de 60 %.

### **7.3 Líneas futuras**

Los resultados de este trabajo son bastante satisfactorios, contando con una tasa de aciertos considerablemente alta, que otorga una gran información al usuario.

Las líneas futuras de este proyecto son la incorporación de sus resultados en una base de datos, para así poder utilizar la predicción del estado en otro algoritmo de predicción del valor exacto de ese consumo. De esta manera sirve como elemento para mejorar la gestión de una Smart Grid en distintos proyectos que se están llevando a cabo en el Instituto de Tecnología Energética.

Por otro lado, también se puede continuar y seguir desarrollando modos de análisis de distintas variables. Aquí se han contemplado 4 de ellas, las más significantes, y 3 escenarios típicos del frigorífico y otros electrodomésticos, pero si se disponen de otras bases de datos de consumos, se puede añadir y caracterizarlo en función de las condiciones climatológicas, de la temperatura, de la humedad, o incluso de la previsión de ocupación.

---

**Desarrollo, optimización e  
implementación de un algoritmo de  
predicción de consumo de energía de  
una vivienda**

---

**PRESUPUESTO**

**GRADO EN INGENIERÍA DE LA ENERGÍA**

Razvan Valentin Stoica

## 1 Necesidad de presupuesto

El objetivo de este apartado es el análisis económico del trabajo realizado, dado que tanto la realización del trabajo, como su puesta en funcionamiento ha supuesto y supondrá un gasto económico para el usuario.

## 2 Presentación de presupuesto

El presupuesto contiene los distintos gastos asociados al Trabajo de Fin de Grado.

Primero se analizan y cuantifican las horas necesarias para el aprendizaje, desarrollo y puesta en marcha del programa, incluyendo las distintas pruebas efectuadas. También se incluye el tiempo de uso de licencia en tal programa.

A continuación, aunque el trabajo se ha realizado con una base de datos perteneciente al Instituto Tecnológico de la Energía, se analiza el coste de la implementación de un sistema de adquisición de datos en una vivienda para así poder obtener esa información de manera continua y almacenarla en la nube.

Finalmente, también se incluye un coste asociado al mantenimiento de tal programa e instalaciones.

## 3 Contenido

**3.1 Coste desarrollo.** El alumno Razvan Valentin Stoica ha realizado este trabajo durante sus prácticas de empresa en el Instituto tecnológico de la Energía. Esto corresponde a su periodo de aprendizaje en la empresa, y al desarrollo, puesta en marcha, pruebas y ejecución del programa en cuestión.

En el contrato, el alumno ha realizado prácticas curriculares por un valor de 18 créditos, es decir, a 450 horas, de las cuales se considera un 70 % dedicadas a la realización de tal aplicación.

Esto corresponde a un total de 315 horas, que en el caso de un ingeniero recién titulado se retribuyen a 8.5€ la hora, por lo que el coste económico asociado es de 2677.5€.

Además, si se contabiliza la licencia del programa utilizado, se tiene un coste de 41 € mensuales. El trabajo se ha realizado a lo largo de 6 meses, por lo que el valor total es de 246 €.

**3.2 Coste adquisición datos.** Tal y como se menciona en la memoria, el programa se ha creado a partir de una base de datos de consumos de diferentes electrodomésticos a lo largo de un mes.

Esta base de datos es minutal, y pertenece al ITE. No obstante, el programa se puede adaptar fácilmente a otras bases con otros periodos de muestreo si se necesitara.

En una vivienda, adquirir esa información tiene un coste, que se detalla a continuación:

La toma de datos se puede realizar de diferentes maneras: Una de ellas es mediante pinzas amperimétricas conectadas a todos los aparatos de la vivienda, que deben de estar comunicadas a un PLC y de ahí recibir y tratar y almacenar toda la información.

Esta es una manera profesional de realizarlo, no obstante, si se trata información proveniente de usuarios domésticos, es más factible el uso de soluciones comerciales, como puede ser el caso de enchufes inteligentes, que ya incluyen el software necesario para el tratamiento de datos y destacan por ser fáciles y muy intuitivos para el usuario de a pie.

Sin embargo, a fin de corroborar la información global, también se incluye en este presupuesto el uso de un Smart meter en la cabecera de la instalación.

Por lo que corresponde a los enchufes inteligentes, se incluye el ejemplo del TP-Link HS110, considerado referente en el mercado, que tiene un precio de 35 € por sistema. Por lo tanto, en una vivienda en la que se quieren tener datos del orden de 10 electrodomésticos, supone una inversión de 350 €.

Por lo que respecta a la cabecera de la instalación, el coste de instalar un Smart meter es de 121 €.

A continuación, se requiere de un ordenador que capte toda la información y la almacene en la red. En este presupuesto no se ha contabilizado su precio, debido a que esto puede realizarse con el sistema doméstico.

Finalmente, si se desea almacenar la información generada en la nube, es necesario disponer de una tarifa de espacio, la cual se contabiliza a 10 € mensuales.

**3.3 Coste mantenimiento.** Es importante contar con un mantenimiento, tanto del programa, comprobar periódicamente su correcto funcionamiento, como la observación de los datos adquiridos a lo largo del tiempo.

Es por ello por lo que el coste de mantenimiento se ha asociado a 2 horas semanales de trabajo de un personal cualificado a precio de ingeniero, por lo que supone 8 horas mensuales, un total de 68 €/mes.

Finalmente, en la siguiente tabla se incluye el resumen de los costes justificados en los apartados anteriores:



Desarrollo, optimización e implementación de un algoritmo de predicción de consumo de energía de una vivienda

---

<b>CONCEPTO</b>	<b>COSTE FIJO €</b>	<b>COSTE VARIABLE €/mes</b>
Desarrollo algoritmo	2677.5	x
Licencia programa	246	41
Enchufe inteligente	350	x
Smart meter	121	x
Espacio almacenamiento nube	x	10
Mantenimiento/Ajuste	x	68
Total	3394.5	119

Tabla 12: Resumen del presupuesto

## ANEJO 1

### 1.Resultados simulaciones Regresión Logística

#### 1.1 Escenario 1

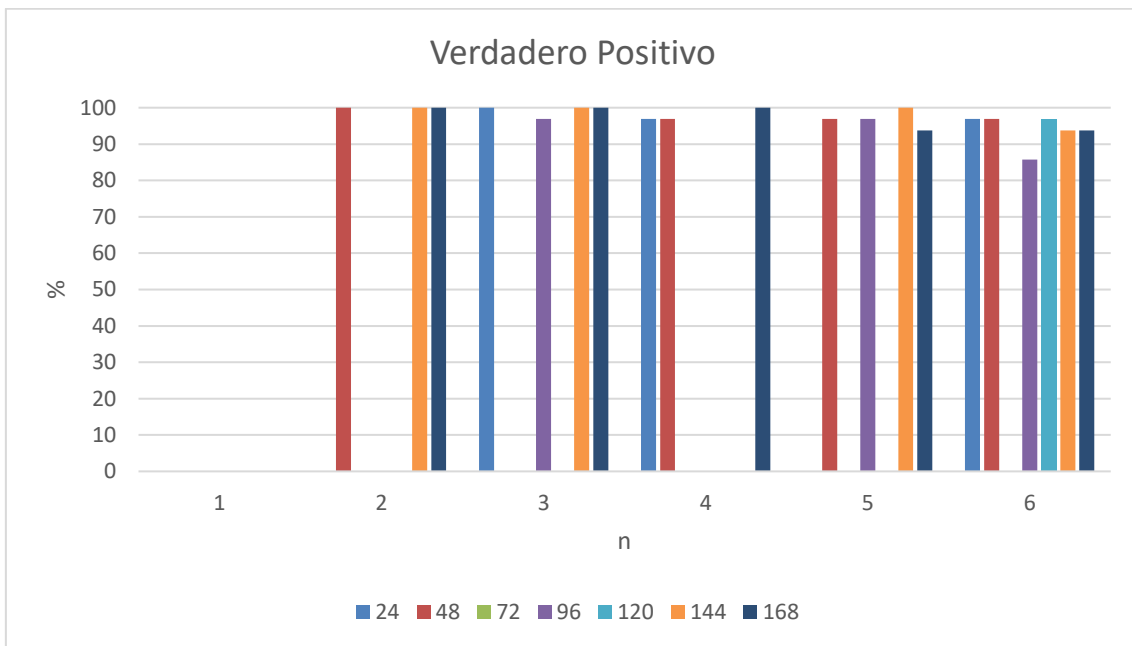


Fig.20: Verdadero Positivo Regresión Logística sin variables

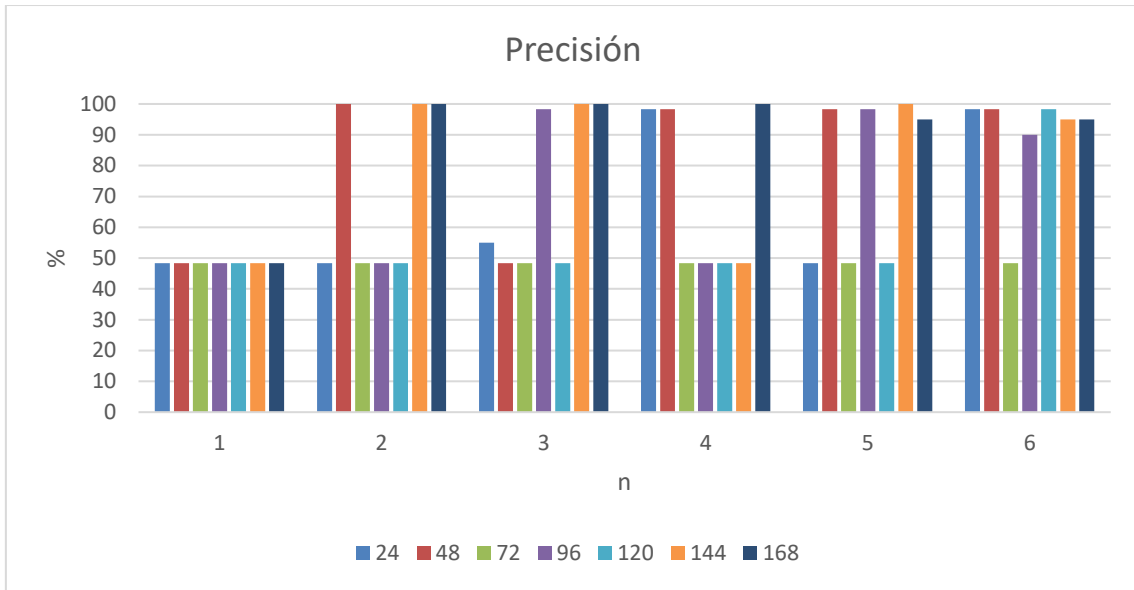


Fig.: 21 Precisión Regresión Logística sin variables

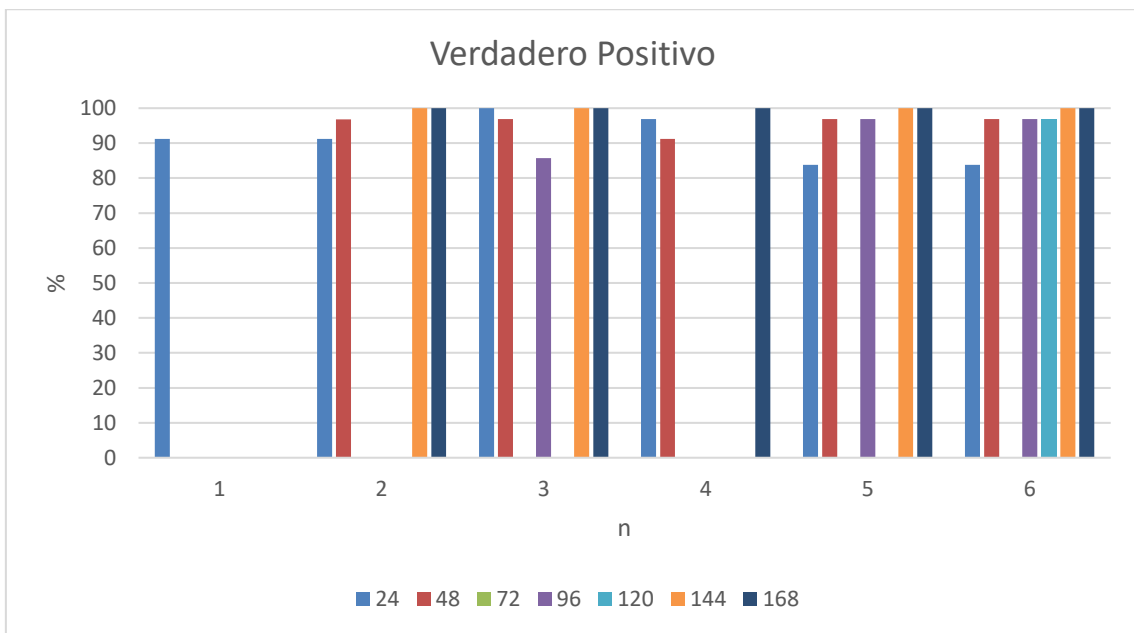


Fig.22: Verdadero Positivo Regresión Logística con "time"

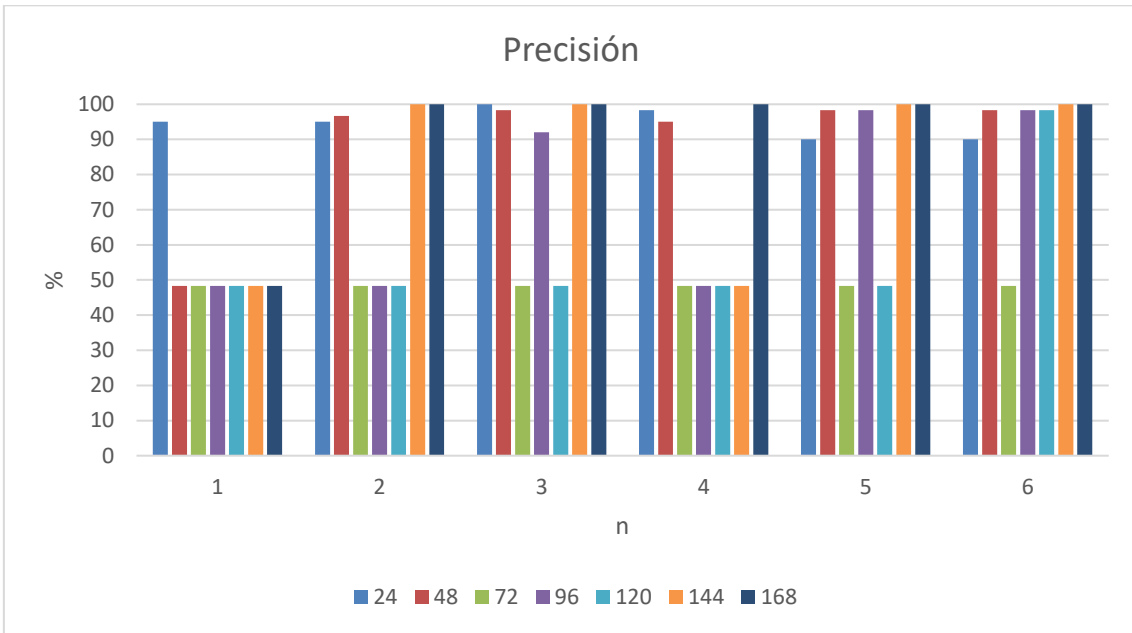


Fig.23: Precisión Regresión Logística con "time"

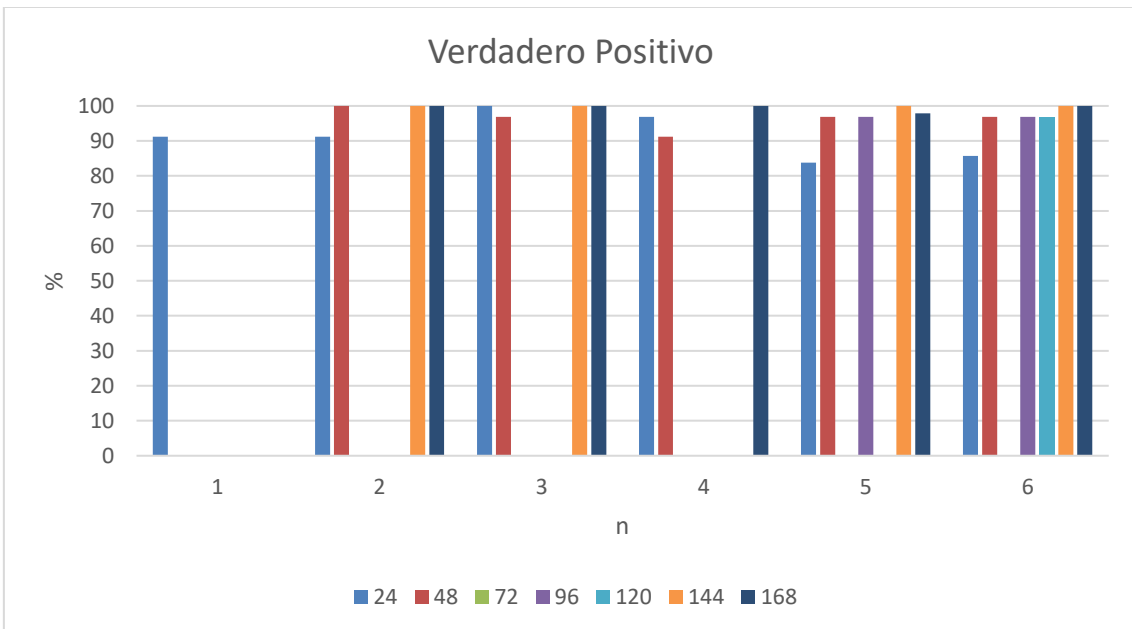


Fig.24: Verdadero Positivo Regresión Logística con "time" y "estado"

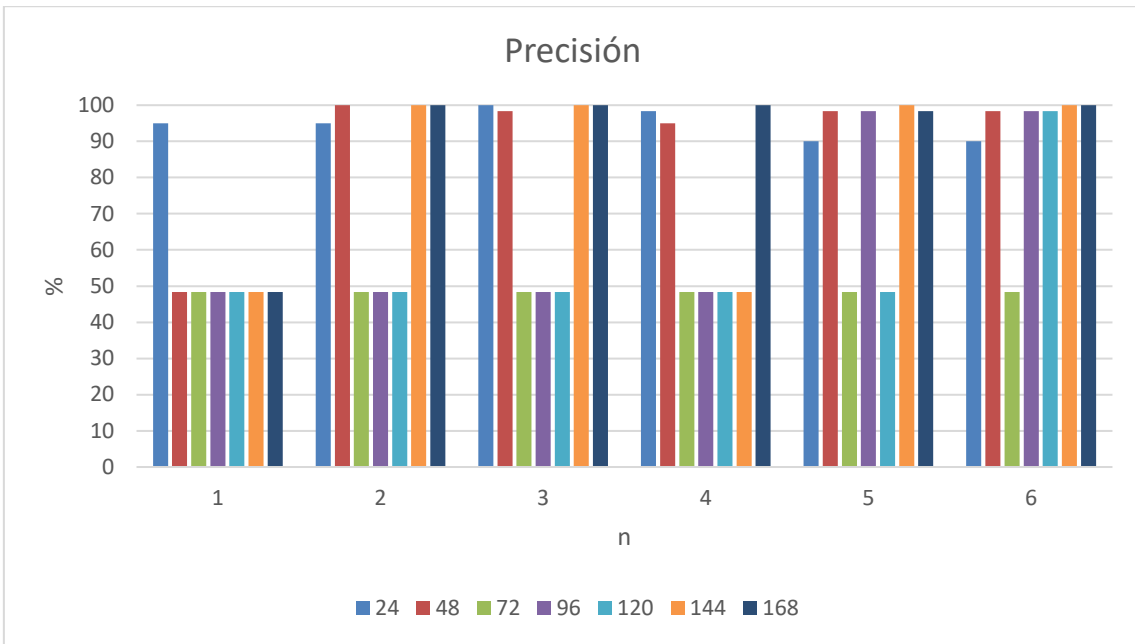


Fig. 25: Precisión Logistic Regression con "time" y "estado"

1.2 Escenario 2

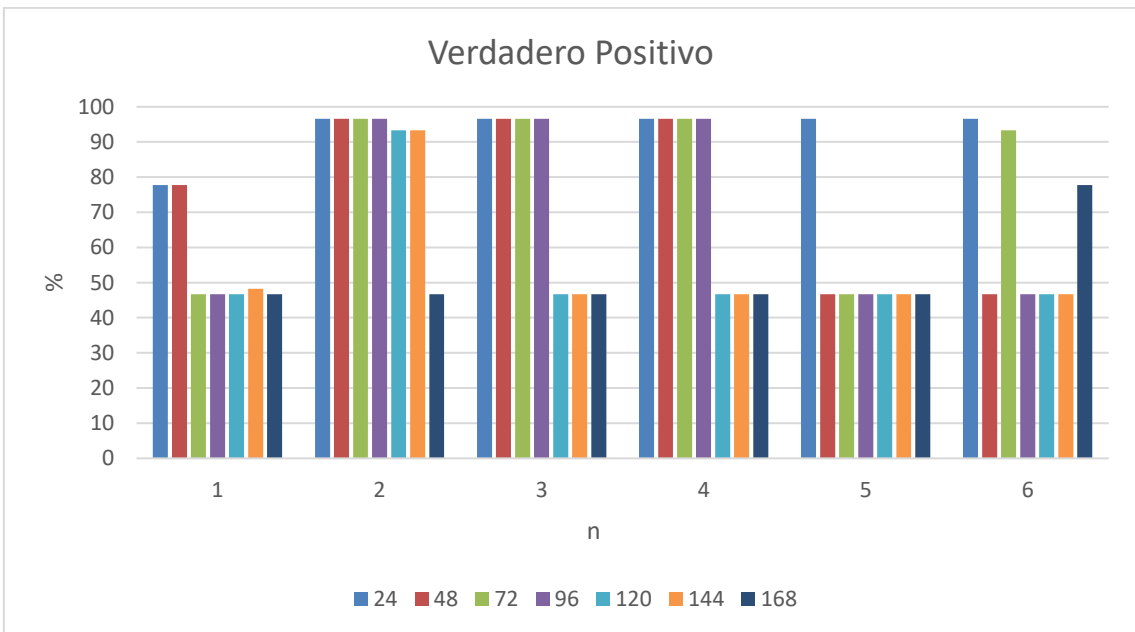


Fig.26: Verdadero Positivo Regresión Logística sin variables

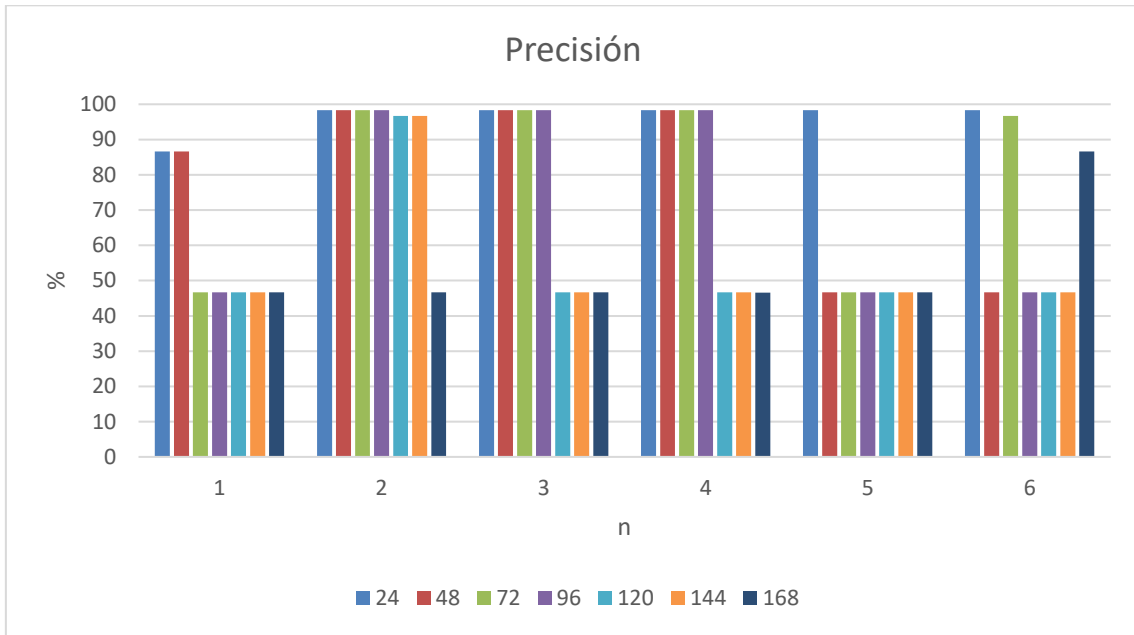


Fig.27: Precisión Regresión Logística sin variables

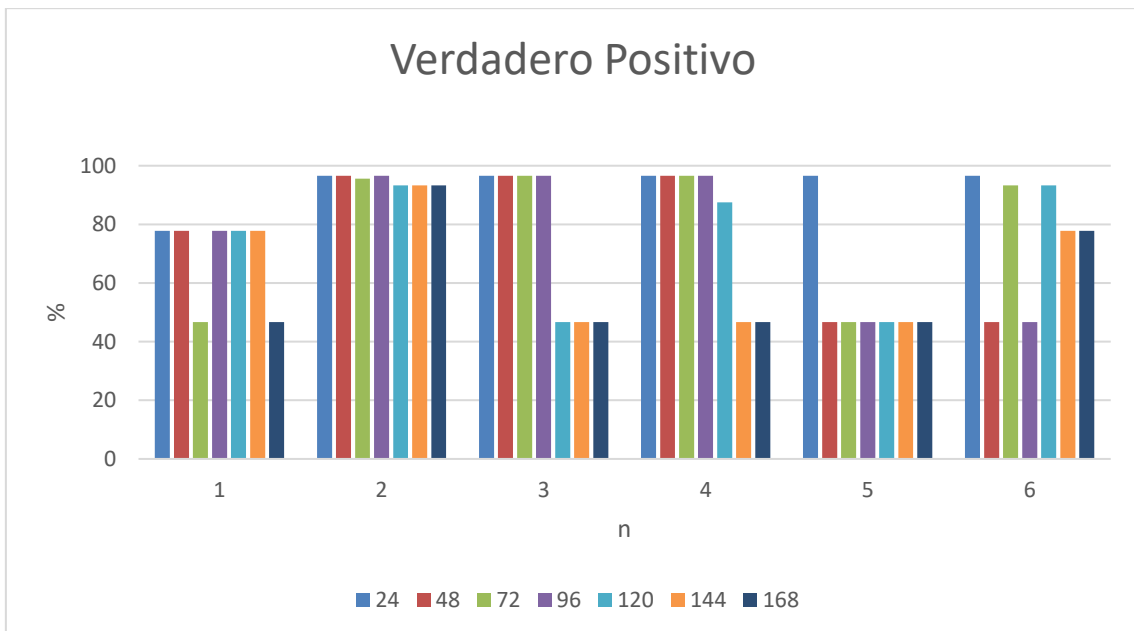


Fig.28: Verdadero Positivo Regresión Logística con "time"

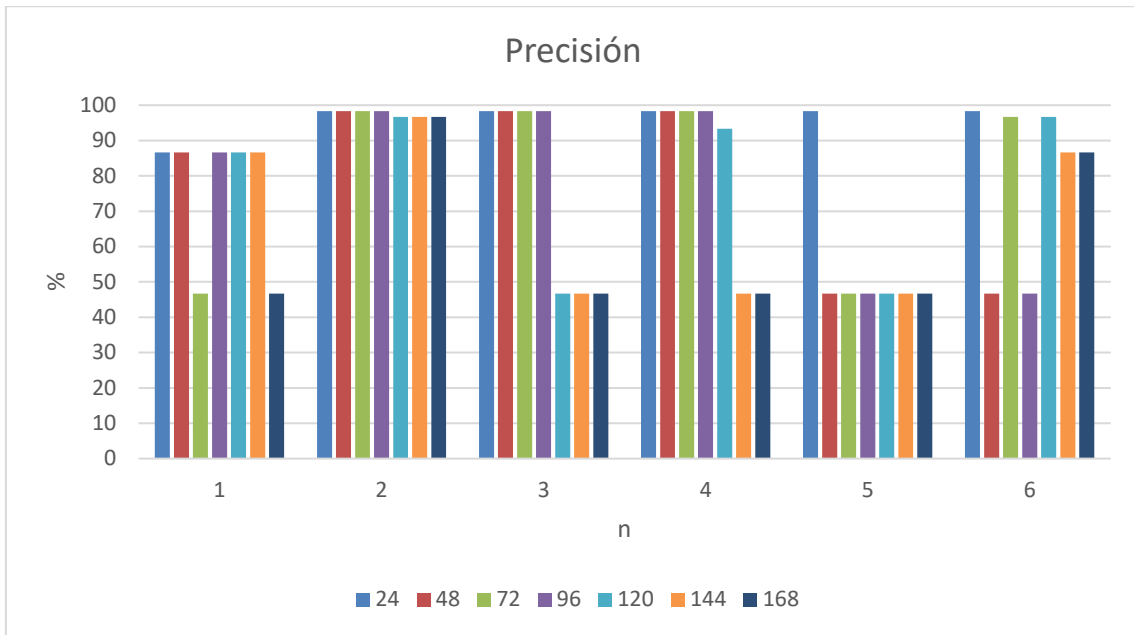


Fig.29: Precisión Regresión Logística con "time"

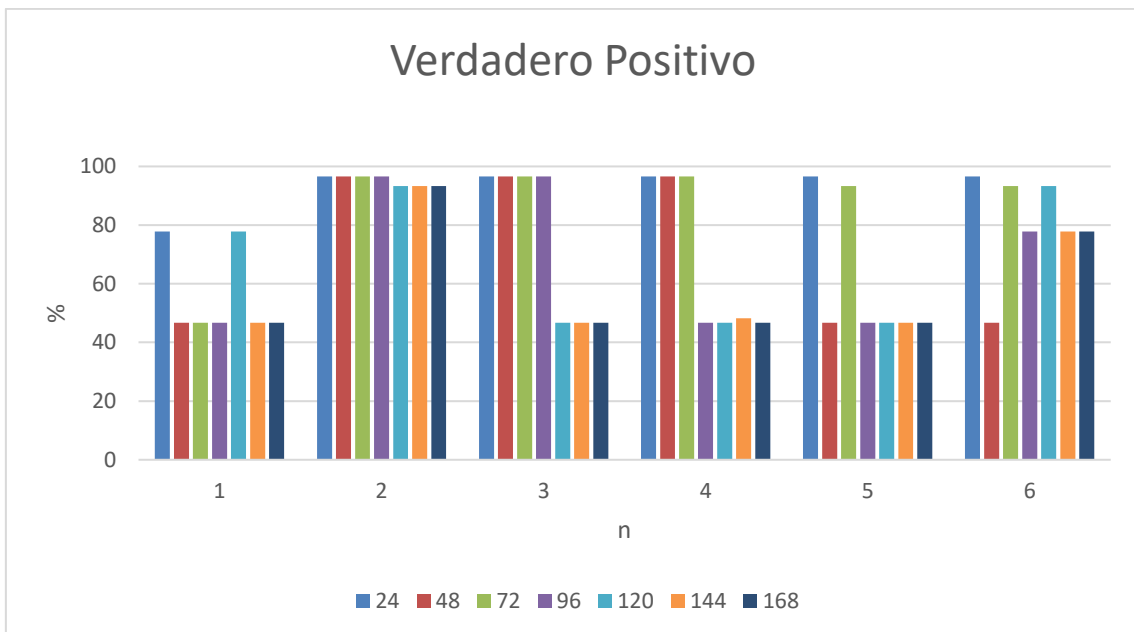


Fig.30: Verdadero Positivo Regresión Logística con "time" y "estado"

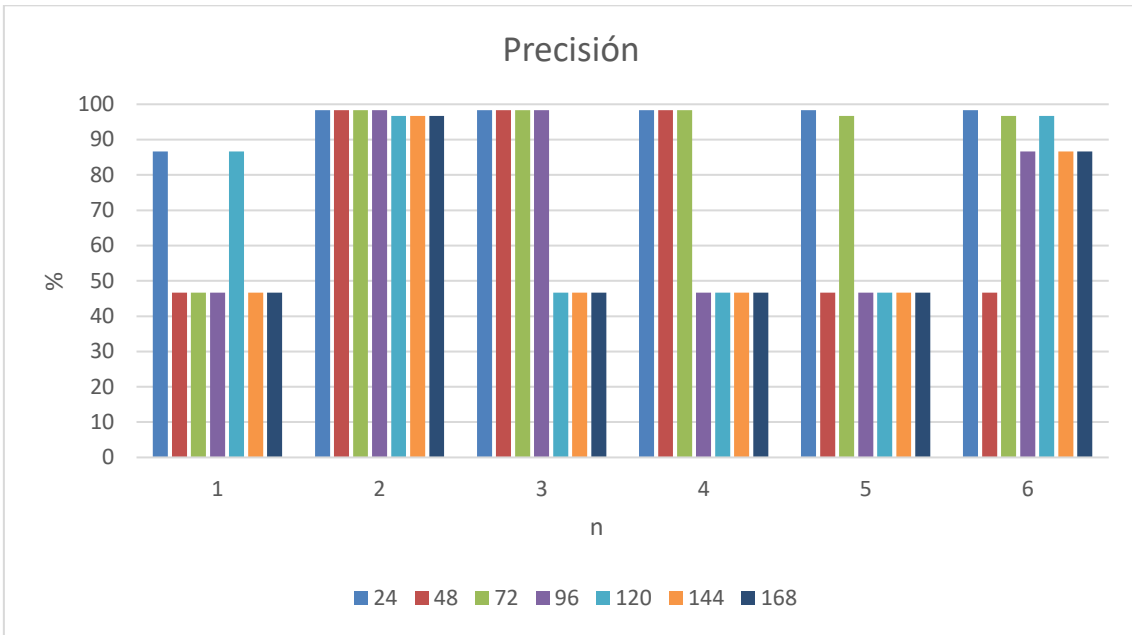


Fig.31: Precisión Regresión Logística con "time" y "estado"

### 1.3 Escenario 3

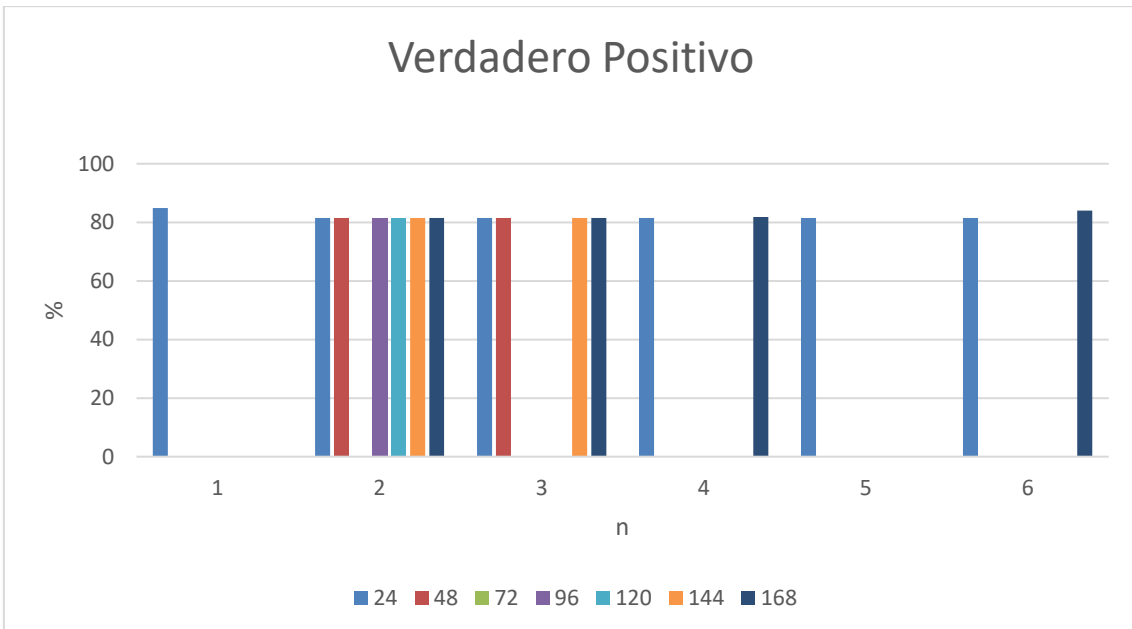


Fig.32: Verdadero Positivo Regresión Logística sin variables



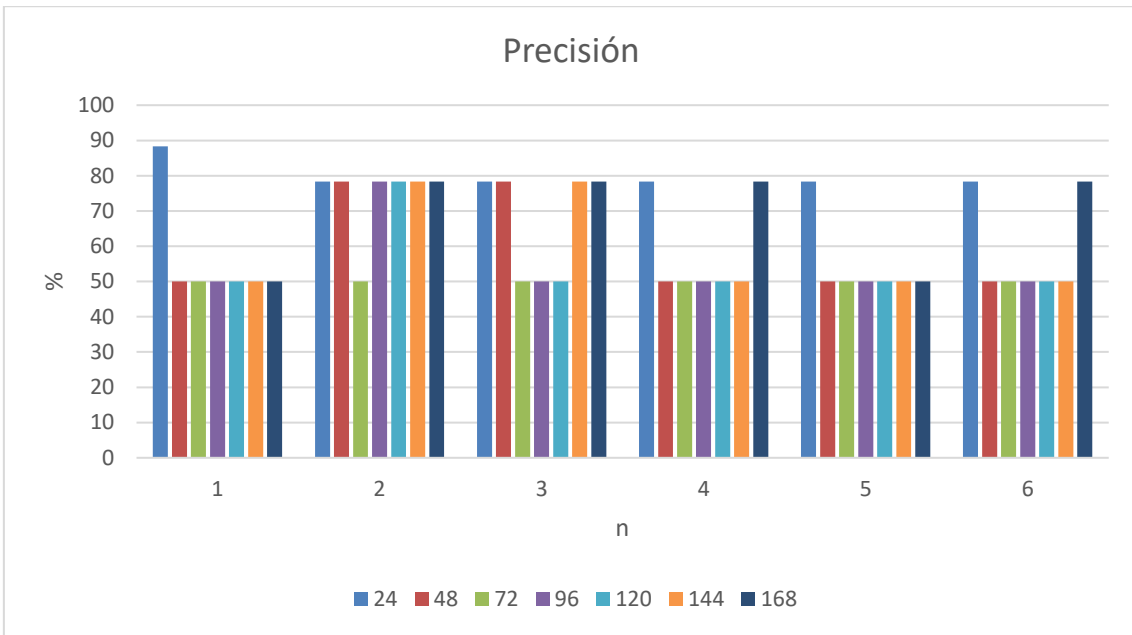


Fig.33: Precisión Regresión Logística sin variables

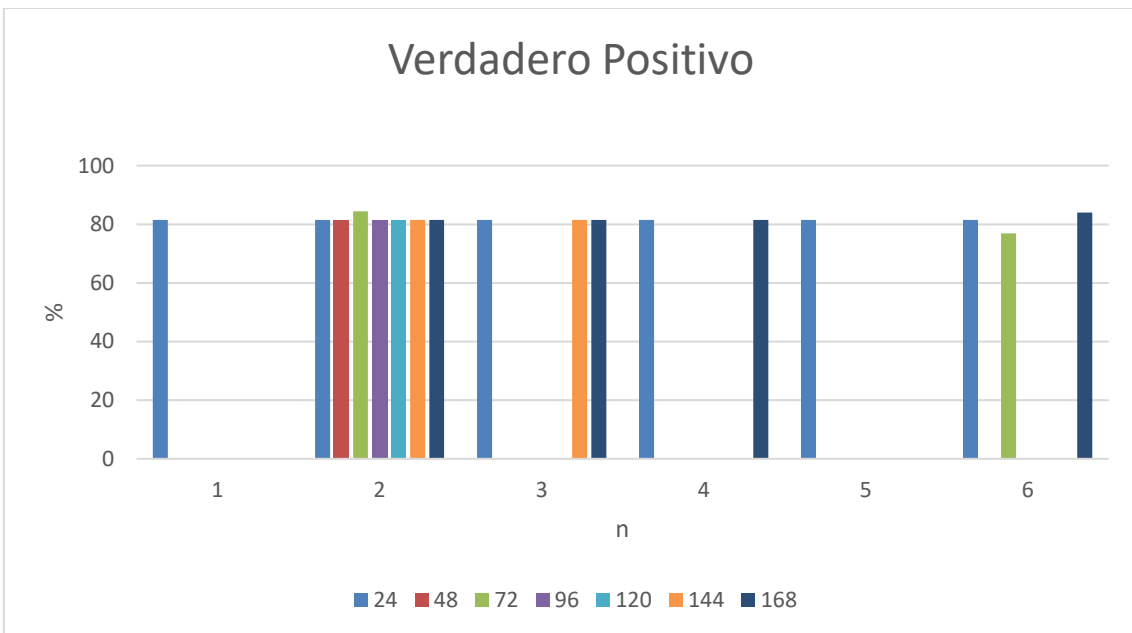


Fig.34: Verdadero Positivo Regresión Logística con "time"

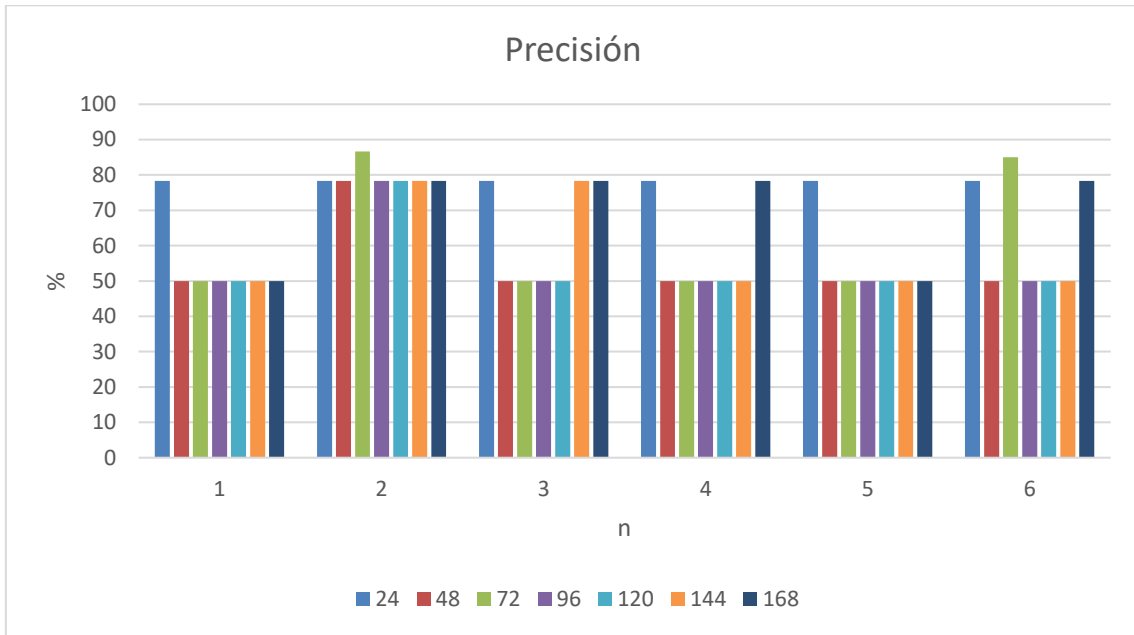


Fig.35: Precisión Regresión Logística con "time"

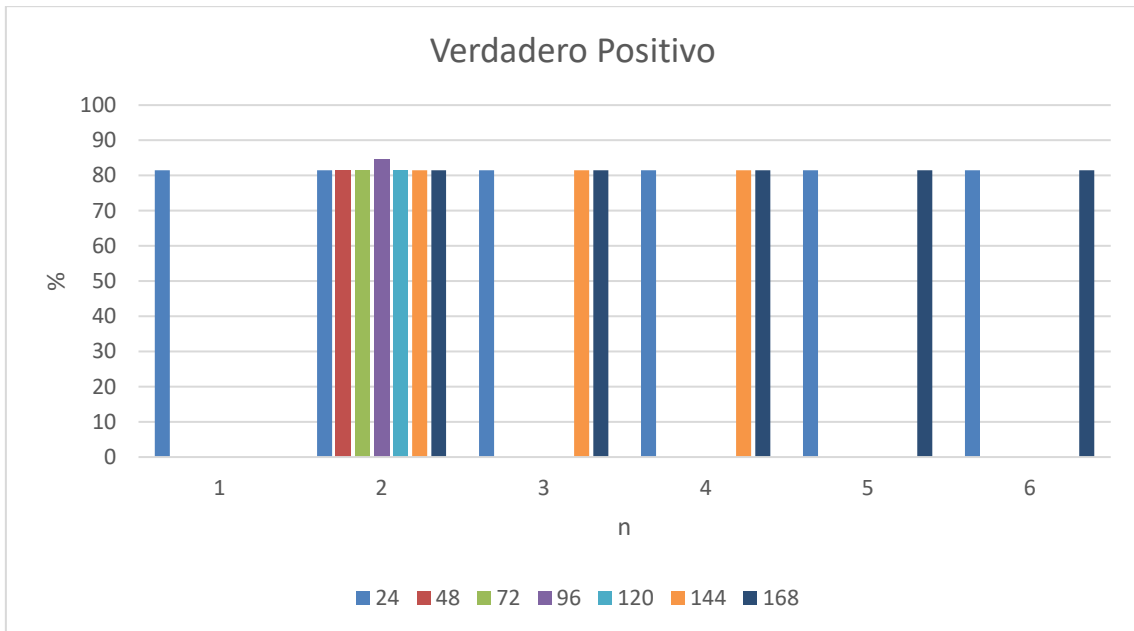


Fig.36: Verdadero Positivo Regresión Logística con "time" y "estado"

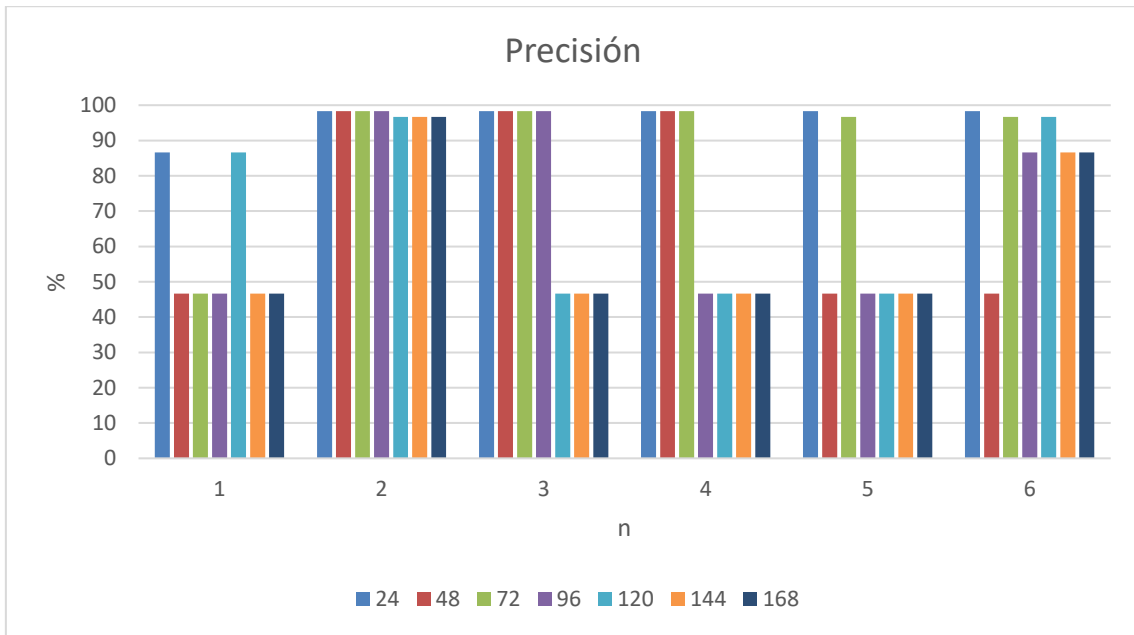


Fig.37: Precisión Regresión Logística con "time" y "estado"

## 2. Resultados simulaciones Márkov

### 2.1 Escenario 1

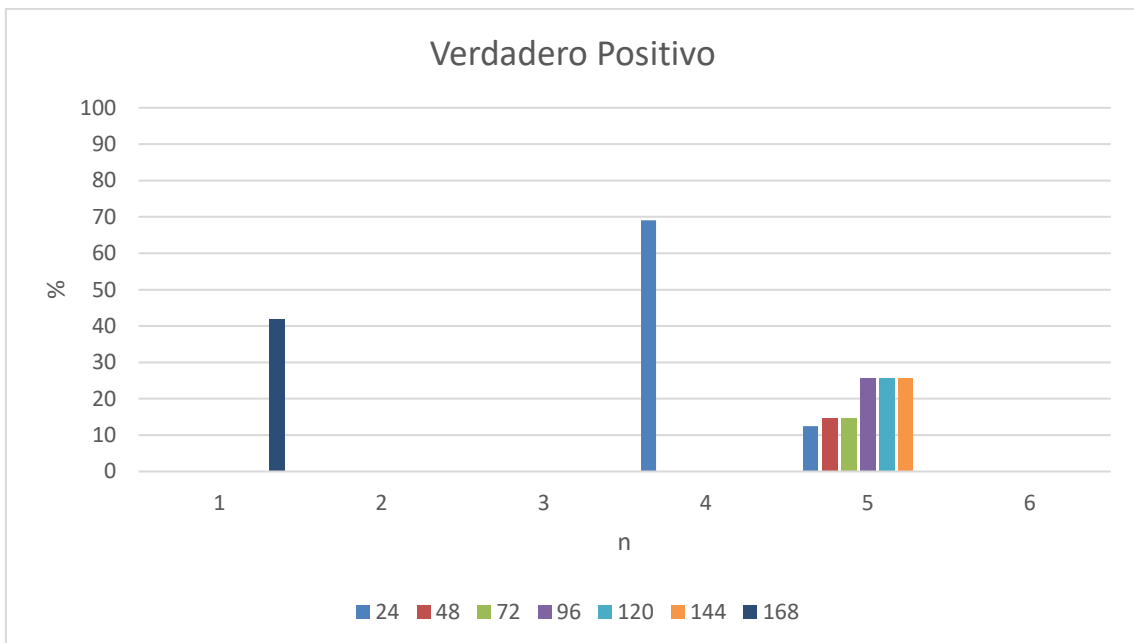


Fig.38: Verdadero Positivo Márkov sin variables

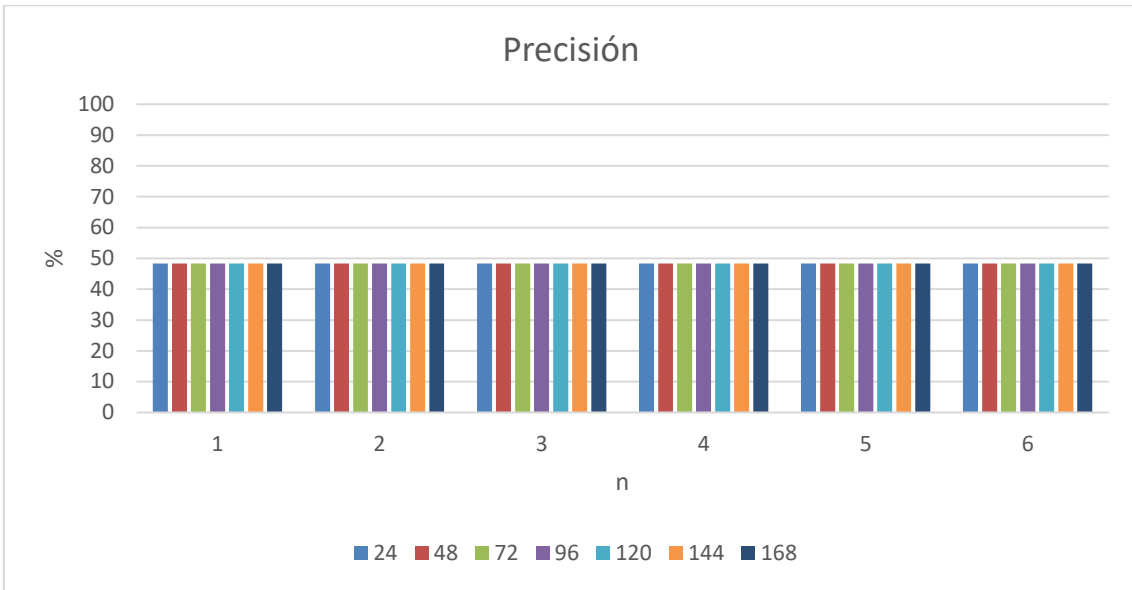


Fig.39: Precisión Márkov sin variables

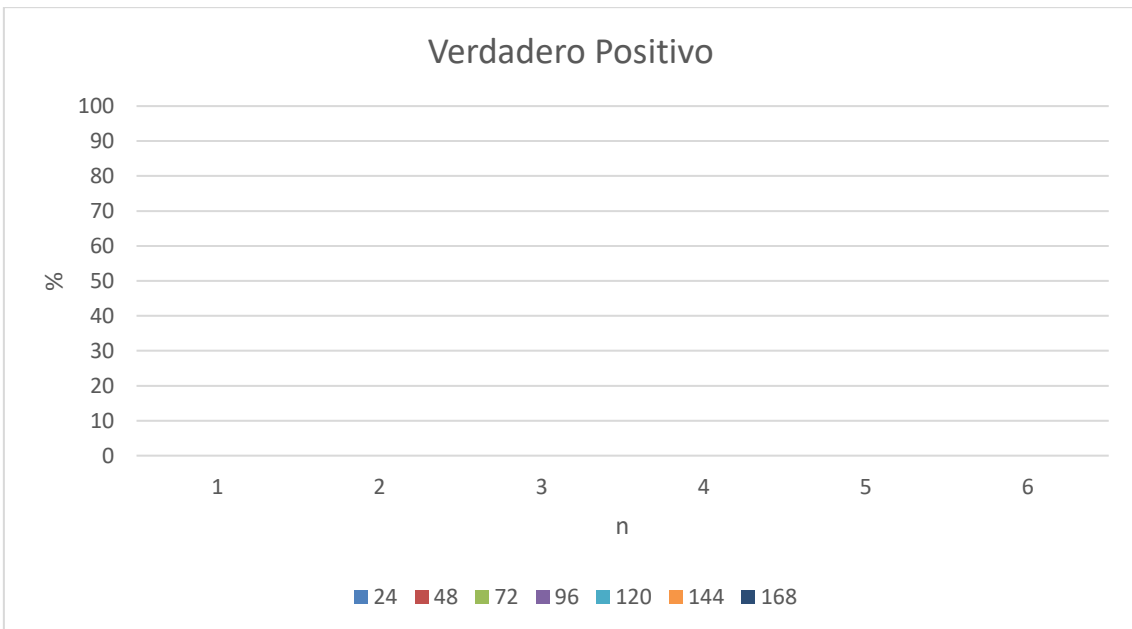


Fig.40: Verdadero Positivo Márkov con "time"

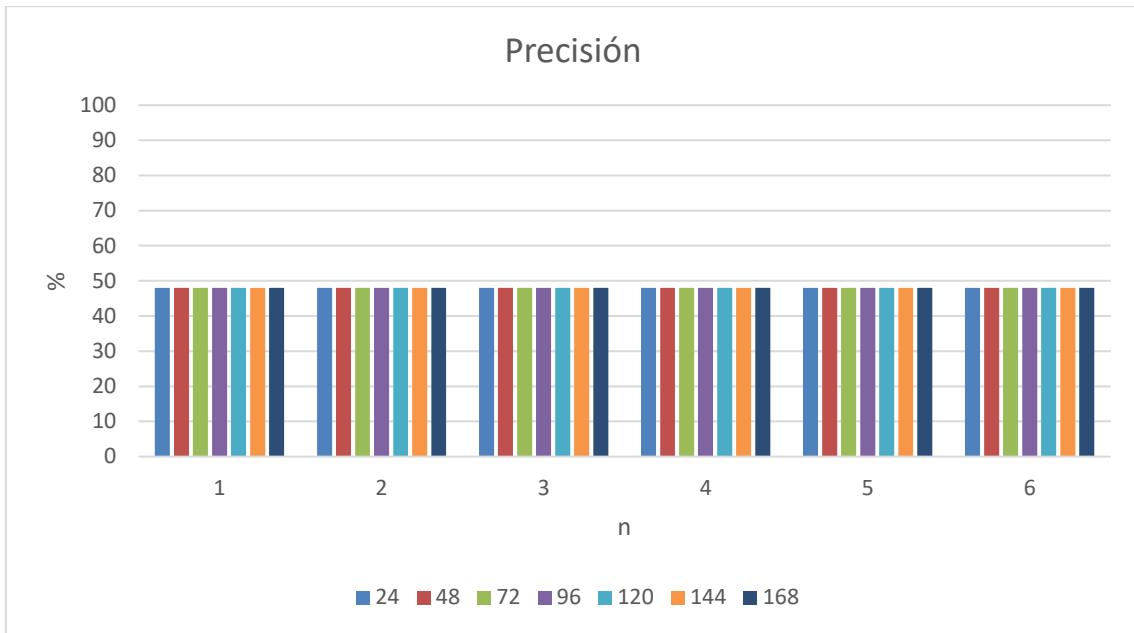


Fig.41: Precisión Márkov con "time"

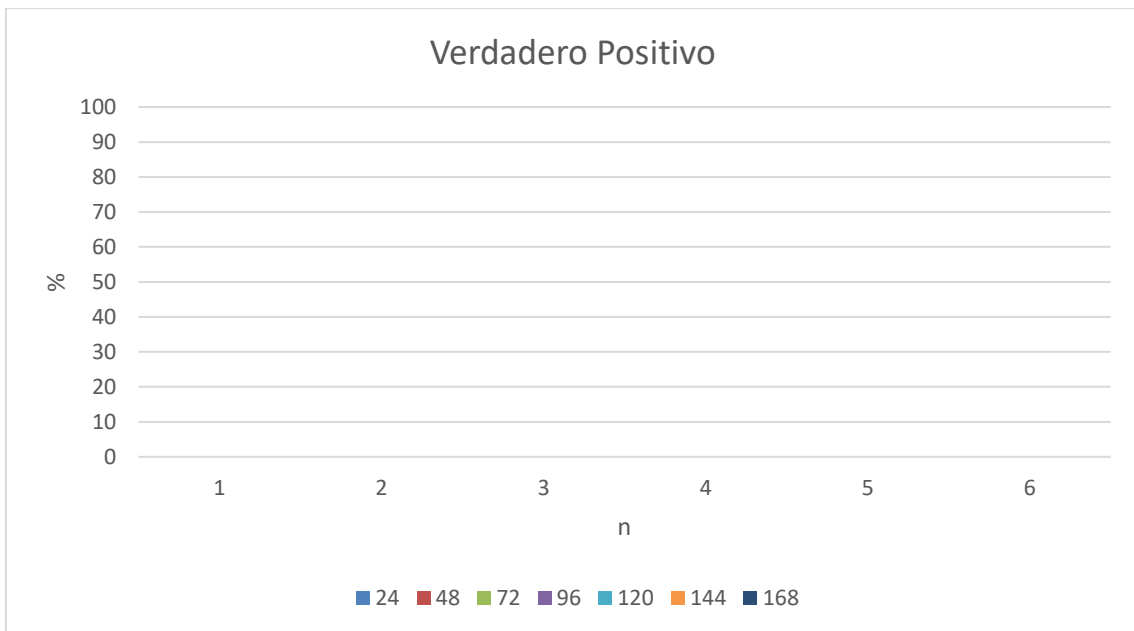


Fig.42: Verdadero Positivo Márkov con "time" y "estado"

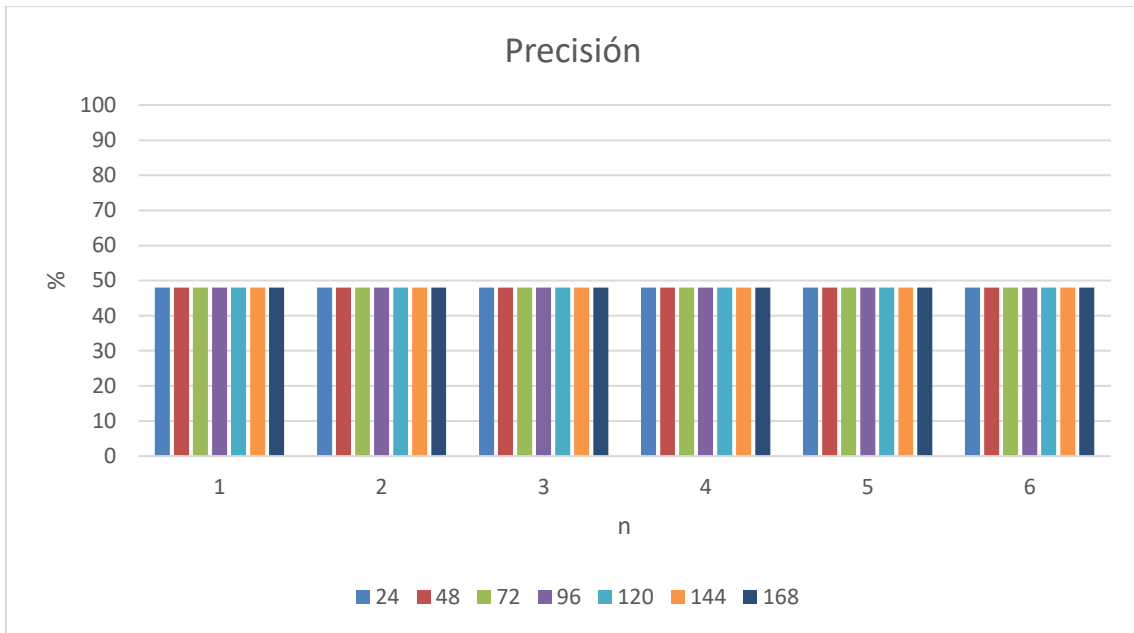


Fig.43: Precisión Márkov con "time" y "estado"

## 2.2 Escenario 2

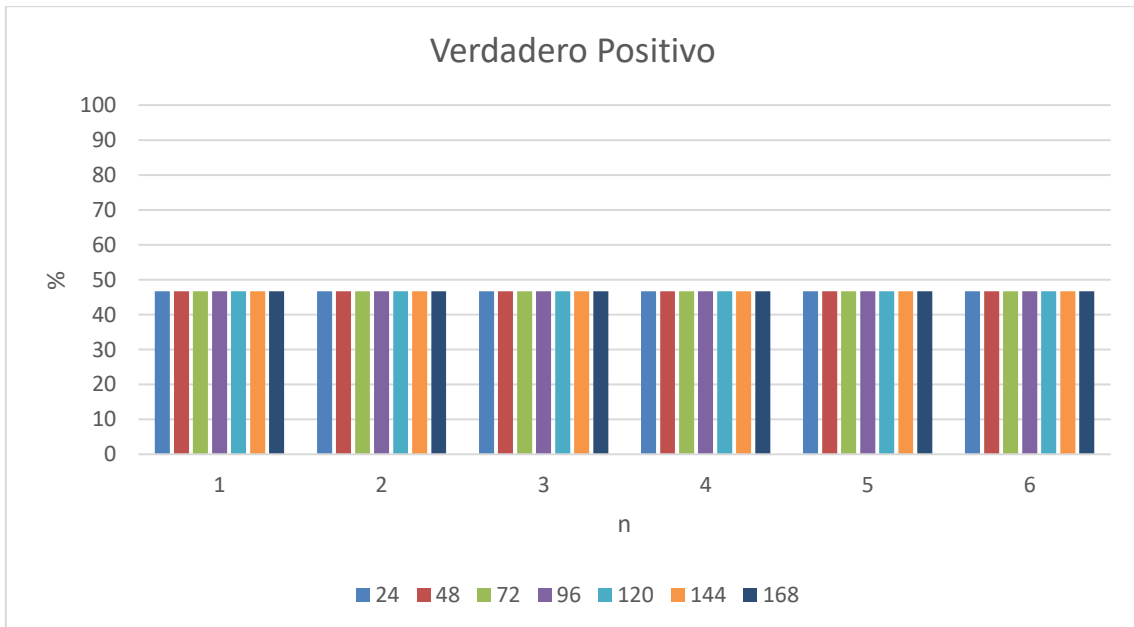


Fig.44: Verdadero Positivo Márkov sin variables

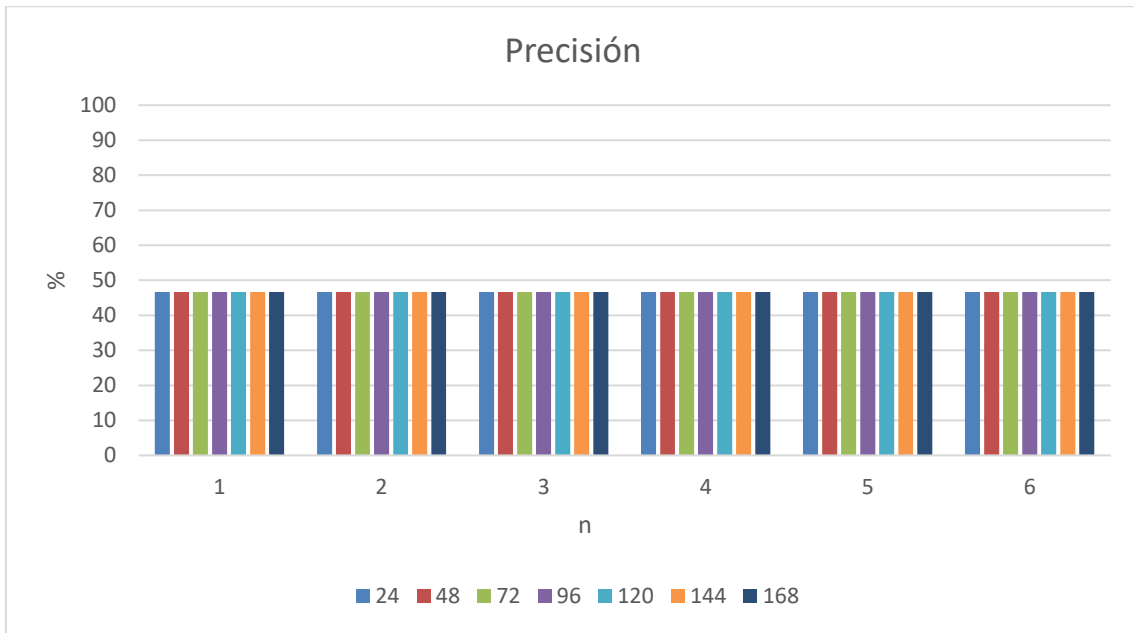


Fig.45: Precisión Márkov sin variables

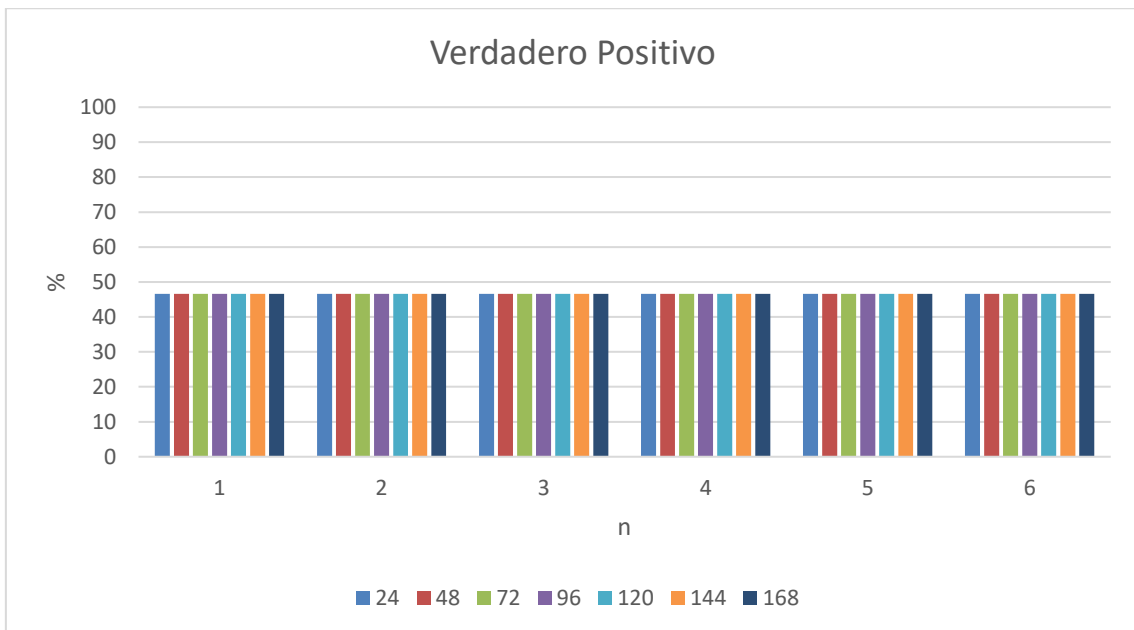


Fig.46: Verdadero Positivo Márkov con "time"

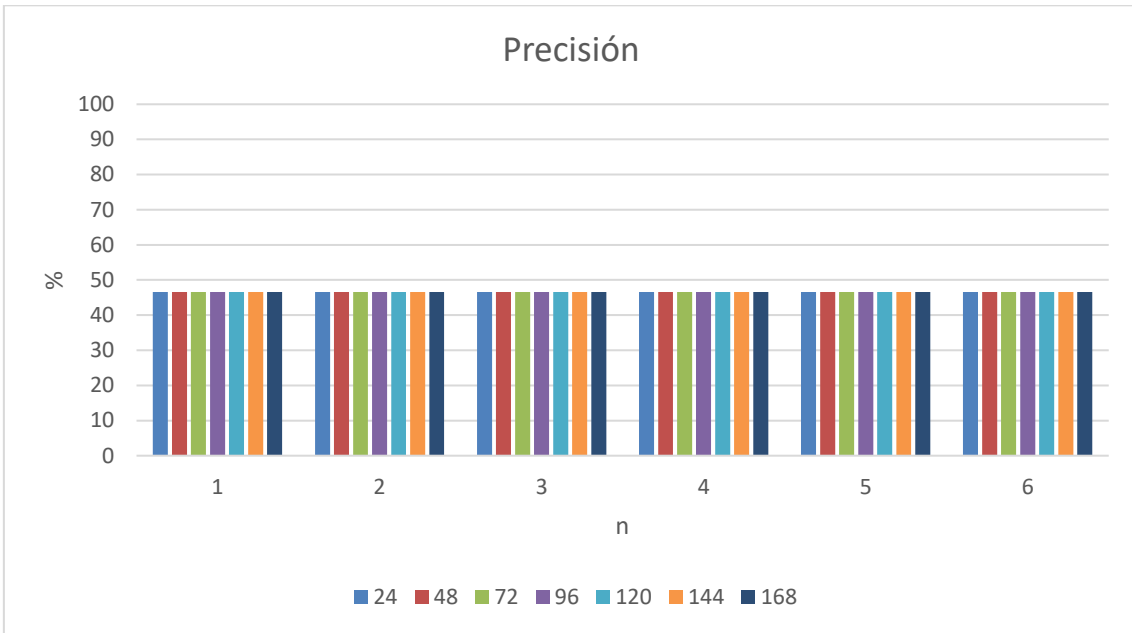


Fig.47: Precisión Márkov con "time"

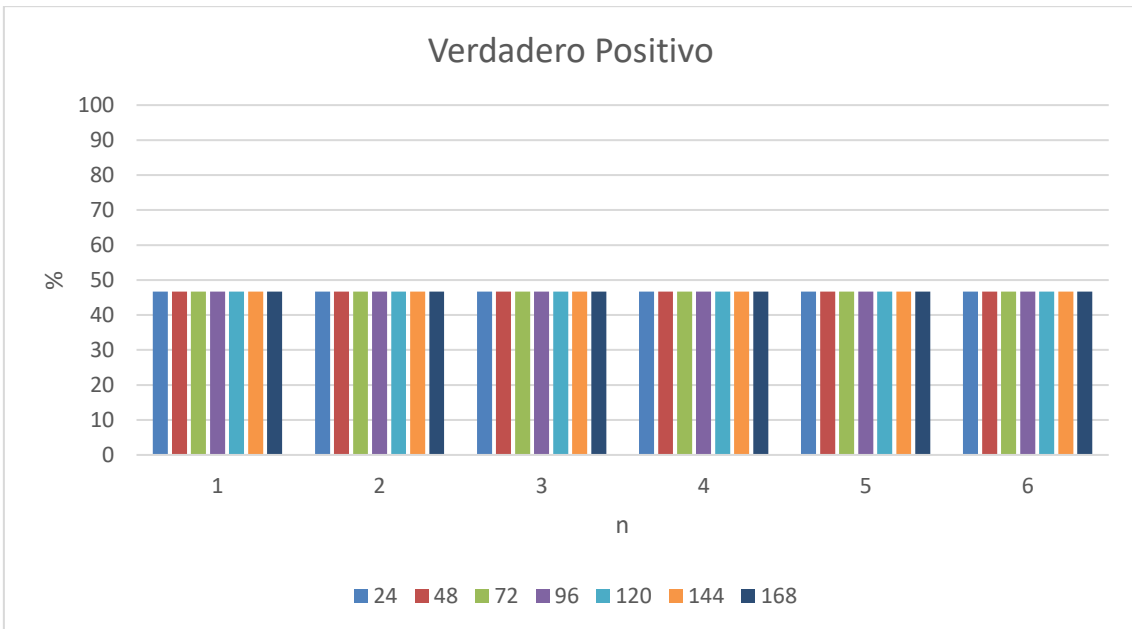


Fig.48: Verdadero Positivo Márkov con "time" y "estado"



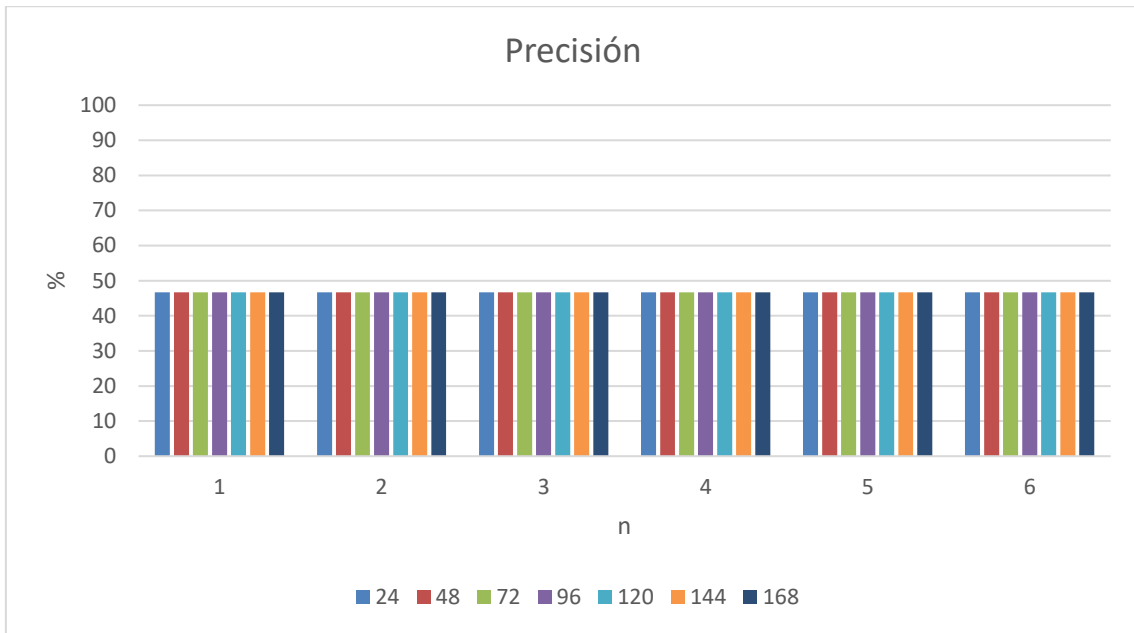


Fig.49: Precisión Márkov con "time" y "estado"

### 2.3 Escenario 3

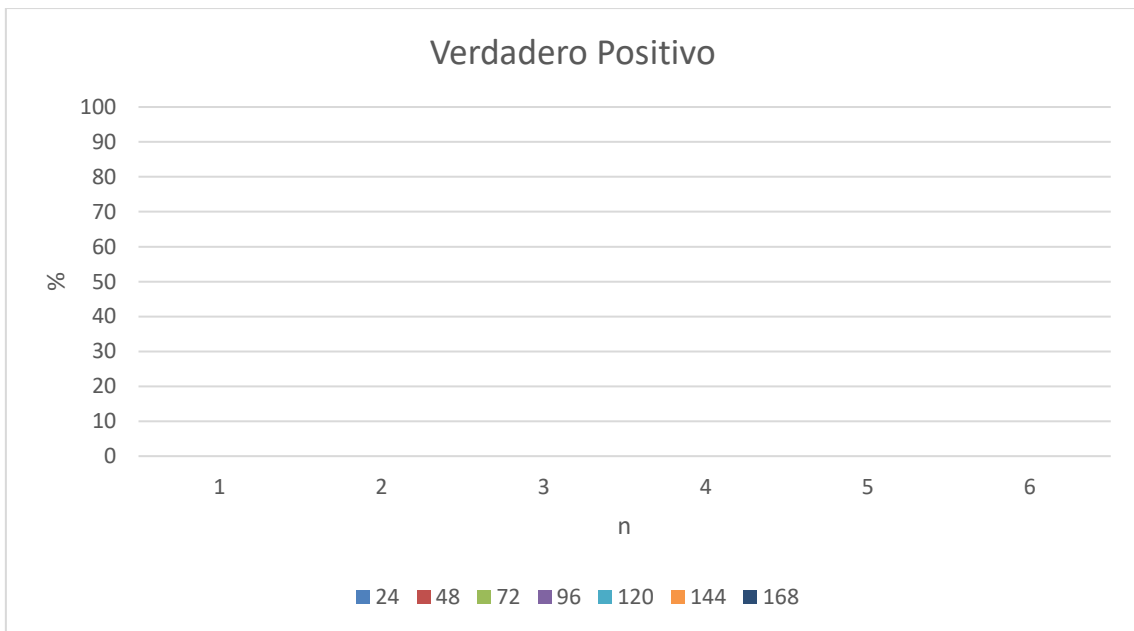


Fig.50: Verdadero Positivo Márkov sin variables

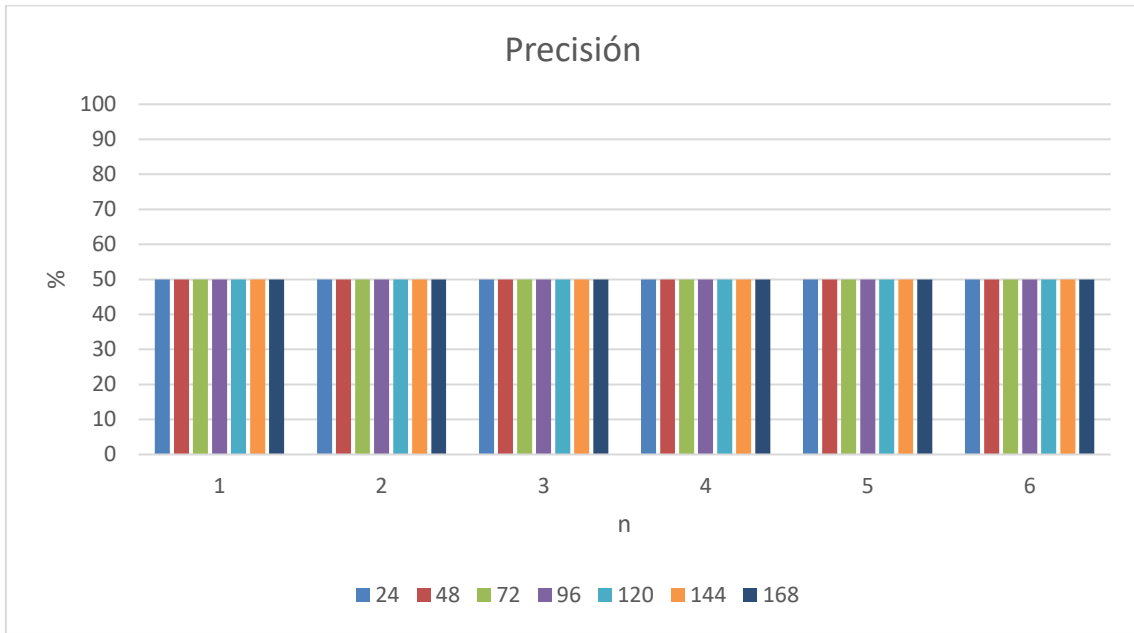


Fig.51: Precisión Márkov sin variables

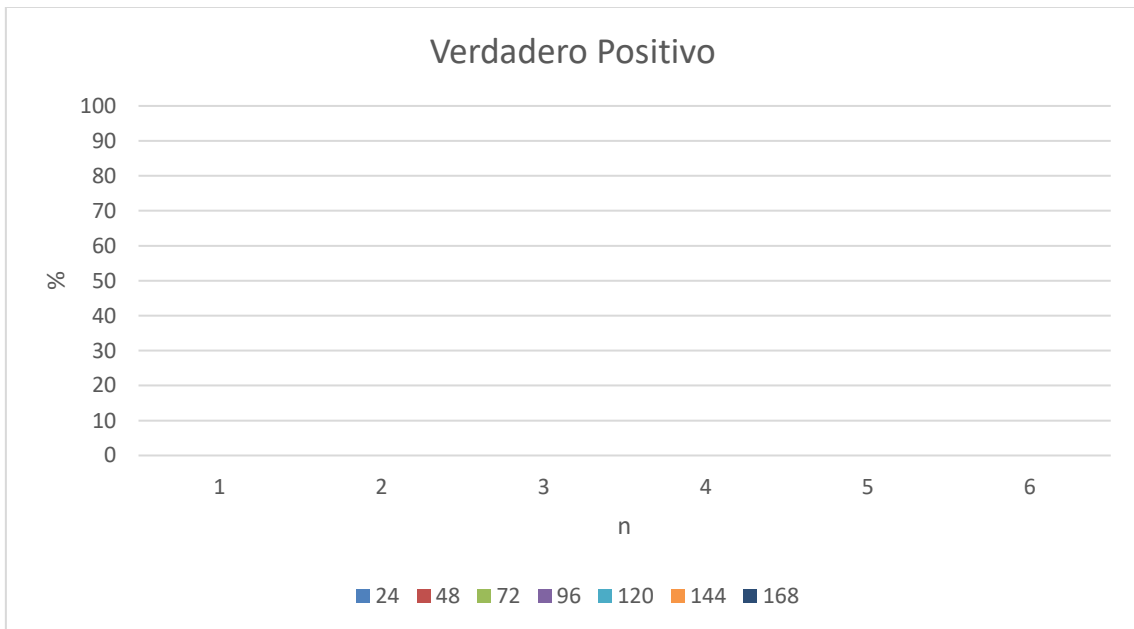


Fig.52: Verdadero Positivo Márkov con "time"

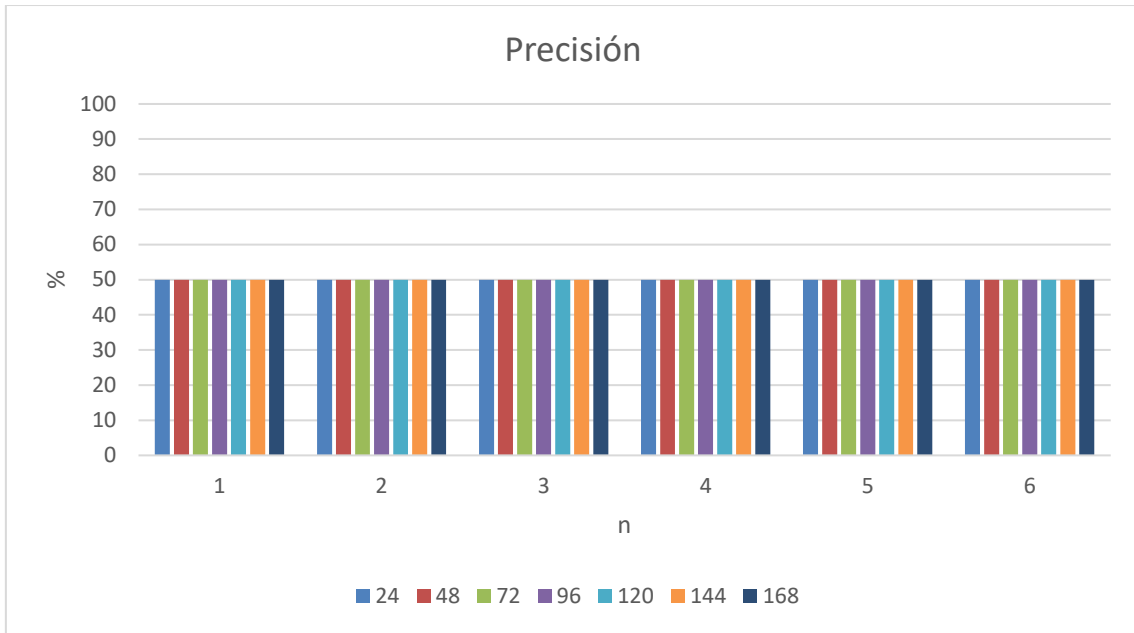


Fig.53: Precisión Márkov con "time"

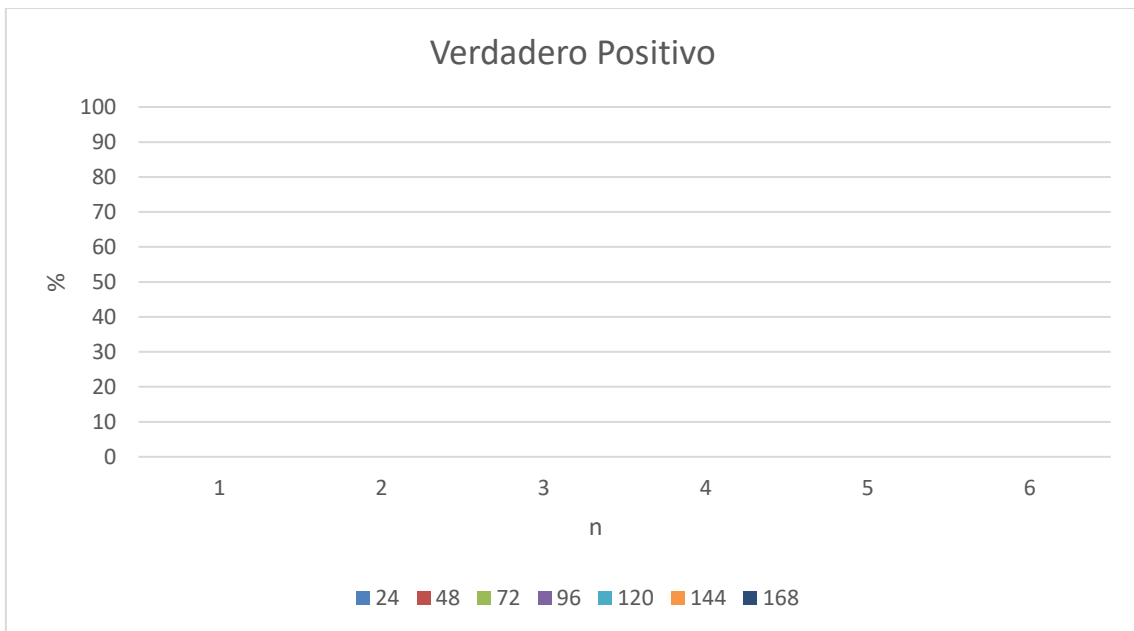


Fig.54: Verdadero Positivo Márkov con "time" y "estado"

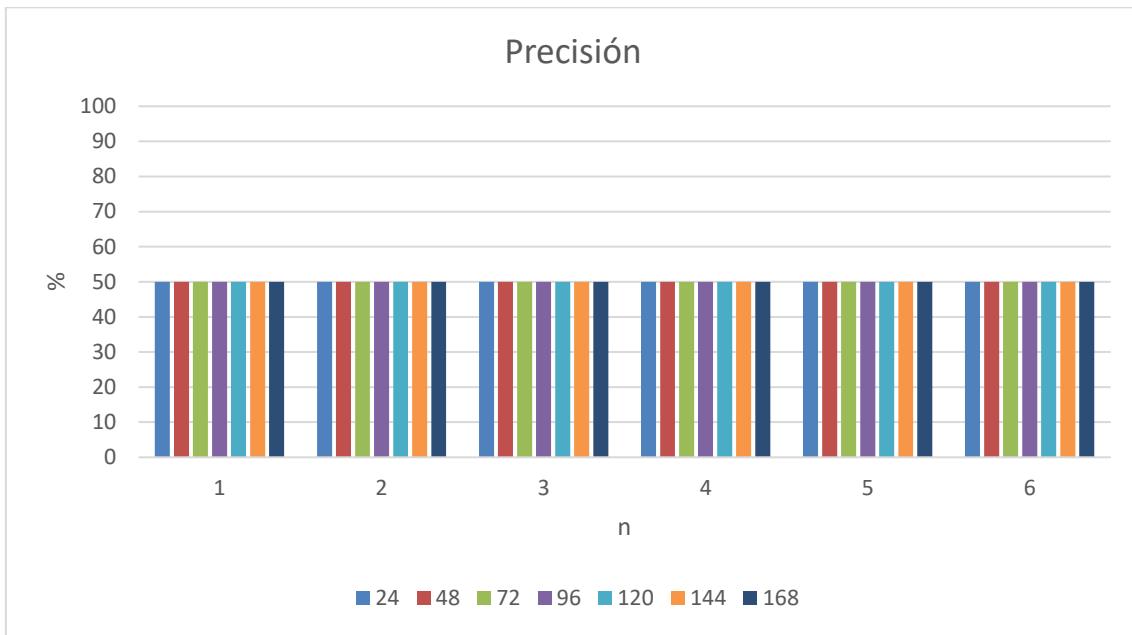


Fig.55: Precisión Márkov con “time” y “estado”

### 3. Resultados simulaciones Árbol de Decisiones

Tal y como se menciona en la memoria, para cada escenario, los resultados son los mismos sin importar la combinación de variables.

#### 3.1 Escenario 1

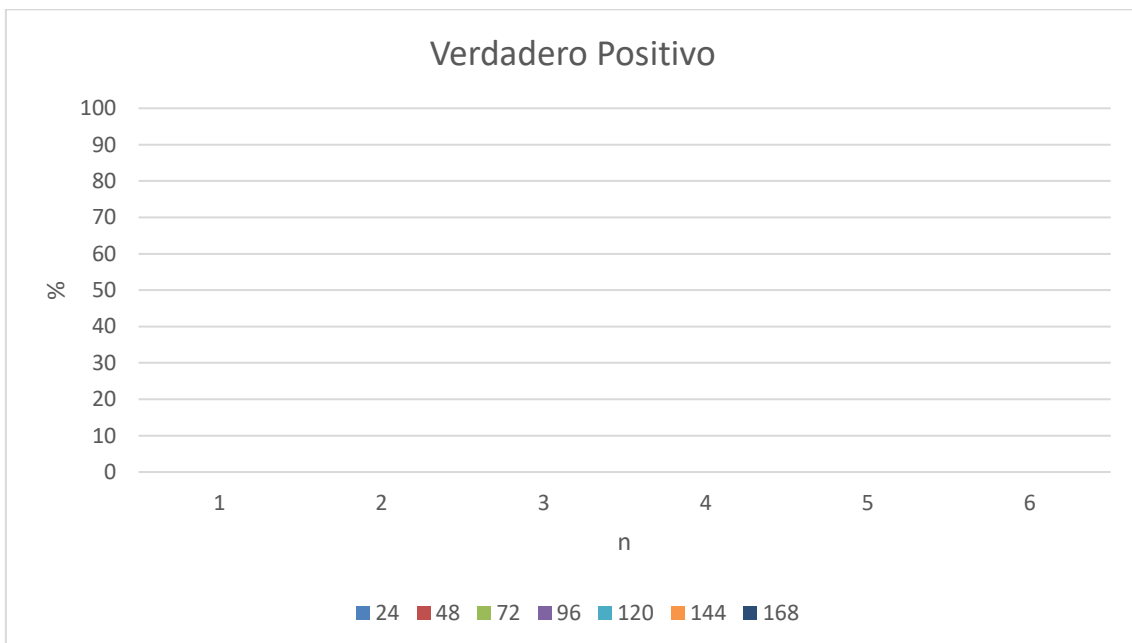


Fig.56: Verdadero Positivo Árbol de Decisiones

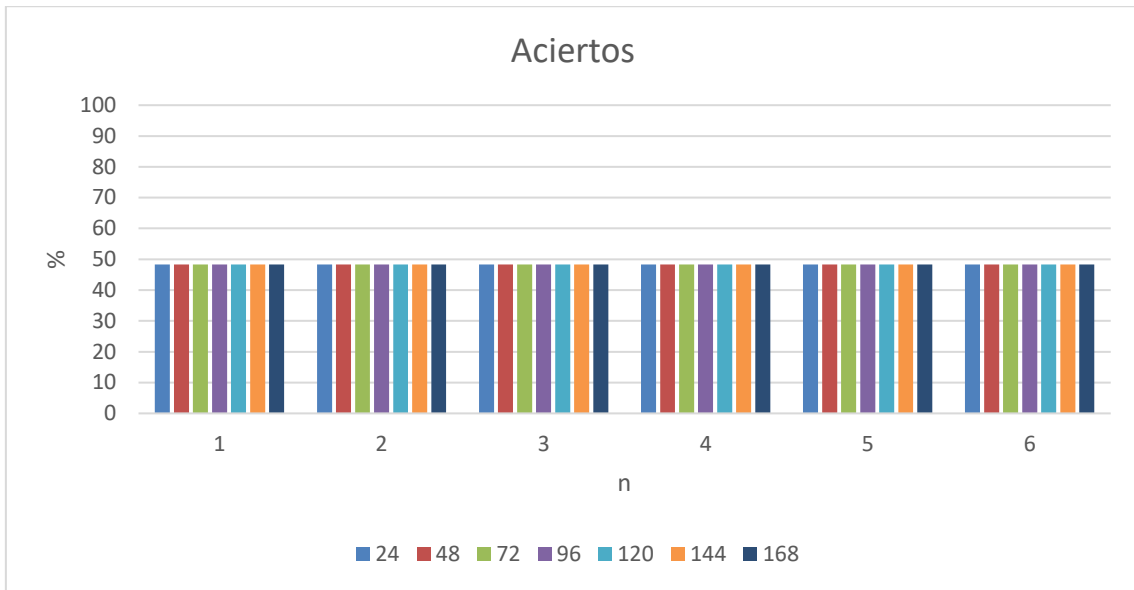


Fig.57: Aciertos Árbol de Decisiones

### 3.2 Escenario 2

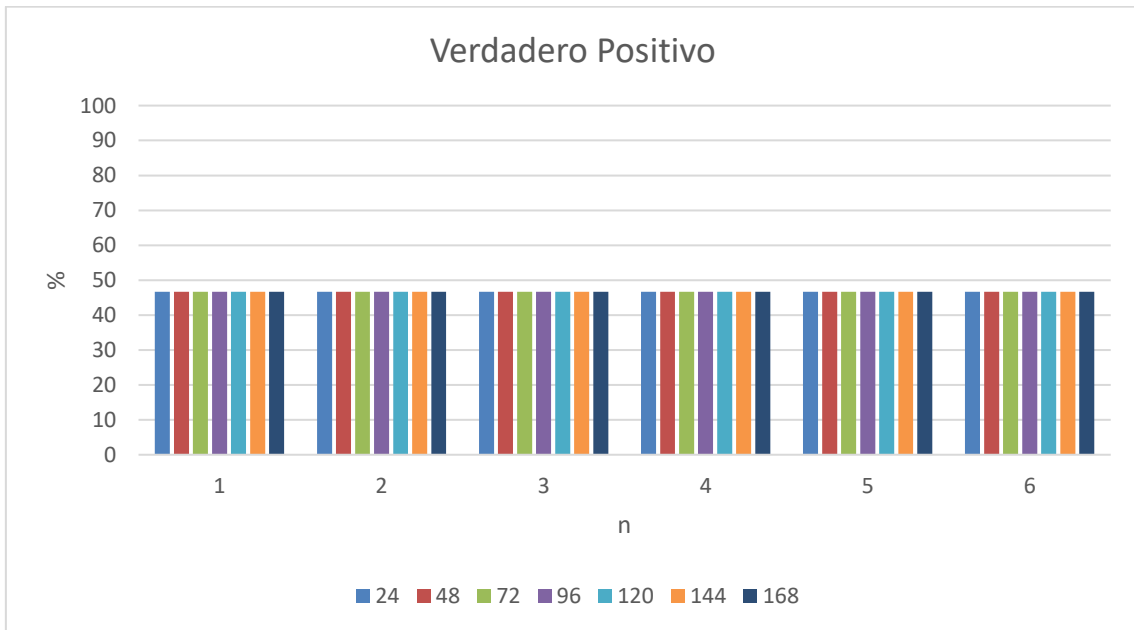


Fig.58: Verdadero Positivo Árbol de Decisiones

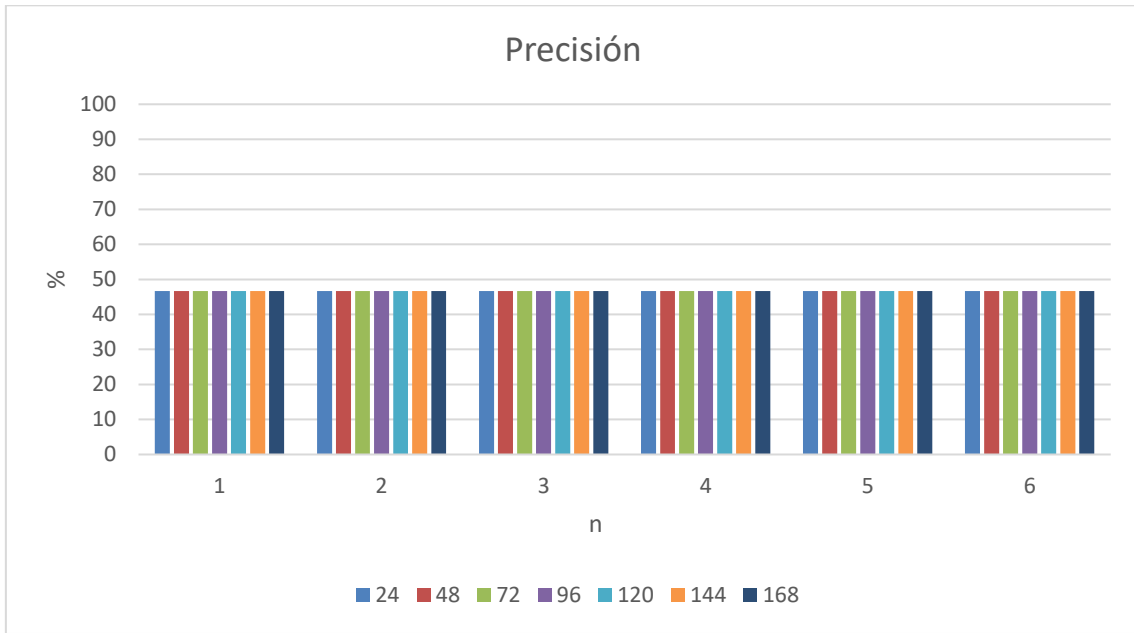


Fig.59: Precisión Árbol de Decisiones

### 3.3 Escenario 3

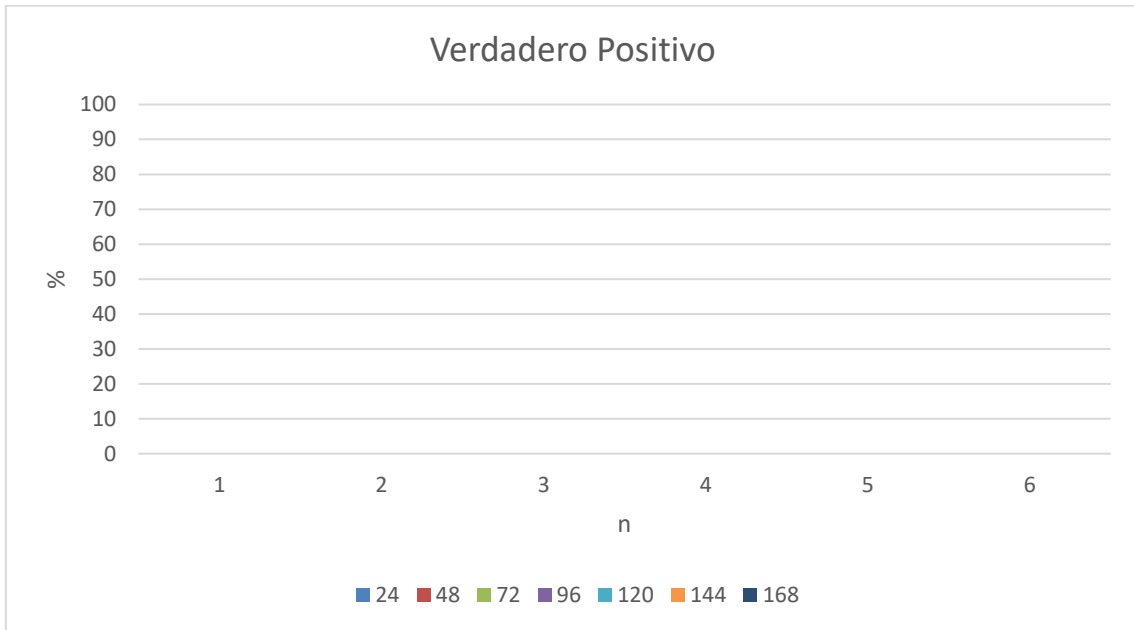


Fig.60: Precisión Árbol de Decisiones

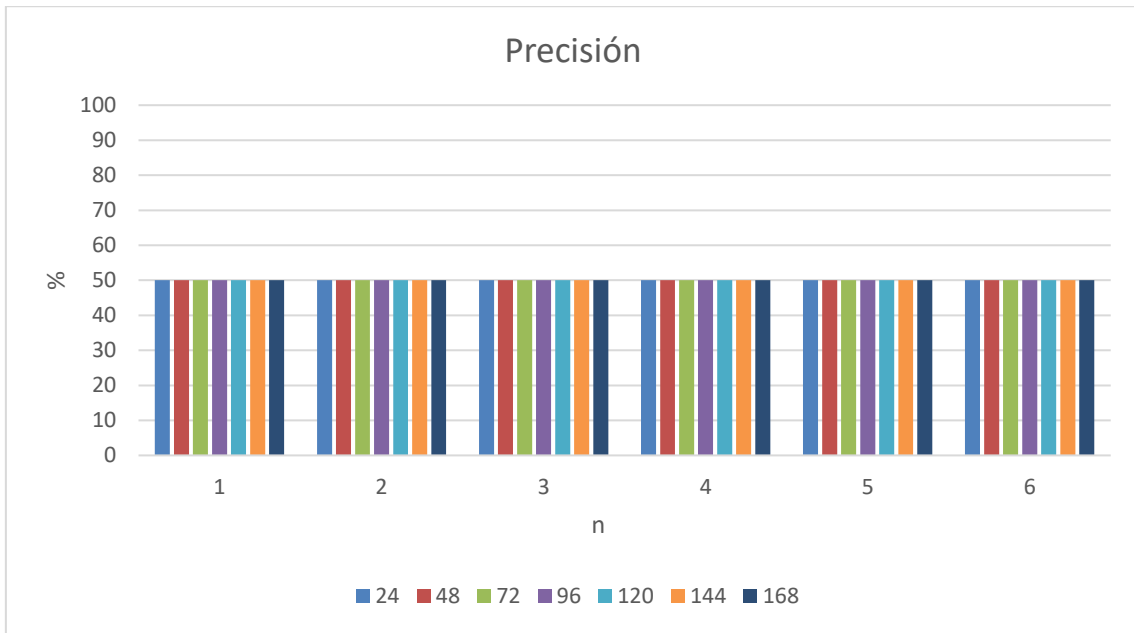


Fig.61: Precisión Árbol de Decisiones

## **BIBLIOGRAFÍA**

- [1] *Hyndman, R. J. & Athanasopoulos, G. (2013), Forecasting. Principles and Practice.*
- [2] Generous et al.2014, Global Disease Monitoring and Forecasting with Wikipedia.
- [3] Azadeh & Faiz 2011, A meta-heuristic framework for forecasting household electricity consumption.
- [4] Hernández et al.2014, Probabilistic Backpropagation for Scalable Learning of Bayesian Neural Networks.
- [5] The M3-Competition: results, conclusions and implications” (Makridakis & Hibon 2000).
- [6] Patrick Huber\*, Mario Gerber, Andreas Rumsch and Andrew Paice, Prediction of domestic appliances usage based on electrical consumption.
- [7] Nicoleta Arghira a,b,\*, Lamis Hawarah a, Stéphane Ploix a, Mireille Jacomino, Prediction of appliances energy use in smart homes.
- [8] Marc Wenninger, Jochen Schmidt and Toni Goeller, Appliance Usage Prediction for the Smart Home with an Application to Energy Demand Side Management.
- [9] Youguo Li, Haiyan Wu, A Clustering Method Based on K-Means Algorithm
- [10] Kaustav Basua, Lamis Hawaraha, Nicoleta Arghiraa,b,\*, Hussein Joumaaa, Stephane Ploixa, A prediction system for home appliance usage.
- [11] Dr. Jon Starkweather and Dr. Amanda Kay Moske, Multinomial Regresión Logística
- [12]. Abdalla M. EL-HABIL Application on Multinomial Regresión Logística Model
- [13] Weicong Kong, Student Member, IEEE, Zhao Yang Dong, Senior Member, IEEE ,A Hierarchical Hidden Márkov Model Framework for Home Appliance Modeling
- [14] ÁRBOLES DE CLASIFICACIÓN Y REGRESIÓN, Universidad de Valenci









