



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Universitat Politècnica de València

Departamento de Informática de Sistemas y Computadores

Diseño de un Sistema Predictivo de la Demanda de Energía Eléctrica Basada en Internet de las Cosas (IoT)

TRABAJO DE FIN DE MÁSTER

Máster Universitario en Ingeniería de Computadores y Redes

Autor

Juan José Sáenz Peñafiel

Tutor

PhD. Lenin Lemus

Co-Tutor

PhD. Jorge Luzuriaga

5 de septiembre de 2018

Agradecimientos

Agradezco al Gobierno de mi país Ecuador que financió mis estudios de este Máster mediante el programa de Becas "Convocatoria Abierta 2016" impulsado por la Secretaria de Educación Superior, Ciencia y Tecnología (Senescyt). Gracias a estas becas muchas personas podemos continuar con nuestros estudios en las mejores universidades a nivel mundial en beneficio de nuestro país.

Agradezco a mi esposa Vane que ha recorrido este camino a mi lado siendo mi apoyo incondicional en todo momento y ha sido mi fuerza en los momentos complicados y mi compañía en las alegrías.

Agradezco a mi madre Saida que a pesar de la distancia ha estado en todo momento junto a mi y que me ha dado sus mejores enseñanzas a lo largo de mi vida.

Agradezco a toda mi familia que siempre está pendiente de mí y envían su apoyo en todo momento, en especial a mi hermano Marcelo, mi tía Sonia, mi primo Sebas, mi suegro Marcelo, mi suegra Rosita y mi sobrina Analía.

Al grupo ICTvsCC por permitirme trabajar con ellos y brindarme sus enseñanzas, especialmente a Lenin y Jorge que han sido mis tutores y guías durante este trabajo.

Agradezco a mis amigos en Ecuador y a mis amigos y compañeros de Máster por cada consejo y ayuda recibida durante mis estudios y mi vida.

Resumen

La gestión de los Recursos Energéticos es una tarea importante para los administradores de municipios y ayuntamientos. Conocer cuánta energía es consumida por los edificios a su cargo es el primer paso para intentar reducir su consumo, así como promover la utilización de energías alternativas distintas a las provenientes de combustibles fósiles. Si se conoce cuanta energía es necesaria para suplir todas las instalaciones se puede pensar en métodos de predicción de demanda de energía como una alternativa para la optimización de dicho recurso que cubra dicha necesidad.

En este trabajo se presenta el desarrollo de una alternativa de predicción de demanda de energía eléctrica basada en un modelo de regresión lineal y un modelo autorregresivo de media móvil. Además, se presenta una arquitectura de red para la captura de datos basado en IoT que alimentan al sistema de predicción. El sistema predictor es presentado en cuatro etapas, empezando por la captura de datos, luego la transformación de dichos datos en formatos que alimentan los modelos, seguido por la etapa de análisis de los datos transformados en la que se incluye la predicción. Finalmente la presentación de resultados con una interfaz de usuario.

La arquitectura de red (basada en una nube de servicios) se plantea para mejorar la captura de datos que alimentan al sistema de predicción considerando los equipos actuales encargados de toma de datos de energía. Con esta arquitectura propuesta se pretende mejorar la calidad de los datos a utilizarse en la predicción.

Se presenta además varios trabajos relacionados a la predicción que nos han dado una pauta en la utilización de modelos predictivos relacionados con demanda energética.

Palabras claves: Internet de las Cosas (IoT), Predictor, Energía, Demanda.

Abstract

The management of energy resources is an important task for the decision makers, heads of municipalities and town halls. To know how much energy is consumed by the buildings in their charge is the first step to try to reduce their consumption, as well as to promote the use of alternative energy sources to fossil fuel. If we know how much energy is necessary for all installations, it will be possible to think about to optimize the resources that cover that needs, like methods to predict their energy demand.

In this document, we present the development of an alternative to predict the electric energy demand based on linear regression and autoregressive moving average models. In addition, a network architecture is presented for data acquisition based on IoT, whose data feed the forecast system. The forecast system is presented in four stages: First the data capture, then the data transformation into valid formats used to feed the forecasting models, followed by the data analysis stage in which the forecast is made. Finally, the results' presentation by a user interface.

The network architecture based in IoT devices in conjunction with different cloud services was proposed to improve the current data acquisition system and their quality in order to obtain accurate forecasts.

The document also presents a short review of several works related to the use of predictive models used in the forecasting of energy demand that gave us a guideline to follow in this work.

Keywords: Internet of Things (IoT), Predictor, Energy, Demand

Índice general

Índice de figuras	iii
Índice de tablas	v
1 Introducción	1
1.1 Motivación	1
1.2 Planteamiento del Problema	2
1.3 Objetivos	2
1.4 Metodología	3
1.5 Estructura del documento	4
2 Estado del Arte	7
2.1 Modelos de Predicción de Demanda	7
2.2 Agrupación de Datos (Clusterización)	8
2.3 Análisis de Datos	9
2.4 Metodologías de Desarrollo de Software	10
3 Diseño y Metodología de Desarrollo	11
3.1 Captura	11
3.1.1 <i>Herramientas Utilizadas</i>	12
3.1.2 <i>Problemas Encontrados</i>	12
3.2 Transformación	13
3.2.1 <i>Transformación de Datos de facturas en formato PDF</i>	13
3.2.1.1 <i>Herramientas Utilizadas</i>	13
3.2.1.2 <i>Problemas Encontrados</i>	15
3.2.2 <i>Transformación de Datos de facturas en formato fichero XML</i>	15
3.2.2.1 <i>Herramientas Utilizadas</i>	15
3.2.2.2 <i>Problemas Encontrados</i>	17
3.2.3 <i>Relación de los campos de los presentes en ficheros PDF, XML con la base de datos</i>	17

3.3	Análisis de datos	19
3.3.1	<i>Análisis de datos históricos</i>	20
3.3.1.1	<i>Herramientas Utilizadas</i>	20
3.3.1.2	<i>Problemas Encontrados</i>	21
3.3.2	<i>Análisis de datos de predicción</i>	21
3.3.2.1	<i>Herramientas Utilizadas</i>	21
3.3.2.2	<i>Problemas Encontrados</i>	25
3.4	Presentación de datos	25
3.4.1	<i>Herramientas Utilizadas</i>	26
4	Resultados y Discusión	31
5	Conclusiones y Trabajo Futuro	37
5.1	Conclusiones	37
5.2	Trabajo futuro	38
Anexos		39
5.3	Anexo 1 Tipos de Facturas	39
5.4	Anexo 2 Presentación	42
	Lista de Acrónimos	46
	Bibliografía	47

Índice de figuras

1	Diagrama General del Flujo de Trabajo	3
2	Diagrama de proceso para descarga de facturas	12
3	Diagrama de proceso para extracción de datos de PDF	14
4	Diagrama de la base de datos	16
5	Diagrama de proceso para extracción de datos de XML	17
6	Diagrama de procesos de análisis de datos	19
7	Ejemplo del consumo de energía eléctrica del <i>Campo de Tiro</i> en un rango específico de fechas (noviembre 2017-marzo 2018).	20
8	Diagrama de proceso de predicción de demanda	22
9	Distribución de las medias de los consumos	23
10	Perfiles de Consumo en un día determinado basados en la clasificación	23
11	Valores reales y valores de predicción	25
12	Diagrama de Proceso de Presentación de Datos	26
13	Partes de las Gráficas de consumo y demanda	27
14	Gráfica ejemplo de históricos de consumo para un determinado contrato	28
15	Gráfica ejemplo de históricos de demanda para un determinado contrato	28
16	Tabla ejemplo de demanda energética	29
17	Ejemplo de Comparativa de datos reales y predicción de 48 horas	32
18	Arquitectura propuesta de la red	35
19	Diagrama de procesos de la arquitectura planteada	36
20	Factura Tipo 1 Parte 1	39
21	Factura Tipo 1 Parte 2	40
22	Factura Tipo 2 Parte 1	40
23	Factura Tipo 2 Parte 2	41
24	Factura Tipo 3 Parte 1	41

25	Factura Tipo 3 Parte 2	42
26	Vista de la mitad superior del sistema	43
27	Vista de la mitad inferior del sistema	43
28	Gráficas de consumo en el sistema	44
29	Selección de contratos en el sistema	44
30	barra de navegación en el sistema	44
31	Datos de predicción en el sistema	45

Índice de tablas

3.1	Relación entre tipos de facturas	14
3.2	Campos comunes en los distintos tipos de facturas	18
4.1	Comparativa de resultados estadísticos de los modelos de predicción	31
4.2	Datos de demanda disponibles	32

Capítulo 1

Introducción

1.1 Motivación

Administradores de municipios y ayuntamientos de las ciudades tienen como una de sus tareas la gestión de recursos, como el suministro de energía, agua, entre otros. Sin embargo, ya sea por falta de recursos o un sin número de motivos, puede darse el caso de que los encargados de las entidades antes mencionadas desconozcan la información relativa (cantidad o distribución) a los consumos efectuados dentro de las locaciones que están a su cargo. En este trabajo consideramos que este conocimiento es una base fundamental para desarrollar iniciativas relacionadas con una adecuada gestión, específicamente la gestión energética. Por ejemplo, el coste que genera el consumo de la energía en edificios públicos supone un gasto bastante alto en los presupuestos de los ayuntamientos [1]. Para intentar reducir este inconveniente se buscan alternativas para disminuir el consumo o en su defecto el coste al que se adquiere la energía, de tal manera, que la distribución de consumo dentro de un tiempo establecido, denominada curva de carga¹, sea menor que a la de consumos anteriores.

Considerando, que la demanda energética tiene un comportamiento similar en relación a las estaciones del año y que el mercado eléctrico está basado en la oferta y demanda energética [2] se pueda realizar la una predicción de dicha demanda, que facilite el proceso de compra de energía eléctrica con un mejor coste en el mercado eléctrico español.

Con toda la información que se posee del comportamiento del consumo ener-

¹<https://www.eresenergia.com/es/public/curvas-de-carga-comercializadora-electrica-contador-inteligente-comercializador-contador-telegestionado-ac-67>

gético en relación a los edificios públicos considerados para este trabajo, se pretende presentar al usuario una interfaz interactiva que ayude al usuario, en calidad de decisor público, a tomar decisiones para la compra de energía eléctrica necesaria, obtenida en base a la predicción. Así mismo, observar los históricos de consumo de tal manera que pueda ejecutar medidas ante la posible disminución de consumo dentro de los edificios que tenga a cargo, o en su defecto analice alternativas para la utilización de energías renovables que cubran la total demanda en sus edificios.

1.2 Planteamiento del Problema

Gracias a tecnologías como redes de sensores (WSN) y el internet de las cosas² (Internet of Things (IoT)) hoy en día se dispone de grandes volúmenes de datos. Si dichos datos no se analizan ni tampoco se tratan, una organización o los entornos relativos a la misma pierden una oportunidad tanto económica como de generación de conocimiento. Es por ello por lo que en este trabajo nos centramos en que la información generada de consumos o de demanda de energía debe ser capturada mediante una infraestructura de red basada en IoT y posteriormente procesada con herramientas de Big Data³. Además, ofrecemos la posibilidad de predecir la demanda energética dentro de un periodo determinado para que el usuario final pueda mejorar la toma de decisiones con la información procesada y las predicciones realizadas mediante una interfaz intuitiva y fácil de manejar.

1.3 Objetivos

El objetivo principal es diseñar una arquitectura de red basada en dispositivos y tecnologías IoT que permita conocer la demanda real de energía eléctrica de una organización administrativa (como un ayuntamiento, municipio, o alcaldía) para diseñar un sistema capaz de predecir la demanda futura de energía eléctrica de dicha organización en días, semanas, meses e incluso años.

Para conseguir el objetivo principal, hemos definido los siguientes objetivos específicos:

- Diseñar la arquitectura de red para obtener datos de la demanda de energía eléctrica.

²Internet of Things (IoT). Conexión de objetos de uso cotidiano a Internet.

³Almacenamiento y tratamiento de grandes volúmenes de datos.

- Trabajar con grandes cantidades de datos (procedentes de IoT, así como datos oscuros y no estructurados⁴), para que sean datos útiles que alimenten los algoritmos del sistema de predicción.
- Utilizar un motor de predicción para inferir la demanda energética para un intervalo de tiempo específico.
- Presentar de una manera ordenada que facilite la interpretación de los resultados de la predicción y de los históricos de datos.

1.4 Metodología

La metodología utilizada en este trabajo de fin de máster (TFM) parte de una definición clara de los objetivos hacia la transición energética, línea de investigación abierta dentro del grupo *Information and Communication Technologies versus Climate Change (ICTvsCC)* del instituto ITACA.

Para abordar esta problemática los métodos utilizados se presentan en el flujo de trabajo de la Figura 1, y se describen a continuación.

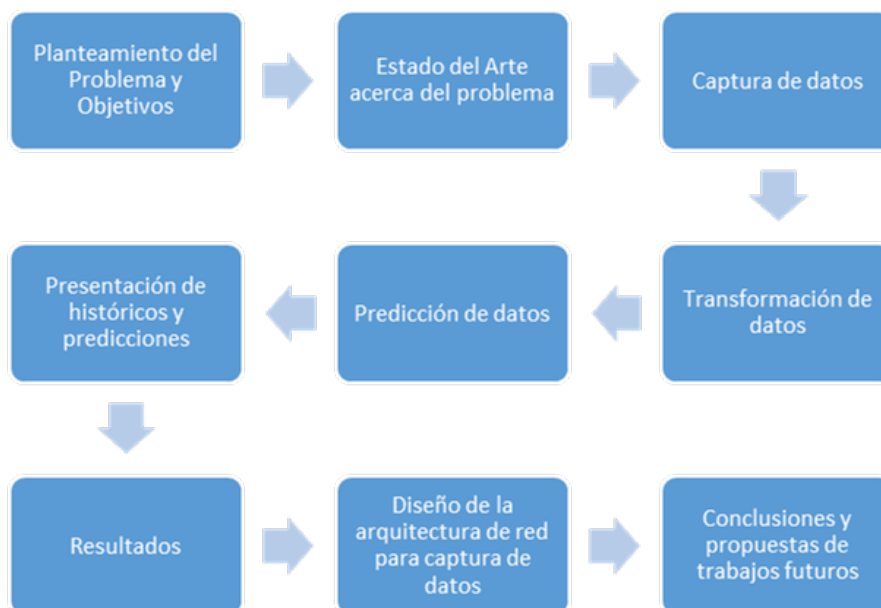


Figura 1: Diagrama General del Flujo de Trabajo

⁴gartner.com/it-glossary/dark-data

Planteamiento de problema y objetivos - Se determina los motivos de la realización de este trabajo al igual que los alcances de este, delimitados por los objetivos.

Estado del arte acerca del trabajo - Búsqueda y compilación de los trabajos relacionados y herramientas útiles que puedan servir para el caso concreto.

Captura de datos – Uso de la infraestructura para la toma de datos que estén disponibles hasta la fecha.

Transformación de datos – De los formatos obtenidos en la captura se extrae los datos necesarios en archivos de texto o bases de datos de acuerdo al caso, de tal manera que se puedan usar en los siguientes pasos con facilidad.

Predicción de datos – Desarrollo de la herramienta de predicción de demanda de energía con los datos obtenidos en la transformación.

Presentación de históricos y predicciones – Mostrar gráficamente al usuario los datos capturados, así como los de las predicciones en forma de cuadro de mando.

Resultados – Presentación de resultados obtenidos en la predicción con los cambios de parámetros y tipos de predictores utilizados.

Diseño de la arquitectura de red para captura de datos – Diseñar una alternativa para la captura automática de los datos de demanda de energía.

Conclusiones y propuestas de trabajos futuros – Se obtienen las conclusiones del trabajo y se presenta las opciones encontradas para extender el trabajo realizado.

1.5 Estructura del documento

Este documento consta de 5 capítulos presentados de la siguiente manera:

- El presente capítulo se ha utilizado como introducción y definición del problema a tratar y la forma a tratar. También presenta los pasos que se usan para cumplir con los objetivos planteados en este trabajo.

- El **capítulo 2** presenta una revisión bibliográfica de trabajos relacionados a demanda energética en España, así como predicción de demanda y presentación de datos energéticos, de tal manera que se tiene una base teórica sustentable del trabajo realizado.
- En el **capítulo 3** se describe paso a paso el proceso realizado en la captura, transformación, predicción y presentación de los datos de la demanda, así como los formatos y descripción de las herramientas que son utilizadas específicamente en cada proceso.
- El **capítulo 4** compara los predictores utilizados y analiza sus resultados para mostrar el mejor predictor para el caso específico. Además presenta la arquitectura planteada para la captura de datos.
- El **capítulo 5** se presentan las conclusiones obtenidas en relación a los objetivos planteados para un caso específico, así como los trabajos futuros que pueden complementar o mejorar el trabajo realizado.

Capítulo 2

Estado del Arte

En este capítulo se presenta un análisis del estado del arte que ha permitido la elaboración de este trabajo. Primero, se muestran trabajos relacionados a modelos de predicción de demanda energética utilizados y las técnicas que han generado resultados destacados en este ámbito. Después, técnicas de agrupación de datos que sirven para los datos de predicción. Luego, las técnicas de análisis de datos que serán utilizadas para la demanda. El capítulo finaliza con las metodologías de desarrollo de software que servirán para realizar el sistema de predicción.

2.1 Modelos de Predicción de Demanda

En muchos municipios y ayuntamientos, sus administradores están trabajando en alternativas energéticas que ayuden a cumplir con las normativas europeas tanto vigentes como futuras para el consumo adecuado y la eficiente de energía eléctrica. Concretamente en España se trabaja cumpliendo las normativas del mercado eléctrico Español que basa su utilización en la Demanda y Consumo energético [2].

En cuanto a la demanda energética, varios estudios han buscado dar alternativas de predicción de demanda, de tal manera que los encargados de utilizar el mercado eléctrico puedan realizar solamente la compra de la energía necesaria basados en la predicción obtenida. Para ello, muchas técnicas y modelos han sido utilizados, desde modelos matemáticos basados en las probabilidades como el propuesto en [3] que utiliza el modelo probabilístico para la predicción de demanda y también de precio ; o en otros casos comparando modelos estadísticos, tanto lineares como estocásticos para la predicción de demanda [4]. También se han implementado modelos basados en redes neuronales para la predicción de demanda

como es el caso de [5] que busca entrenar varios modelos de redes neuronales para predicción a largo plazo. O basados en lógica difusa relacionando los valores de la demanda mediante agrupación de dichos valores por fechas semejantes a las que se van a predecir y dando un valor semejante al grupo [6].

Adicionalmente, se han propuesto alternativas basadas en series de tiempo como es el caso de [7] que hacen la predicción de la demanda por intervalos de tiempo para datos de dos países distintos. Además, para las series de tiempo, [8] ha utilizado y comparado vectores autorregresivos, es decir que las salidas son directamente relacionadas con el vector lineal de sus datos predecesores, y un modelo de red neuronal multicapa, dando un resultado de un acierto mayor en el modelo autorregresivo.

Otra alternativa a las series de tiempo son los modelos autorregresivos que trabajan con intervalos de tiempo para mejorar la predicción, como es el caso de ARIMA⁵, que como lo demuestran en [9] y [10] trabaja con muy buenas prestaciones con series de tiempo y demanda energética.

Cuando se refiere a consumo energético han existido muchos trabajos relacionados a su predicción, teniendo resultados notables con el uso de redes neuronales como en [11] que realiza la predicción de consumo con varias técnicas, dando resultados positivos y porcentajes de predicción con error menor al 4 %. O en otros casos con resultados comparativos con regresiones, arboles de decisión y redes neuronales, como proponen en [12], que como resultado obtienen valores muy semejantes en las técnicas mencionadas anteriormente.

De igual manera que en demanda, también los modelos ARIMA han dado buenos resultados en la predicción de consumo energético, como son los casos de [13] y [14] que obtienen resultados de predicción con dicho modelo y también lo comparan con modelos propios, dando resultados cercanos entre ellos.

2.2 Agrupación de Datos (Clusterización)

La clusterización de datos, definida en [15] basado en agrupación de datos que tienen un criterio similar, nos ayuda para la predicción de series de tiempo con datos semejantes en las columnas que lo contengan, como en [16] donde los autores concluyen que de manera exploratoria la clusterización de datos es óptima para facilitar el análisis de dichos datos.

⁵Autoregressive Integrated Moving Average

Las técnicas de clusterización varían de acuerdo al tipo de datos que se maneja; en el caso específico de series de tiempo la clusterización existen técnicas basadas en la jerarquía de los datos, que agrupan los datos basados en la importancia de los mismos [17], o técnicas basadas en la reubicación del centro de predicción como es el caso de [18] que realizan clusterización de datos de energía moviendo el centroide de datos para conseguir grupos en relación a la distancia a dicho centroide.

Además, para series de tiempo trabajos como [19] o [20] utilizan la técnica de *k-means* que separa los datos en k grupos de tal manera que los datos se clasifiquen en el grupo con la media de datos más cercana. En [19] se usa exitosamente la clusterización con *k-means* en datos de series de tiempo multimedia para agrupar los datos por tiempo pero generando intervalos de tiempo propios. Mientras que en [20] utilizan la agrupación por *k-means* para realizar predicción de datos de precios.

2.3 Análisis de Datos

Para el análisis de datos se han manejado diversas metodologías tanto estadísticas, computacionales o visuales, que varían su uso de acuerdo con el tipo de datos y resultados que se desean obtener en el estudio. En el caso concreto de datos con modelos estadísticos se utilizan técnicas con modelos autorregresivas a las series de tiempo, de correlación, o modelos gráficos como lo detallan en [21], especificando el tipo de análisis correcto para cada tipo de serie de tiempo.

En cuanto a modelos computacionales para series de tiempo, técnicas como la minería de datos han sido ampliamente usadas como lo indica [22] en su revisión del uso de las técnicas, destacando ejemplos de clusterización y agrupación de datos, además de herramientas que han usado la minería de datos mediante visualizadores.

En casos concretos de análisis de datos de energía eléctrica, [23] realizan el análisis de predicción de precios mediante el modelo autorregresivo ARIMA obteniendo valores de error de hasta un 10 % con datos de precios de energía.

2.4 Metodologías de Desarrollo de Software

En cuanto al desarrollo de software, se debe separar en relación al tipo de aplicación que se va a desarrollar, siendo metodologías complejas de software orientadas a software empresarial como Rational Unified Process (RUP) que es una guía para construcción de software basada en un conjunto de procesos adaptables para software [24]; y Service Oriented Architecture (SOA) que orienta la construcción de software basándose en servicios reutilizables, como lo ejemplifican los autores de [25].

A su vez, las metodologías ágiles para desarrollo de software experimental como la basada en procesos iterativos e incrementales denominada Dynamic Systems Development Method (DSDM)⁶; o metodologías también basadas en procesos iterativos pero tolerantes a cambios como Adaptive Software Development⁷, que han sido parte de análisis comparativo en [26].

Como una metodología adaptable a software experimental también se encuentra SCRUM⁸ que ha sido aplicada a diversos proyectos de predicción como en [27] que usa SCRUM para predicción de fallos en análisis de datos y presenta estadísticas de fallos del uso de la metodología.

Con esta revisión de la literatura en las secciones previas, hemos notado que los modelos de predicción basados en regresiones lineales son ampliamente utilizados para predicciones energéticas, por su capacidad de trabajar con datos de entrada con series temporales como es el caso de datos de demanda energética. De igual manera, para facilitar la creación de modelos predictivos, se recomienda agrupar los datos mediante técnicas de agrupación como *k-means* de tal manera que se pueda generalizar un modelo de predicción para diferentes datos de entrada. En cuanto a metodologías de desarrollo de software, hemos visto que SCRUM es una metodología que cumple con los parámetros para ser usadas en proyectos de software experimental.

⁶<https://www.agilebusiness.org/what-is-dsdm>

⁷<http://www.adaptivesd.com/>

⁸<https://www.scrum.org/>

Capítulo 3

Diseño y Metodología de Desarrollo

Para el desarrollo de la aplicación se requería de una metodología que itere en cada etapa y que en cada iteración el grupo de trabajo pueda realizar los cambios pertinentes. Por esta razón se seleccionó la metodología SCRUM por su capacidad de adaptación a cambios de requisitos e iteraciones dentro del proceso.

Como parte de la metodología indicada, en la fase de requerimientos se definieron cuatro etapas dentro del desarrollo de la aplicación. La primera etapa la captura de los datos de energía, posteriormente enviados a la transformación en formatos necesarios en la aplicación (segunda etapa); luego, los datos transformados son enviados a la etapa de análisis de datos, sea estos para históricos o para predicción, para finalmente llevar los resultados a la etapa de presentación para el usuario final. En cada etapa se realizaron reuniones de control de la aplicación para que las iteraciones realizadas con cambios o mejoras cumplan con las necesidades de la metodología.

3.1 Captura

Para la captura se utilizó la técnica de Web Scraping⁹ que consiste en automatizar la extracción de datos desde una determinada página web y almacenarlos en nuestro repositorio, sea este una base de datos o ficheros separados por coma. La técnica de scraping fue utilizada para simular el comportamiento de un usuario en la descarga de los ficheros PDF y XML de consumo de datos de la comercializadora de la página de Iberdrola¹⁰, realizando el proceso de automatización de navegación web que se muestra en el diagrama de la Figura 2.

⁹<https://sitelabs.es/web-scraping-introduccion-y-herramientas/>

¹⁰<https://www.iberdrola.es/>

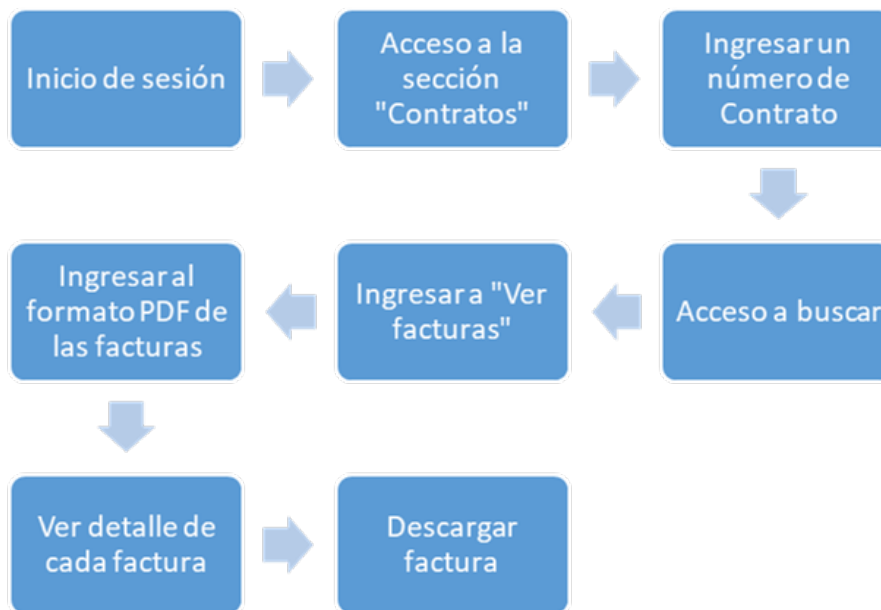


Figura 2: Diagrama de proceso para descarga de facturas

Este proceso obtiene como resultado un conjunto de ficheros PDF correspondientes a facturas para cada contrato. El proceso debe iterarse uno a uno en todos los contratos requeridos para la aplicación.

3.1.1 Herramientas Utilizadas

Para esta etapa se usó el lenguaje de programación *Python*¹¹, mediante la librería *BeautifulSoup*¹², la cual analiza las etiquetas del lenguaje HTML de una página web indicada para extraer los datos concretos en dichas etiquetas.

3.1.2 Problemas Encontrados

En el desarrollo de esta etapa los problemas encontrados fueron:

- Cambios en la presentación y programación de la página web que modificaba el script de scraping.
- Dificultad de permisos de acceso.

¹¹<https://www.python.org>

¹²<https://www.crummy.com/software/BeautifulSoup/>

3.2 Transformación

El objetivo de la etapa de transformación es extraer, organizar y almacenar los datos transformados en base de datos y ficheros en formato CSV para que posteriormente sean útiles en las siguientes etapas del desarrollo. El método utilizado en esta etapa es la transformación secuencial de cada tipo de dato, empezando por los datos de ficheros PDF, después de ficheros XML y al final relacionar los datos dentro de la base de datos según la información que es común.

3.2.1 Transformación de Datos de facturas en formato PDF

Con el objetivo de obtener los datos provenientes de las facturas emitidas y descargadas de la web de Iberdrola Comercializadora, se realizaron varios scripts que se encargan desde la organización de datos crudos hasta la carga en la base de datos; dicha base de datos ha sido elaborada en relación con los datos disponibles.

Los PDF obtenidos en el proceso de captura corresponden al periodo comprendido entre los años 2015 y 2018. Para procesar estos ficheros primero fueron separados por años y luego agrupados basándose en el tipo de factura de cada fichero como se explica posteriormente.

3.2.1.1 Herramientas Utilizadas

Para la transformación de datos se utilizó la técnica *PDF scraping* de datos mediante la librería de Python llamada Pdfminer¹³, la cual convierte el PDF en secuencia de textos, específicamente buscando, eliminando y/o reemplazando palabras. Con esta herramienta se procedió a separar los documentos en tres tipos considerando las especificidades de cada factura.

En la Tabla 3.1 se puede observar las diferencias más marcadas e importantes que caracterizan a cada tipo de factura y en el anexo 1, se muestra un ejemplo de cada factura señalando gráficamente las diferencias indicadas.

Posterior a la separación de las facturas, se realizó la extracción de los datos mediante los pasos descritos en el diagrama de procesos de la Figura 3.

¹³<https://pypi.org/project/pdfminer/>

Tipo 1	Tipo 2	Tipo 3
Fondo blanco en títulos	Fondo verde en títulos	Fondo verde en títulos
Gráficas Historial de consumo	Gráficas de evolución de consumo pueden o no estar presentes	Gráficas de evolución de consumo
Sin Gráfica Recargos	Gráfica Recargos	Gráfica Recargos
Texto: Referencia contrato	Texto: Referencia contrato suministro	Texto: Referencia contrato suministro
Texto: "Nº factura"	Texto: "Número de factura"	Texto: "Número de factura"
Texto: "Tarifa ATR"	Texto: "Peaje de acceso a la red (ATR)"	Texto: "Peaje de acceso a la red (ATR)"
Texto: "Precios B.O.E"	Texto: "Precios de peajes de acceso"	Texto: "Precios de peajes de acceso"
Sin detalle de llano, valle y pico en energía y potencia	Sin detalle de llano, valle y pico en energía y potencia	Con detalle de pico llano y valle en energía y potencia
Un rango de fechas en detalle de factura	Un rango de fechas en detalle de factura	Dos rangos de fechas en detalle potencia y energía

Tabla 3.1: Relación entre tipos de facturas

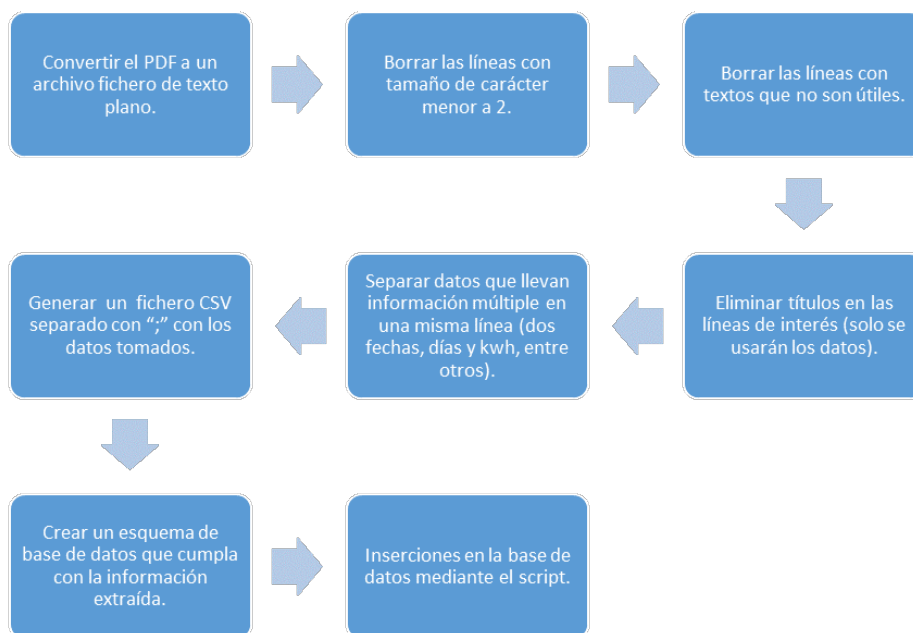


Figura 3: Diagrama de proceso para extracción de datos de PDF

3.2.1.2 *Problemas Encontrados*

Los siguientes problemas fueron identificados en el transcurso del proceso de transformación:

- Inicialmente se manejaba un solo formato de factura, por lo que el primer script se basaba en dicho formato, posteriormente analizando otras facturas se identificó otros formatos distintos, de manera que se cambió la base de datos y el script.
- Caracteres especiales dentro de la factura, con símbolos como el de euros (€) que impedía que el script se ejecutara normalmente. Para solucionarlo se cambiaron los formatos de lectura de cada fichero.
- Tamaños variables en datos por facturas. Para solucionarlo se realizaron consultas de líneas genéricas y se eliminaron caracteres que no son relevantes.
- Una vez obtenidos los datos desde los ficheros PDF, se observó variación en los campos de un PDF a otro y en consecuencia un desfase en los datos obtenidos, esto debido a que la técnica de scraping toma los datos en posiciones del texto muy específicas y los formatos de ficheros PDF eran muy variables generando problemas cuando los detalles de las facturas aumentaban o disminuían. Esto causaba que los tipos de facturas en formato PDF llegaran a ser incontables por la relación de número de detalles y combinación de estos.

3.2.2 *Transformación de Datos de facturas en formato fichero XML*

Por los problemas encontrados en la transformación de ficheros PDF y aprovechando que la web de Iberdrola activó la descarga de ficheros en formato XML, se realizó la transformación de datos de facturas en ficheros XML.

3.2.2.1 *Herramientas Utilizadas*

Para el procesamiento de datos de los ficheros XML, se utilizó también el lenguaje de programación Python con la librería *Xmltodict*¹⁴, la cual convierte el fichero XML en un diccionario¹⁵ como si se tratase de una matriz de datos, en la que se tiene etiquetas de los nombres con los que se identifican los valores que contiene la

¹⁴<https://pypi.org/project/xmltodict/>

¹⁵<https://docs.python.org/2/tutorial/datastructures.html>

factura. Con las etiquetas mencionadas se puede acceder a los datos directamente independientemente de su posición. Esto soluciona el problema anteriormente mencionado que teníamos con los PDF.

Los datos obtenidos son almacenados en dos tablas en una base de datos relacional. Una primera tabla con datos generales de cada factura: número, dirección, coste total, entre otros... (como se puede observar en la Figura 4) y una segunda tabla con el detalle de los valores a pagar para esa factura, separando detalles de potencia, energía e impuestos.

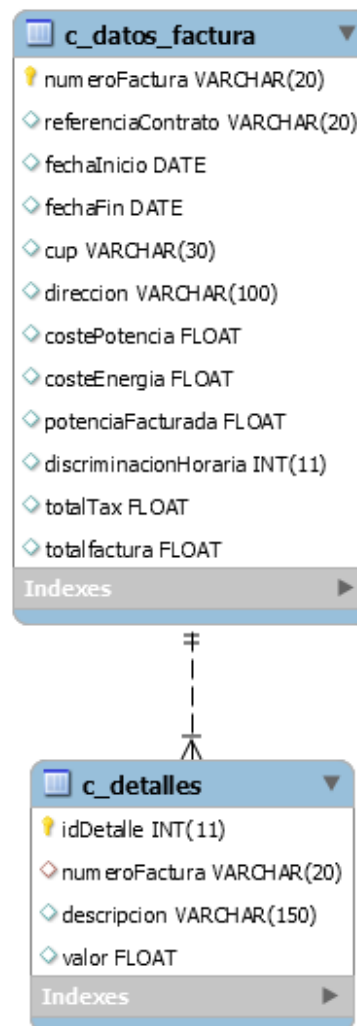


Figura 4: Diagrama de la base de datos

El proceso se puede observar en el diagrama de procesos de la Figura 5.

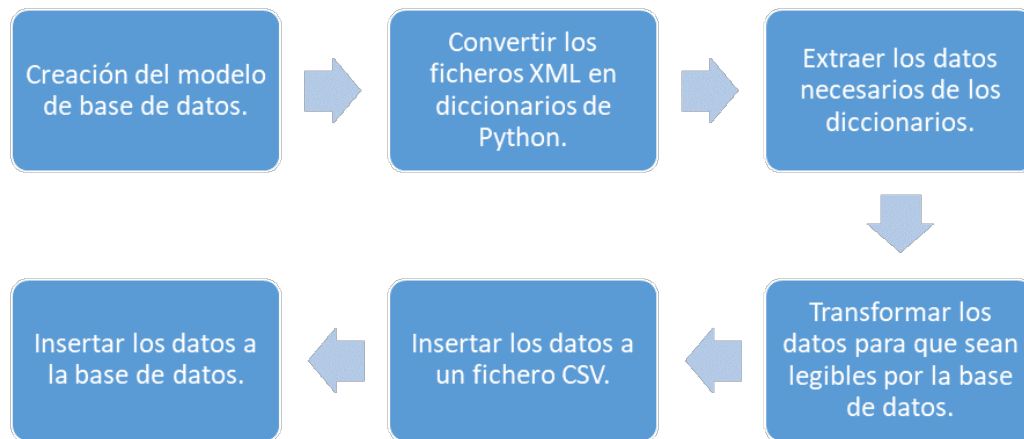


Figura 5: Diagrama de proceso para extracción de datos de XML

3.2.2.2 Problemas Encontrados

Los siguientes problemas fueron identificados en el transcurso de este proceso:

- 138 facturas de un total de 2824 no contenían la etiqueta “DebitReconciliationReference”, en la que se junta el número de factura y la referencia de contrato. Dichas facturas fueron indicadas en un fichero de log para ser transformadas con otro método no implementado en este trabajo.
- Los datos de las facturas que incluían discriminación horaria tuvieron que ser extraídos mediante separación de los textos del detalle de la factura, debido a que ninguna etiqueta los contiene.

3.2.3 Relación de los campos de los presentes en ficheros PDF, XML con la base de datos

Una vez obtenidos los datos tanto de los XML y los PDF y haberlos insertado en la base de datos, es importante presentar los datos comunes entre cada uno de ellos y su presencia en la base. Para ello la Tabla 3.2 indica que datos están presentes en cada documento y cuales se encuentran en la base de datos.

Campo	PDF	XML	Base de Datos
Fecha Inicio	X	X	X
Fecha Final	X	X	X
Número de Factura	X	X	X
Tipo de Lectura	X		X
Referencia Contrato de Suministro	X	X	X
Dirección	X	X	X
CUP	X	X	X
Tipo de Discriminación Horaria	X	X	X
Potencia Contratada	X		
Peaje de Acceso a la Red (ATR)	X		
Precios de Peaje de Acceso	X		
Potencia facturada	X	X	X
Energía facturada	X	X	X
Coste Potencia	X	X	X
Coste Energía	X	X	X
Cobros Extras	X	X	X
Impuestos	X	X	X

Tabla 3.2: Campos comunes en los distintos tipos de facturas

3.3 Análisis de datos

Luego de transformados y organizados los datos es necesario analizarlos; para ello nos basamos en metodologías comunes utilizadas en publicaciones e investigaciones referentes al análisis de datos descritas anteriormente en el capítulo del estado del arte (sección 2.3). Básicamente seguimos el esquema que se muestra en la Figura 6 y se describe a continuación:

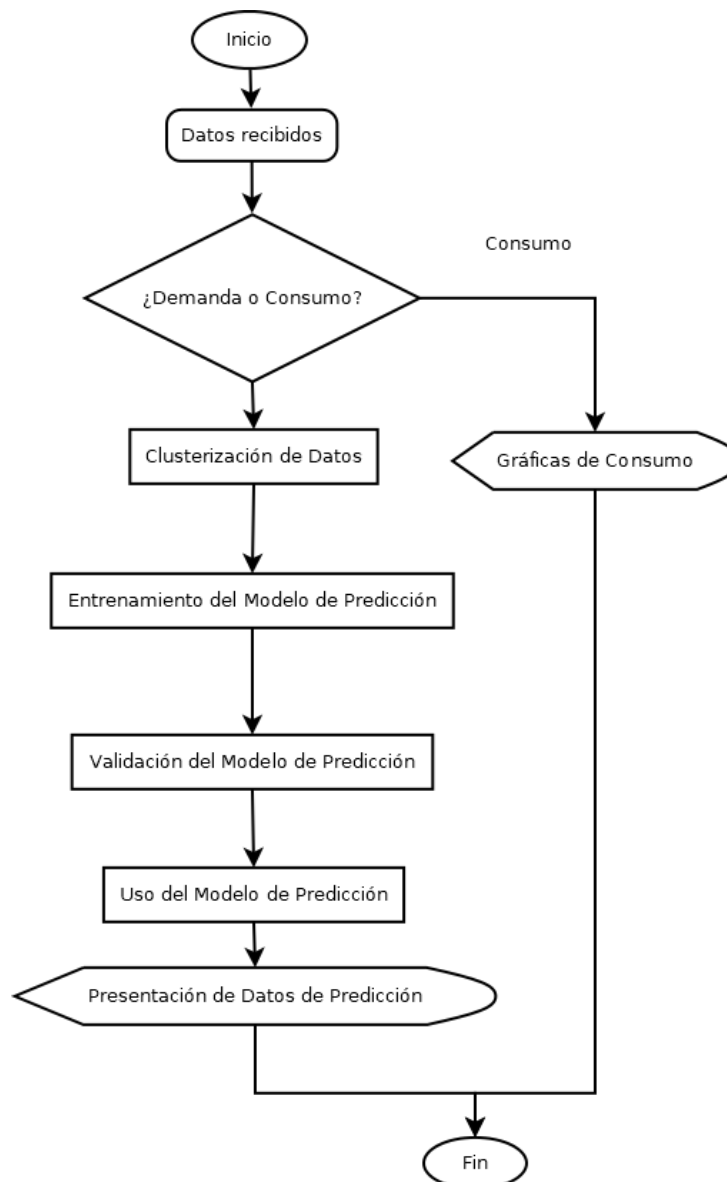


Figura 6: Diagrama de procesos de análisis de datos

Por una parte, tenemos los **datos históricos de consumo** que nos sirven para observar y analizar el comportamiento en años, meses, días y horas de cada contrato. Por otro lado, los **datos de predicción de demanda energética** con los cuales se hará el entrenamiento del modelo de predicción y las pruebas correspondientes.

3.3.1 *Análisis de datos históricos*

Para el análisis de datos históricos se usan los datos obtenidos de la fase de transformación y almacenados en la base de datos relacional.

Para ver la evolución de la demanda se usan los valores de consumo en kilovatios (kW), estos datos son presentados al usuario sin ningún tratamiento, ofreciendo la posibilidad de seleccionar un rango de fechas y mediante comparativas de consumo o mirando gráficamente valores extremos, el usuario pueda tomar decisiones, basado en gráficas similares a la de la Figura 7.

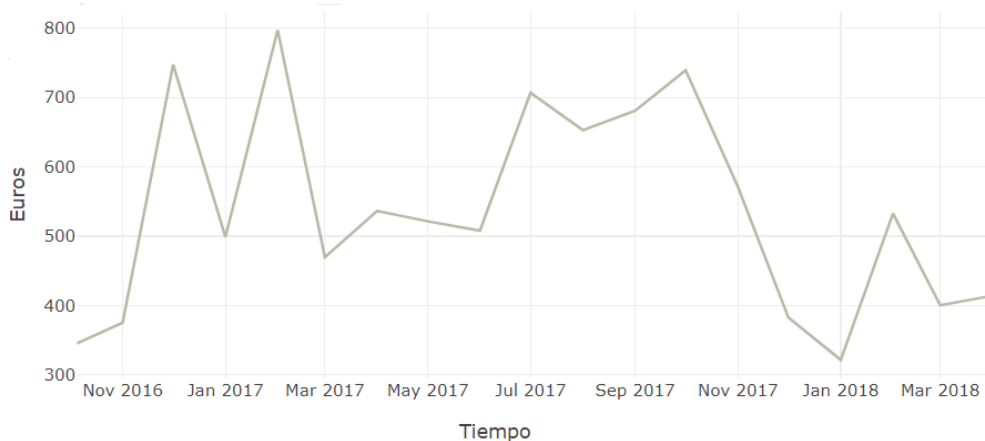


Figura 7: Ejemplo del consumo de energía eléctrica del *Campo de Tiro* en un rango específico de fechas (noviembre 2017-marzo 2018).

3.3.1.1 *Herramientas Utilizadas*

Para las gráficas de históricos se usó el lenguaje de programación Python con la herramienta Pyplot¹⁶ perteneciente a la librería Matplotlib¹⁷.

¹⁶https://matplotlib.org/api/pyplot_api.html

¹⁷<https://matplotlib.org/>

3.3.1.2 *Problemas Encontrados*

Durante el análisis gráfico de datos históricos se encontraron los siguientes problemas:

- Los gráficos presentan vacíos por los datos nulos encontrados.
- Los rangos de fechas en los datos son variables en cada contrato.

3.3.2 *Análisis de datos de predicción*

Analizando los datos obtenidos en la etapa de transformación, se observó que los datos de los años de 2015 y 2016, además del mes de enero de 2017 presentan datos que no han sido registrados ni capturados. Para la predicción de demanda de energía se ha preparado un set de datos compuesto por un año y un mes, comprendido desde el 1 de febrero 2017 hasta el 28 de febrero 2018. Las unidades de los datos son vatios y la periodicidad de cada lectura de estos datos es horaria.

De los 95 contratos eléctricos que se poseen, los que cumplen el requerimiento anterior son un total de 30 contratos. Dentro de estos 30 contratos se observa que existen 2 tipos de consumos, un primer grupo que identifica el consumo de edificios de oficinas públicas o instalaciones que ofrecen algún servicio a la comunidad (como por ejemplo piscinas públicas) y un segundo compuesto por el alumbrado público. Estos perfiles de consumo son muy importantes porque caracterizan el comportamiento de cada uno de los contratos dentro de una determinada tipología que se verá reflejado en su demanda energética.

3.3.2.1 *Herramientas Utilizadas*

Para el desarrollo se configuró un entorno basado en el lenguaje programación Python, con las librerías Sklearn¹⁸ para la clasificación y predicción de datos de acuerdo a su comportamiento, Pandas¹⁹ para graficar datos y resultados, Statsmodels²⁰ para estadísticas y resultados, Matplotlib²¹ para gráficos estadísticos, Math para funciones matemáticas y para agrupar los tipos de datos basado en su comportamiento se utilizó la librería Kmeans²² incluida dentro de Sklearn.

¹⁸<http://scikit-learn.org>

¹⁹<https://pandas.pydata.org/>

²⁰<https://www.statsmodels.org/>

²¹<https://docs.python.org/3.0/library/math.html>

²²<http://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

Siguiendo el proceso de predicción y debido a la ausencia de datos de consumo y datos nulos dentro de las series de datos antes mencionadas, se procedió a solventar/**rellenar** mediante la técnica de propagación hacia delante que utiliza el promedio de los valores de una hora antes y una hora después, con ello se evita datos extremos que dañen las predicciones.

Para la predicción de los datos se seleccionaron dos modelos de los antes mencionados en el estado del arte, un modelo lineal simple y un modelo autorregresivo ARIMA que serán descritos durante el desarrollo del proceso de predicción que se presenta en el diagrama de procesos de la Figura 8.

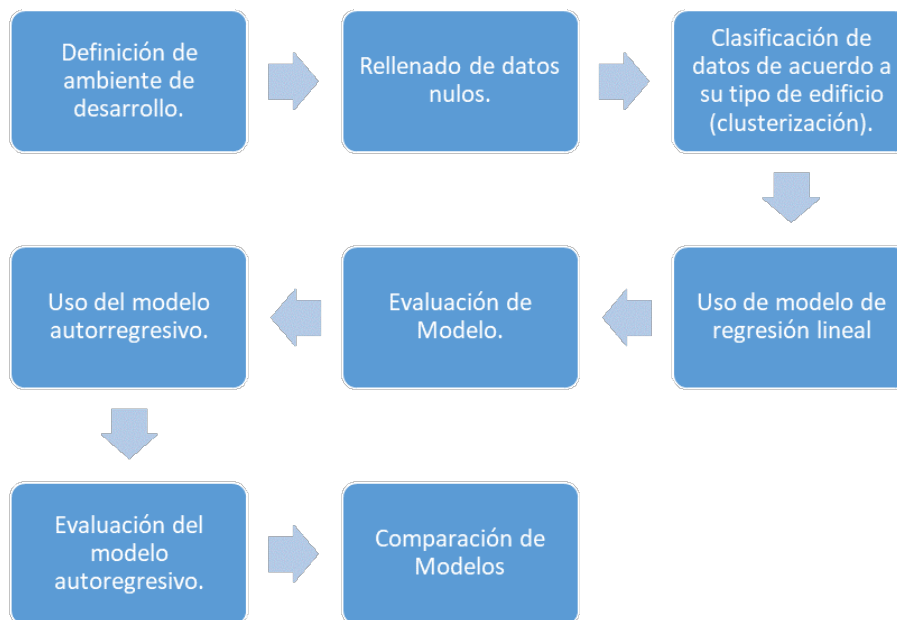


Figura 8: Diagrama de proceso de predicción de demanda

Como siguiente paso, para la **clasificación/clusterización** de datos de acuerdo a su comportamiento se separó los datos en 2 grupos basándose en el valor medio más cercano a la observación (*K-means*²³) [28]. En el caso específico se separa cada contrato tomando su valor medio y comparando entre sí para al final generar 2 perfiles de consumo que serán utilizados para crear el modelo de predicción. En la Figura 9 se puede observar la dispersión de las medias de los consumos a partir de los datos de demanda, con lo que se distingue dos grupos marcados para clasificación.

²³K-medios, método de ordenamiento de datos basado en su media

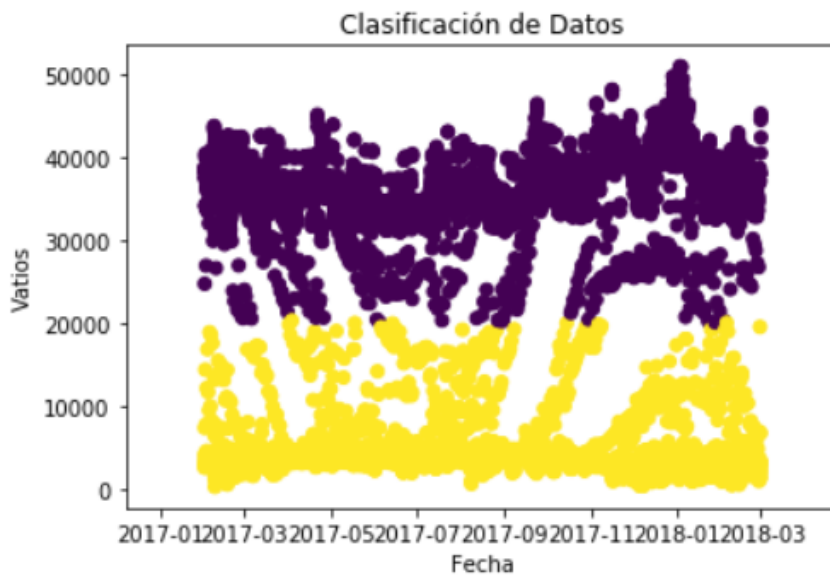


Figura 9: Distribución de las medias de los consumos

A partir de la clasificación en grupos que se puede observar en la Figura 9 los perfiles de consumo en un día específico se pueden observar en la Figura 10.

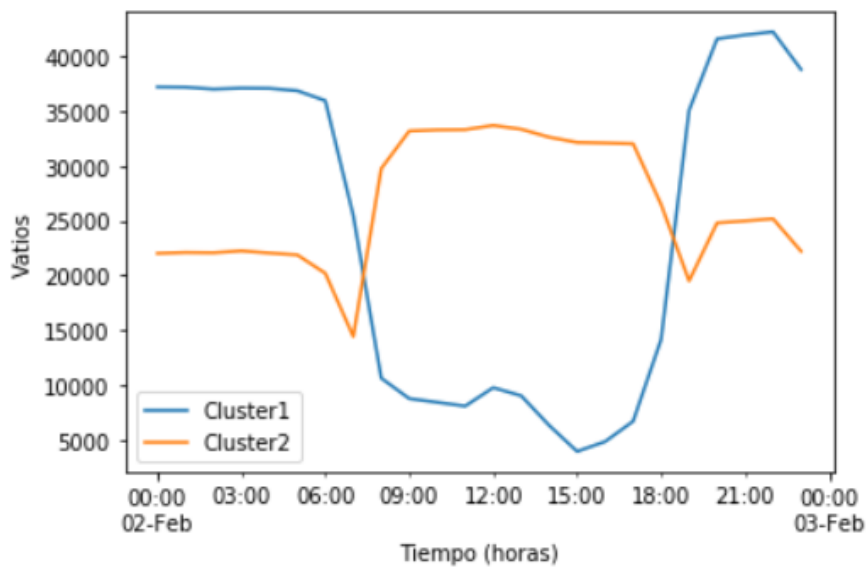


Figura 10: Perfiles de Consumo en un día determinado basados en la clasificación

Como se puede observar en la Figura 10 el perfil de Clúster 1, que corresponde al alumbrado público, tiene un comportamiento encendido en horarios nocturnos

y apagado durante el día; mientras que el Clúster 2 es de un perfil de oficinas en el que en horarios de trabajo sube su demanda energética, pero en la noche mantiene una demanda baja.

Una vez clasificados los datos, continuamos con el proceso de predicción y para ello se busca una primera aproximación de **predicción** basándose en un **modelo de regresión lineal simple** de pronóstico progresivo, es decir en base a los datos actuales predecir el siguiente paso, agregarlo al set de datos y volver a realizar el procedimiento de predicción del siguiente y así sucesivamente hasta obtener el total de los datos necesarios en la predicción [29]. Para este primer acercamiento se usan el 50 % del set total para entrenar y posteriormente evaluarlo con la siguiente mitad de los datos.

A continuación, para la **evaluación** del modelo de regresión lineal simple se usa el error cuadrático medio (Root Mean Square Error (RMSE)) de tal manera que los valores de error extremos tengan un peso más grande dentro de la evaluación del modelo. Adicional a ello el RMSE nos da un valor en vatios lo cual facilita la lectura al usuario. También se evalúa con el error absoluto medio (Mean Absolute Error (MAE)) para no dar un peso excesivo a los valores extremos y el error porcentual absoluto medio (Mean Absolute Percentage Error (MAPE)), para manejar un porcentaje de error sin relación a las unidades de medida utilizadas (vatios).

El modelo de regresión lineal simple nos da un punto de partida para la predicción, pero puede ser mejorado mediante parámetros o configuraciones que nos den mejores resultados, como lo es el caso de los **modelos lineales autorregresivos** como ARIMA [30]. ARIMA tiene la fortaleza en predicción usando series temporales [30]. Para el modelo ARIMA parametriza al modelo según el número de saltos que queremos dar en la regresión del modelo o el número de medias que queremos tener dentro del conjunto total de datos.

Para obtener el mejor resultado se debe realizar la combinación de parámetros de ARIMA de manera secuencial uno a uno hasta conseguir bajar el error de la predicción. Para lograrlo se ejecutó el modelo durante 15 veces, combinando cada uno de los parámetros con valores de 0 a 4 como lo recomienda Grillenzoni [31]. A partir de los parámetros conseguidos se debe evaluar el modelo mediante el uso de RMSE , MAE y MAPE y en caso de tener predicciones muy erróneas volver a configurar los parámetros antes mencionados hasta tener unos estadísticos bajos y el modelo pueda ser usado como definitivo.

Una vez configurado el modelo se evalúa los resultados gráficamente, usan-

do parte de los datos y graficando valores reales contra valores de la predicción. Como se puede observar en la Figura 11, donde en azul se observan los valores reales y en rojo los valores de la predicción.

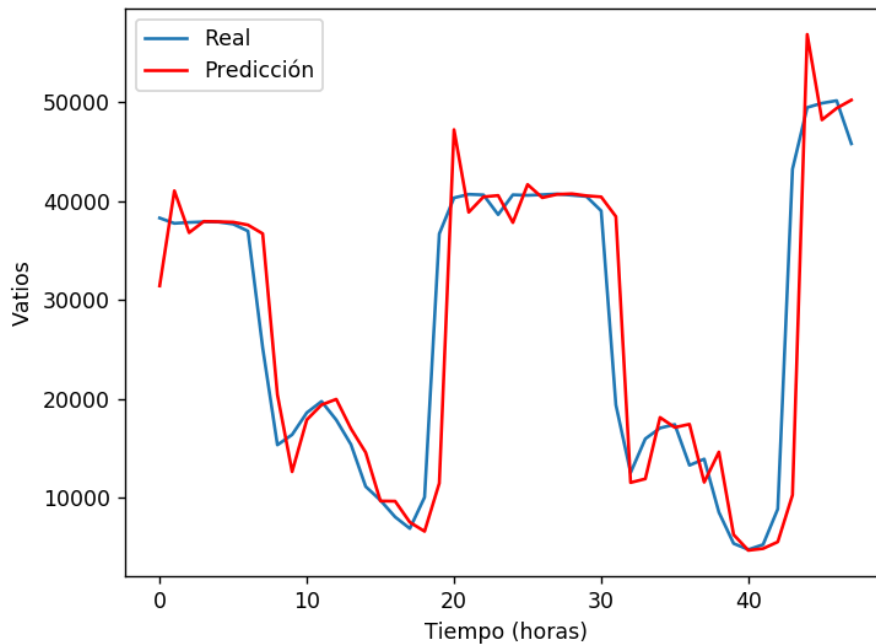


Figura 11: Valores reales y valores de predicción

3.3.2.2 Problemas Encontrados

En el proceso de análisis de datos de predicción se encontraron los siguientes problemas:

- Valores extremos que complicaban el entrenamiento del modelo.
- Falta de datos de más años para poder entrenar el modelo de mejor manera.

3.4 Presentación de datos

Los datos de las etapas de transformación y predicción son presentados de manera ordenada, interactiva y fácil de interpretar en una interfaz que ayuda al usuario a la toma de decisiones. Para ello se siguió el proceso de trabajo descrito en el diagrama de la Figura 12.

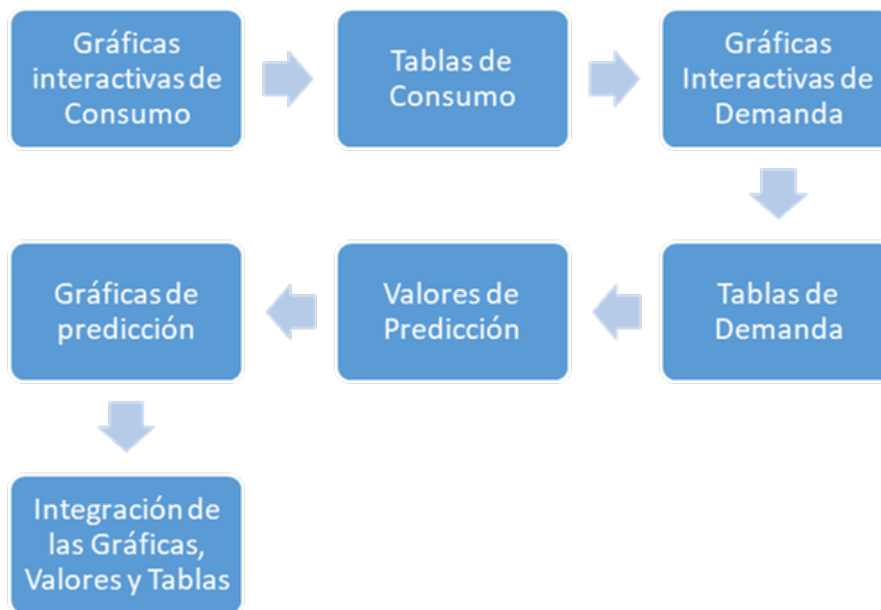


Figura 12: Diagrama de Proceso de Presentación de Datos

3.4.1 Herramientas Utilizadas

El entorno configurado con lenguaje Python. Se usó la librería Plotly²⁴ para las gráficas interactivas y tablas de los históricos tanto de consumo como demanda y predicción. En tanto que para la integración de las gráficas, valores y tablas se utiliza el cuadro de mando de Plotly offline²⁵.

Las gráficas permiten al usuario seleccionar intervalos de tiempo o el número de contrato del cual quiere observar los datos graficados, además permite ver el valor exacto en un punto determinado.

Las gráficas tanto de históricos de consumo, como de históricos de demanda constan de las siguientes partes (Figura 13):

1. Botones de selección de intervalos de tiempo. Estos botones son variables en consumo y demanda porque los intervalos son distintos de acuerdo al tipo de dato (horario o mensual).
2. La opción para seleccionar el contrato para mostrar su gráfica a partir de una lista desplegable con todos los contratos disponibles.

²⁴<https://plot.ly/>

²⁵<https://plot.ly/python/offline/>

3. Barra de navegación para selección de intervalos de fechas en un rango específico, como en los botones de selección de intervalos. Una barra que pueda seleccionar un rango de fechas arrastrando la barra en el tiempo que se requiera.
4. El espacio con la gráfica resultante con las selecciones indicadas, en las que acercando el ratón podemos observar el valor específico de la gráfica en el punto seleccionado.

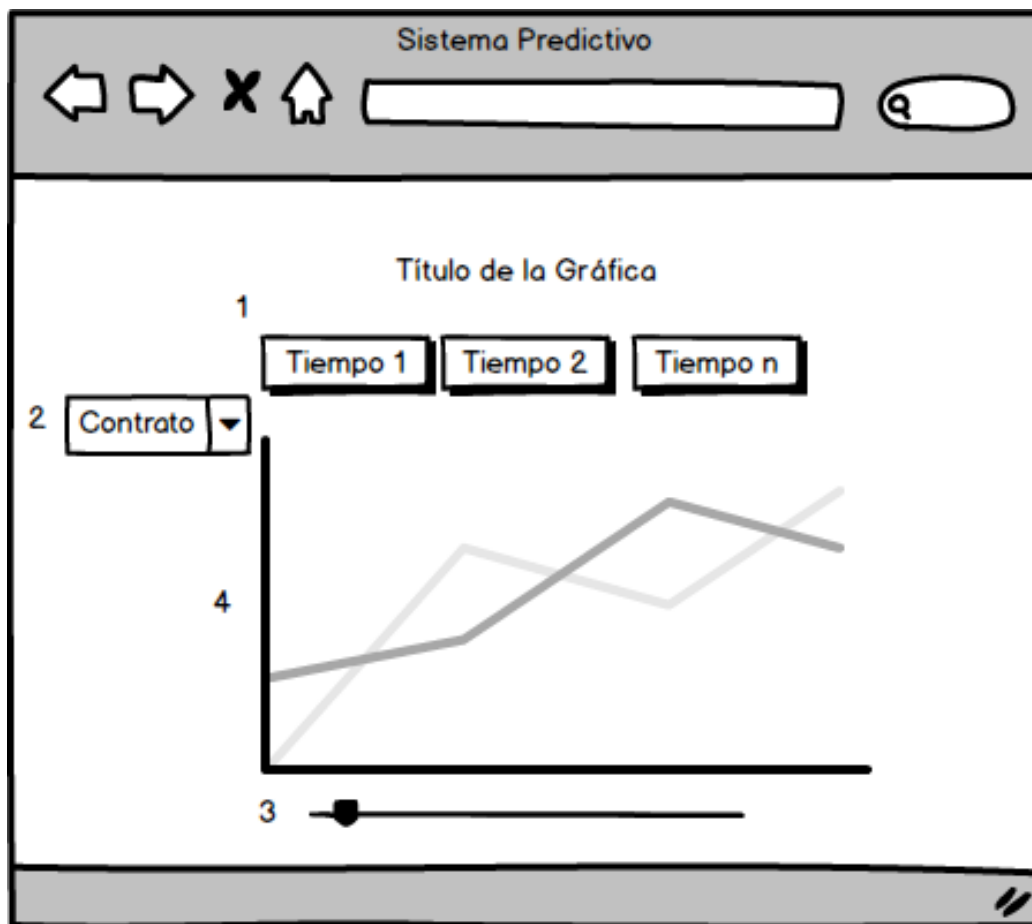


Figura 13: Partes de las Gráficas de consumo y demanda

En la Figura 14 se puede observar un ejemplo de gráfica de histórico de consumo por mes, mientras que en la Figura 15 se puede observar la gráfica de histórico de demanda. Se debe considerar que los datos de demanda son horarios, mientras que los de consumo son mensuales por lo que los intervalos de selección serán distintos entre una y otra gráfica.

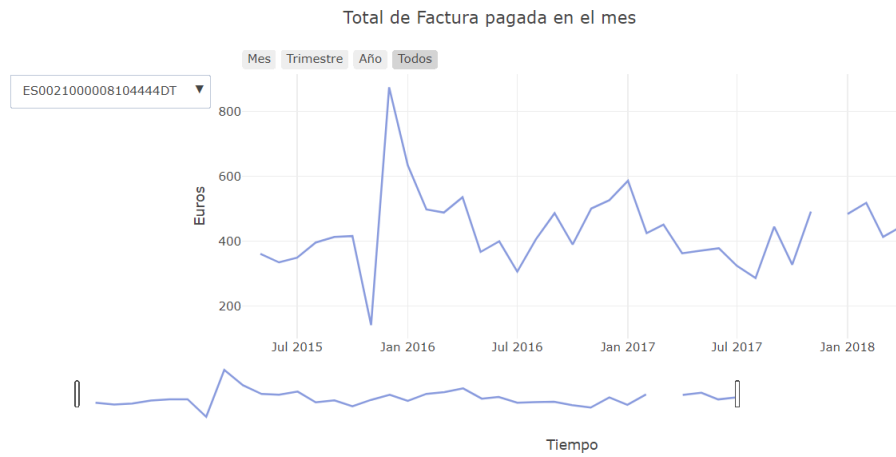


Figura 14: Gráfica ejemplo de históricos de consumo para un determinado contrato

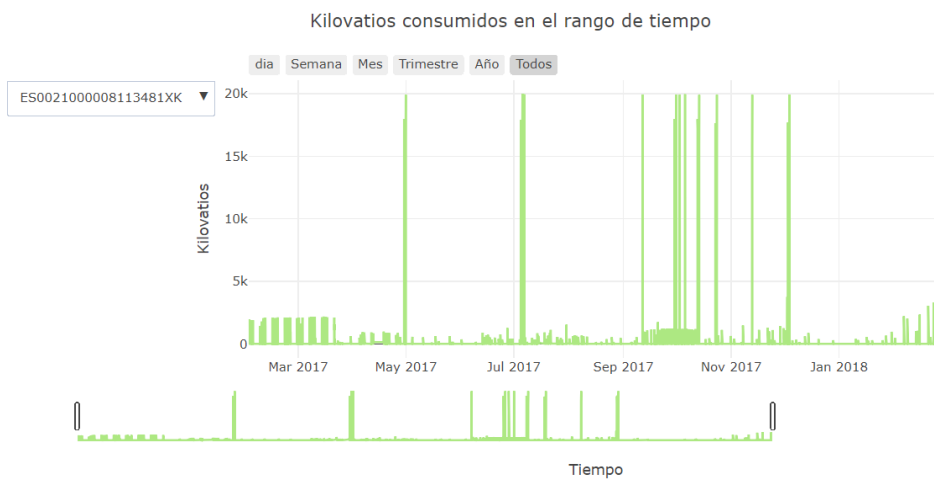


Figura 15: Gráfica ejemplo de históricos de demanda para un determinado contrato

Las tablas tanto de consumo y demanda se presentan en dos columnas, una con la fecha y otra con el valor en la fecha indicada, para cada contrato por separado, como indica en la Figura 16.

2017-02-01 01:00:00	16813
2017-02-01 02:00:00	16735
2017-02-01 03:00:00	16757
2017-02-01 04:00:00	16571
2017-02-01 05:00:00	16509
2017-02-01 06:00:00	16684
2017-02-01 07:00:00	17901
2017-02-01 08:00:00	5194

Figura 16: Tabla ejemplo de demanda energética

En el anexo 2 se puede observar la interfaz del sistema.

Capítulo 4

Resultados y Discusión

Para la obtención de la predicción se usaron dos modelos, regresión lineal simple y un modelo ARIMA. Para la evaluación de los mismos se utilizó RMSE (Root Mean Square Error), MAPE (Mean absolute percentage error) y MAE (Mean Absolute Error), los cuales son presentados en la Tabla 4.1.

	RMSE (kW)	MAE (kW)	MAPE (%)
Regresión Lineal Simple	5,888	4,187	14,24 %
ARIMA	6,275	3,547	10,74 %

Tabla 4.1: Comparativa de resultados estadísticos de los modelos de predicción

Teniendo en cuenta estos resultados se puede observar que el modelo lineal trabaja mejor para predicción de valores extremos porque RMSE valora más este tipo de valores y nos da un valor menor a los 6 Kilovatios (kW), pero tanto MAE como MAPE nos da un valor menor en ARIMA por lo que en general dicho modelo trabaja mejor para la mayor cantidad de valores exceptuando los extremos antes mencionados. Esto se puede mirar en la Figura 17 con la predicción de 48 horas con los datos utilizados.

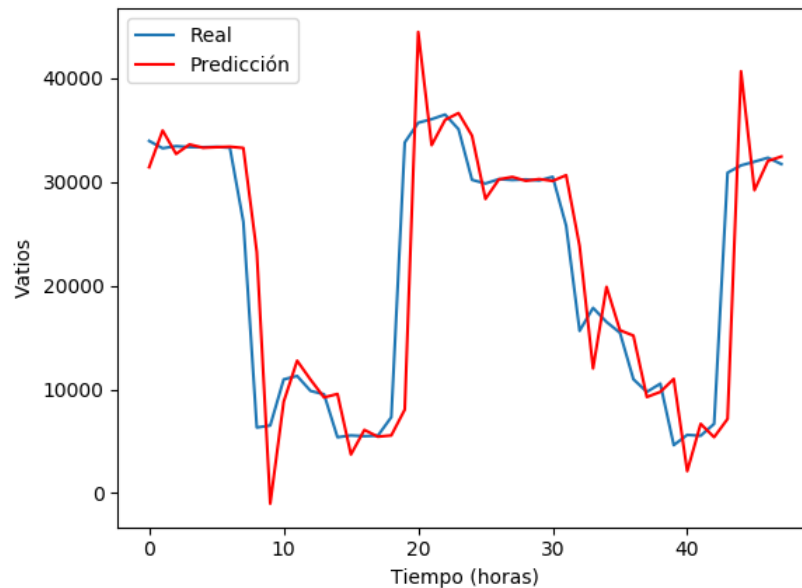


Figura 17: Ejemplo de Comparativa de datos reales y predicción de 48 horas

Los resultados de las predicciones fueron elaborados con los datos de históricos de demanda obtenidos de las etapas de desarrollo, pero como se mencionó en el capítulo 3, los datos no estaban completos, por lo que la calidad del dato para predicción fue muy deficiente con un porcentaje de 31,58 % de datos utilizables como se presenta en la Tabla 4.2, siendo necesario rellenar datos para completar un set y poder realizar la predicción.

	#
Contratos Totales	95
Contratos con Datos Demanda	54
Contratos con datos de demanda de 1 año completo	30
% Datos Utilizables	31,58 %

Tabla 4.2: Datos de demanda disponibles

Por este motivo se plantea una alternativa para automatizar la toma de datos con una arquitectura IoT, de manera que los datos capturados tengan un mayor porcentaje para ser utilizados.

La arquitectura planteada se basa en una arquitectura IoT en conjunto con una

nube de servicios con componentes dentro de ella separados en capas como se muestra en la Figura 18 y consta de los siguientes componentes/capas:

- La **Capa de Aplicación** encargada de presentar al usuario la interfaz y las opciones disponibles para el manejo de los datos mediante la aplicación desarrollada. A su vez es la encargada de dar la seguridad de accesos con registros de usuarios y controles de inicio de sesión a la aplicación, además de separar la visualización para diferentes tipos de usuarios determinados. La capa de aplicación consta de dos tipos de vistas:
 - La primera vista de **históricos** con gráficas y tablas para el usuario de datos que se tiene hasta el momento de la consulta y pueden ser observados de manera interactiva.
 - La segunda vista del **predictor**, que utiliza el componente desarrollado de los modelos de predicción para los datos de demanda y presenta al usuario los resultados.
- La **Capa Intermedia de Separación y Calidad**, nombrada como intermedia porque es la encargada de ordenar los datos recibidos de las capas inferiores y enviarlos de manera ordenada y que sean utilizables para la capa superior de aplicación. Para ello tiene dos componentes principales:
 - **El Separador de Datos** será el encargado de recibir los datos y ordenarlos como se establece en la aplicación para después almacenarlos para ser usados. El separador recibe datos en formato CSV y envía en formato de consultas a la base de datos.
 - **El Controlador de Calidad de Datos** será el encargado de verificar que el dato recibido sea el correspondiente al tiempo en el que lo recibe, caso contrario enviará una alerta bidireccional tanto a la aplicación como a las capas inferiores a manera de aviso para comprobar o reenviar el dato. Una función adicional del controlador de calidad es el relleno de datos en caso de que en un intervalo de tiempo determinado por el usuario no se reciba el dato requerido y el mismo no se pueda reenviar por los sensores de la capa de control.
- Junto a la capa de aplicación y la capa intermedia, se encuentra el **almacén de datos**, el cual garantiza la persistencia de los datos ya organizados y servirá como acceso para la aplicación.
- La **Capa de Controlador** es la encargada de ensamblar los diferentes tipos de nodos y los diferentes tipos de datos que envían los nodos, extrayendo

solamente los datos necesarios de los mismos. Para ello debe tener una base de tipos de controladores necesarios para la conexión y comunicación con la diversidad de nodos finales. La capa de controlador establece que los datos recibidos por cada tipo de nodo sean transformados en ficheros con formato CSV para enviar a la capa intermedia.

- Para la seguridad de la arquitectura se cuenta con el **Firewall**, encargado de controlar los accesos de los nodos y evitar intrusos en las conexiones y en la arquitectura.
- Los **nodos** finales con conexión inalámbrica a la red son los contadores digitales que ya son homologados por la empresa proveedora del servicio²⁶ que disponen de telegestión, es decir son autónomos y poseen conexión directa a la red a la que sean conectados.

²⁶<https://www.iberdroladistribucion.es/redes-inteligentes/contador-telegestionado/contadores-digitales-homologados#1>

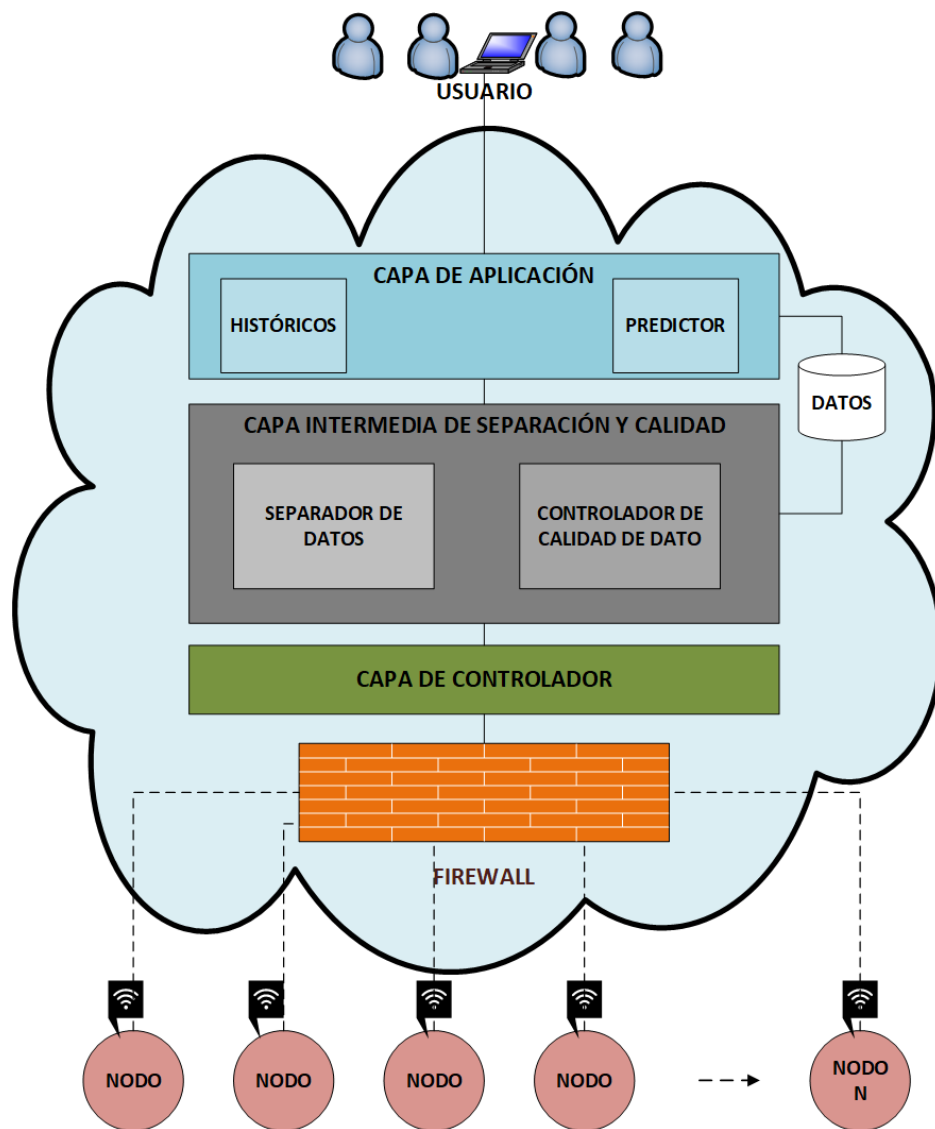


Figura 18: Arquitectura propuesta de la red

Esta arquitectura es posible gracias a las características de los contadores digitales que disponen de servicios de conexión sea mediante ethernet o mediante conexión inalámbrica a la red local. Esta arquitectura a su vez está conectada. Con dicha arquitectura se debe generar un flujo de trabajo secuencial que garantice que el dato final presentado al usuario sea correcto, este flujo se puede observar en la Figura 19.

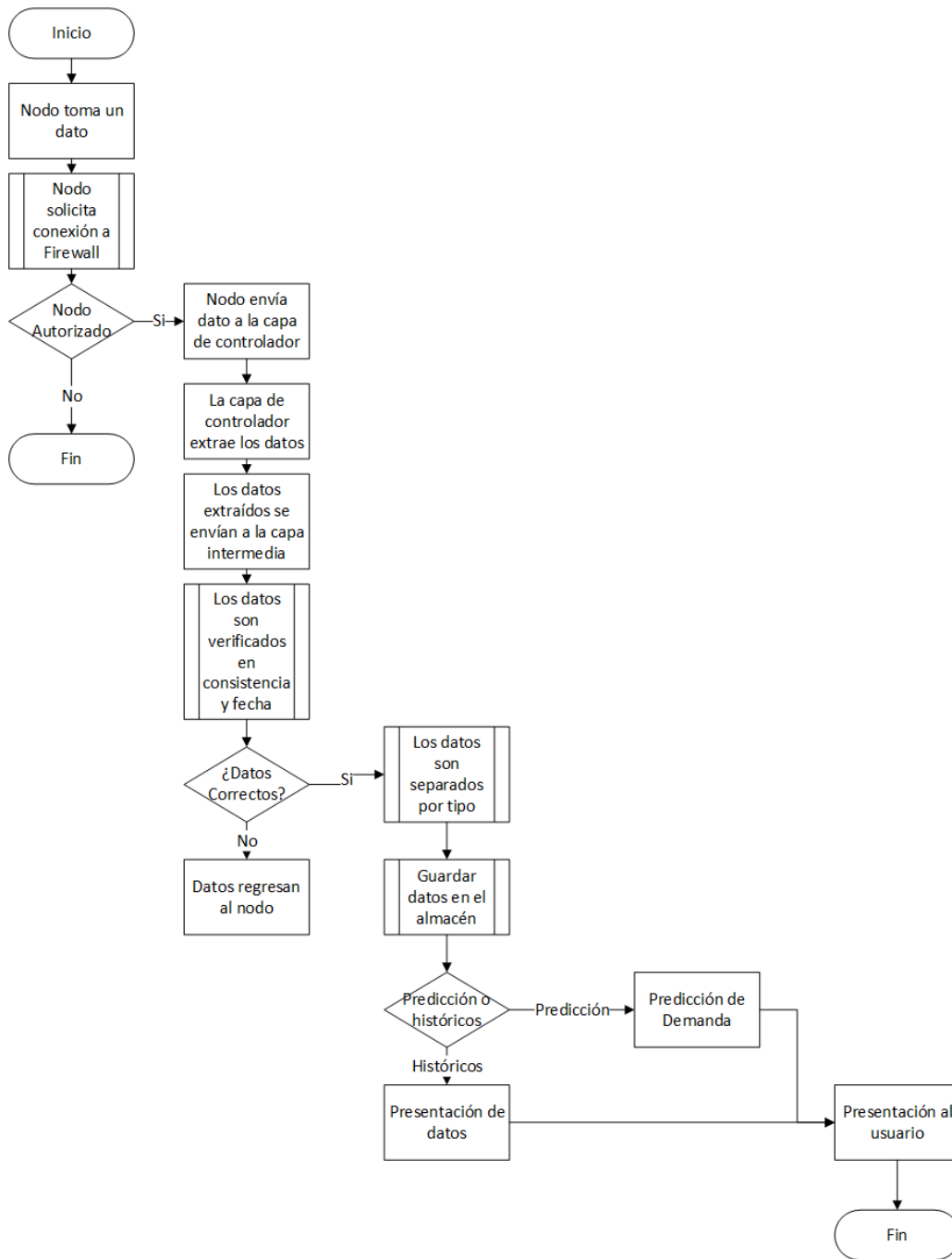


Figura 19: Diagrama de procesos de la arquitectura planteada

Capítulo 5

Conclusiones y Trabajo Futuro

5.1 Conclusiones

Este trabajo de Fin de Máster presenta un sistema para la predicción de demanda energética basado en la recolección de datos de diferentes fuentes. Se ha comprobado que los datos proporcionados por la compañía eléctrica tienen una calidad deficiente por la cantidad de datos vacíos y nulos que presentan; por este motivo se realizó el diseño teórico de una arquitectura de red basada en dispositivos del internet de las cosas (IoT) para obtener datos directamente desde las fuentes de información de energía como son los contadores digitales inteligentes.

Para el tratamiento de grandes cantidades de datos de demanda y consumo energético se utilizaron herramientas escritas en lenguaje de programación Python. Los datos que antes eran oscuros para el usuario fueron organizados mediante herramientas visuales, específicamente a manera de cuadro de mando para que el usuario final pueda conocer el consumo que presentan los diferentes contratos eléctricos y pueda actuar de acuerdo a la información en la toma de decisiones.

Las grandes cantidades de datos históricos de demanda energética son utilizados como una entrada para nuestro sistema de predicción. La predicción, sea de demanda o consumo energético, tiene muchas alternativas para ser trabajadas y con el paso del tiempo genera mayores variantes que son influenciadas por normativas vigentes o por capacidades de los equipos que generan o administran energía; es por ello por lo que la actualización de capturadores de datos y predictores debe ser constante para mejorar la calidad de los datos y las predicciones.

Para la predicción de la demanda se usaron dos modelos. El modelo ARIMA es una alternativa bastante buena para la predicción de la demanda con series de

tiempo. El modelo ARIMA cumple con las necesidades de predicción y aproxima su predicción a los datos reales incluyendo los valores extremos.

5.2 Trabajo futuro

Como trabajo futuro se plantea la implementación física de la arquitectura de red para la captura de datos basándonos en el diseño de la plataforma planteada en este trabajo, de manera tal que se mejorará la calidad de los datos que alimentan al sistema de predicción. También para obtener mejores resultados al momento de predecir series de tiempo, resultaría interesante utilizar modelos de predicción alternativos como redes neuronales o modelos variantes al modelo ARIMA.


Finalmente, se plantea como trabajo futuro buscar una manera con la que se pueda ayudar a los encargados de ayuntamientos/municipios a saber cuál es el mejor momento para la compra de la energía en el mercado eléctrico, realizando la predicción no solamente de la demanda sino también del precio basándose en los datos históricos disponibles acerca del coste de la energía en el mercado energético.

Anexos

5.3 Anexo 1 Tipos de Facturas

Ejemplo de cada tipo de factura detectado

IBERDROLA GENERACIÓN, S.A.U. Hoja número 1 / 3



1 DATOS DEL CLIENTE

CIF: [REDACTED]
 Dirección de suministro: [REDACTED]
 46160 LLIRIA (VALENCIA)

FACTURA DE ELECTRICIDAD

Referencia contrato: [REDACTED]
 Período de facturación 28/11/2013 - 31/01/2014
 Fecha factura 27 de febrero de 2014
 N° factura: [REDACTED]

IMPORTE FACTURA 3.606,50 €

IBERDROLA GENERACIÓN, S.A.U. P.O. BOX 1000, 5 46100 LLIRIA
 IN Nº. NC. 00615637 D.S.
 [REDACTED]
 46160 LLIRIA (VALENCIA)

Dirección fiscal P.O. BOX 1000 46100 LLIRIA

2 FACTURACIÓN EUROS

ENERGÍA		
Potencia facturada 01/11/2013-31/01/2014	13,2 kW x 64 días x 0,115129 €/kW	97,26
Energía facturada 01/11/2013-31/01/2014	17.061 kWh x 0,168601 €/kWh	2.876,50
TOTAL ENERGÍA		2.973,76
SERVICIOS Y OTROS CONCEPTOS		
Alquiler equipos medida (0,015 €/día x 64 días)	64 días x 0,106521 €/día	6,82
TOTAL SERVICIOS Y OTROS CONCEPTOS		6,82
IMPORTE TOTAL		2.980,58
IVA	21 % s./2.980,58 €	625,92
TOTAL IMPORTE FACTURA		3.606,50

Inscripción en el Registro Mercantil de Valencia al Tomo 3603, Folio 176, Sección 1ª, Inscripción 1ª - 4

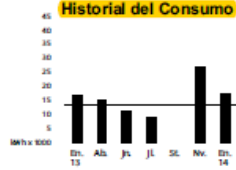
Figura 20: Factura Tipo 1 Parte 1

3 DATOS DEL CONTRATO

Referencia contrato [REDACTED]
 CUPS [REDACTED]
 CNAE [REDACTED]
 Fecha Fin Contrato 01/02/2014
 Potencia contratada 13,2 kW
 Tarifa ATR 2.1 A Precios B.O.E. del 03/08/2013
 Número de póliza del contrato de acceso [REDACTED]

4 CONSUMO

Historial del Consumo



Su consumo medio diario en este último periodo facturado ha sido: 56,35 €
 Su consumo medio diario en los últimos 12 meses ha sido: 46,79 €

Nº contador	Periodo horario	Desde	Lectura	Hasta	Lectura	Consumo/Potencia
0021345444	TOTAL	28/11/2013	003047	31/01/2014	020108	17.061 kWh
0099574698	REACTIVA	28/11/2013	108451	31/01/2014	108458	7 kvArh

Registro Mercantil de Baleares al Tomo 3803, Folio 199, Página 01-27039, Inscripción 1ª - 1

Figura 21: Factura Tipo 1 Parte 2

IBERDROLA CLIENTES, S.A.U.
 CIF A- [REDACTED]

DATOS DE FACTURA

Periodo de facturación 17/03/2015 - 09/04/2015
 Número de factura [REDACTED]
 Fecha de emisión de factura 29 de abril de 2015
 Fecha prevista de cargo 01/06/2015
 Factura con lectura real
 Título [REDACTED]
 CIF titular [REDACTED]
 Referencia contrato suministrado [REDACTED]

OP. COMP. [REDACTED]
 OPER. [REDACTED]
 OPER. [REDACTED]

TOTAL IMPORTE FACTURA: 952,28 €

RESUMEN DE FACTURACIÓN

ENERGÍA	784,56 €
SERVICIOS Y OTROS CONCEPTOS	2,45 €
IVA 21% s/787,01 €	165,27 €
TOTAL A PAGAR	952,28 €

FACTURA DE ELECTRICIDAD Página 1/2

IBERDROLA CLIENTES, S.A.U. Apartado de Correos 61 015 20180 Madrid
 N.º de identificación fiscal: [REDACTED] IREG: 038467 03491 201504

**Plza MAYOR, 1
 AYUNTAMIENTO DE LLIRIA
 46160 LLIRIA (VALENCIA)**

Dirección de suministro: [REDACTED]

Figura 22: Factura Tipo 2 Parte 1

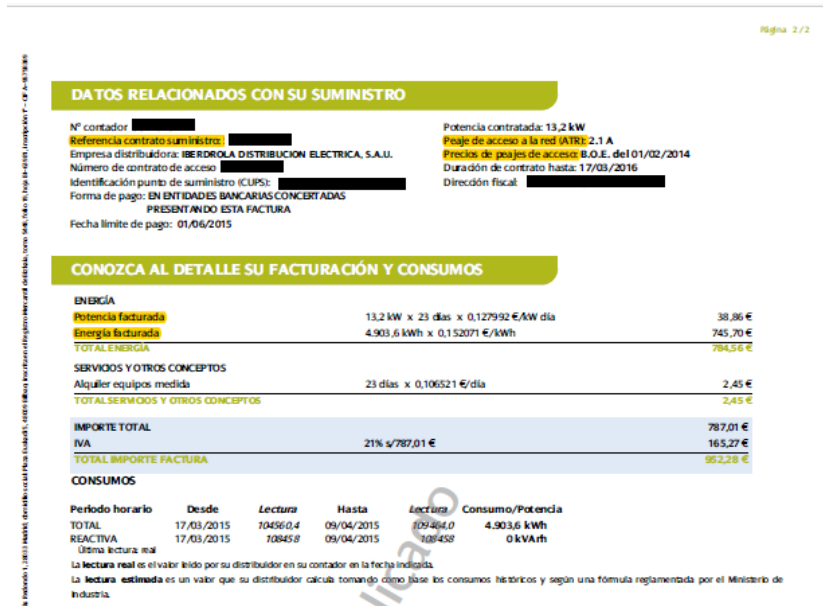


Figura 23: Factura Tipo 2 Parte 2



Figura 24: Factura Tipo 3 Parte 1

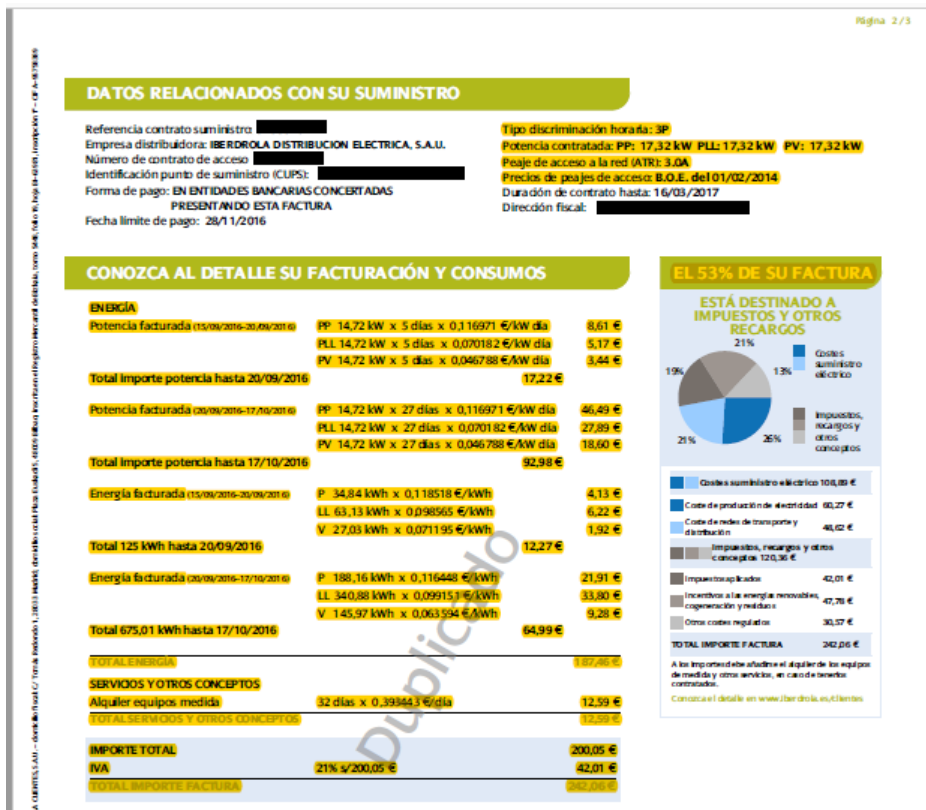


Figura 25: Factura Tipo 3 Parte 2

5.4 Anexo 2 Presentación

Imágenes de la interfaz de usuario.

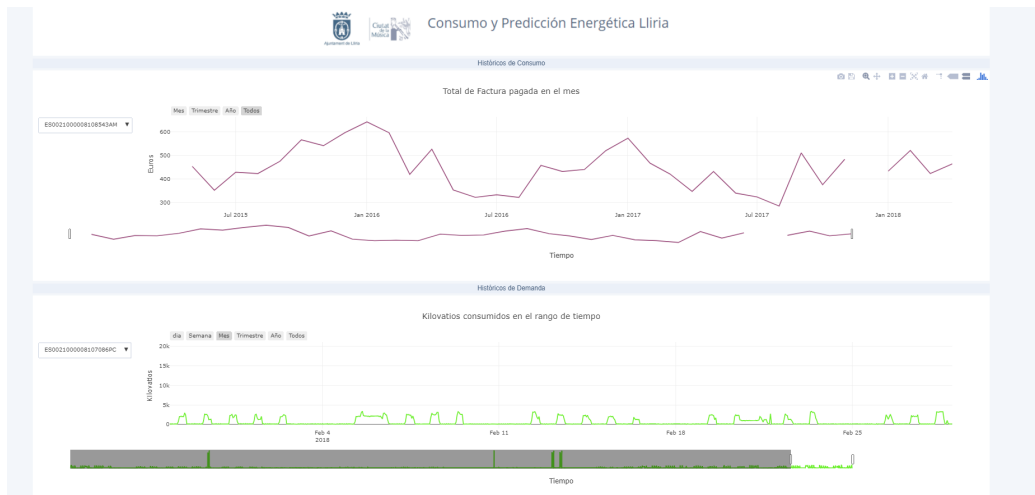


Figura 26: Vista de la mitad superior del sistema

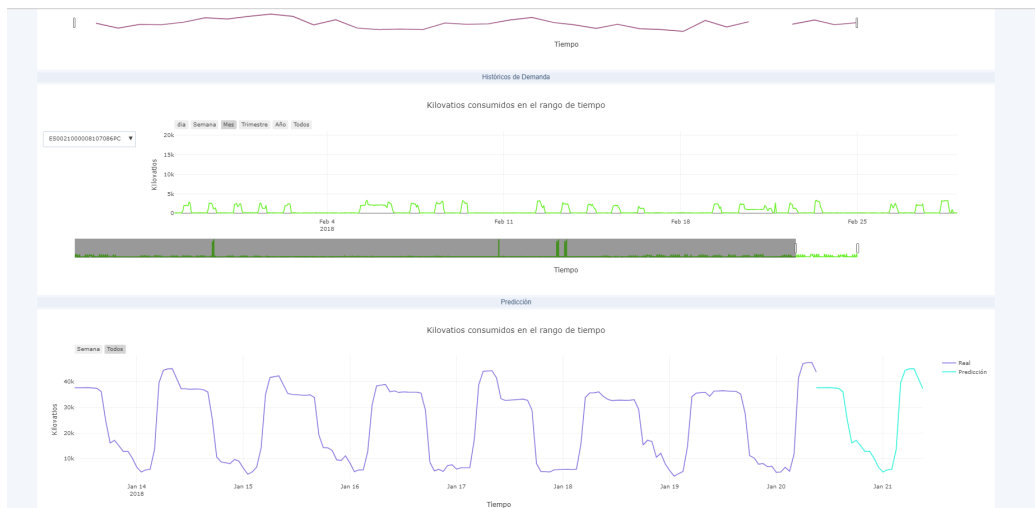


Figura 27: Vista de la mitad inferior del sistema

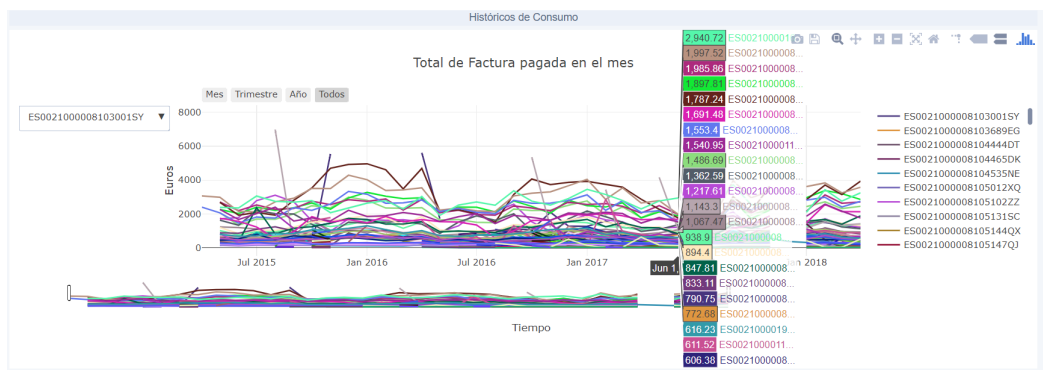


Figura 28: Gráficas de consumo en el sistema

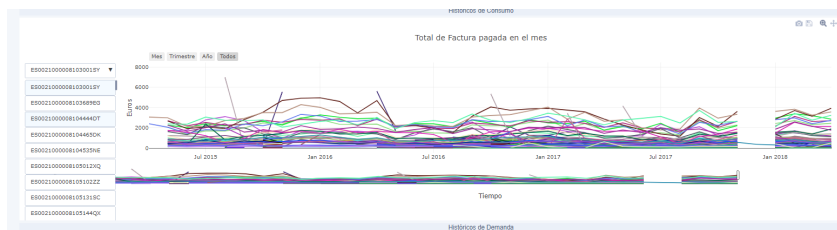


Figura 29: Selección de contratos en el sistema



Figura 30: barra de navegación en el sistema

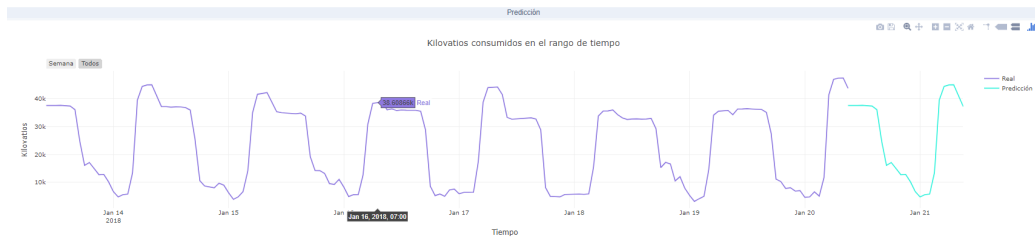


Figura 31: Datos de predicción en el sistema

Lista de Acrónimos

IoT Internet of Things

kW Kilovatios

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

RMSE Root Mean Square Error

RUP Rational Unified Process

SOA Service Oriented Architecture

Bibliografía

- [1] E. R. Díez and M. G. Pérez, “Políticas y medidas contra la pobreza energética; a quién le corresponde?” *Areas. Revista Internacional de Ciencias Sociales*, no. 36, pp. 153–165, 2017.
- [2] A. G. Alonso, “Análisis del mercado eléctrico español,” p. 41, 2014.
- [3] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, “Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 896–913, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.ijforecast.2016.02.001>
- [4] H. Verdejo, A. Awerkin, C. Becker, and G. Olguin, “Statistic linear parametric techniques for residential electric energy demand forecasting. A review and an implementation to Chile,” *Renewable and Sustainable Energy Reviews*, vol. 74, no. January, pp. 512–521, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.rser.2017.01.110>
- [5] M. Kankal and E. Uzlu, “Neural network approach with teaching–learning–based optimization for modeling and forecasting long-term electric energy demand in Turkey,” *Neural Computing and Applications*, vol. 28, no. s1, pp. 737–747, 2017.
- [6] S. Kucukali and K. Baris, “Turkey’s short-term gross annual electricity demand forecast by fuzzy logic approach,” *Energy Policy*, vol. 38, no. 5, pp. 2438–2445, 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.enpol.2009.12.037>
- [7] P. Koprinkova-Hristova, *Artificial Neural Networks Methods and Applications in Bio-/Neuroinformatics*, 2014.
- [8] C. García-Ascanio and C. Maté, “Electric power demand forecasting using interval time series: A comparison between VAR and iMLP,” *Energy Policy*, vol. 38, no. 2, pp. 715–725, 2010.

- [9] A. A. Mati, M. Eng, B. G. Gajoga, D. Ph, B. Jimoh, and A. Adegoby, "Electricity Demand Forecasting in Nigeria using Time Series Model ." vol. 10, no. 2, pp. 479–485, 2009.
- [10] V. Ediger and S. Akar, "ARIMA forecasting of primary energy demand by fuel in Turkey," *Energy Policy*, vol. 35, no. 3, pp. 1701–1708, 2007.
- [11] Y. C. Hu, "Electricity consumption prediction using a neural-network-based grey forecasting approach," *Journal of the Operational Research Society*, vol. 68, no. 10, pp. 1259–1264, 2017.
- [12] G. K. Tso and K. K. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761–1768, 2007.
- [13] H. T. Pao and C. M. Tsai, "Modeling and forecasting the CO₂emissions, energy consumption, and economic growth in Brazil," *Energy*, vol. 36, no. 5, pp. 2450–2458, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.energy.2011.01.032>
- [14] C. Yuan, S. Liu, and Z. Fang, "Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model," *Energy*, vol. 100, pp. 384–390, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.energy.2016.02.001>
- [15] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification," *New York: John Wiley, Section*, p. 69, 1995.
- [16] V. Estivill-Castro, "Why so many clustering algorithms - A Position Paper," *ACM SIGKDD Explorations Newsletter*, vol. 4, no. 1, pp. 65–75, 2002. [Online]. Available: <http://dl.acm.org/citation.cfm?id=568574.568575>
- [17] P. P. Rodrigues, J. Gama, and J. P. Pedroso, "Hierarchical Clustering of Time Series Data Streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 615–627, 2008.
- [18] A. Lavin and D. Klabjan, "Clustering time-series energy data from smart meters," *Energy Efficiency*, vol. 8, no. 4, pp. 681–689, 2015.
- [19] V. Niennattrakul and C. A. Ratanamahatana, "On clustering multimedia time series data using k-means and dynamic time warping," *Multimedia and Ubiquitous Engineering, 2007. MUE'07. International Conference on*, pp. 733–738, 2007.

- [20] N. N. Astakhova, L. A. Demidova, and E. V. Nikulchev, “Forecasting method for grouped time series with the use of k-means algorithm,” *Applied Mathematical Sciences*, vol. 9, no. 97, pp. 4813–4830, 2015. [Online]. Available: <http://www.m-hikari.com/ams/ams-2015/ams-97-100-2015/55391.html>
- [21] G. M. L. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4, Ed. Wiley, 2015.
- [22] T. C. Fu, “A review on time series data mining,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.engappai.2010.09.007>
- [23] J. Contreras, R. Espínola, F. J. Nogales, and A. J. Conejo, “ARIMA models to predict next-day electricity prices,” *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014–1020, 2003.
- [24] Paper Rational Software White, “Rational Unified Process Best Practices for Software,” *Development*, pp. 1–21, 2004. [Online]. Available: <http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Rational+Unified+Process+Best+Practices+for+Software{#}4>
- [25] L. González and F. Piedrabuena, “Reporte Técnico RT 06-16 Desarrollo de aplicaciones con enfoque SOA Andrea Delgado,” 2006.
- [26] P. Letelier, M. Canós, E. Sánchez, and M. Penadés, “Metodologías Ágiles en el Desarrollo de Software,” *Valencia, Valencia, España*, pp. 1–8, 2003. [Online]. Available: [http://www.carlosfau.com.ar/nqi/nqifiles/XP{_\]Agil.pdf](http://www.carlosfau.com.ar/nqi/nqifiles/XP{_]Agil.pdf)
- [27] F. A. Batarseh and A. J. Gonzalez, “Predicting failures in agile software development through data analytics,” *Software Quality Journal*, vol. 26, no. 1, pp. 49–66, 2018.
- [28] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, “Constrained K-means Clustering with Background Knowledge,” *International Conference on Machine Learning*, pp. 577–584, 2001.
- [29] K. H. Zou, K. Tuncali, and S. G. Silverman, “Correlation and Simple Linear Regression,” *Radiology*, vol. 227, no. 3, pp. 617–628, 2003. [Online]. Available: <http://pubs.rsna.org/doi/10.1148/radiol.2273011499>
- [30] T. Jakasa, I. Androcec, and P. Sprcic, “Electricity price forecasting - ARIMA model approach,” *2011 8th International Conference on the European Energy Market (EEM)*, no. May, pp. 222–225, 2011.

- [31] C. Grillenzoni, "Arima Processes With Arima Parameters," *Journal of Business and Economic Statistics*, vol. 11, no. 2, pp. 235–250, 1993.