

Trabajo Fin de Máster

Sistemas de Energía Eléctrica

Aplicación de técnicas de Inteligencia Artificial para definir una metodología de predicción de carga en transformadores MT/BT no observables

Autor: Rubén Carmona Pardo

Tutor: Antonio de la Villa Jaén, Esther Romero Ramos

Dpto. Ingeniería Eléctrica
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2022



Trabajo Fin de Máster
Máster en Sistemas de Energía Eléctrica

Aplicación de técnicas de Inteligencia Artificial para definir una metodología de predicción de carga en transformadores MT/BT no observables

Autor:

Rubén Carmona Pardo

Tutores:

Antonio de la Villa Jaén

Profesor titular

Esther Romero Ramos

Catedrática de Universidad

Dpto. de Ingeniería Eléctrica
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla
Sevilla, 2022

Trabajo Fin de Máster: Aplicación de técnicas de Inteligencia Artificial para definir una metodología de predicción de carga en transformadores MT/BT no observables

Autor: Rubén Carmona Pardo

Tutor: Antonio de la Villa Jaén, Esther Romero Ramos

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2022

El Secretario del Tribunal

Agradecimientos

En primer lugar, querría agradecer a mis tutores Esther Romero y Antonio de la Villa la amabilidad y el tiempo que han dedicado a lo largo del desarrollo del proyecto para ayudarme y enseñarme. De la misma forma, agradecer al resto de miembros del Dpto de Ingeniería Eléctrica que de una forma u otra han conseguido ayudarme o inspirarme.

Igualmente, quiero agradecer eternamente a mi familia todo el apoyo que me brindan. A mi pareja, Macarena, por su confianza, por soportar las horas dedicadas a la investigación de este TFM, las situaciones de frustración y el tiempo que no puedo dedicarle. A mis padres, porque sin su ayuda nunca habría llegado a ser el profesional que soy a día de hoy. Y a mis hermanos, por todos sus consejos y la confianza que me han brindado.

Finalmente, me gustaría dar las gracias a todos los compañeros que desde que empecé la carrera me han ayudado a crecer como persona e ingeniero, a todos mis compañeros de Ingelectus, a Susana Carillo y Javier Leiva por guiarme en primeros pasos en este sector, y a la Escuela de Ingenieros de Málaga y Sevilla.

Rubén Carmona Pardo

Sevilla, 2022

Resumen

En este Trabajo Fin de Máster (TFM) se presenta el desarrollo de una metodología de predicción de curvas de potencia en transformadores MT/BT instalados en las redes de distribución eléctrica y para los que no existe ningún equipo de medida ni sensor instalado en él o en el Centro de Transformación donde se aloje, es decir, transformadores no observables.

En primer lugar, se hace una breve introducción sobre las redes de distribución eléctrica, los transformadores de potencia MT/BT y una revisión del estado del arte de las diferentes técnicas de predicción de potencia o carga relacionadas con transformadores MT/BT. Posteriormente se describe el caso de uso principal del TFM y el entorno del Smartcity Málaga Living Lab cuyos datos se han utilizado. Asimismo se detalla en profundidad todos los datos que se han considerado.

En segundo lugar, se introducen los conceptos básicos de Inteligencia Artificial (IA), y se describen con detalle los modelos evaluados en este TFM así como otras técnicas de analítica de datos y Machine Learning que se han utilizado.

A continuación se presentan los resultados obtenidos en las dos pruebas realizadas, una con el conjunto de datos original y otra con datos procesados con clustering. Todos los resultados de estas pruebas se han evaluado tanto gráfica como numéricamente utilizando las métricas MAE (del inglés Mean Average Error) y MSE (del inglés Mean Squared Error).

Finalmente, se comparan los resultados obtenidos con los que se obtendrían con otros procedimientos más deterministas. Además, se aplica esta misma metodología en la predicción de potencia a futuro y se presentan los resultados.

Abstract

This Master's Thesis presents the development of a novel methodology that pursues the prediction of active power values in distribution network power transformers for whom there is no electrical behaviour data available, namely, non-observable distribution network power transformer.

In the first place, it is given a brief description about power distribution networks and transformers, and the state-of-art about prediction techniques on this field is reviewed. Afterwards it is described the main use case of this Thesis and the Smartcity Málaga Living Lab, whose data have been used in this work. Likewise, it is deeply detailed all the variables used from the data.

Secondly, it is introduced the basis about Artificial Intelligence (AI) and it has been explained in greater detail the different models evaluated in this work, as well as other Data Analytics and Machine Learning techniques applied.

Then, all the results are presented, split in two tests carried out: one using the whole set of the original data, and the second after processing it. These results have been evaluated graphically, plotting some charts in the document, and numerically using MAE and MSE metrics.

Finally, prior results are compared with those delivered by other deterministic approaches. Also, it is presented the results of the methodology here developed applied to forecast active power.

Índice

Agradecimientos	vii
Resumen	ix
Abstract	xi
Índice	xiii
Índice de Tablas	xv
Índice de Figuras	xvii
1 Introducción	1
2 Casos de uso y datos utilizados	5
2.1. <i>Casos de Uso</i>	5
2.2. <i>Datos reales del Smartcity Málaga Living Lab</i>	8
2.2.1 Alcance	10
2.2.2 Medidas históricas de potencia activa instantánea en transformadores	11
2.2.3 Datos topológicos	14
2.2.4 Datos meteorológicos	16
2.2.5 Otros datos	17
3 Inteligencia Artificial y descripción modelos de predicción utilizados	19
3.1 <i>Decision Tree Regressor</i>	21
3.2 <i>Gradient Boosting Regressor</i>	22
3.3 <i>XGBoost</i>	23
3.4 <i>Redes Neuronales</i>	23
3.5 <i>Otras técnicas de procesamiento de datos empleadas</i>	27
3.5.1 Codificación de datos de entrada	27
3.5.2 Escalado de datos	27
3.5.3 Clustering	28
3.5.4 Validación cruzada	28
4 Resultados algoritmos de predicción	31
4.1 <i>Resultados iniciales</i>	32
4.1.1 Decision Tree Regressor	32
4.1.2 Gradient Boosting Regressor	34
4.1.3 XGBoost	35
4.1.4 Redes Neuronales	36
4.1.5 Conclusiones	38
4.2 <i>Resultados tras clustering</i>	39
4.2.1 Decision Tree Regressor	41
4.2.2 Random Forest Regressor	42
4.2.3 XGBoost	43
4.2.4 Redes Neuronales	44
4.2.5 Conclusiones pruebas tras clustering	46

4.3	<i>Comparativa con otros métodos</i>	49
4.3.1	Curvas normalizadas de potencia activa normalizadas (PAN)	49
4.3.2	Procedimiento publicado en BOE en términos de energía	50
4.4	<i>Otros alcances de predicción</i>	51
5	Conclusiones y trabajo futuro	55
	Referencias	57

ÍNDICE DE TABLAS

Tabla 1 Estado de carga global de los transformadores del alcance	13
Tabla 2 Resumen de la información topológica utilizada en como dato de entrada en los modelos de predicción	14
Tabla 3 Ejemplo de estructura de datos utilizada en los modelos de predicción	20
Tabla 4 Métricas para los resultados de la prueba inicial con el modelo DTR	34
Tabla 5 Métricas para los resultados de la prueba inicial con el modelo GBR	35
Tabla 6 Métricas para los resultados de la prueba inicial con el modelo XGBoost	36
Tabla 7 Métricas para los resultados de la prueba inicial con el modelo RNP 1	37
Tabla 8 Métricas para los resultados de la prueba inicial con el modelo RNP 2	38
Tabla 9 Agrupación de transformadores por clusters	40
Tabla 10 Métricas para los resultados del algoritmo DTR con datos del cluster 3	42
Tabla 11 Métricas para los resultados del algoritmo GBR con datos del cluster 3	43
Tabla 12 Métricas para los resultados del algoritmo XGBoost con datos del cluster 3	44
Tabla 13 Métricas para los resultados del algoritmo RNP 1 con datos del cluster 3	45
Tabla 14 Métricas para los resultados del algoritmo RNP 2 con datos del cluster 3	46
Tabla 15 Comparativa de las métricas de evaluación inicial y con datos clusterizados	47
Tabla 16 Comparativa de métricas entre datos clusterizados y método con potencia activa normalizada	49
Tabla 17 Comparativa de métricas en términos de energía entre modelos estudiados y procedimiento BOE	51
Tabla 18 Comparativa de métricas en el caso de uso de predicción	53

ÍNDICE DE FIGURAS

Figura 1. Tipo1 - Estimación de carga de en transformador basada en perfiles de carga estándar. Fuente: Elaboración propia.	5
Figura 2. Tipo 2 - Estimación de perfil de carga con medidas agregadas en subestación primaria (load allocation) y/o medidas de contadores inteligentes. Fuente: Elaboración propia.	6
Figura 3. Tipo 3 - Predicción de carga en transformador basada en sus propios históricos. Fuente: Elaboración propia.	7
Figura 4. Caso de Uso abordado: estimación de carga en transformador basándose en la carga recogida por sensores en otros transformadores. Fuente: Elaboración propia.	7
Figura 5 Área de alcance del Smartcity Málaga Living Lab	9
Figura 6 Esquema unifilar de los anillos de MT 1 y 4 a izquierda y derecha, respectivamente, indicando los equipos de medida (cuadrado verde)	10
Figura 7 Mapa con la distribución de Centros de Transformación considerados dentro del alcance del TFM.	11
Figura 8 Ejemplo de serie temporal de potencia activa de uno de los transformadores	12
Figura 9 Diagrama de cajas con la distribución de potencia activa por transformador	13
Figura 10 Diagrama de pares de los datos topológicos de los transformadores considerados en el alcance del TFM	15
Figura 11 Diagrama de cajas de variables meteorológicas utilizadas para la predicción	16
Figura 12 Esquema descriptivo de la Inteligencia Artificial, el Machine Learning y el Deep Learning. Fuente: https://medium.com/@experiencia18/diferencias-entre-la-inteligencia-artificial-y-el-machine-learning-f0448c503cd4	19
Figura 13 Esquema de la lógica de generación de un árbol de decisión. Fuente: Elaboración propia.	21
Figura 14 . Esquema representativo del funcionamiento del modelo GBR. Fuente: https://www.researchgate.net/figure/Flow-diagram-of-gradient-boosting-machine-learning-method-The-ensemble-classifiers_fig1_351542039	22
Figura 15 Esquema de funcionamiento de un perceptron o neurona artificial. Fuente: Wikipedia.	24
Figura 16 Esquema de la Red Neuronal 1 evaluada	26
Figura 17 Esquema de la Red Neuronal 2 evaluada	26
Figura 18 Esquema de funcionamiento del proceso de validación cruzada. Fuente: https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation	29
Figura 19 Resultados prueba inicial con algoritmo DTR	33
Figura 20 Resultados prueba inicial con algoritmo GBR	34
Figura 21 Resultados prueba inicial con algoritmo XGBoost	35
Figura 22 Resultados prueba inicial Redes Neuronales Profundas 1	36
Figura 23 Resultados prueba inicial Redes Neuronales Profundas 2	37
Figura 24 Esquema de las diferentes técnicas para balancear set de datos. Fuente: https://www.researchgate.net/figure/Proposed-texonomy-for-the-review-of-imbalanced-class-problems-in-data-mining_fig2_331522710	39
Figura 25 Gráficos de dispersión representativos de los clústers obtenidos – Ejemplo 1.	40

Figura 26 Resultados algoritmo DTR con datos del cluster 3	41
Figura 27 Resultados algoritmo GBR con datos del cluster 3	43
Figura 28 Resultados algoritmo XGBoost con datos del cluster 3	43
Figura 29 Resultados algoritmo RNP 1 con datos del cluster 3	44
Figura 30 Resultados algoritmo RNP 2 con datos del cluster 3	46
Figura 31 Evaluación de varias estadísticas del error de predicción horario	48
Figura 32 Resultados métricas para curvas de potencia de activa normalizadas	50
Figura 33 Comparativa de las predicciones a tres días vista con los modelos DTR, XGB, NN1 y NN2	52

1 INTRODUCCIÓN

Deje que el futuro diga la verdad y evalúe a cada uno de acuerdo con su trabajo y sus logros. El presente es de ellos; el futuro, para el que realmente he trabajado, es mío.

Nikola Tesla

Las metodologías de predicción aplicadas a conocer el estado futuro de carga de los transformadores de potencia en redes de distribución es, como se verá en el análisis del Estado del Arte más adelante, una cuestión que ha sido ampliamente abordada en la literatura en los últimos años. Esto se debe al interés que supone conocer la carga futura de los transformadores para los desafíos a los que se enfrentan estas redes de distribución en el medio y largo plazo.

A continuación se listarán los más destacados:

- Seguridad y continuidad del suministro:

Las redes de distribución se han convertido en el centro de la transición energética puesto que son en ellas donde conectan la mayoría de los nuevos agentes eléctricos, gran parte de ellos en las redes de Baja Tensión: vehículo eléctrico, generación distribuida, aplicaciones power-to-X, etc. Y éstas redes, en buena parte de su extensión, tienen una infraestructura que pudiera no estar preparada para albergarlos, siendo el transformador uno de los elementos más críticos a evaluar. Conocer la carga futura a medio y largo plazo de los transformadores permitirá llevar un correcto mantenimiento de la red y dirigir la planificación de ésta acorde a los escenarios futuros.

- Operación de redes:

La operación activa de las redes de distribución, especialmente en redes de Baja Tensión, es una asignatura pendiente por parte de las operadoras de la red. La predicción de la carga de transformadores puede un suponer un habilitador para herramientas de monitorización y observabilidad en redes, el Dynamic Transformer Rating o la operación de interruptores seccionadores para topología de Media Tensión, entre otras.

- Conocimiento futuro de la capacidad de acceso:

Las empresas distribuidoras están obligadas a publicar sus mapas de capacidad de acceso y conexión a la red para que los usuarios puedan solicitar nuevas conexiones, ya sea de demanda o generación, o repotenciar las actuales. Si las distribuidoras dispusieron de herramientas que le permitan predecir la

carga futura en transformadores de potencia MT/BT les permitiría hacer una estimación mucho más precisa de la capacidad de acceso a su red y la ciudadanía en general se vería beneficiada de ello.

- Flexibilidad en redes de Baja Tensión:

La predicción del estado de carga de los transformadores es el punto de partida para conocer las futuras situaciones de riesgo en la operación del mismo (sobrecargas, puntos calientes, etc). Con la flexibilidad que se espera que esté disponible en las redes de Baja Tensión se podrá evitar llegar a estos escenarios críticos de operación, siendo la predicción de la carga la señal que se necesita para activar estos recursos de flexibilidad.

De esta manera, se demuestra que proporcionar información sobre el estado de operación de los transformadores de potencia, ya sea a pasado, presente o futuro a corto, medio o largo plazo, supone un gran valor añadido para los diferentes agentes del sector eléctrico. Sin embargo, las metodologías con las que este problema ha sido abordado pueden llegar a ser muy diferentes entre sí, tanto en el enfoque como en el alcance o la información de partida.

Por un lado, se encuentran las técnicas clásicas de perfilado y modelado de curvas de carga ampliamente utilizadas en los sistemas eléctricos de potencia y que, generalmente, proporcionan curvas horarias históricas o futuras de la carga de un activo de red. Esta carga puede hacer referencias a diferentes variables eléctricas aunque en la literatura se suele hacer referencia a la corriente eléctrica, potencia aparente o potencia activa.

Estas técnicas, que suelen generar diversos perfiles normalizados, están complementadas con otras que buscan asociar los activos eléctricos a su perfil o modelado más adecuado para, posteriormente, escalarlo y obtener su curva de carga. Es posible encontrar algunos artículos que abordan estas técnicas aplicadas a los transformadores de potencia MT/BT, predominando el uso de lógica difusa para caracterizar la pertenencia de los transformadores a uno o varios perfiles de carga, como es el caso de [1] y [2].

De igual forma y como evolución de las técnicas anteriores, se pueden encontrar otro tipo de técnicas que, además de utilizar perfiles de carga, se caracterizan por utilizar una medida agregada, principalmente recogida por sensores instalados en las salidas de la subestación primaria, para posteriormente distribuirla por todas las cargas que cuelguen de esa línea.

Estas técnicas, conocidas como load allocation o distribución de cargas, utilizan los perfiles estándares de carga y las medidas a nivel de subestación primaria para ajustar los parámetros de estimadores desarrollados que, una vez optimizados, se usarán para estimar la carga en transformadores de potencia MT/BT. Buen ejemplo de ello son las contribuciones [3] y [4] donde, además de medidas de las salidas de MT, utilizan información estructural de la red alimentada por el transformador MT/BT a estimar como es la potencia contratada total y el número total de suministros.

Similar al caso anterior, pero con un enfoque diferente, se pueden encontrar otras metodologías de estimación de cargas en transformadores MT/BT que hacen uso de medidas procedentes de los Smart Meters que estos alimentan. Como ejemplo el artículo [5] que requiere, además, del conocimiento de la topología de red y las impedancias de línea.

Estas técnicas se utilizan principalmente, no para hacer predicciones a futuro del elemento analizado, sino para caracterizar su comportamiento pasado.

Por otro lado, existen tecnologías más centradas en la predicción de la carga futura de transformadores de potencia MT/BT basándose en técnicas de Machine Learning o Deep Learning, cuyo uso se ha vuelto viral en el sector eléctrico en los últimos años. Dentro de estas técnicas de aprendizaje existen multitud de modelos diferentes según se trate de Aprendizaje Supervisado (Supervised Learning Models) o Aprendizaje No Supervisado (Unsupervised Learning Models).

Dentro de los primeros, se pueden destacar los modelos regresivos como Vectores Soporte Máquina (Support Vector Machine), modelos basados en árboles de decisión y ensambladores como Random Forest, AdaBoost o Gradient Descent, y modelos no lineales basados en redes neuronales como las Redes Neuronales Convolucionales (CNN, Convolutional Neural Network) o Redes Neuronales de memoria a corto-largo plazo (LSTM, Long-Short Term Memory).

En cuanto al segundo grupo, los modelos por excelencia utilizados en la literatura son los modelos de clustering utilizando la técnica de los vecinos cercanos (kNN, k-Nearest Neighbors).

El uso de estas técnicas está bastante extendido en la literatura tanto para la predicción de perfiles carga eléctrica en general, como es el caso de [6] y [7], como específicamente aplicada a los transformadores de potencia MT/BT como se puede apreciar en [8]- [9]. De estos últimos, lo primero que se puede destacar es la aplicación de varios de los modelos anteriores de forma individual (Redes Neuronales [10], Regresiones Lineales [11], Random Forest [9]) pero también combinando modelos de aprendizaje supervisado y no supervisado, como es el caso de [8] y [12], donde se combinan técnicas de clustering con modelos basados en Descenso de Gradiente y Redes Neuronales Profundas, respectivamente.

No obstante, lo realmente llamativo de todas estas contribuciones, y de la mayoría de literatura que el autor de este TFM ha encontrado sobre este tema, es que todas ellas se sirven de las medidas de los sensores instalados en el transformador o en la red suministrada por el transformador cuya carga se quiere predecir. Es decir, el alcance de éstos consiste en la predicción a corto, medio o largo plazo de la carga de un transformador basándose en sus históricos de medida y, en muchos casos, de otro conjunto de variables exógenas. Esto es posible gracias a las ambiciosas estrategias de digitalización de activos de red que están siguiendo las compañías de distribución eléctrica con la instalación de smart metering y de supervisión avanzada en Centros de Transformación.

Igualmente, existen procedimientos oficiales para calcular el perfil de consumo de los activos eléctricos en redes de distribución como el publicado en el artículo 9 del Real Decreto 1435/2002, de 27 de diciembre, por el que se establece un método de cálculo del perfil de consumo de contadores a efectos de la liquidación de su consumo eléctrico por parte de las empresas comercializadoras, en función de la tarifa de acceso contratada, para contadores tipo 4 y tipo 5.

Para ello, todos los años se publica en el BOE una tabla con valores de referencia de la Demanda de Referencia y Perfiles Iniciales propuesto por la Comisión Nacional de los Mercados y Competencia. Este procedimiento será utilizado en este Trabajo Fin de Máster para obtener datos de carga de transformadores y compararlo con los perfiles de consumo calculados a partir de los modelos aquí entrenados y poder, así, evaluarlo a partir de una herramienta oficial de uso común en España.

Con esto es importante destacar que los resultados que devuelven los modelos de predicción evaluados en este TFM no son predicciones de consumo de energía de contadores o transformadores. El resultado que devuelven es la curva de carga cinco-minutal del transformador bajo estudio que, para compararlo con los resultados que se obtendrían con el procedimiento publicado en el BOE, se convertirán a curvas de consumos horarios.

Todas estas aportaciones son de gran utilidad para la explotación del transformador y tienen una aplicación directa en muchos de los puntos que se han destacado al principio de esta introducción. Sin embargo, la utilización de la ingente cantidad de datos que se generan y almacenan actualmente en la redes de distribución para dar observabilidad a aquellos transformadores que no tienen instalados ningún tipo de equipo de medida o que, por el contrario, sí que lo tienen pero presentan problemas, se ha abordado mínimamente.

En cuanto al primero de los casos, a pesar de los esfuerzos que las compañías distribuidoras de electricidad están haciendo en pos de la digitalización de sus activos, es muy complicado tanto desde un punto de vista técnico como de rentabilidad económica sensorizar el 100% del parque de transformadores. En cuanto al segundo caso, las averías, problemas en la comunicación de los equipos y la falta de mantenimiento u obsolescencia afectan significativamente a las medidas recibidas.

Ante esta situación, hay que ver la digitalización de las redes de distribución y toda la información disponible actualmente, que será mayor en el futuro, como una oportunidad para complementar las soluciones predictivas que existen actualmente con otras nuevas cuyo enfoque permite paliar estos problemas de observabilidad. Es aquí donde se centra este Trabajo Fin de Máster.

En el caso de este Trabajo Fin de Máster, el alcance propuesto, los datos utilizados y las técnicas evaluadas tratan un caso de uso de predicción en transformadores diferente a los que se han analizado anteriormente. Éste, como se detallará en profundidad en los próximos apartados, aprovecha el ecosistema de datos del que actualmente las distribuidoras disponen por el proceso de digitalización que están viviendo para predecir la carga de operación de un transformador de potencia residencial.

Así, se han probado diferentes metodologías de predicción actuales basadas en Aprendizaje Máquina y Aprendizaje Profundo (del inglés Machine Learning y Deep Learning, respectivamente), y se propone una nueva

metodología de fácil implementación para las empresas de distribución eléctrica para maximizar la observabilidad de la red sin necesidad de instalar equipos de medida en el 100% de su parque de transformadores.

Finalmente, en la exposición de resultados, se verá como aplicando técnicas de aprendizaje no supervisado como clustering, para preprocesado de datos previo al entrenamiento de los modelos, los resultados originales mejoran significativamente.

2 CASOS DE USO Y DATOS UTILIZADOS

¿Qué es el alma? Es como la electricidad, realmente no sabemos lo que es, pero es una fuerza que puede encender una habitación.

- Ray Charles -

Como se ha adelantado en la introducción, el objetivo de este Trabajo Fin de Máster es investigar un caso de uso de predicción de carga de potencia activa en transformadores de potencia MT/BT no teledados que ha sido poco abordado en la literatura y que supondría un beneficio indirecto de la digitalización de activos para las compañías de distribución eléctrica en cuanto al uso de la ingente cantidad de datos de la que empiezan a disponer.

En este apartado se presentará una breve contextualización de los principales enfoques de predicción abordados en la literatura junto con el enfoque aquí planteado y, además, se describirá con todo detalle la información que se ha utilizado en el desarrollo de la metodología propuesta en este Trabajo Fin de Máster.

2.1. Casos de Uso

En la revisión del Estado del Arte, se ha puesto de manifiesto que existen tres enfoques principales claros en cuanto al cálculo/estimación/predicción de la carga de un transformador:

- *Tipo 1.* Hacer uso de perfiles de carga estándar para inferir el del transformador bajo estudio y para diferentes ventanas de predicción: Imagen 1.

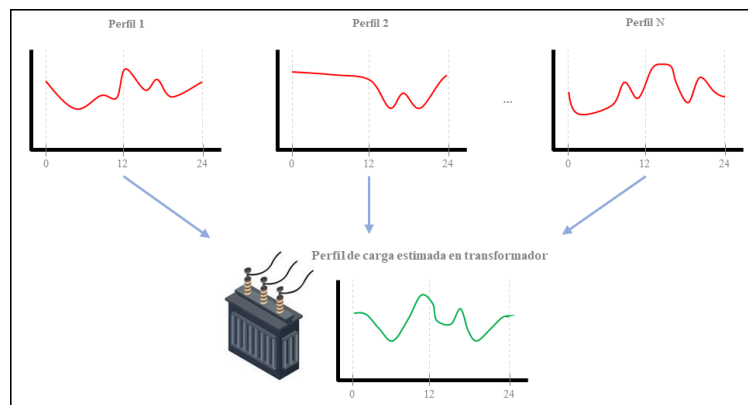


Figura 1. Tipo1 - Estimación de carga de en transformador basada en perfiles de carga estándar. Fuente:

Elaboración propia.

- *Tipo 2.* Utilizar medidas agregadas a nivel de subestación primaria o medidas desagregadas de contadores inteligentes para obtener la carga del transformador, también para diferentes ventanas de predicción: Imagen 2.

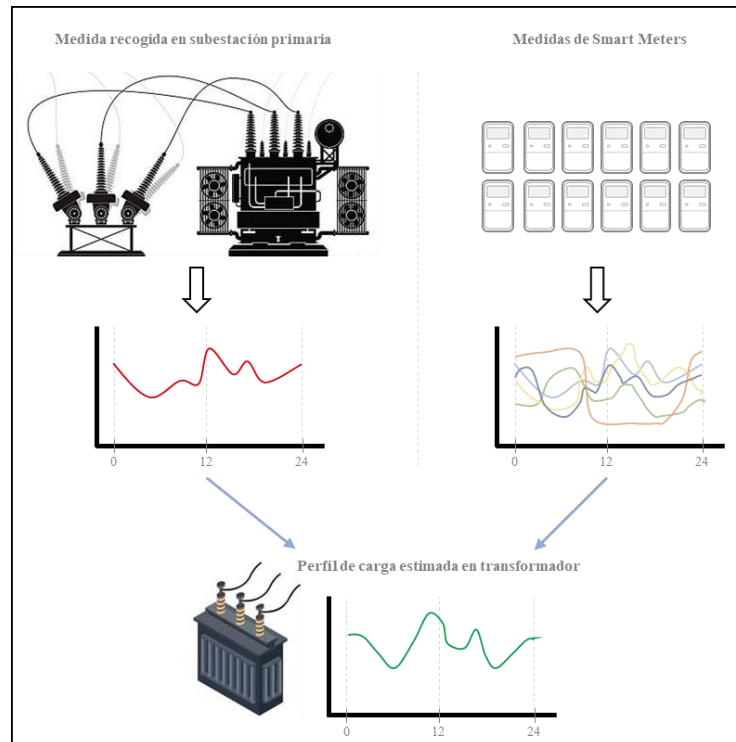


Figura 2. Tipo 2 - Estimación de perfil de carga con medidas agregadas en subestación primaria (load allocation) y/o medidas de contadores inteligentes. Fuente: Elaboración propia.

- *Tipo 3.* Basarse en medidas procedentes de sensores instalados en el Centro de Transformación (celda de protección, puente de Baja Tensión SABT, telemida en concentrador secundario, etc) donde se encuentra el transformador alojado para predecir la carga futura a corto o medio plazo: Imagen 3. Este enfoque ha sido abordado previamente por el autor de este Trabajo Fin de Máster en una tesis [13] en la que se desarrollaron diversos algoritmos de predicción a una hora y a un día vista de la intensidad que circula a través de un a Línea de Baja Tensión utilizando datos de Supervisión Avanzada de Baja Tensión (SABT). En este se desarrollaron algoritmos basados en árboles de decisión y en Redes Neuronales, más concretamente Redes Convolucionales y Redes LSTM.

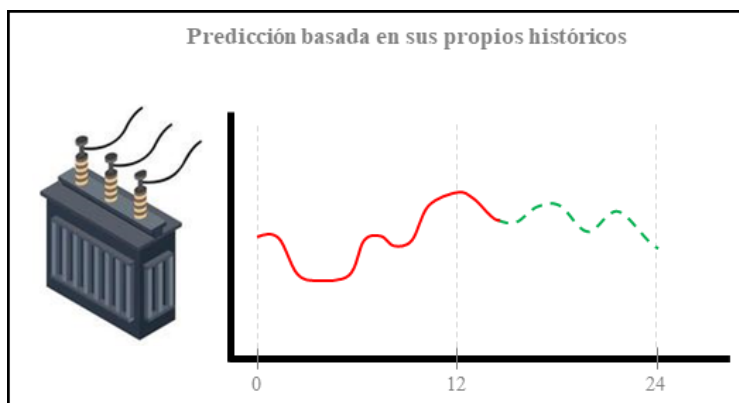


Figura 3. Tipo 3 - Predicción de carga en transformador basada en sus propios históricos. Fuente: Elaboración propia.

En este Trabajo Fin de Máster el enfoque planteado es una adaptación de los tres anteriores, de manera que se quiere utilizar los datos de potencia activa en aquellos transformadores con sensores para predecir la potencia activa de un transformador para el que no se tiene ninguna medida disponible, ni en tiempo real ni medidas históricas. Este caso de uso se ha representado en la Figura 4.

Este contexto podría ser común para las compañías distribuidoras en aquellos transformadores que no tienen equipos de medidas instalados por dificultad física (difícil acceso y compatibilidad en transformadores en poste de zonas rurales, falta de espacio disponible en el habitáculo del Centro de Transformación o problemas de cobertura para comunicaciones), o porque su instalación todavía no ha sido abordada o por falta de rentabilidad económica.

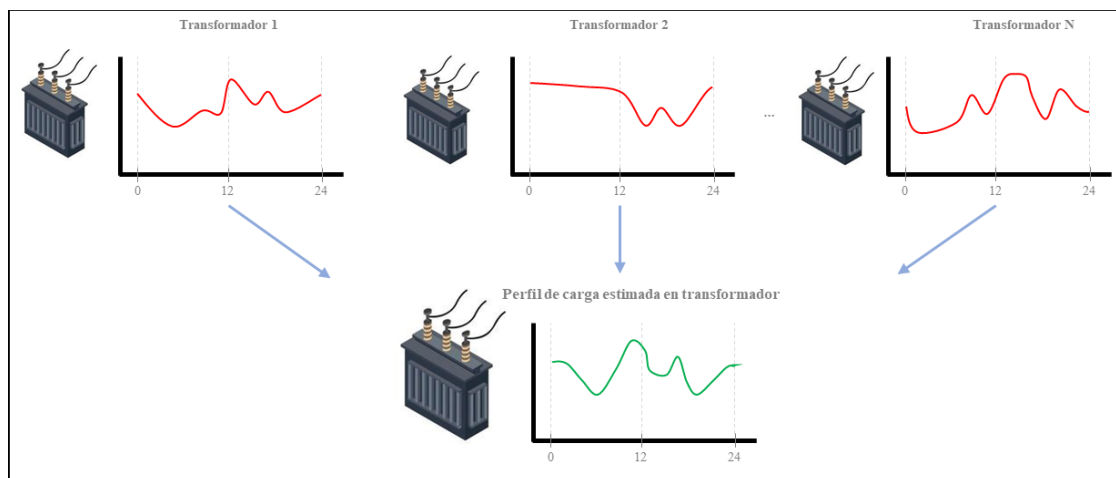


Figura 4. Caso de Uso abordado: estimación de carga en transformador basándose en la carga recogida por sensores en otros transformadores. Fuente: Elaboración propia.

Para ello, se ha tomado como caso base la predicción de carga de un transformador de potencia MT/BT de tipo residencial, ubicado en Málaga capital, del que se conocen sus medidas históricas de potencia activa pero éstas se utilizarán únicamente para compararlos con los resultados de las predicciones de potencia de los modelos probados y, así, evaluar su efectividad.

Estos datos se completarán con un conjunto mucho mayor de datos históricos de potencia activa de diversos transformadores que sí se utilizarán para entrenar los modelos de predicción. La ventana de tiempo que se considerará será el año 2019 completo, del que será conocida la temperatura ambiente.

Además se utilizarán otros datos topológicos o estructurales del transformador y su red suministrada que suelen ser conocidos por parte de las distribuidoras:

- **Potencia nominal** del transformador: normalmente aparece en la placa de características del transformador y suele ser un dato típico conocido por la distribuidora.
- Posición geográfica del transformador: información GIS con la **latitud y longitud** de la ubicación del transformador.
- Número de **líneas de baja tensión** que salen del transformador o del cuadro de baja tensión.
- Número de **Puntos de Conexión a la Red (PCR)** suministrados. Este es un punto ficticio de la red que se suele utilizar para separar la red que pertenece a la distribuidora y la parte de la red que pertenece al cliente/s. Normalmente, suele ubicarse antes de la Caja General de Protección (CGP).
- **Número de clientes suministrados** conectados aguas abajo del transformador.
- **Potencia contratada** total, máxima y media de los clientes suministrados anteriores para cada uno de los transformadores. Estas estadísticas se consideran indicadores significativos del conjunto de suministros.

Toda esta información, excepto la potencia nominal del transformador y la posición geográfica del mismo, es dinámica y puede variar significativamente en el tiempo, en efecto, al tratarse de redes de distribución eléctrica reales donde las actuaciones en campo, conexiones y desconexiones de suministros suelen ser frecuentes. Sin embargo, hacer un seguimiento exhaustivo de su evolución es realmente complicado en la práctica. Es por ello que esta información se ha considerado invariable en el periodo de tiempo para el que se han usado los datos, a pesar de las imprecisiones que pueda ocasionar en las predicciones de los modelos evaluados en este TFM y que se detallarán más adelante. No obstante, los modelos utilizados permitirían incorporar estos cambios como entrada de datos mejorando así los resultados que se obtendrían.

Con toda esta información (curvas históricas de potencia, información estructural y ubicación, caracterización básica de la red suministrada) lo que se pretende es encontrar modelos lineales y no lineales que sean capaces de generalizar el comportamiento eléctrico de un conjunto de transformadores conocidos para, posteriormente, extrapolarlo a otros que no lo son. Esto tendría grandes beneficios para la observabilidad de la red, la inteligencia de las mismas y las consecuencias que esto tendrían en su explotación, planificación y transición.

Para este caso de uso, se ha evaluado el siguiente escenario: predecir la potencia activa instantánea del transformador para el año completo de 2019.

2.2 Datos reales del Smartcity Málaga Living Lab

Toda la información que se ha utilizado para los desarrollos de este Trabajo Fin de Máster es información real de la compañía distribuidora e-distribución (Grupo Enel), de la que se han utilizado datos reales procedentes del Smartcity Málaga Living Lab.

El Smartcity Málaga Living Lab es un laboratorio real de Redes Inteligentes situado en Málaga capital donde, desde 2009, se han realizado multitud de proyectos de desarrollo e innovación en los que se han probado la repercusión que las soluciones orientadas a las Smart Grids tienen en las redes.

La particularidad de este laboratorio de redes es que se extiende en una red de distribución real en la que se ha realizado un importante despliegue de sensorística en los Centros de Transformación. En concreto, tal y como se ve en la Imagen 5, el área que abarca el Smartcity Málaga Living Lab comprende dos anillos de MT compuesto por cuatro líneas que suministran un total de 48 Centros de Transformación, con 59 transformadores MT/BT propiedad de la compañía distribuidora, y 11 clientes de MT con transformador MT/BT propio.

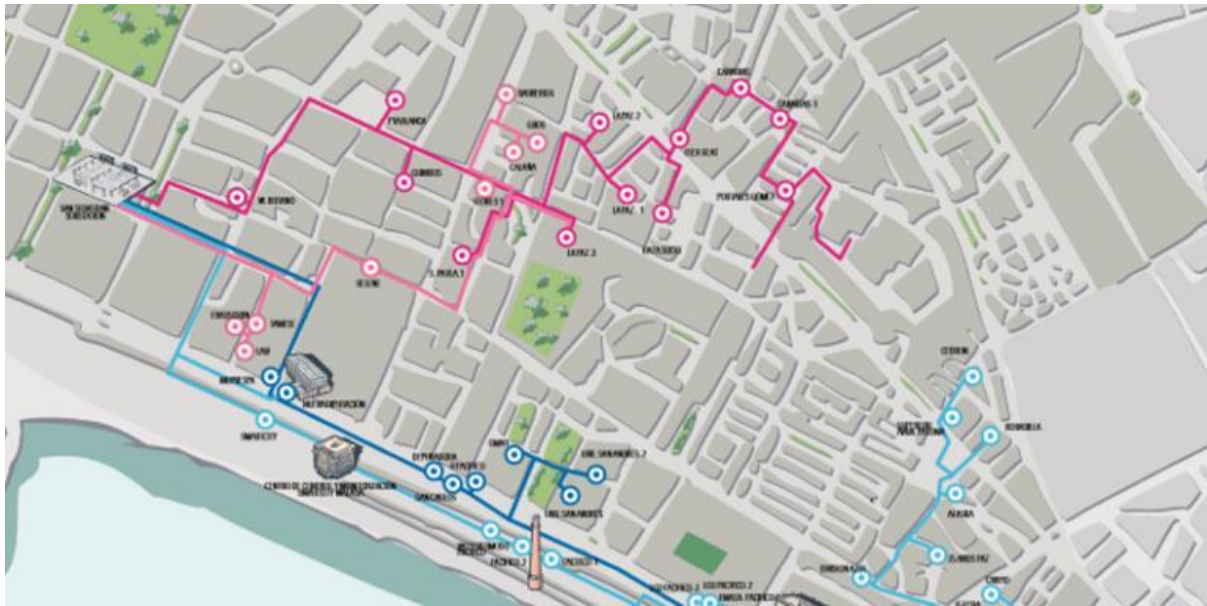


Figura 5 Área de alcance del Smartcity Málaga Living Lab

El planteamiento de colocación de equipos de medida en la red que se hizo fue, por un lado, sensorizar completamente uno de los anillos (Anillo 1) y parcialmente otro (Anillo 4). De esta forma, en el primer anillo completamente sensorizado, se instalaron equipos de medida en celdas de entrada y salida, equipos de medida en celdas de protección del transformador, equipos de medida en el puente de Baja Tensión del transformador y equipos de medida en los cuadros de Baja Tensión. Algunos Centros de Transformación tienen toda la tipología anterior de sensores instalados y, otros, una parte, pero en el Anillo 1 todos los Centros de Transformación disponen de un mínimo de observabilidad. En el caso del segundo anillo, la tipología de equipos de medida considerada es la misma que en el caso anterior pero, como diferencia, sólo una parte de los Centros de Transformación del mismo tienen instalado al menos uno de estos equipos. Todo esto se puede ver de forma esquematizada en la Figura 6.

Como se puede observar, el despliegue de sensores en el Anillo 1 permite una investigación e implementación de soluciones de red mucho más amplia que en el caso del Anillo 4. Sin embargo, el planteamiento de éste último es mucho más realista que el primero y refleja mucho mejor el estado de digitalización de activos de las redes de distribución en general.

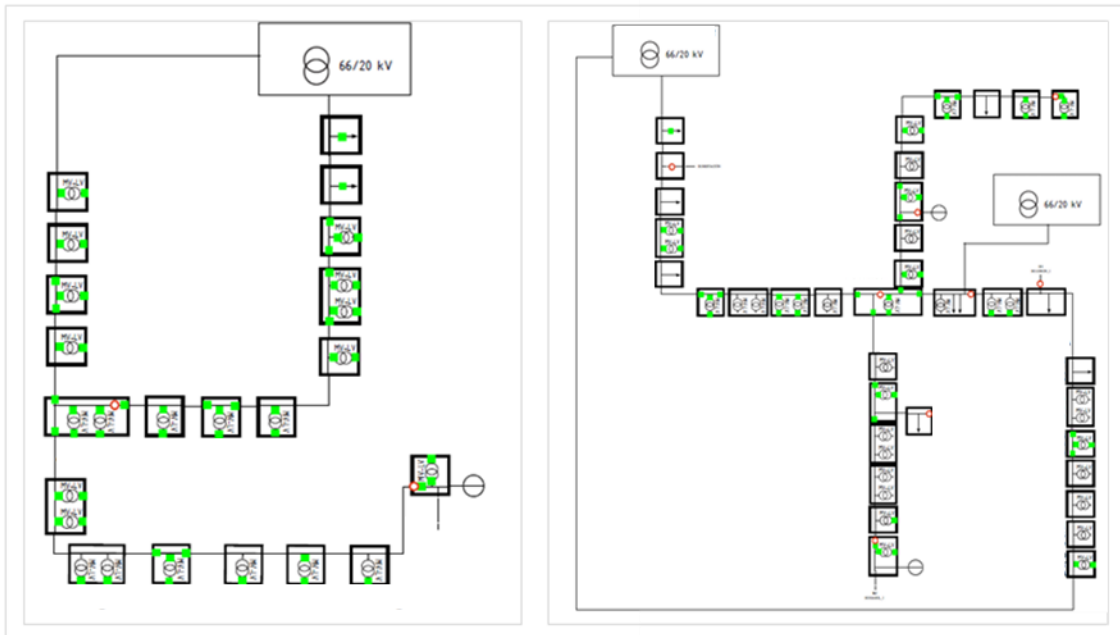


Figura 6 Esquema unifilar de los anillos de MT 1 y 4 a izquierda y derecha, respectivamente, indicando los equipos de medida (cuadrado verde)

Este despliegue ha generado una cantidad de datos reales ingente, los cuales poseen un gran valor para la operación, planificación, optimización e investigación de redes inteligentes.

Por ello, para este Trabajo Fin de Máster se ha querido sacar partido de todo este volumen de datos para resolver una problemática actual de las redes de distribución, como es la observabilidad, y enfocando su aplicación a un escenario realista y más representativo de las redes de distribución. Es decir, se han utilizado los datos del Smartcity Málaga Living Lab para hacer que las redes con una infraestructura digital similar a la del Anillo 4 puedan llegar a ser operadas como las redes análogas a las del Anillo 1.

A continuación se muestra con más detalle los datos que finalmente se han utilizado.

2.2.1 Alcance

El alcance, desde el punto de vista de transformadores, abarca un total de 40 transformadores de potencia ubicados en 33 Centros de transformación diferentes dentro de los dos anillos 1 y 4 presentados en el apartado anterior. En la Figura 7 se puede observar un mapa donde se indica la localización de todos ellos.

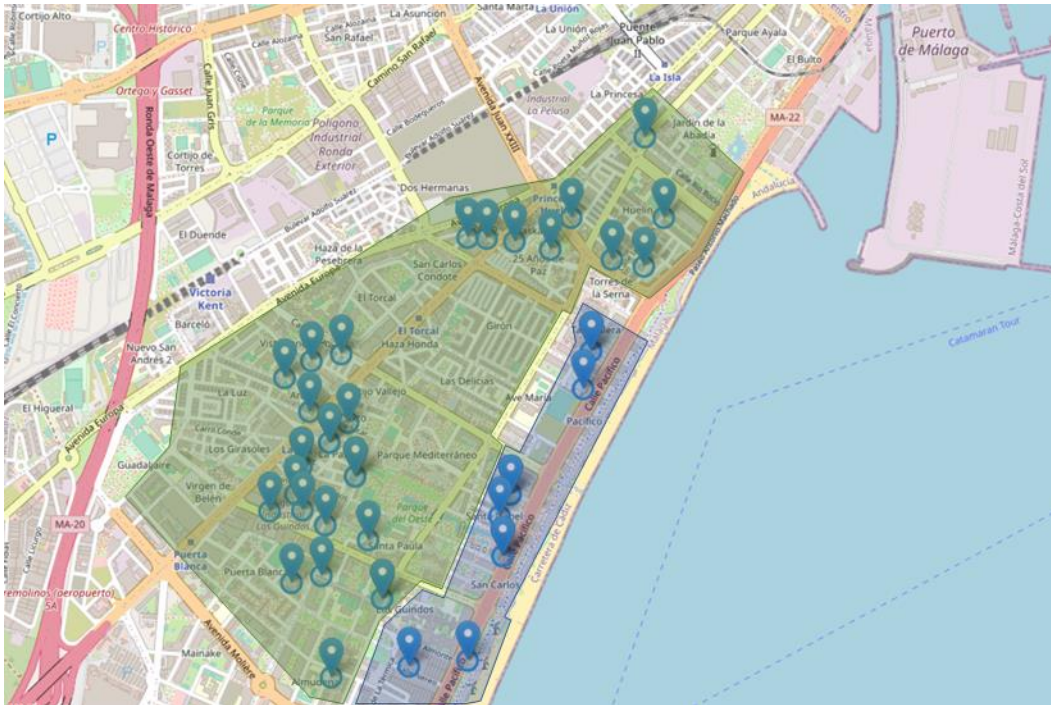


Figura 7 Mapa con la distribución de Centros de Transformación considerados dentro del alcance del TFM.

Como se puede observar en la imagen anterior, la mayoría de los Centros de Transformación considerados en los datos se distribuyen por zonas de carácter principalmente residencial y comercial (zona resaltada en verde), aunque también hay transformadores que suministran consumos más industriales en polígonos y paseo marítimo (zona resaltada en azul).

Los datos recopilados de todos estos transformadores son los que se van a utilizar para entrenar los modelos de predicción. De todos ellos, se escogerá uno para el que se hará la predicción de potencia activa cuyos datos no se utilizarán en el entrenamiento, sino que servirán para evaluar la validez de los modelos.

2.2.2 Medidas históricas de potencia activa instantánea en transformadores

De cada uno de los transformadores introducidos en el apartado anterior se dispone del histórico de medidas de potencia activa trifásica para todo el año 2019. La Figura 8 es un ejemplo de la serie temporal de potencia activa consumida de uno de los transformadores del alcance del estudio, con frecuencia 5-minutal, con la que se quiere reflejar una muestra de información eléctrica utilizada.

En esta misma imagen se puede apreciar también como, para un periodo de dos semanas de junio, hay un intervalo de tiempo del que no se recibieron medidas del sensor (destacado en naranja), por lo que la serie temporal se representa con una línea recta. Este fenómeno, que ocurre en casi la totalidad de los transformadores considerados en este TFM, no se ha evitado ni se han aplicado técnicas de rellenado de huecos para atenuar su efecto con el objetivo de evaluar los modelos estudiados y la metodología desarrollada en un contexto lo más realista posible.

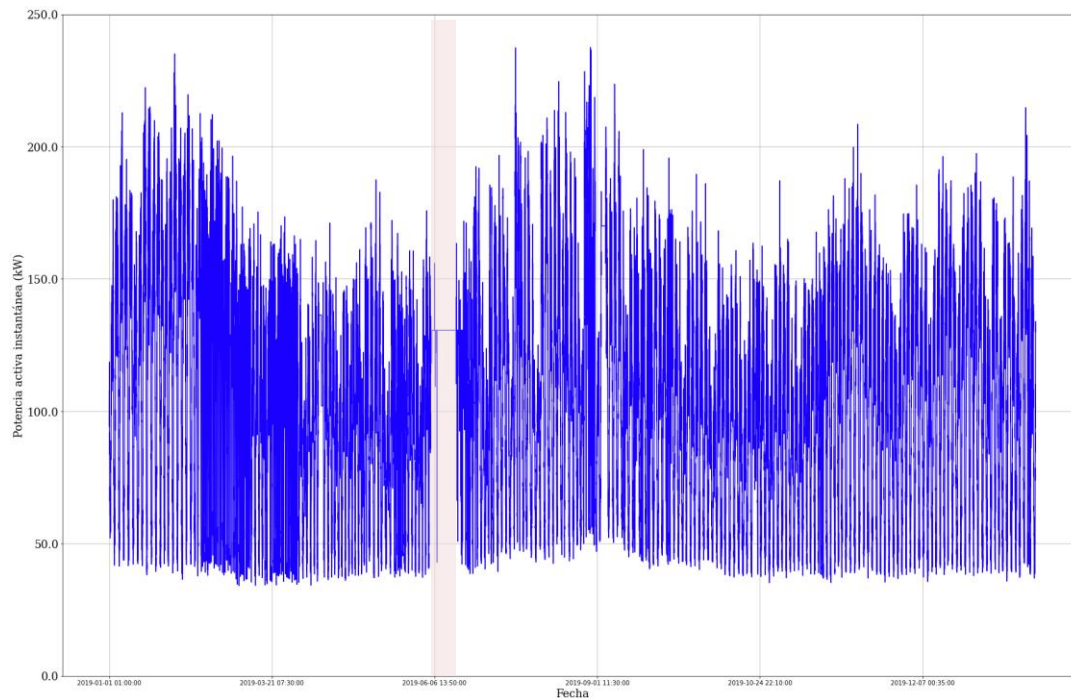


Figura 8 Ejemplo de serie temporal de potencia activa de uno de los transformadores

Con el objetivo de presentar un vistazo general de las medidas utilizadas de todos los transformadores, se ha representado en la Imagen 9 de forma matricial, y para cada uno de ellos, un diagrama de cajas que recoge la distribución de la potencia activa total registrada durante 2019 de cada uno y, como título, la potencia nominal del transformador en kVA. Estos diagramas de cajas se utilizan para representar gráficamente un conjunto de datos a través de sus cuartiles. Con este tipo de representación se puede apreciar a simple vista los siguientes componentes:

- Rango de valores del conjunto de datos.
- Datos atípicos u outliers.
- Rango intercuartil.
- Cuartiles (Q1, Q2 y Q3).
- Mediana (Q2).
- Valor mínimo y valor máximo.

Como se puede apreciar, en cada imagen se pueden distinguir los valores mínimos, máximo, percentil 25 y percentil 75 de potencia activa. Los dos primeros hacen referencia a la línea azul horizontal inferior y superior, respectivamente, y los dos últimos hacen referencia a la parte inferior y superior, respectivamente, de la caja y que recogen el 50% de la muestra. Dentro de esta se puede ver también una línea roja horizontal que representa la mediana de la muestra de potencias activa.

En la Tabla 1 se muestra un resumen del estado de carga global de los transformadores:

Tabla 1 Estado de carga global de los transformadores del alcance

Estadística	Estado de carga respecto a la potencia nominal S_n (%)
Media	21.08 %
Mínimo	0.0 %
Percentil 25	11.16 %
Mediana	19.19 %
Percentil 75	29.30 %
Máximo	84.63 %

A la vista la tabla anterior, se puede determinar que el parque de transformadores bajo estudio está muy descargado (percentil 75 por debajo del 30%) y con elevados picos de carga (>80%) en tres transformadores principalmente en los meses de agosto y enero.

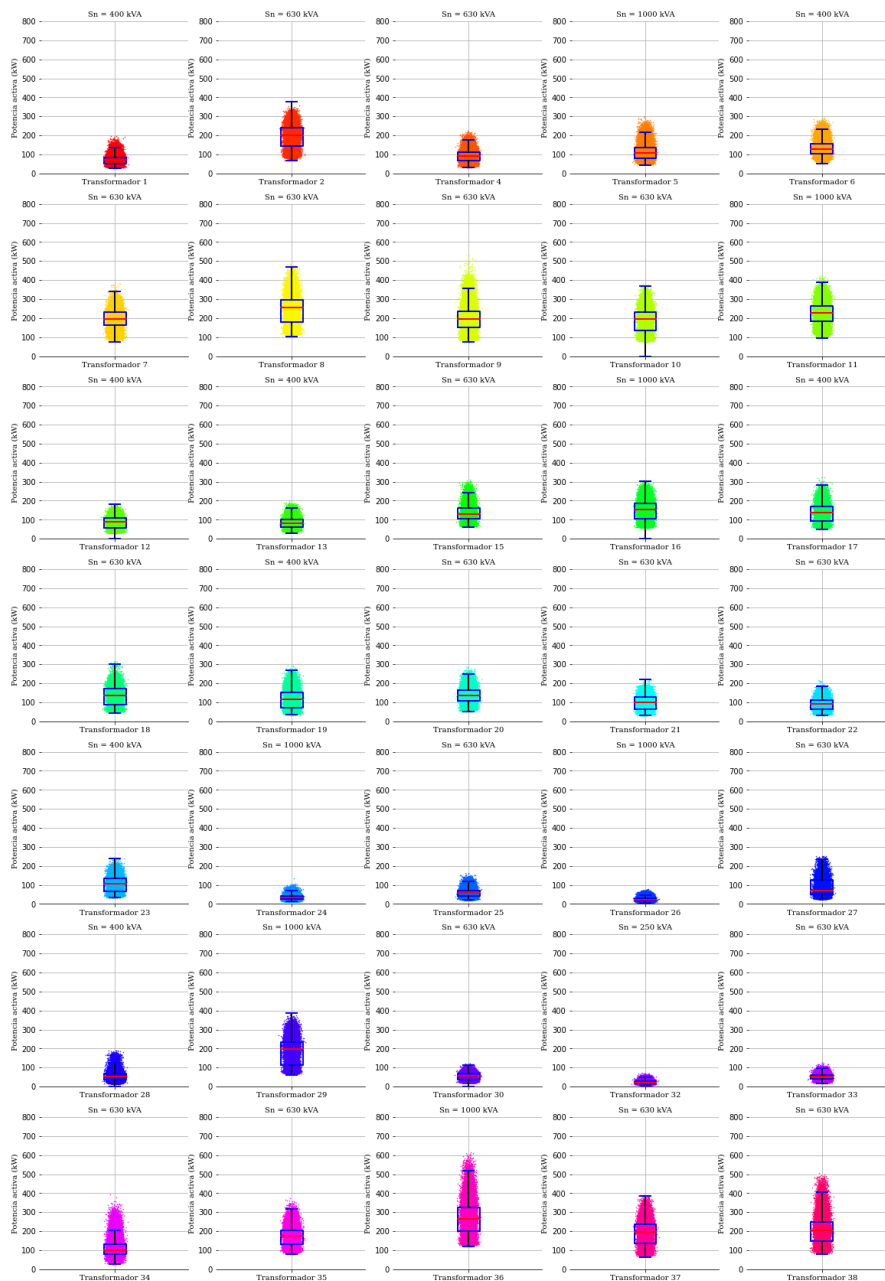


Figura 9 Diagrama de cajas con la distribución de potencia activa por transformador

2.2.3 Datos topológicos

Como ya se introdujo en el apartado de Caso de Uso, además de las medidas históricas de potencia activa, los modelos de predicción se han entrenado también basándose en información topológica de los transformadores. Esta información topológica se encuentra resumida en la Tabla 2 donde se detalla el valor de estos parámetros para cada uno de los transformadores.

Tabla 2 Resumen de la información topológica utilizada como dato de entrada en los modelos de predicción

Transformador	Nº LBTs	Nº PCR _s	Nº CUPS	Pc total (W)	Pc mediana (W)	Pc máxima (W)	Sn (kW)	Latitud	Longitud
Transformador 1	4	11	163	2284022	4600	8800	400	36.69528	-4.440919
Transformador 2	8	42	541	2258454	3300	4600	630	36.700142	-4.4482189
Transformador 3	7	56	340	2462900	3450	5500	630	36.704128	-4.4419351
Transformador 4	4	18	296	2212427	3450	6750	630	36.704163	-4.4427642
Transformador 5	8	14	155	2545924	5750	21200	1000	36.698914	-4.4377855
Transformador 6	8	36	517	2001546	2200	3450	400	36.703738	-4.4391963
Transformador 7	7	26	423	2015133	3300	5700	630	36.691643	-4.4464589
Transformador 8	7	24	637	2849338	3300	4600	630	36.693700	-4.4470459
Transformador 9	8	48	535	2552817	3300	4600	630	36.70339	-4.4365143
Transformador 10	7	17	511	2464983	3300	4600	630	36.69588	-4.4476243
Transformador 11	8	204	519	2701600	3300	4600	1000	36.704858	-4.4382654
Transformador 12	7	22	526	2252027	3300	4600	400	36.696243	-4.4499038
Transformador 13	4	15	213	1173283	4400	6836.4	400	36.693130	-4.4412596
Transformador 14	7	55	124	1208607	3300	13644.2	630	36.689097	-4.4445980
Transformador 15	12	33	395	1538244	3300	4600	630	36.692449	-4.4490708
Transformador 16	4	14	301	1729178	3300	4936	1000	36.697078	-4.4487323
Transformador 17	5	49	518	1689852	3300	4400	400	36.697764	-4.4479056
Transformador 18	8	16	169	1205841	3300	7122	630	36.695275	-4.4501760
Transformador 19	6	13	309	1558967	3300	4600	400	36.698140	-4.4495806
Transformador 20	7	37	412	1717637	3300	4580	630	36.704008	-4.4407350
Transformador 21	4	33	240	1389010	3300	5251.4	630	36.694192	-4.4488796
Transformador 22	4	35	215	1580218	4400	6242	630	36.694517	-4.4413750
Transformador 23	8	19	488	1795333	3300	5260	400	36.697078	-4.4487323
Transformador 24	12	12	105	666192	4600	5750	1000	36.700225	-4.4374517
Transformador 25	4	10	128	944092	4600	6928	630	36.69528	-4.4409193
Transformador 26	11	11	115	518992	4600	5750	1000	36.700225	-4.4374517
Transformador 27	4	8	32	322394	3450	29492.8	630	36.691643	-4.4464589
Transformador 28	3	24	25	885052	15100	126000	400	36.69588	-4.4476241
Transformador 29	3	9	227	890943	3300	5221.6	1000	36.696243	-4.4499032
Transformador 30	4	14	149	576928	3300	4600	630	36.69473	-4.4498637
Transformador 31	3	2	2	187928	93964	163592.8	400	36.694676	-4.4512841
Transformador 32	3	11	89	310561	3300	4600	250	36.688940	-4.4486132
Transformador 33	4	7	92	757940	4930	10392	630	36.693130	-4.4412596
Transformador 34	8	21	20	489690	6928	52300	630	36.689516	-4.4427367
Transformador 35	8	106	376	3035638	3450	6928	630	36.703146	-4.4351664
Transformador 36	10	34	999	3524466	3300	3450	1000	36.699281	-4.4506654
Transformador 37	7	25	605	3018690	3300	4600	630	36.699881	-4.4495013
Transformador 38	7	54	856	3812605	3300	4400	630	36.692218	-4.4503424
Transformador 39	7	153	514	3791484	3450	5750	630	36.704851	-4.4343025
Transformador 40	8	68	752	3254352	3300	4600	630	36.707645	-4.4351046

Toda esta información, al igual que las medidas históricas de potencia activa de los transformadores, son datos reales relativos a transformadores de la red de distribución de Endesa en Málaga y se han recogido en el ámbito del proyecto PASTORA [14].

Para mostrar gráficamente los datos de la tabla anterior, en la Figura 10 se ha utilizado un diagrama de pares o *pairplot* que es una herramienta gráfica comúnmente utilizada en estadística y analítica de datos para, por un lado, analizar la distribución de cada variable representada y, por otro lado, comprender la relación por pares entre diferentes variables de un conjunto de datos, en este caso, de los datos topológicos. Para representar este gráfico se ha utilizado la librería *seaborn* [15] de *Python*.

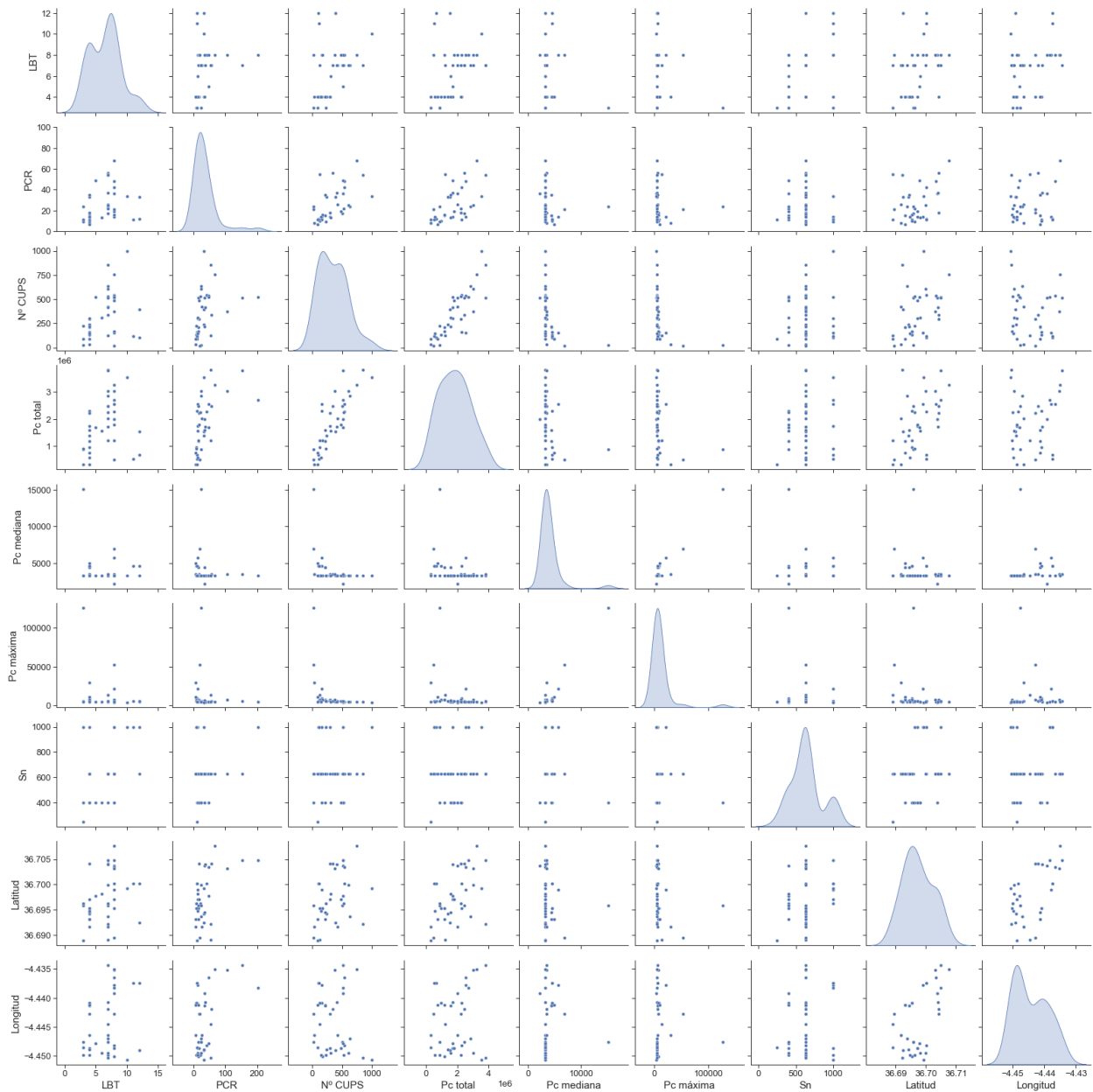


Figura 10 Diagrama de pares de los datos topológicos de los transformadores considerados en el alcance del TFM

En el diagrama de pares anterior, la distribución de cada variable se muestra en forma de histograma a lo largo de los cuadros diagonales, y todos los demás cuadros muestran un diagrama de dispersión de la relación entre cada combinación de variables por pares. Por ejemplo, la caja superior izquierda muestra la distribución del número de LBTs presente en la muestra de datos y la caja justo debajo muestra un diagrama de dispersión de valores entre el número de LBTs y el número de PCRs.

Este tipo de representaciones son de gran utilidad cuando se empieza a trabajar con un conjunto de datos grande, como es el caso de este TFM, puesto que permite obtener información de la relación entre variables de forma visual, ayuda a comprobar cómo afecta a una variable los cambios producidos en otra estableciendo relaciones de causa/efecto y permite corroborar, por medio de las correlaciones, si el conjunto de datos tiene o no sentido.

De igual forma, las gráficas de la diagonal permiten, de un solo vistazo, identificar por cada variable la presencia de *outliers* que podrían alterar el ajuste de los modelos de predicción y, también, conocer si el conjunto de datos están más o menos balanceado.

Por ejemplo, analizando con detenimiento los histogramas de la diagonal, destaca significativamente como las variables “potencia contratada mediana” y “potencia contratada máxima” presentan para algunos transformadores valores muy extremos, por encima de los 100.000 W, muy alejados del valor normal presente en la muestra, en torno a los 3.300 W. Esto podría deberse bien a errores en los datos de potencia contratada recogidos por la distribuidora para ciertos puntos de suministro o bien a que hay en la muestra puntos de suministro muy diferentes al resto, como puntos de suministro trifásicos de gran consumo de carácter industrial o servicios.

Otro ejemplo, en este caso relativa a las gráficas de dispersión, puede ser la alta correlación positiva entre el “número de CUPS” y la “potencia contratada total” del transformador o, también llamativo, la falta de correlación positiva entre la “potencia nominal del transformador” (S_n) y el “número de CUPS” o “potencia contratada total”, síntoma de una posible estrategia de sobredimensionamiento de la distribuidora y que debe tenerse en cuenta en los modelos de predicción.

Este tipo de relaciones entre variables topológicas, junto con el resto de datos que se han utilizado, son las que se busca que los algoritmos de predicción evaluados utilicen para generalizar sus modelos y hacer estimaciones de potencia válidas.

2.2.4 Datos meteorológicos

Para el mismo periodo de tiempo de las medidas de potencia, año 2019, se ha accedido a los datos meteorológicos de la ciudad en la que se localizan los transformadores del alcance de este TFM para utilizarlo, además de los datos anteriores, en el ajuste de los modelos de predicción. Estos datos meteorológicos proceden de BBDD públicas, en particular se trata de datos procedentes de AEMET, y están compuestos por las siguientes variables:

- Temperatura ambiente.
- Presión atmosférica.
- Velocidad del viento.
- Dirección del viento.

En la Figura 11 se ha representado la distribución de estas cuatro variables en forma de diagrama de cajas.

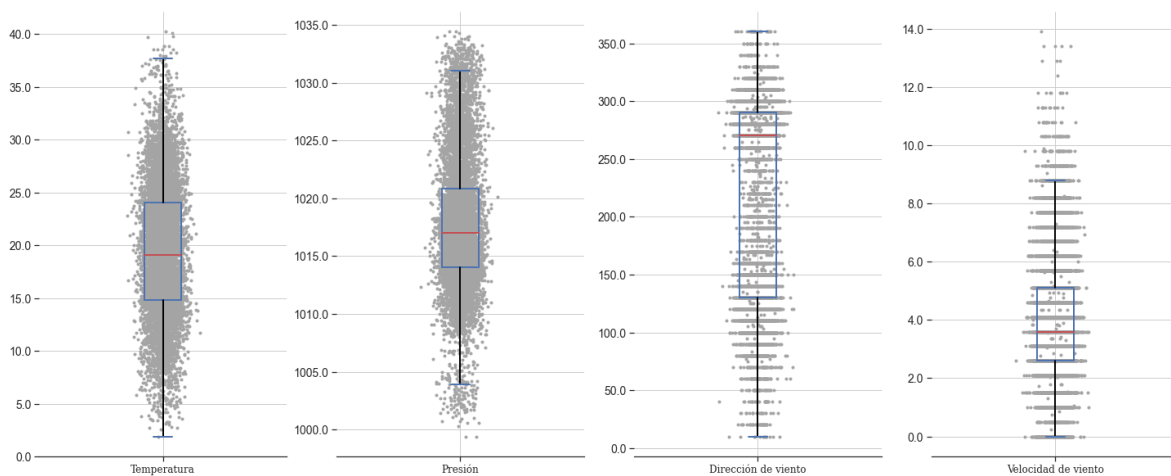


Figura 11 Diagrama de cajas de variables meteorológicas utilizadas para la predicción

En la Figura 11 se muestra el diagrama de cajas de las cuatro variables meteorológicas utilizadas. Para cada imagen se ve en el eje de ordenadas el rango de valores de cada variable, en gris todos los puntos del conjunto de datos de cada variable, en rojo la mediana y en azul la caja acotada por los cuartiles Q1 y Q3 y que contiene

el 50% de la muestra.

Estos datos son de gran importancia para la predicción de potencia activa en transformadores puesto que la meteorología influye significativamente el consumo final de los clientes suministrados aguas abajo de estos transformadores.

2.2.5 Otros datos

Además de los históricos de potencia activa, de los datos topológicos de los transformadores y de los datos meteorológicos para el mismo periodo de los históricos, en los modelos de predicción se ha incluido información relativa a cada una las fechas consideradas en el histórico para las que se ha identificado:

- Si se trata de un día de la semana o de fin de semana.
- Si se trata de un periodo vacacional o no.
- Si se trata de un día festivo o día laborable.

3 INTELIGENCIA ARTIFICIAL Y DESCRIPCIÓN MODELOS DE PREDICCIÓN UTILIZADOS

Hay una fuerza motriz más poderosa que el vapor, la electricidad y la energía atómica: la voluntad.

- Albert Einstein -

La Inteligencia artificial es según McKinsey [16] empresa reconocida como la consultora estratégica más prestigiosa del mundo, la “capacidad de una máquina para realizar funciones cognitivas que asociamos a la mente humana, como percibir, razonar, aprender, interactuar con el entorno y resolver problemas o incluso utilizar la creatividad”. Los sistemas de Inteligencia Artificial se adaptan y toman decisiones en base a lo que aprenden viendo datos. Esta capacidad de aprendizaje se puede conseguir con técnicas de Machine Learning (Aprendizaje Automático) y Deep Learning (Aprendizaje Profundo).

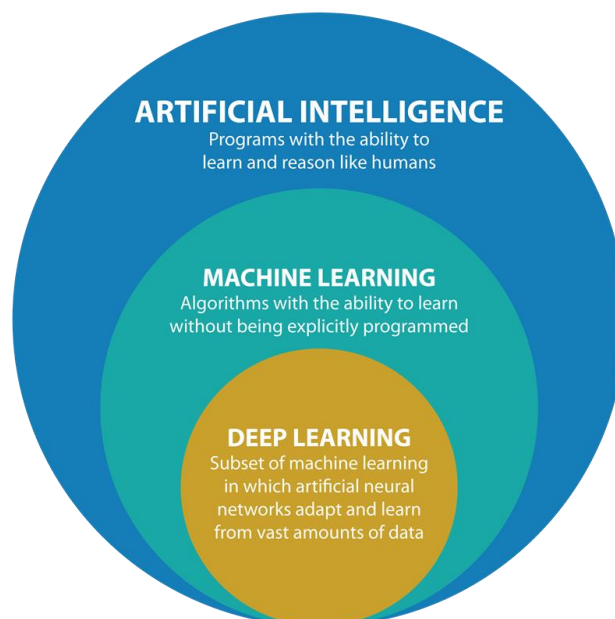


Figura 12 Esquema descriptivo de la Inteligencia Artificial, el Machine Learning y el Deep Learning. Fuente: <https://medium.com/@experiencia18/diferencias-entre-la-inteligencia-artificial-y-el-machine-learning-f0448c503cd4>

Como se puede ver en la Figura 12, la Inteligencia Artificial es un concepto muy amplio formado por el Aprendizaje Máquina (en inglés Machine Learning) y este, a su vez, por el Aprendizaje Profundo (en inglés Deep Learning). En los últimos años se está investigando y avanzando en una nueva técnica comprendida dentro del Aprendizaje Máquina conocida como Aprendizaje por Refuerzo (del inglés Reinforcement Learning).

El primero se define como la capacidad que tienen las máquinas de recibir un conjunto de datos y aprender por sí mismas, cambiando y ajustando los algoritmos a medida que procesan información y conocen el entorno. Es decir, usar algoritmos para analizar y procesar datos, aprender de ellos y luego ser capaces de hacer una predicción o sugerencia sobre algo.

El Aprendizaje Profundo, en su caso, se define como el conjunto de técnicas que buscan el aprendizaje con el ejemplo y que son capaces de reconocer problemas y soluciones complejas.

Por último, el Aprendizaje por Refuerzo es una técnica que trabaja en armonía con el aprendizaje semi-supervisado y supervisado y se emplea cuando no existe un conjunto de datos de los que un algoritmo pueda aprender conductas y sacar conclusiones.

Mientras el Aprendizaje Máquina utiliza algoritmos para analizar datos, aprender y generar resultados o tomar decisiones con base en lo aprendido, el deep learning estructura los algoritmos en capas de redes neuronales que le ayudan a aprender y generar resultados más precisos.

Para poder explotar las capacidades de las técnicas anteriores todo lo posible con los datos que se disponen, de manera que puedan encontrar o inferir el patrón que permita hacer una correcta predicción, requiere separar el conjunto de datos totales disponibles en dos subconjuntos, primero para entrenar el modelo y, posteriormente, para testear o comprobar el algoritmo entrenado. Estos son conocidos popularmente en el ámbito de la inteligencia del dato como conjuntos de train y de test, respectivamente.

Los datos de entrenamiento, o *training data*, son los datos que se utilizan para entrenar el modelo y la calidad del modelo de aprendizaje automático será proporcional a la calidad de los datos y cómo se utilicen en el caso de uso de aplicación. Esto se demostrará a lo largo de los siguientes apartados. Para ello es indispensable el uso de técnicas de limpieza, unificación, consolidación y normalización de datos para que se pueda utilizar y extraer información de valor.

Los datos de prueba, o *testing data*, son el conjunto de datos que se reservan para comprobar que el modelo que se ha entrenado con el training data funciona correctamente. Es importante que el conjunto de datos de prueba tenga un volumen suficiente como para generar resultados estadísticamente significativos, y a la vez, que sea representativo del conjunto de datos global.

Igualmente, es necesario que, para ambos conjuntos, el formato, así como la estructura de los datos, sean exactamente iguales. La estructura de columnas tiene la forma de la siguiente tabla:

Tabla 3 Ejemplo de estructura de datos utilizada en los modelos de predicción

LBT	PCR	CUPS	Ptot	Pmed	Sn	Pmax	Mes	Día	Hora	Minuto	...	Día festivo	P. Activa (kW)
12	33	395	1538244	3450	630	5500	1	1	0	0	...	0	131.4
12	33	395	1538244	3450	630	5500	1	1	0	5	...	0	136.5
...
4	35	215	1580218	4400	630	9860	9	29	7	55	...	1	48.01

De la tabla anterior, todas las columnas excepto la última, relativa con la potencia activa, corresponden con los atributos o *features* y la última con la etiqueta o *label*. Los atributos son la entrada de información al modelo, las variables independientes a partir de las cuales el algoritmo debe inferir las relaciones existentes entre las mismas para generar una salida. Dicha salida corresponde con la etiqueta, el objetivo de predicción del algoritmo.

La base del Machine Learning, y la relación entre los datos de entrada y las variables de salida, se puede resumir con la siguiente expresión:

$$y = f(X)$$

donde:

- y : salida, valor de predicción.
- f : función matemática, modelo de predicción o algoritmo
- X : entrada, conjunto de variables independientes conocidas.

De forma sencilla, los algoritmos de Machine Learning pueden describirse como las funciones (f) que mejor combinan las variables de entrada (X) para generar unas salidas (y).

Todos los algoritmos que se han probado en este Trabajo Fin de Máster se han programado utilizando el lenguaje de programación Python. Para ello, entre otras, hay dos librerías de modelos de predicción que se han utilizado: *sklearn* y *tensorflow*. La primera para evaluar modelos de predicción de Machine Learning clásicos y la segunda para construir Redes Neuronales, tanto Redes Neuronales simples compuestas por una única capa de neuronas como Redes Neuronales Profundas o *Deep Learning* compuestas por más de una capa de neuronas. Los algoritmos que se han probado y evaluado en este Trabajo Fin de Máster son los siguientes:

- Decision Tree Regressor (DTR).
- Gradient Boosting Regressor (GBR).
- Extreme Gradient Boosting (XGBoost).
- Redes Neuronales Profundas (DNL).

A continuación se describe con más detalle cada uno de ellos.

3.1 Decision Tree Regressor

Los algoritmos basados en árboles de decisión son uno de los modelos de regresión más utilizados tanto en el ámbito científico como en el empresarial. Estos algoritmos destacan por dos características principalmente:

- Los árboles de decisión son modelos predictivos que combinan conjuntos de reglas binarias para calcular el valor objetivo.
- Cada árbol es un modelo sencillo que se define por tres variables: número de ramas, número de nudos y número de hojas.

Proceso del árbol de decisión

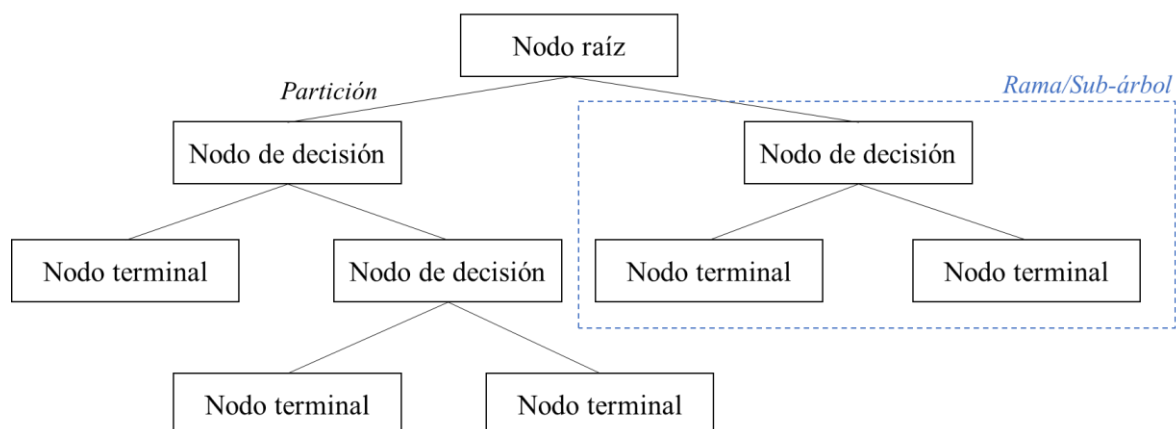


Figura 13 Esquema de la lógica de generación de un árbol de decisión. Fuente: Elaboración propia.

Un árbol de decisión llega a una estimación haciendo una serie de preguntas a los datos, cada pregunta reduce los valores posibles hasta que el modelo está lo suficientemente seguro como para hacer una predicción válida. El orden de las preguntas, así como su contenido, están determinados por la parametrización que se le aplique al modelo. Además, las preguntas formuladas están todas en formato Verdadero/Falso.

La decisión de hacer divisiones estratégicas para ir creando sub-árboles afecta en gran medida a la precisión de

árbol en su total. Los árboles de decisión normalmente hacen estas divisiones utilizando indicadores como el error cuadrático medio (MSE). Con esto, deciden dividir un nodo en dos o más subnodos.

Para su evaluación, se ha utilizado el modelo `DecisionTreeRegressor` [17] de la librería `sklearn` para el que se han analizado los parámetros principales:

- Criterio de división
- Profundidad máxima
- Atributos máximos a tener en cuenta para la división
- Mínimo número de muestras para la división

Los algoritmos de predicción basados en árboles de decisión se han convertido en uno de los referentes dentro del ámbito predictivo debido a los buenos resultados que generan en problemas muy diversos. Su popularidad y la extensión en su uso se deben principalmente a que:

- Pueden aplicarse tanto a predictores numéricos (el caso de este TFM) como categóricos
- No es necesario que el conjunto de datos cumpla con una distribución específica
- No se ven muy influenciados por outliers
- Son fácilmente interpretables.

3.2 Gradient Boosting Regressor

La técnica de *Boosting* en *Machine Learning* es una forma de combinar múltiples modelos sencillos o débiles, conocidos como *weak learners*, trabajando como un único modelo de predicción. Esta técnica es también conocida como “modelos aditivos” puesto que crea un modelo de conjunto mediante la combinación secuencial de varios modelos débiles, asignando ponderaciones a la salida individual de cada uno de ellos. El término *Gradient* se debe al hecho de que en el algoritmo *Gradient Boosting* se utiliza la técnica de descenso de gradiente para minimizar la función de pérdida.

En el Gradient Boosting Regressor (GBR) los modelos débiles utilizados son árboles de decisión entrenados de forma secuencial de manera que cada nuevo árbol trata de mejorar la predicción del árbol anterior, como se puede ver en la Figura 14.

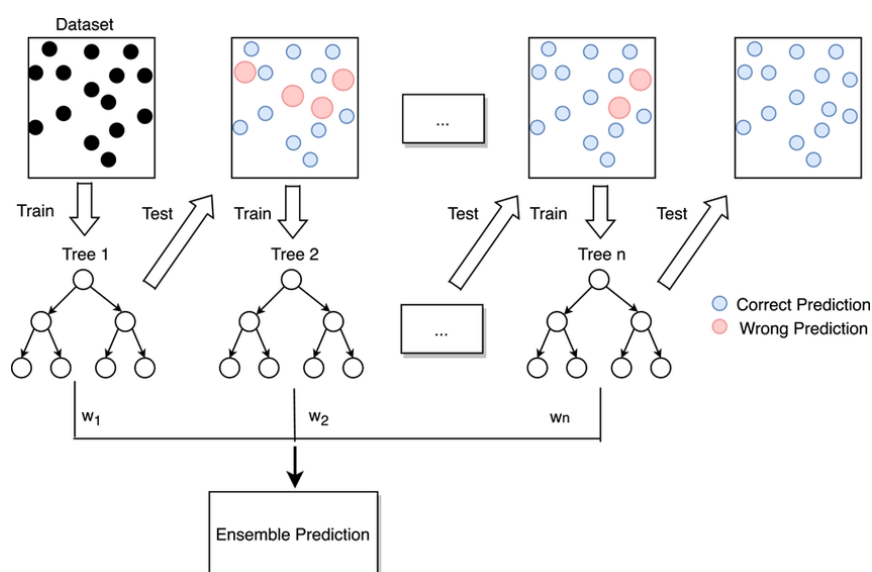


Figura 14 . Esquema representativo del funcionamiento del modelo GBR. Fuente: https://www.researchgate.net/figure/Flow-diagram-of-gradient-boosting-machine-learning-method-The-ensemble-classifiers_fig1_351542039

Los modelos GBR calculan la diferencia entre la predicción actual hecha con modelo global y el valor objetivo correcto conocido. Esta diferencia se llama residual. Después de eso, el GBR entrena un modelo débil que mapea las características del set de datos a ese residual. Este residuo predicho por un modelo débil se agrega a la entrada del modelo existente y, por lo tanto, este proceso reorienta al modelo hacia el objetivo correcto. Repetir este paso una y otra vez mejora la predicción general del modelo.

Los principales parámetros que definen los modelos GBR y que se deben ajustar para obtener la mejor predicción posible son:

- Número de estimadores, es decir, el número de árboles de decisión considerados en el modelo.
- Máxima profundidad de los árboles de decisión.
- Learning rate o ratio de aprendizaje que determine la velocidad de búsqueda de óptimos para minimizar la función de pérdida.
- Criterio de división.
- El tipo de función de pérdida.

Los modelos GBR heredan todas las ventajas del uso de árboles de decisión como modelos base, exceptuando el inconveniente de que, al combinar múltiples árboles, se pierde la interpretabilidad que tienen los modelos basados en un único árbol.

3.3 XGBoost

XGBoost, o eXtreme Gradient Boosting, es una librería software de código abierto que tiene implementados algoritmos de Aprendizaje Automático distribuidos basado en árboles de decisión que funcionan bajo el marco de la Potenciación de Gradiente o Gradient Boosting. Este marco, como se introdujo en el apartado anterior, consiste en una técnica de que produce un modelo predictivo en forma de un conjunto de modelos de predicción débiles, lo hace de forma distribuida y los generaliza permitiendo la optimización de su predicción. Esta tecnología tiene como origen una publicación de Jerome Friedman en el año 2000 [18].

Esta librería utiliza aproximaciones mucho más precisas para encontrar el mejor modelo de árboles de decisión, los cuales se construyen en paralelo en lugar de secuencialmente como los Gradient Boosting Decision Trees, siguiendo una estrategia de búsqueda por niveles a través de los gradientes. Entre los trucos que otorgan un gran rendimiento a los algoritmos XGBoost destacan:

- Utiliza gradientes de segundo orden, derivadas parciales de segundo orden de la función de pérdidas (similar al método de Newton), lo que le proporciona más información sobre la dirección del gradiente facilitando, así, la minimización de la función de pérdidas.
- Tiene implementadas técnicas de regularización avanzadas, regularización Lasso L1 y Ridge L2 [17], que consiste principalmente en añadir una penalización a la función de coste, simplificando los modelos y consiguiendo que estos generalicen mejor puesto que evita su sobreajuste.

Además de esto, XGBoost entrena y ajusta sus parámetros muy rápido y permite realizar esta tarea en paralelo o distribuido en cluster de servidores, lo que le otorga una ventaja significativa con respecto a otros modelos en cuanto a su aplicación.

Por estas razones, esta librería se ha vuelto muy popular en los últimos años tanto en competiciones de Machine Learning como en su aplicación a nivel profesional. Esto hace que una amplia lista de científicos de datos de todo el mundo está contribuyendo en el desarrollo de esta librería de código abierto haciéndola mucho más eficiente, flexible y adaptable.

3.4 Redes Neuronales

Las técnicas de machine learning convencionales se encuentran limitadas en su habilidad de procesar datos naturales en bruto. Durante décadas, construir un sistema de reconocimiento de patrones o machine learning requería de ingeniería de detalle y un dominio experimentado en la extracción de características, "features", con

el objetivo de transformar los datos en bruto de manera adecuada para que sirvieran de entrada a los subsistemas de aprendizaje, a menudo clasificadores, de manera que estos pudieran detectar o clasificar patrones.

En este contexto, surge el concepto de aprendizaje profundo o deep learning. En comparación a lo anterior, los algoritmos de deep learning pueden llevar a cabo la extracción de features de manera automática, lo que supone que los desarrolladores no necesiten ser expertos en el dominio del problema en cuestión y reduce al mínimo el esfuerzo humano necesario. Así, las técnicas de deep learning, como se refleja en la Figura 12, son un subconjunto de métodos de machine learning que pueden aprender patrones de manera jerárquica, lo que permite obtener conceptos complejos a partir de otros más simples, construyendo arquitecturas profundas.

En estos algoritmos se definen capas de distintos niveles de representación del conocimiento, donde las características de capas de niveles más altos se extraen a partir de las de las capas de niveles más bajos. Este tipo de arquitecturas se inspiran originalmente en el funcionamiento del cerebro humano, dando lugar a una familia de métodos que recibe el nombre de redes neuronales artificiales.

Al igual que el cerebro humano, las redes neuronales están compuestas por un gran número de unidades de procesamiento, llamadas neuronas, y cada una de ellas está conectada a otras muchas. Las neuronas operan en paralelo y se transfieren información entre ellas a través de sinapsis.

El objetivo de las redes neuronales artificiales es imitar el cerebro humano y sus neuronas, a través de una unidad fundamental: el perceptrón o neurona. Típicamente, el perceptrón cuenta con muchas entradas y una única salida. Cada entrada tiene asociada un peso, que multiplica a la señal de entrada, ponderando su valor. Estas entradas ponderadas se suman junto con un término llamado bias o sesgo.

La salida de la neurona se calcula aplicando a la suma anterior una función de activación que añade una no linealidad al resultado. Una función de activación es una función que transmite la información generada por la combinación lineal de los pesos y las entradas, es decir, estas constituyen la manera de transmitir la información por las conexiones de salida. Como lo que generalmente se busca es que la red sea capaz de resolver problemas cada vez más complejos, las funciones de activación generalmente harán que los modelos sean no lineales. Todo esto está representado en la Figura 15.

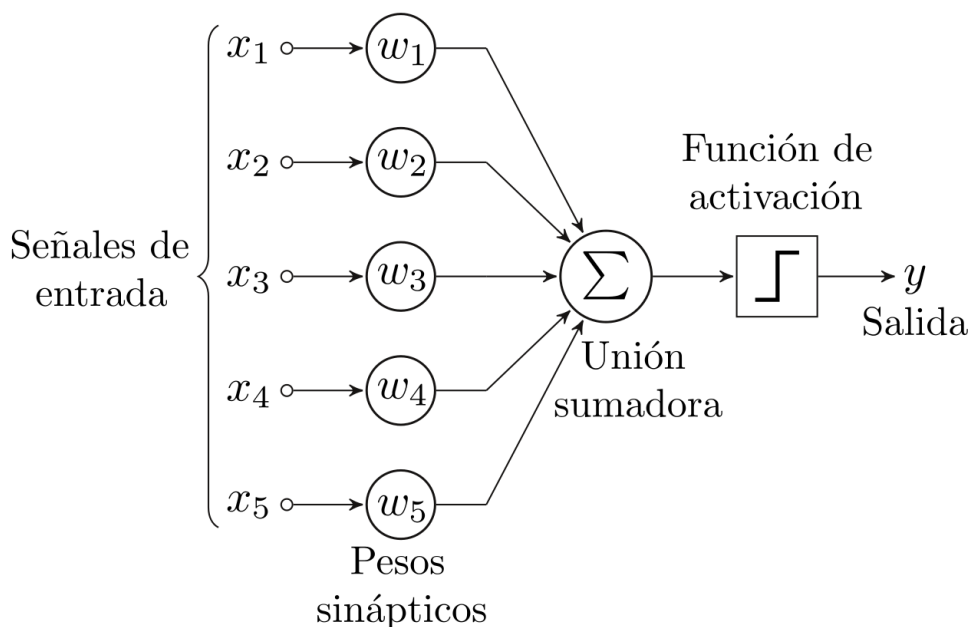


Figura 15 Esquema de funcionamiento de un perceptrón o neurona artificial. Fuente: Wikipedia.

Entre las funciones de activación más conocidas o más usadas se encuentran:

- Función escalón.
- Función sigmoideal.
- Función rectificadora (ReLU).

- Función tangente hiperbólica.
- Funciones de base radial (gaussianas, multicuadráticas, multicuadráticas inversas, ...)

Las neuronas pueden disponerse en formando capas, donde todas las neuronas de una capa toman como entrada las salidas de las neuronas de la capa previa, calculan sus valores de salida en paralelo, y con estos valores alimentan a todas las neuronas de la siguiente capa. Esta arquitectura recibe el nombre de perceptrón multicapa o red neuronal feedforward. La superposición de muchas funciones simples no lineales permite al perceptrón multicapa aproximar funciones extremadamente no lineales, lo que lo convierte en una excelente herramienta en tareas de detección de patrones, clasificación y regresión.

El algoritmo que se usa principalmente para el entrenamiento de las Redes Neuronales se conoce como *Backpropagation*. Este consiste en un método de ajuste fino de los pesos de las redes neuronales basándose en los errores obtenidos en las iteraciones previas. Ser capaz de hacer un ajuste apropiado de los pesos de la Red Neuronal permite reducir el ratio de error del modelo y hacerlo mucho más fiable incrementando su capacidad de generalización.

El *Backpropagation* en Redes Neuronales es una abreviatura de *backward propagation of errors*, que quiere decir *propagación hacia atrás de los errores*, y hace referencia a la capacidad que tiene este método de “enseñar” a las capas previas de neuronas como de válidos son los pesos asignados en una iteración a la hora de predecir. De esta manera se consigue calcular el gradiente de la función objetivo con respecto a los pesos de la Red Neuronal.

En este Trabajo Fin de Máster se han evaluado diferentes estructuras de redes neuronales profundas, y los resultados que se van a presentar son relativos a dos tipos de redes neuronales.

Los parámetros principales que se han ido ajustando a lo largo de las pruebas son:

- Función de activación: función limitadora o umbral, encargada de modificar el input recibido antes de proseguir a otra neurona.
- Epoch: este es el número de veces que se ejecutaran los algoritmos de forwardpropagation y backpropagation. En cada ciclo (epoch) todos los datos de entrenamiento pasan por la red neuronal para que esta aprenda sobre ellos.
- Batch size: es el número de datos que tiene cada iteración de un ciclo (epoch), esto es útil porque la red neuronal actualiza los pesos y el bias más veces. También ayuda a limitar la cantidad de información que se procesa en cada ciclo, pues cuando se tienen grandes cantidades de datos se necesitan computadoras con más memoria y la red neuronal tarda más en ejecutar cada ciclo.

Los dos últimos son de gran importancia para evitar que el algoritmo llegue a sobreajustarse.

Para evaluar el comportamiento de este tipo de modelos de predicción en el caso de uso abordado, en este Trabajo Fin de Máster se han evaluado dos configuraciones de Redes Neuronales diferentes:

- Red Neuronal 1

La primera RN evaluada consiste en tres capas, una capa de entrada con 24 neuronas, una capa oculta con 24 neuronas y una capa de salida con 1 neurona. La Figura 16 muestra una representación esquemática de la estructura de la Red Neuronal 1.

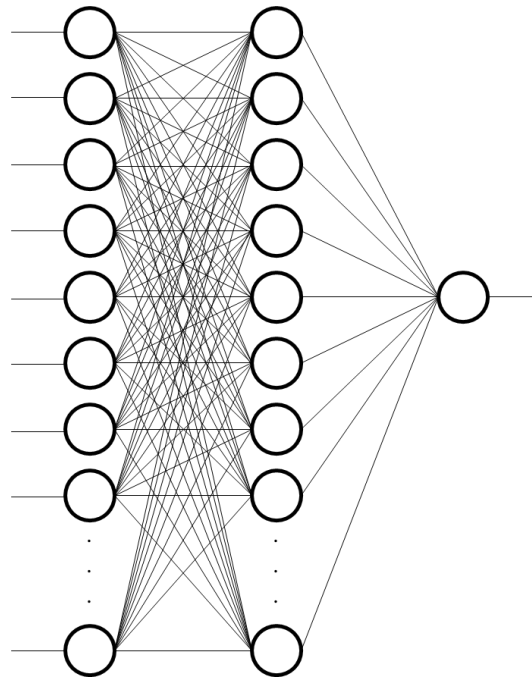


Figura 16 Esquema de la Red Neuronal 1 evaluada

- Red Neuronal 2

La segunda RN evaluada consiste en cuatro capas, una capa de entrada con 24 neuronas, dos capas ocultas con 24 neuronas cada una y una capa de salida con 1 neurona. La Figura 17 muestra una representación esquemática de la estructura de la Red Neuronal 1.

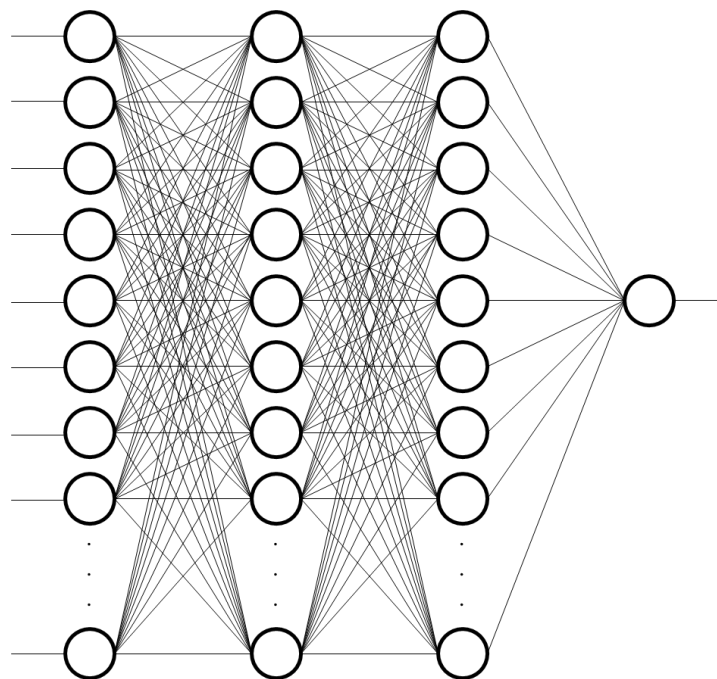


Figura 17 Esquema de la Red Neuronal 2 evaluada

En ambos casos, las capas de neuronas están intercaladas por capas intermedias de Dropout, un método de regularización que desactiva un número de neuronas de las capas previas de forma aleatoria. En cada iteración de la red neuronal, el Dropout desactivará diferentes neuronas, las neuronas desactivadas no se tendrán en cuenta para el *forwardpropagation* ni para el *backwardpropagation*, lo que obliga a las neuronas cercanas a no depender tanto de las neuronas desactivadas. Este método ayuda a reducir el *overfitting*, ya que las neuronas cercanas suelen aprender patrones que las relacionan y estas relaciones pueden llegar a formar un patrón muy específico con los datos de entrenamiento. Con Dropout esta dependencia entre neuronas es menor en toda la

red neuronal, de esta manera las neuronas necesitan trabajar mejor de forma solitaria y no depender tanto de las relaciones con las neuronas vecinas.

Por último, hay que destacar que el proceso de entrenamiento se realizó utilizando como función de activación la función *relu* [19] y como **función de pérdida el *Mean Squared Error***

3.5 Otras técnicas de procesamiento de datos empleadas

A continuación se describirán algunas de las principales técnicas de procesamiento de datos que se han empleado en los desarrollos de este TFM.

3.5.1 Codificación de datos de entrada

Como se ha presentado en el Apartado 2.2, en el conjunto de datos que se han utilizado para el entrenamiento de algoritmos hay, en su mayoría, características de categoría numérica. Sin embargo, se ha considerado también otra información de tipo categórico como la clasificación de las fechas según días festivos, días entre semana y fin de semana, y periodos vacacionales. Hay algunos algoritmos de Machine Learning como los árboles de decisión que pueden trabajar con datos categóricos, pero muchos otros no pueden operar directamente con las etiquetas de estas categorías. Esto ocurre especialmente en los algoritmos basados en Redes Neuronales como el Deep Learning que necesita preprocesar esta información categórica y convertirla a atributos numéricos.

Para cumplir este objetivo, hay dos estrategias diferentes que se pueden seguir:

- a) Codificación ordinal mediante enteros.
- b) Codificación one-hot.

En el primer caso, para cada una de las categorías, se le asigna un valor entero único a cada una de sus etiquetas. Por ejemplo, en la categoría “Color” compuesta por tres etiquetas “Rojo”, “Amarillo” y “Azul”, éstas se codificarían con los siguientes identificadores: 1, 2 y 3, respectivamente. Esta técnica es fácilmente aplicable y reversible, y para algunos algoritmos puede ser suficiente. Sin embargo, ésta adolece del problema que los números asignados en su codificación tienen una relación entre sí y los algoritmos de Machine Learning pueden ser capaces de aprender estas relaciones y aprovecharlas en su lógica de predicción. En muchos casos, como el de este TFM, dicho comportamiento debe evitarse.

En contraposición al anterior, la segunda estrategia se aplica en aquellos casos en los que las etiquetas de cada variable categórica no tienen una relación de orden entre sí. En estos casos, se aplica una codificación especial con la que se agrega una nueva variable binaria, con valores 0 y 1, para cada etiqueta posible. Siguiendo el ejemplo anterior, con la codificación one-hot, se añadirían tres nuevas columnas al conjunto de datos (“Rojo”, “Amarillo” y “Azul”) en las que todas sus filas serán 0, excepto en aquellas a las que pertenecen los datos, a la que se asigna un 1.

Ésta última es la técnica que se ha utilizado en este TFM.

3.5.2 Escalado de datos

El escalado de datos es un paso dentro del preprocesamiento de datos que se aplica a categorías numéricas y tiene por objetivo establecer el rango de valores de las variables numéricas dentro de una escala similar. Éste es generalmente el último paso en todo el preprocesamiento de datos y se realiza justo antes de los algoritmos de aprendizaje automático.

Este paso es de gran importancia puesto que hay muchos algoritmos, como los basados en descenso de gradiente, regresión lineal o logística, que son especialmente sensibles a la escala de las variables de entrada, haciendo que aquellas con mayor rango dominen con respecto a las que tienen un rango menor e incluso afectando al tiempo de convergencia.

Existen diversos métodos para escalar las variables de entrada: estandarización, normalización media, escalado

a valor máximo, escalado a cuantiles y mediana o escalado a valores máximos y mínimos. Éste último, también conocido como MinMaxScaler, es la técnica que se ha utilizado en el preprocesamiento de datos de este TFM y consiste en reescalar el rango de características para escalar el rango en $[0, 1]$ o $[-1, 1]$, en este caso se ha utilizado el rango $[0, 1]$.

3.5.3 Clustering

El clustering es la técnica de aprendizaje no supervisado más extendida y consiste en detectar potenciales grupos entre los datos de entrada. Clustering es una técnica de clasificación de los datos crudos de manera razonable, de manera que permita encontrar patrones ocultos que puedan existir en los conjuntos de datos. Este proceso permite agrupar datos en grupos (clusters) inicialmente inconexos de manera que los datos pertenecientes a un mismo cluster sean similares mientras que aquellos que pertenezcan a clusters distintos difieran entre sí.

Entre los algoritmos de clustering más extendidos destacan K-Means, Correlation clustering, Análisis de Componentes Principales o Análisis Factorial. El algoritmo que se ha utilizado en este TFM es el primer, el K-Means, por su simpleza y rapidez. Consiste en clasificar los datos de entrada en K clusters distintos a través de un proceso iterativo, convergiendo a un mínimo local, dando como resultado clusters compactos e independientes.

El algoritmo consta de dos fases: la primera consiste en seleccionar K centroides de manera aleatoria, donde el valor de K es fijo y definido previamente. Una vez hecho esto, se asocia cada dato del set de datos al centro más cercano (normalmente se emplea la distancia euclídea). La primera fase queda completada cuando ya no queden datos sin asociar a un centroide, y de esta forma queda conformado un agrupamiento inicial. El siguiente paso consiste en recalcular los K nuevos centroides como el baricentro de los clusters generados en la primera fase. Este proceso se repite hasta que al iterar no se producen cambios en las posiciones de los centros o se alcanza algún otro criterio de convergencia.

K-means [17] tiene como ventajas ser un algoritmo muy rápido, simple y computacionalmente eficiente, sin embargo es difícil determinar el valor de K.

3.5.4 Validación cruzada

Cada uno de los modelos de predicción o estimadores que se han evaluado en este TFM está compuesto por un conjunto de hiper-parámetros que definen un modelo, por ejemplo la profundidad máxima en los árboles de decisión. Como ya se ha explicado previamente, cuando se evalúan diferentes conjuntos de hiper-parámetros de un modelo para determinar su efectividad, para evitar el sobreajuste se recomienda hacerlo dividiendo el conjunto de datos original en dos, uno para el entrenamiento y el segundo para probarlo. Aun así, la efectividad del modelo puede depender en gran medida de cómo se haya hecho esa división entre datos de entrenamiento y de prueba y, por lo tanto, puede ser significativamente diferente en función de cómo ésta se haga manteniendo así el riesgo de sobreajuste.

Este problema podría resolverse haciendo una división adicional del conjunto de datos, llamada conjunto de validación, de manera que el modelo se entrenaría con el conjunto de entrenamiento, posteriormente se realizaría su evaluación con el conjunto de validación y, si en éste último resulta tener éxito, hacer una evaluación final sobre el conjunto de prueba.

Sin embargo, dividir el conjunto de datos disponible en tres subconjuntos reduce drásticamente, especialmente en este caso, el número de muestras que se usan para el entrenamiento del modelo. Debido a estas carencias, aparecen procedimientos de entrenamiento como la validación cruzada o cross-validation. Dentro de este procedimiento se mantiene el conjunto de pruebas para la evaluación final pero el conjunto de validación dejaría de existir. En su planteamiento más sencillo, como se puede apreciar en la Figura 18, consiste en dividir el conjunto de datos de entrenamiento en k conjuntos más pequeños o folds, entrenar el modelo con $k-1$ folds y validarlo con el fold restante.

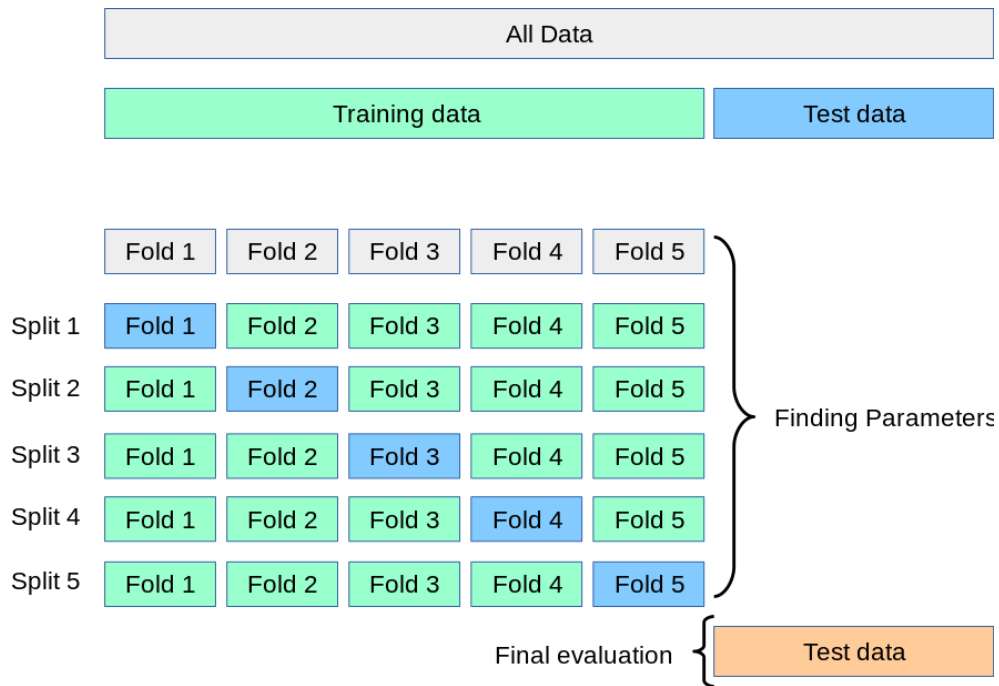


Figura 18 Esquema de funcionamiento del proceso de validación cruzada. Fuente: https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation

Este proceso se repite de forma iterativa y el rendimiento final del modelo, para un conjunto de hiper-parámetros dado, será la media del rendimiento en cada una de las iteraciones. Este proceso puede ser computacionalmente caro, pero favorece la evaluación de los modelos de predicción en aquellos casos en los que el número de muestras no es elevado, como es el de este TFM.

4 RESULTADOS ALGORITMOS DE PREDICCIÓN

El lenguaje de la experiencia es más autorizado que cualquier otro razonamiento: los hechos pueden destruir nuestros raciocinios, pero no viceversa

Alessandro Volta

Una vez descritos el grupo de algoritmos que se han evaluado, se van a presentar los resultados relativos a las pruebas iniciales que se hicieron con cada uno de ellos y el dataset completo, es decir, con la información de los 40 transformadores de los que se disponen datos.

Para estas pruebas se han utilizado, como se ha introducido y descrito previamente, los datos topológicos, meteorológicos, días de semana, fin de semana, vacaciones y festivos (codificados con one-hot encoding) y series temporales de potencia activa cinco-minutal de 39 transformadores para entrenar los modelos predictivos. Con el modelo entrenado y el input de datos del transformador restante, de carácter principalmente residencial, se ha predicho el histórico de potencia activa completo de 2019.

A continuación se presenta, para cada uno de los modelos, una evaluación de los resultados obtenidos comparando la predicción hecha por el algoritmo con las medidas reales que se disponen del transformador.

Esta comparativa se ha hecho utilizando cuatro gráficas diferentes (véase Figura 19 como ejemplo):

- **Histograma del error de predicción** (arriba-izquierda): considerando el error como la diferencia, en términos de potencia activa en kW, entre el dato predicho y el dato real para todo 2019. La distribución esperada en estos histogramas será de tipo normal centrada en cero y desviación estándar pequeña.
- **Gráfico de dispersión potencia activa real vs potencia activa predicha** (arriba-derecha): originalmente se usa para representar la correlación entre dos variables, en este caso se utilizará para ver cómo de alejado está el modelo de la predicción perfecta (correlación perfecta). Esta predicción perfecta vendrá representada por una recta diagonal en color rojo donde $y_{predicha} = y_{real}$.
- **Serie temporal comparativa** (abajo-izquierda): representación de la serie temporal real de la potencia activa del transformador (azul) y la potencia activa predicha (rojo) para todo el año.
- **Serie temporal comparativa, zoom** (abajo-derecha): zoom sobre la gráfica anterior que abarca un mes.

Por último, en los siguientes apartados, además de gráficamente, se calcularán dos métricas típicamente utilizadas para estimar el rendimiento y evaluar el ajuste de un modelo:

- Error Absoluto Medio, más conocido como MAE (Mean Average Error): este error se calcula como un promedio de diferencias absolutas entre los valores objetivo y las predicciones. En esta métrica todas las diferencias individuales se pondrán por igual en el promedio.

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^N |y_i - \hat{y}_i| \quad (4-1)$$

Siendo N el número de muestras, y_i el valor real e \hat{y}_i el valor predicho.

- Error Media Cuadrático, más conocido como MSE (Mean Squared Error): es la métrica más común para la evaluación de las regresiones y mide el error cuadrado promedio de las predicciones hechas por un modelo. Esta métrica es útil para modelos que deban ajustarse correctamente a predecir valores extremos pero, a su vez, tiene un comportamiento particularmente problemático si se utilizan datos ruidosos.

$$MSE = \frac{1}{N} \cdot \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (4-2)$$

Siendo N el número de muestras, y_i el valor real e \hat{y}_i el valor predicho.

4.1 Resultados iniciales

En los siguientes apartados se presentan los resultados de la prueba inicial. Éstos consisten en la evaluación de los modelos descritos en el apartado 3 tras haber sido entrenados con el conjunto de datos completo. Tras este el entrenamiento, se han considerado conocidos los siguientes datos del transformador objetivo:

- Fecha objeto de predicción
- Potencia nominal del transformador
- Número de LBTs.
- Número de PCRs.
- Número de clientes.
- Potencia contratada total de clientes.
- Potencia contratada media de clientes.
- Potencia contratada máxima de clientes.
- Temperatura ambiente de la fecha y hora objeto de predicción.
- Latitud.
- Longitud.
- Día entre semana o fin de semana (one hot encoded).
- Periodo vacacional (one hot encoding).
- Día festivo (one hot encoding).

Al final de este apartado, en el 4.1.5, se han incluido algunas conclusiones de los resultados de prueba.

4.1.1 Decision Tree Regressor

En la Figura 19 se presentan los resultados de la predicción de la potencia activa del transformador objetivo con todo el set de datos disponible y el modelo de predicción regresivo basado en árboles de decisión. Como se puede apreciar a simple vista, los resultados son ciertamente desfavorables y el entrenamiento realizado sobre este modelo genera unas malas predicciones.

Si se presta atención a las dos imágenes inferiores, relativas a la comparativa entre la serie temporal real y la serie temporal predicha, se pueden observar distintos fenómenos.

Por un lado, las dos series temporales se encuentran muy desfasadas una de la otra, puesto que el modelo entrenado genera unas predicciones (en rojo) cuya serie oscila en torno a una media mucho mayor que la real (en azul).

Por otro lado, en el proceso de entrenamiento el modelo ha aprendido ligeramente la tendencia del patrón de potencia activa estacional de la serie (imagen inferior izquierda) y mínimamente la tendencia del patrón de consumo semanal de la serie (imagen inferior derecha). En ambos casos, la predicción se mantiene más estable con una media prácticamente constante, especialmente en los patrones semanales, por lo que el algoritmo no predice correctamente valores de potencia activa elevados entre semana y su reducción en los fines de semana, típico de un transformador de carácter residencial.

Este comportamiento se hace evidente si se analizan los gráficos que contienen el histograma del error y la dispersión. En el primero (imagen superior izquierda) se puede apreciar que la distribución no se parece a una normal, no está centrada en cero y la mayor parte de los errores se encuentran por encima de los 50 kW (en valor absoluto). En el segundo, el gráfico de dispersión (imagen superior derecha) se ve claramente que los puntos están situados muy por encima de la línea diagonal de la figura, lo cual representa el que modelo está sobrestimando el valor predicho.

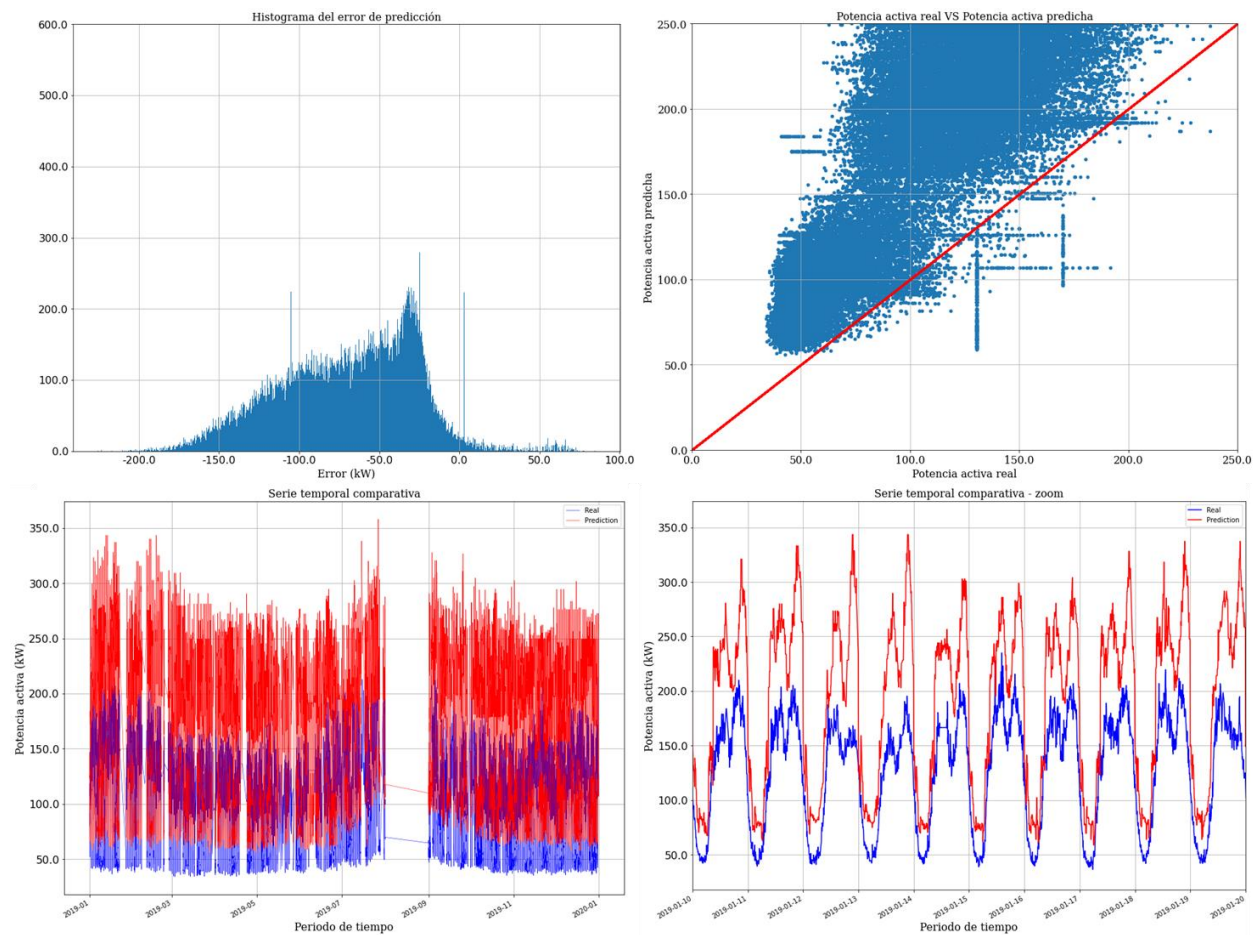


Figura 19 Resultados prueba inicial con algoritmo DTR

Además del análisis anterior, en la Tabla 4 se muestran las métricas calculadas para esta prueba.

Tabla 4 Métricas para los resultados de la prueba inicial con el modelo DTR

MAE (kW)	MSE (kW ²)
70.68	6628.94

4.1.2 Gradient Boosting Regressor

De manera análoga al caso anterior, en la Figura 20 se presentan los resultados de predicción de potencia activa para el modelo de predicción regresivo de Gradient Boosting. Para este modelo, a pesar de tener un comportamiento muy distinto al modelo anterior, las predicciones efectuadas tras el entrenamiento del GBR con todo el set de datos original no ofrece unos resultados que puedan considerarse válidos.

En este caso, analizando las comparativas de las series temporales real y predicha (imágenes inferiores) se puede comprobar que, si bien el modelo ha aprendido que la media con consumo de potencia activa es más elevada que en el anterior, la predicción no se ajusta en nada al comportamiento real del transformador objetivo y tiene unos valores constantes y escalonados. Esta predicción refleja un claro síntoma de sobreajuste u overfitting. Dicho efecto ocurre principalmente cuando, durante el proceso de aprendizaje automático, el modelo aprende demasiado de los datos de entrenamiento, o de datos anómalos, haciendo que el algoritmo tenga problemas para generalizar el comportamiento deseado.

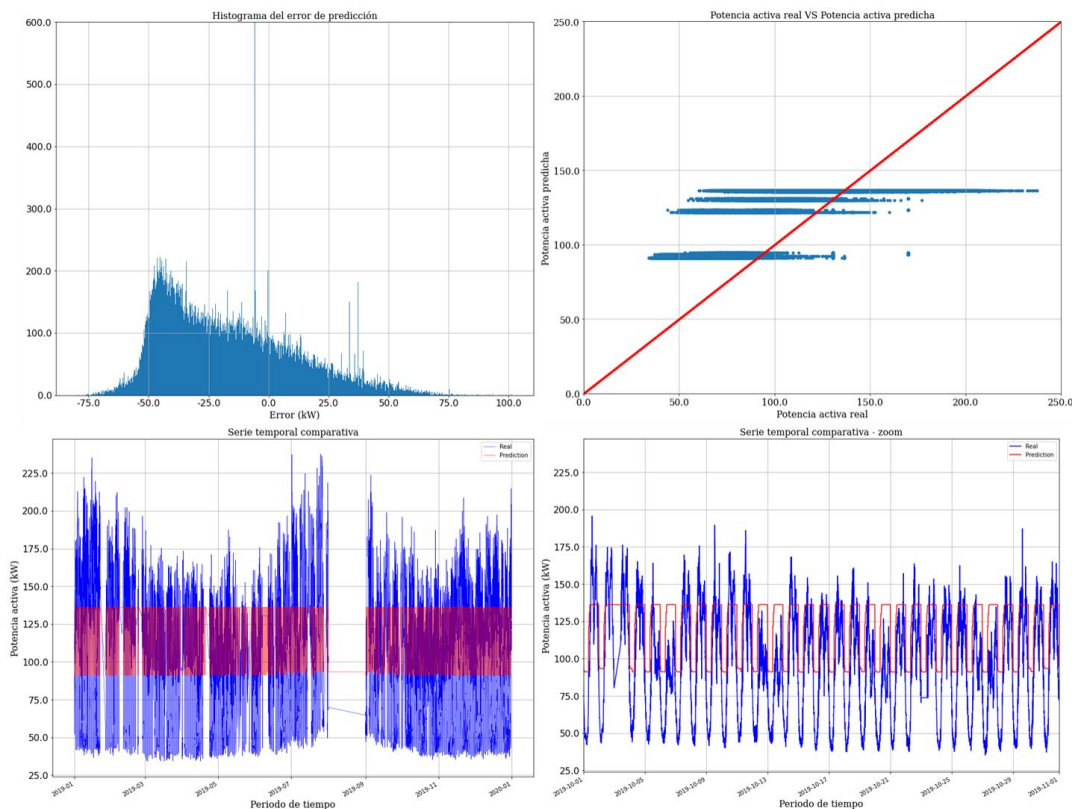


Figura 20 Resultados prueba inicial con algoritmo GBR

Este efecto se puede apreciar claramente en el gráfico de dispersión (imagen superior derecha) donde se puede apreciar como todas las predicciones se ajustan dentro de 4 escalones principales entre los 80 kW y los 140 kW. Todo esto provoca que, evidentemente, el modelo se incapaz de predecir correctamente los patrones estacionales y semanales, y mucho menos los máximos y mínimos de consumo de potencia activa.

Finalmente, en la Tabla 5 se muestran las métricas calculadas para esta prueba. Si se comparan estos valores con los anteriores se puede observar que los mejoran significativamente, aunque esta apreciación sea errónea pues

las imágenes anteriores evidencian el mal comportamiento del modelo Gradient Boosting Regressor. Comparando estas métricas con las anteriores, se pone de manifiesto que éstas, por si solas, no son representativas del buen comportamiento de un modelo de predicción.

Tabla 5 Métricas para los resultados de la prueba inicial con el modelo GBR

MAE (kW)	MSE (kW ²)
27.85	1059.71

4.1.3 XGBoost

En la Figura 21 se presentan los resultados de la predicción de la potencia activa transformador objetivo con todo el set de datos disponible y el modelo de predicción Extreme Gradient Boosting. Como en los dos casos anteriores, las predicciones arrojadas por este modelo no pueden considerarse válidas.

De forma análoga al primer modelo, las predicciones oscilan sobre una media que se encuentra muy desfasada de los valores reales, en este caso, muy por debajo de los mismos. Este comportamiento se hace evidente en el gráfico de dispersión, donde se puede apreciar como el modelo subestima para todo el periodo el valor de potencia activa.

Además, a pesar de reproducir la tendencia estacional del patrón de potencia activa, como se puede apreciar en la imagen inferior izquierda, al igual que con el primer modelo, la predicción reproduce mínimamente el patrón semanal, como se ve en la imagen inferior derecha. En este último, la predicción tampoco repite correctamente los picos y bajadas del patrón semanal, haciéndolo de forma asíncrona.

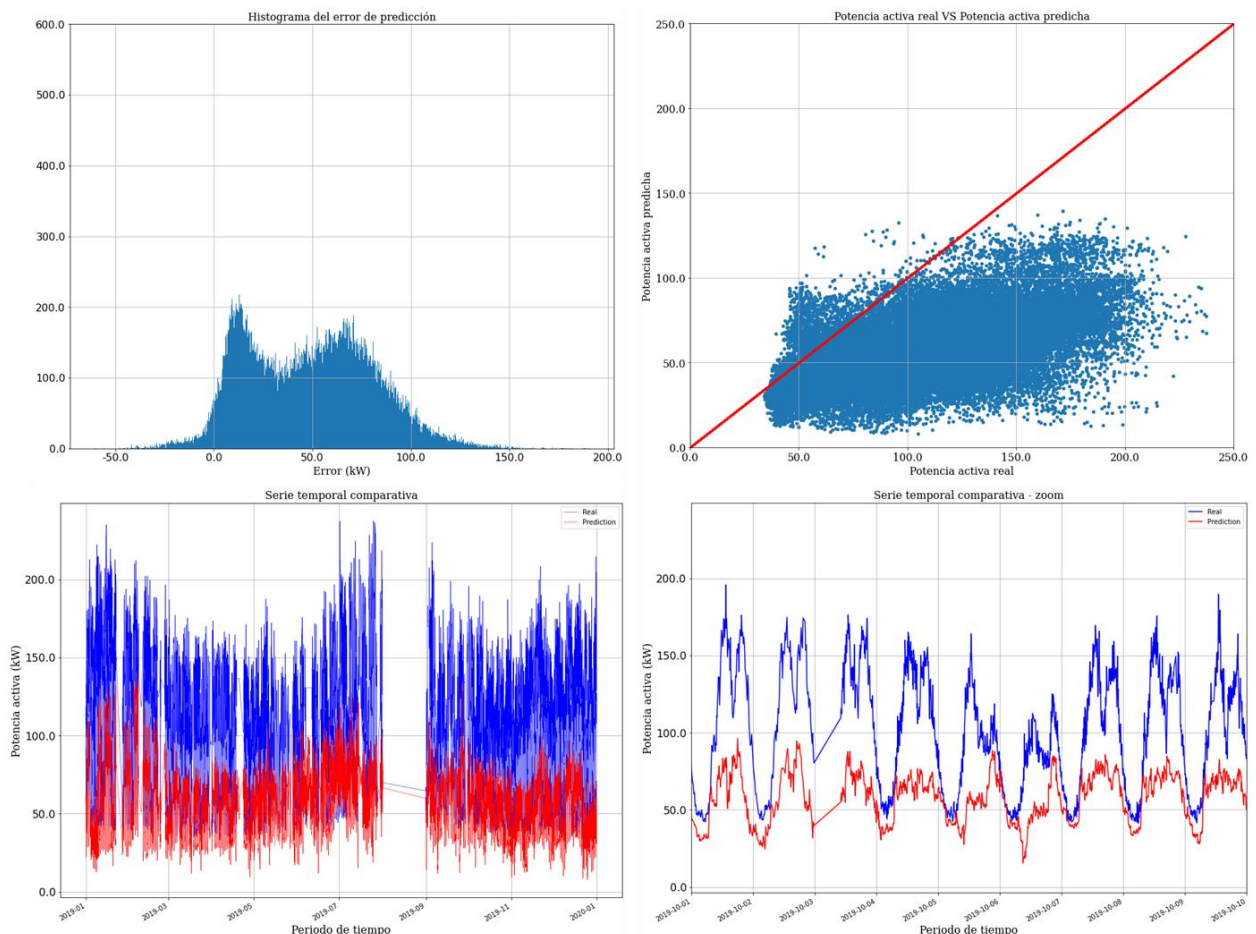


Figura 21 Resultados prueba inicial con algoritmo XGBoost

Finalmente, en la Tabla 6 se muestran las métricas calculadas para esta prueba. Comparadas con las métricas del modelo DTR, éstas mejoran sus resultados pero siguen sin poder reconocerse como unos buenos resultados.

Tabla 6 Métricas para los resultados de la prueba inicial con el modelo XGBoost

MAE (kW)	MSE (kW ²)
49.30	3413.57

4.1.4 Redes Neuronales

4.1.4.1 Red Neuronal 1

Los resultados relativos a la Red Neuronal 1 entrenada en las pruebas iniciales presentan, como se puede ver en Figura 22, un comportamiento diferente al de los modelos anteriores. A primera vista, analizando la comparativa de la serie temporal completa, se puede apreciar como la predicción de potencias mínimas se ajusta a un valor prácticamente constante mientras que, por el contrario, los picos de potencia predichos siguen la tendencia estacional de la serie. Esto es mucho más evidente sobre el zoom de la serie temporal donde se observan claramente este límite mínimo y como el patrón semanal predicho no encaja con el real.

Analizando los gráficos que contienen el histograma del error y la gráfica de dispersión, se confirma que el modelo no genera unos buenos resultados. Especialmente en el primero (imagen superior izquierda), donde se puede apreciar que la distribución no se parece a una normal, no está centrada en cero y la mayor parte de los errores se distribuyen entre los -30 kW y los 40 kW.

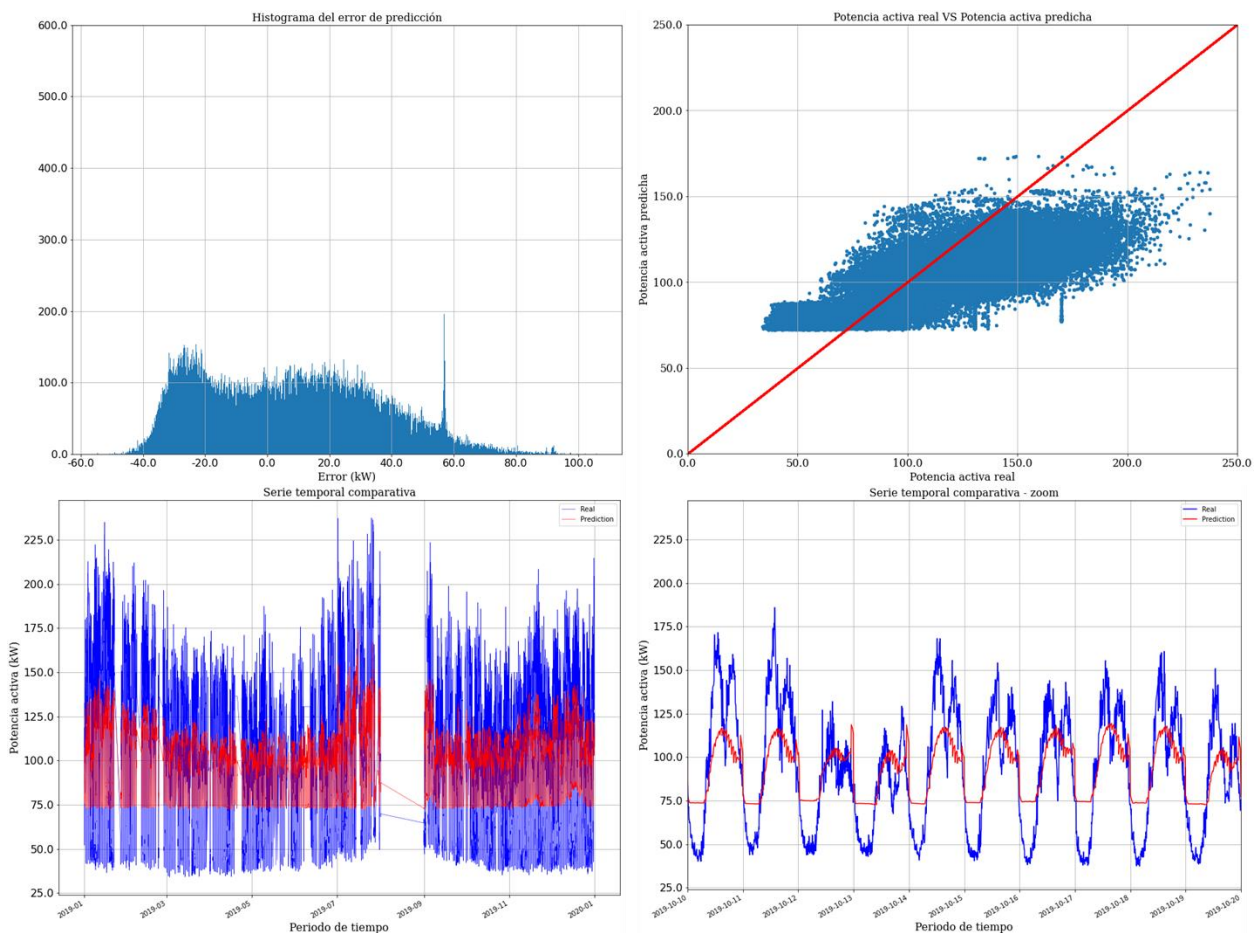


Figura 22 Resultados prueba inicial Redes Neuronales Profundas 1

En la Tabla 7 se muestran las métricas calculadas para esta prueba, cuyos valores son similares aunque ligeramente mejores que los del modelo DTR.

Tabla 7 Métricas para los resultados de la prueba inicial con el modelo RNP 1

MAE (kW)	MSE (kW ²)
24.238	853.65

4.1.4.2 Red Neuronal 2

Los resultados relativos a la Red Neuronal 2, el último modelo evaluado con los datos de esta prueba inicial, como se puede ver en Figura 23, muestran un comportamiento mejor que los modelos anteriores aunque presenta también ciertas deficiencias.

Analizando la comparativa de series temporales completa se observa que, aunque mantiene los valores de la predicción de potencia mínima limitados, la nueva capa oculta de que presenta la RNP 2 respecto a la anterior mejora las predicciones de los picos de potencia manteniendo la tendencia estacional de la serie. Igualmente, sobre el zoom de la serie temporal, se puede apreciar aunque no predice correctamente los patrones diarios sí que lo hace con la tendencia de los patrones semanales.

Reparando en el histograma de error y el gráfico de dispersión se puede ver que, en el primero, la distribución de errores es irregular, debido precisamente a la mala predicción de las potencias mínimas, y está desplazada a la izquierda.

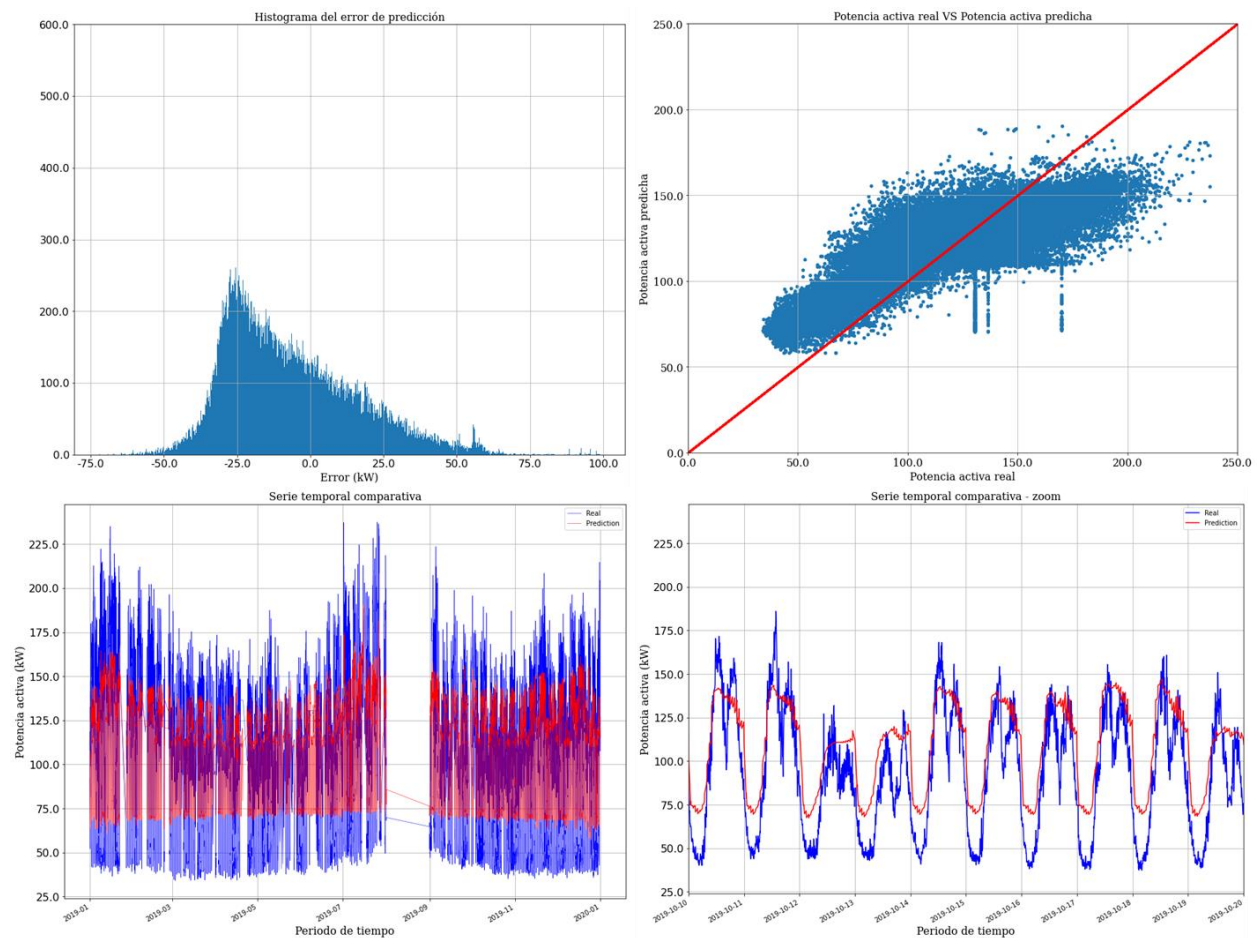


Figura 23 Resultados prueba inicial Redes Neuronales Profundas 2

En la Tabla 8 se muestran las métricas calculadas para esta prueba, cuyos valores reflejan como este modelo es el que presenta mejor comportamiento en la prueba inicial.

Tabla 8 Métricas para los resultados de la prueba inicial con el modelo RNP 2

MAE (kW)	MSE (kW ²)
19.90	558.95

4.1.5 Conclusiones

En los puntos anteriores se ha comprobado que los resultados iniciales, en los que se han evaluado cinco modelos predictivos con el set de datos original, están alejados del rendimiento esperado para un algoritmo de predicción puesto en producción. La razón por la que se están obteniendo dichos resultados, como se pondrá de manifiesto en el próximo apartado, puede ser que el set de datos original no se está utilizando correctamente en el Caso de Uso aquí planteado.

Los datos son la base de las técnicas de Machine Learning e Inteligencia Artificial, por lo que disponer un buen set de datos es clave para el correcto desempeño de las mismas. Y en Aprendizaje Supervisado (Supervised Machine Learning) un buen set de datos se define, en otras, por dos características principales: la **falta de sesgo** y el **balanceo de datos**.

El primero, hace referencia a la propiedad del sesgo muestral, también llamado efecto de selección o error muestral, que es una distorsión que se introduce debido a la forma en la que se selecciona la muestra de datos. Se refiere a la distorsión de un análisis estadístico, debido al método de recolección de muestras. Si el sesgo muestral no es tomado en cuenta, entonces algunas conclusiones propuestas pueden ser erróneas.

El segundo, el balanceo de datos o *resampling*, es una técnica derivada del anterior que consiste en modificar la distribución original de la muestra de datos ya sea eliminando casos o instancias de la clase mayoritaria, *undersampling*, o replicando y creando nuevas instancias de la clase minoritaria, *oversampling*.

Teniendo esto en cuenta, el uso de datos procedentes de Supervisión instalada en Centros de Transformación para predecir consumos en transformadores no observables es un Caso de Uso de naturaleza sesgada. La razón es que, analizándolo desde un punto de vista estadístico, la muestra total completa representativa de todo el espectro de transformadores es realmente amplia puesto que la caracterización del consumo de un transformadores de potencia de las redes de distribución depende de muchos atributos: potencia nominal, número de clientes suministrados, tipo de clientes suministrados, localización, clima, etc.

Esta propiedad, la falta de sesgo, que tan importante es en el Machine Learning y la Inteligencia Artificial, puede llegar a estar presente en las Bases de Datos de las distribuidoras pero, en el caso de los datos disponibles en este Trabajo Fin de Máster, la muestra está significativamente sesgada. Esto se debe principalmente a que el Smartcity Málaga Living Lab, el área que acoge el alcance los datos, es una zona reducida de una ciudad grande en la que la sensorización de los transformadores no se escogió para armonizar una muestra de datos sino para aglutinar la mayoría de los transformadores que la ocupan. Siendo así, es evidente que los 40 transformadores, cuyos datos conforman la muestra disponible, no suponen una muestra representativa de todo el espectro de transformadores de las redes de distribución.

Por todo esto, se ha decidido aplicar técnicas de *resampling* con las que balancear la muestra de datos con la que se están entrenando los modelos predictivos y, así, mejorar los resultados que arrojan. En la Figura 24 se presentan las principales técnicas posibles para ello agrupadas en cuatro tipos:

- Técnicas de preprocesamiento.
- Enfoques algorítmicos.
- Aprendizaje sensible al coste
- Aprendizaje ensamblado

Para el presente Trabajo Fin de Máster, se ha seguido el enfoque algorítmico, concretamente la metodología de agrupación o *clustering*, una técnica muy extendida dentro del Aprendizaje No Supervisado.

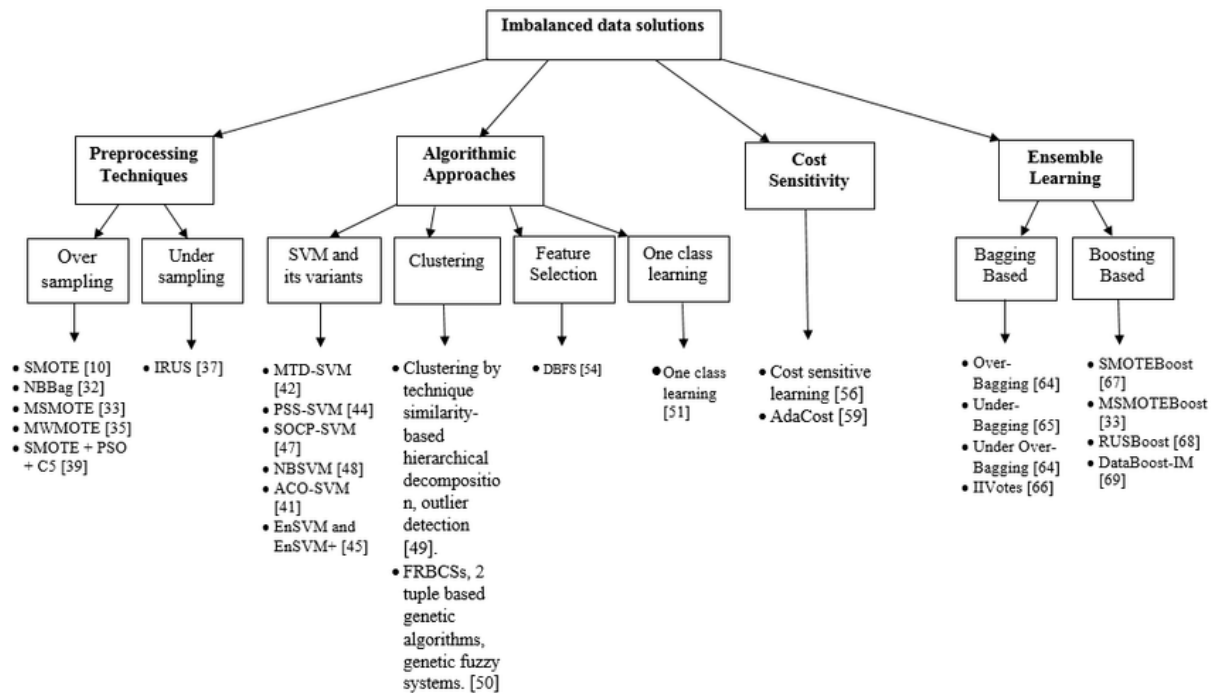


Figura 24 Esquema de las diferentes técnicas para balancear set de datos. Fuente: https://www.researchgate.net/figure/Proposed-texonomy-for-the-review-of-imbalanced-class-problems-in-data-mining_fig2_331522710

En el siguiente apartado se detallará la técnica aplicada, las agrupaciones obtenidas y los resultados conseguidos tras su aplicación.

4.2 Resultados tras clustering

Como se ha introducido en el apartado 3.5.3, las técnicas basadas en clustering, desde un punto de vista más general que aplicado a resampling, consisten en agrupar datos en grupos con características similares y son ampliamente utilizadas en el mundo de la Ciencia de Datos para determinar patrones climáticos, agrupar artículos por temas o para segmentación de clientes. Clusterizar un conjunto de datos tiene por objetivo formar grupos cerrados y homogéneos a partir de un conjunto de elementos que tienen diferentes características o propiedades, pero comparten ciertas similitudes.

Este es el mismo objetivo que se ha perseguido en su aplicación con el set de datos original del Smartcity Málaga Living Lab: crear grupos bien definidos de transformadores, identificar aquellos que pertenecen al mismo grupo que el transformador residencial objetivo de predicción y utilizar únicamente los datos de dicho grupo para reducir el sesgo en las muestras de datos y mejorar las predicciones. Para ello, los datos que se han utilizado para clusterizar los transformadores corresponden con la información topológica de la Tabla 2.

En este caso se ha decidido considerar un número pequeño de clusters, $k = 4$, para evitar que se generen grupos con un número muy bajo de transformadores asociados. En la siguiente tabla se muestra la asignación de cada transformador al cluster 1, 2, 3 o 4 al que pertenece.

Tabla 9 Agrupación de transformadores por clusters

Transformador	Cluster	Transformador	Cluster	Transformador	Cluster	Transformador	Cluster
Transformador 1	2	Transformador 11	2	Transformador 21	3	Transformador 21	1
Transformador 2	2	Transformador 12	2	Transformador 22	3	Transformador 22	1
Transformador 3	2	Transformador 13	3	Transformador 23	3	Transformador 23	1
Transformador 4	2	Transformador 14	3	Transformador 24	1	Transformador 24	1
Transformador 5	2	Transformador 15	3	Transformador 25	1	Transformador 25	4
Transformador 6	2	Transformador 16	3	Transformador 26	1	Transformador 26	4
Transformador 7	2	Transformador 17	3	Transformador 27	1	Transformador 27	4
Transformador 8	2	Transformador 18	3	Transformador 28	1	Transformador 28	4
Transformador 9	2	Transformador 19	3	Transformador 29	1	Transformador 29	4
Transformador 10	2	Transformador 20	3	Transformador 30	1	Transformador 30	4

Esta agrupación se puede visualizar en los siguientes gráficos de dispersión, Figura 25, en las que se han representado diferentes comparativas entre dos atributos del set de datos utilizados para el *clustering* y se ha diferenciado por colores cada grupo al que pertenece. Siendo el color morado el cluster 1, el color verde el cluster 2, el color azul el cluster 3 y el color amarillo el cluster 4. Estas dos imágenes, los gráficos que cada una contiene, ayudan a identificar de forma más clara cómo se han organizado estos clusters.

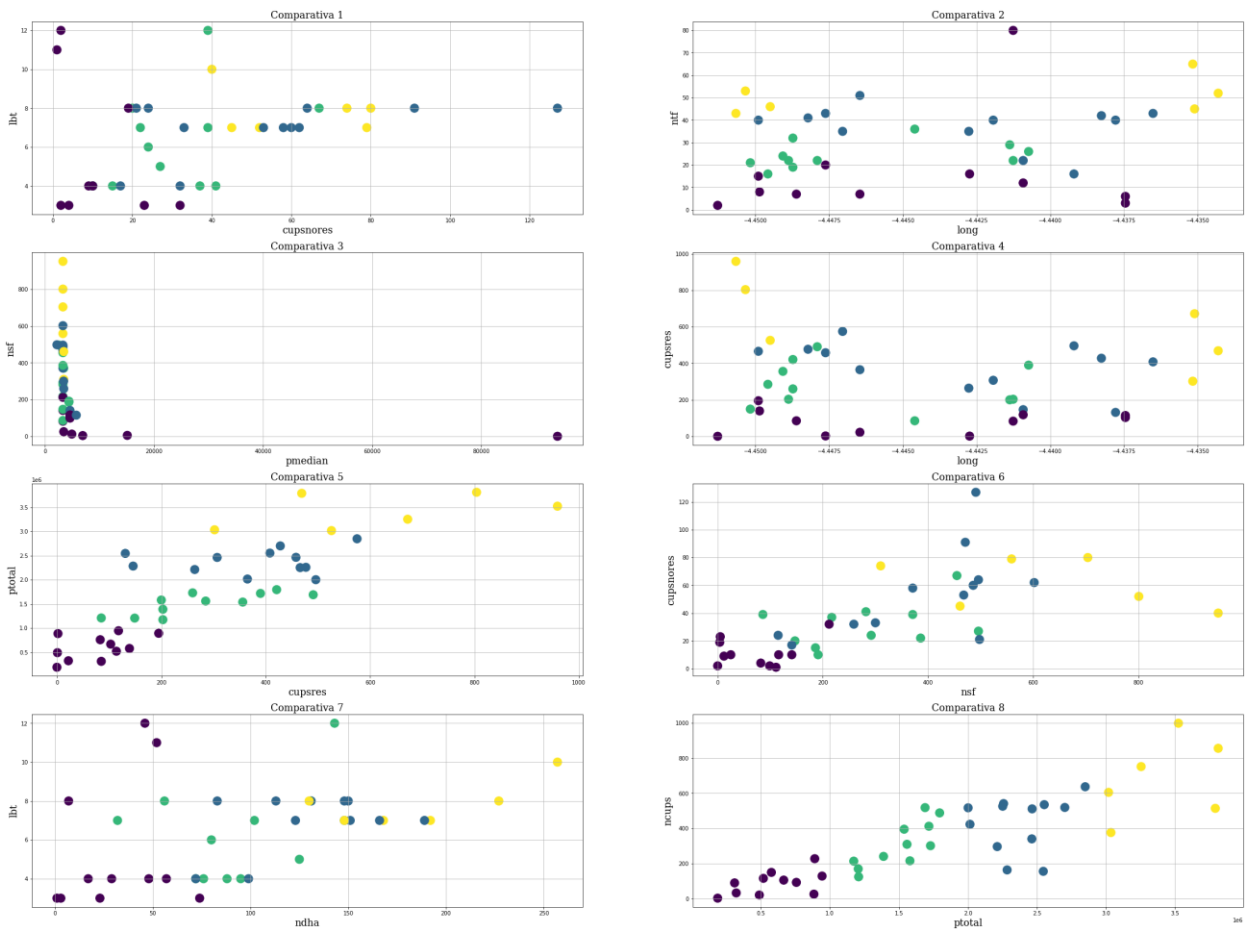


Figura 25 Gráficos de dispersión representativos de los clústers obtenidos – Ejemplo 1.

Un ejemplo de esta distribución por clusters puede interpretarse con la Comparativa 5 de la Figura 25, donde se ha representado la potencia contratada total de los transformadores respecto al número total de CUPS residenciales que suministra. En ella se observa claramente como los transformadores pertenecientes al cluster 1 son aquellos con una menor potencia contratada total y mayor concentración en términos de clientes residenciales, y que el resto de cluster van cada uno teniendo una potencia contratada total mayor y presentan más dispersión en el número de suministros residenciales.

Una vez definidos los clusters y teniendo en cuenta que el transformador residencial objetivo es el Transformador 23, se va a repetir el entrenamiento de los algoritmos considerados en este TFM pero, ahora, utilizando únicamente los datos de los transformadores que pertenecen al cluster 3. Con estos algoritmos se ha repetido la evaluación de las predicciones realizadas y los resultados de este análisis se muestran en los siguientes apartados.

4.2.1 Decision Tree Regressor

En la Figura 26 se muestra, con las mismas 4 gráficas que en las pruebas iniciales, los resultados de las predicciones del algoritmo DTR con los datos del cluster 3.

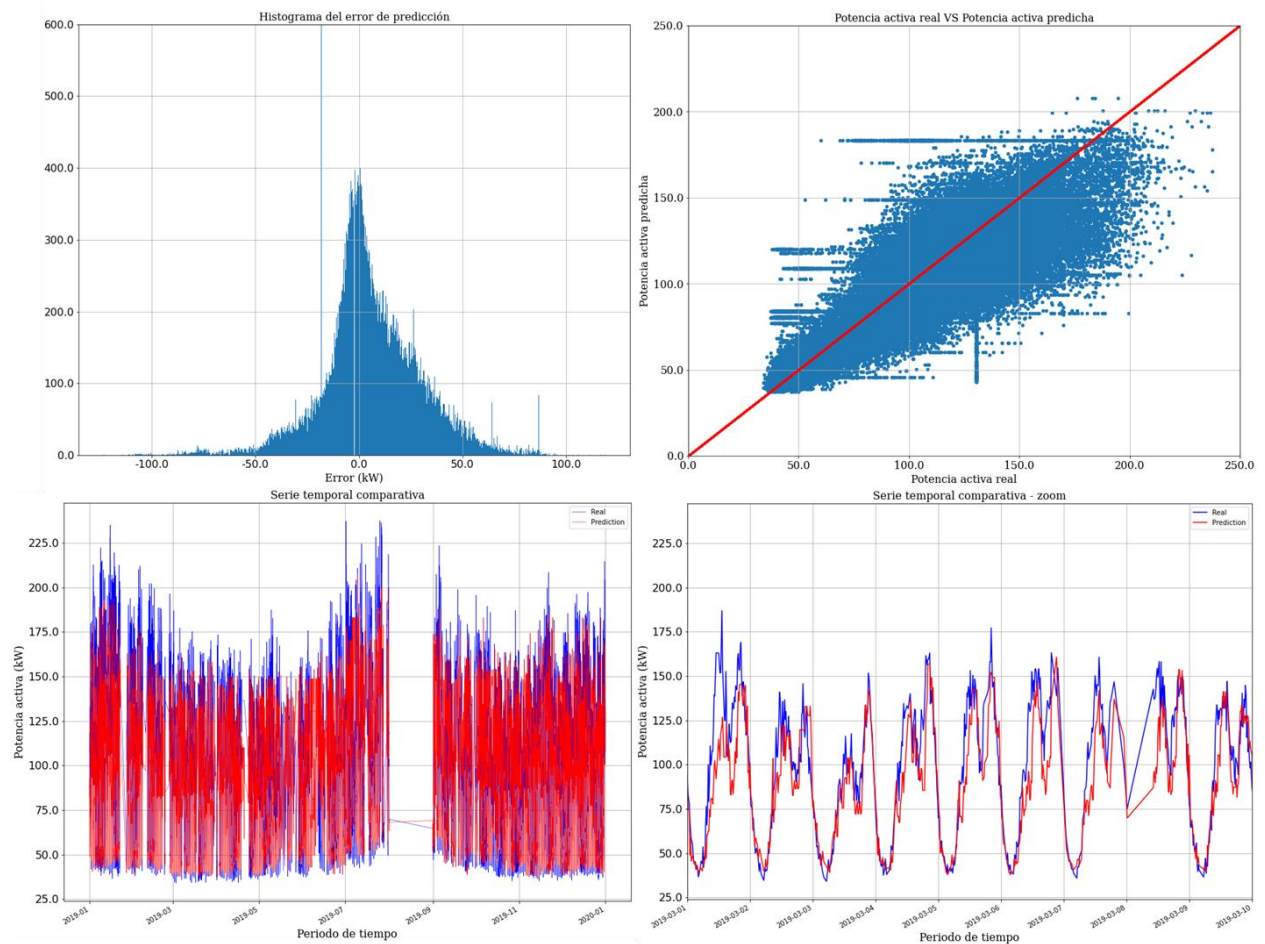


Figura 26 Resultados algoritmo DTR con datos del cluster 3

Analizando, en primer lugar, la comparativa de las series temporales real y predicha de la potencia activa en las dos gráficas inferiores, se puede apreciar como las dos series están prácticamente superpuestas para todo el periodo predicho.

En la imagen de la izquierda se observan como la serie temporal predicha cumple con la tendencia de la serie real, prediciendo correctamente el patrón estacional de consumo de potencia activa del transformador.

En el caso de la imagen de la derecha, se aprecia cómo se predice correctamente el patrón de consumo diario, con un error de predicción ligeramente significativo en los picos de consumo pero se ajusta correctamente en las

horas en las que el consumo es menor. Además, en ambos casos se ve, comparándolo con los resultados de la Figura 19, que ya no están las curvas real y predicha desfasadas.

Por otro lado, poniendo el foco en la distribución de los errores representados en la imagen superior izquierda, se ha conseguido una distribución más parecida a una distribución normal esperada: más simétrica y centrada en cero.

Además, con el gráfico de dispersión se confirma la mejora del comportamiento de la predicción cuyos valores se distribuyen en torno a la línea diagonal. En esta imagen se aprecia también que, para valores de potencia elevados, el algoritmo tiende a subestimar la potencia activa por lo que, en estos niveles, hay más puntos por debajo de la línea diagonal.

Por último, en la Tabla 10 se muestran las métricas calculadas para evaluar el rendimiento de los resultados de esta prueba. Se puede comprobar cómo se han mejorado significativamente los resultados del modelo DTR de la prueba inicial.

Tabla 10 Métricas para los resultados del algoritmo DTR con datos del cluster 3

MAE (kW)	MSE (kW ²)
17.96	616.86

4.2.2 Random Forest Regressor

Los nuevos resultados obtenidos para el algoritmo Gradient Boosting Regressor con los datos resampleados del cluster 3, como se puede apreciar en la Figura 27 y en la Tabla 11, arrojan unos resultados similares a las pruebas iniciales. Puede concluirse que este tipo de modelos no son apropiados para el caso de uso que aquí se plantea.

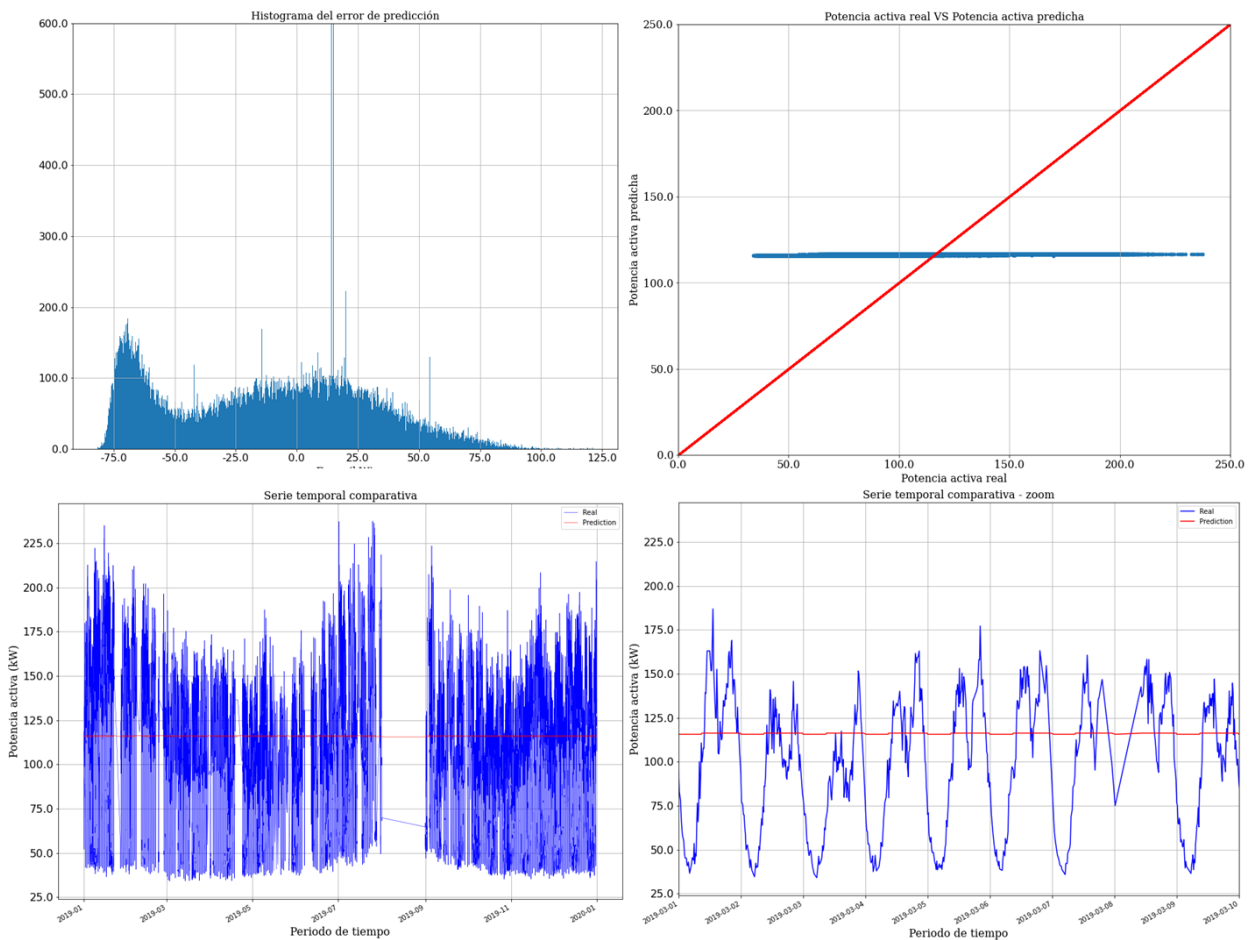


Figura 27 Resultados algoritmo GBR con datos del cluster 3

Tabla 11 Métricas para los resultados del algoritmo GBR con datos del cluster 3

MAE (kW)	MSE (kW ²)
35.57	1826.57

4.2.3 XGBoost

En la Figura 28 se han representado los resultados de la predicción sobre el Transformador 23 utilizando, para el entrenamiento del modelo XGBoost, solo los datos de transformadores pertenecientes al cluster 3.

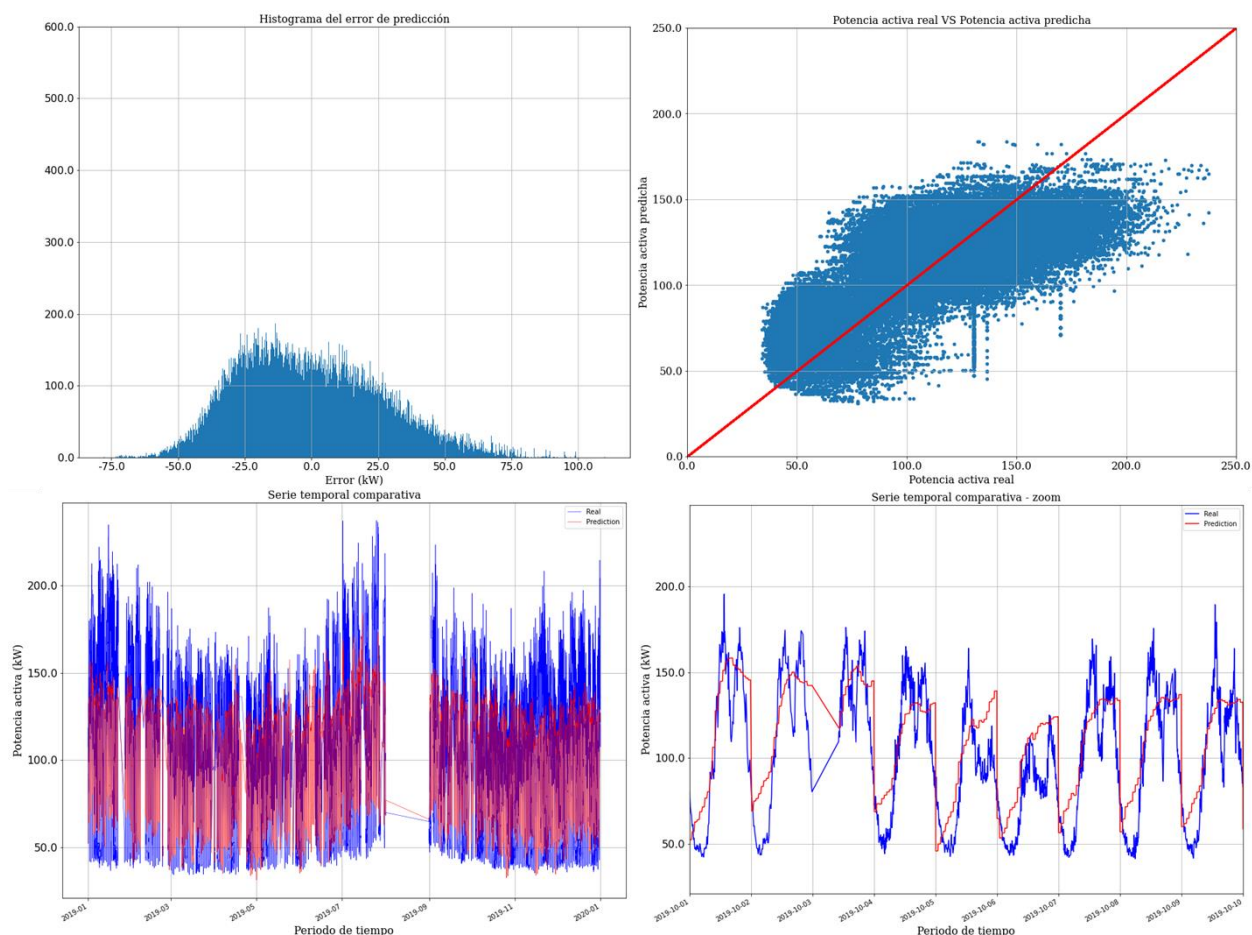


Figura 28 Resultados algoritmo XGBoost con datos del cluster 3

A simple vista se puede apreciar cómo han mejorado significativamente los resultados obtenidos con los datos del conjunto original.

Por un lado, prestando atención a las dos series temporales, se observa como es evidente que la predicción replica la tendencia estacional de la serie y el patrón semanal de la misma, aunque, a diferencia del algoritmo DTR, no deduce correctamente el patrón diario del transformador. Además se aprecia como tiene a subestimar los picos de potencia, síntoma de no predecir correctamente el patrón diario.

Por otro lado, atendiendo al histograma de errores y al gráfico de dispersión, a pesar de no conseguir una distribución parecida a la normal, se ha conseguido centrar la predicción aunque todavía está desviada hacia la

izquierda, concentrándose buena parte del valor del error entre 0 y -25 kW.

Esta mejora es evidente, también, en los valores de MAE y MSE de la Tabla 12 los que, comparados con la prueba anterior, se han reducido significativamente. Estas métricas son mejores, aunque no demasiado, que las obtenidas por el modelo DTR a pesar que las series que reflejan el patrón diario evoquen lo contrario. Esta duda se resolverá en las conclusiones de estas pruebas donde se muestran algunos indicadores globales de error en la predicción del patrón diario.

Tabla 12 Métricas para los resultados del algoritmo XGBoost con datos del cluster 3

MAE (kW)	MSE (kW ²)
15.28	411.29

4.2.4 Redes Neuronales

4.2.4.1 Red Neuronal 1

Los resultados obtenidos con este algoritmo tras su entrenamiento están representados en la Figura 29.

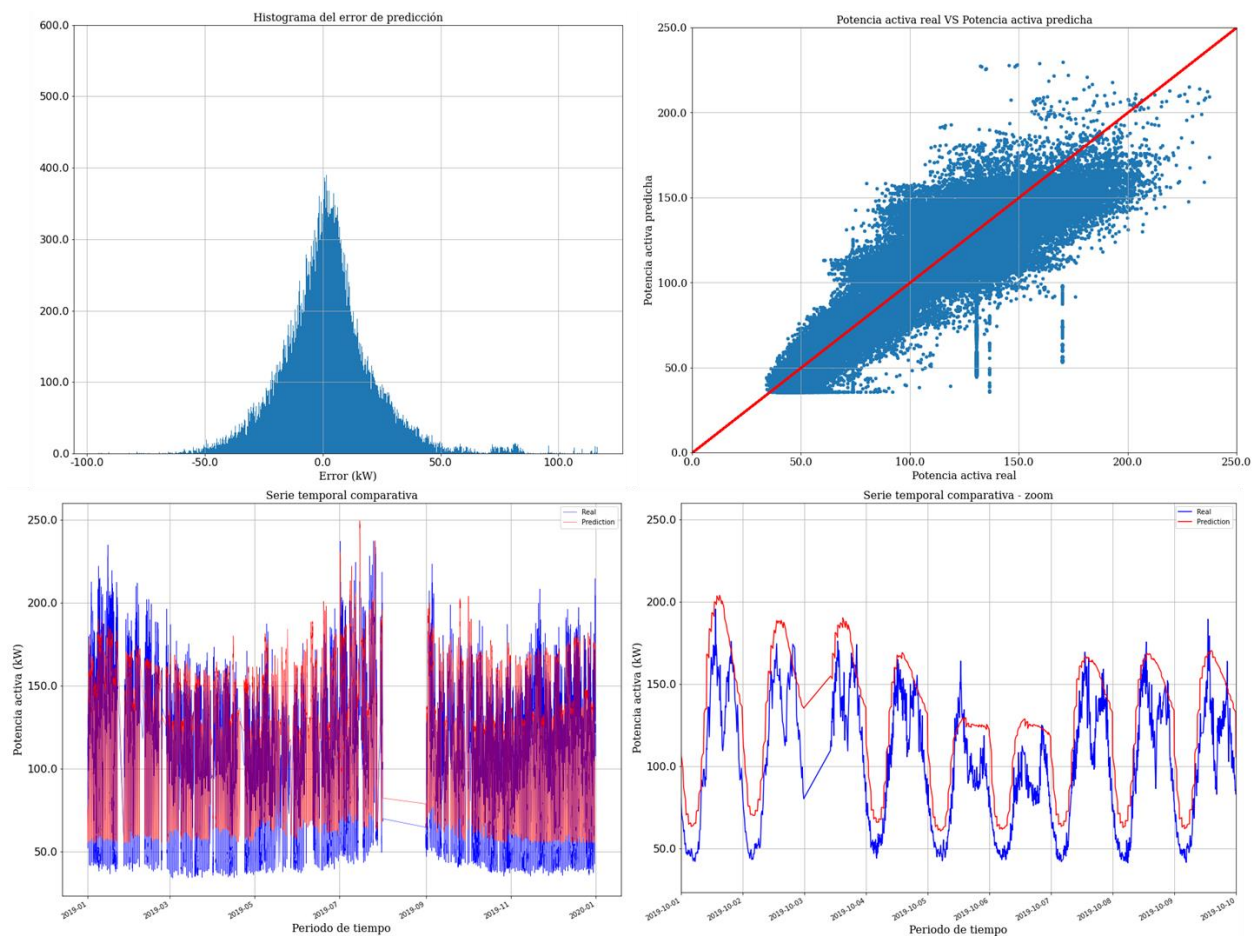


Figura 29 Resultados algoritmo RNP 1 con datos del cluster 3

De forma análoga a los análisis de resultados anteriores, poniendo el foco inicialmente en la comparativa de la serie temporal de potencia real y predicha, se puede apreciar en las dos imágenes inferiores que, al igual que en el DTR y el XGBoost, los resultados en comparación con los iniciales han mejorado significativamente.

Fijándose en la serie temporal completa, se aprecia como la amplitud de la predicción, no sólo mejorando el límite inferior de las predicciones que se conseguía en la prueba inicial sino también los picos de potencia. Esto permite mejorar también la tendencia estacional de la serie aunque sigue teniendo deficiencias.

En el caso del zoom, se aprecia cómo predice muy bien el patrón semanal especialmente la reducción de consumo de los fines de semana. No obstante, se aprecia cómo no predice correctamente los máximos y los mínimos de la serie, sobreestimando en algunos casos el primero y subestimando los segundos. Adicionalmente, a diferencia del DTR, no reproduce correctamente el patrón diario. Este análisis hace referencia al ejemplo de la imagen anterior, en el siguiente apartado se analizará el comportamiento del patrón diario de forma global.

Por otro lado, si se analiza el error en forma de histograma, se puede comprobar que el error sigue una distribución prácticamente normal, centrada en cero y muy simétrica. Este resultado es el esperado al emplear técnicas de predicción.

Además, con el gráfico de dispersión se confirma la mejora del comportamiento de la predicción cuyos valores se distribuyen en torno a la línea diagonal. En esta imagen se aprecia también que cuanto mayor sea la potencia del transformador mayor es el error.

Por último, en la Tabla 13 se muestran las métricas calculadas para evaluar el rendimiento de los resultados de esta prueba. Se puede comprobar cómo se han mejorado significativamente los resultados del modelo RNP 1 de la prueba inicial.

Tabla 13 Métricas para los resultados del algoritmo RNP 1 con datos del cluster 3

MAE (kW)	MSE (kW ²)
14.31	393.53

4.2.4.2 Red Neuronal 2

Finalmente, en la Figura 30 se presentan los resultados de esta prueba con el modelo RNP 2. Gráficamente, presentan unos resultados parecidos a los de la prueba inicial aunque con ligeras diferencias.

Por un lado, en las dos series temporales comparativas se aprecia una mejora en la replicabilidad de la tendencia estacional y semanal aunque se mantienen los errores para predecir el patrón diario. Los valores mínimos de potencia se ajustan mejor tras este entrenamiento aunque los valores máximos siguen sin predecirse correctamente. Esto última es un síntoma de no predecir correctamente el patrón diario.

Por otro lado, el histograma de errores refleja una distribución similar a la de la prueba inicial aunque, en este caso, mucho más centrada en cero. Esto se confirma revisando el gráfico de dispersión en el que se ve que, para los casos de potencia activa elevada, el modelo tiende a subestimar el valor predicho.

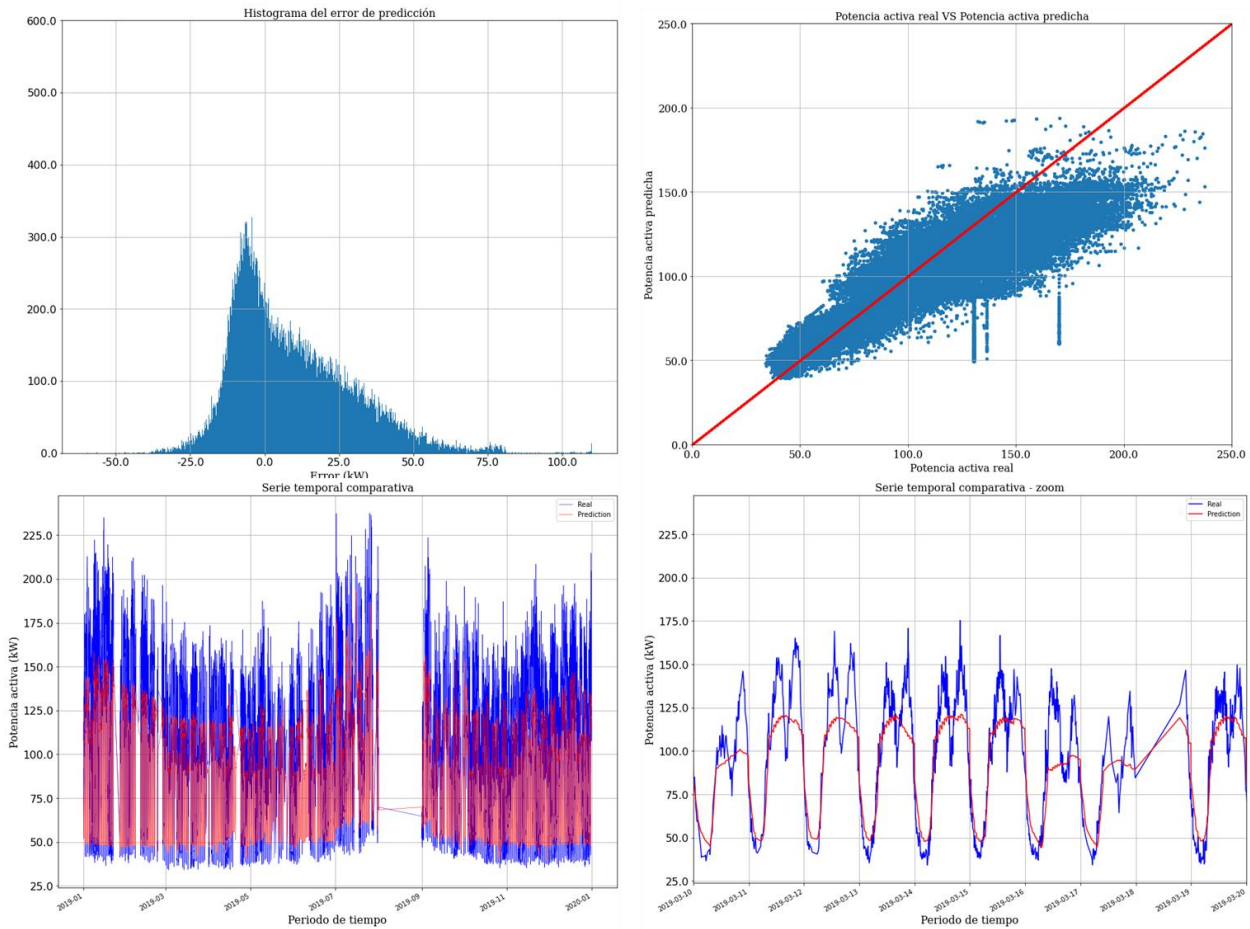


Figura 30 Resultados algoritmo RNP 2 con datos del cluster 3

Por último, en la Tabla 14 quedan recogidas las métricas de esta prueba. Comparadas con la prueba anterior, han mejorado ambos resultados.

Tabla 14 Métricas para los resultados del algoritmo RNP 2 con datos del cluster 3

MAE (kW)	MSE (kW ²)
16.23	490.96

4.2.5 Conclusiones pruebas tras clustering

A la vista de los resultados presentados en el apartado anterior, relativos a las predicciones hechas por los modelos evaluados utilizando datos preprocesados con el método de clustering, se puede comprobar que han mejorado significativamente con respecto a los iniciales.

En la Tabla 15 queda recogido un resumen del valor de las métricas de evaluación de la prueba inicial y la prueba con los datos clusterizados.

Tabla 15 Comparativa de las métricas de evaluación inicial y con datos clusterizados

Modelo	Evaluación inicial		Evaluación con datos clusterizados	
	MAE	MSE	MAE	MSE
DTR	70.68	6628.94	17.96	616.86
GBR	27.85	1059.71	35.57	1826.57
XGBoost	49.30	3413.57	15.28	411.29
RNP 1	24.238	853.65	14.31	393.53
RNP 2	19.90	558.95	16.23	490.96

A partir de la tabla anterior se pone de manifiesto la mejora significativa en todos los modelos, excepto el RNP 2 para el que, aun habiendo mejorado, la mejora no es tan grande como en el resto. Con los valores ahí recogidos se puede determinar que, en general, el modelo que mejor comportamiento tiene es el RNP 1, considerando tanto el error medio como el cuadrático.

No obstante, otro de los puntos que se han puesto de manifiesto a lo largo de la última prueba es que, además de las métricas anteriores, es interesante comparar el comportamiento que cada modelo ha tenido en la predicción de patrones, especialmente el patrón diario que es el más difícil de predecir correctamente.

Para poder analizar los resultados de esta prueba considerando este punto de vista, se han representado las gráficas recogidas en la Figura 31. En esta se han representado cinco gráficas que recogen diferentes estadísticas del error de predicción horario:

- Patrón horario del error medio
- Patrón horario del error mínimo
- Patrón horario del error máximo
- Patrón horario del error absoluto medio
- Patrón horario del error absoluto máximo.

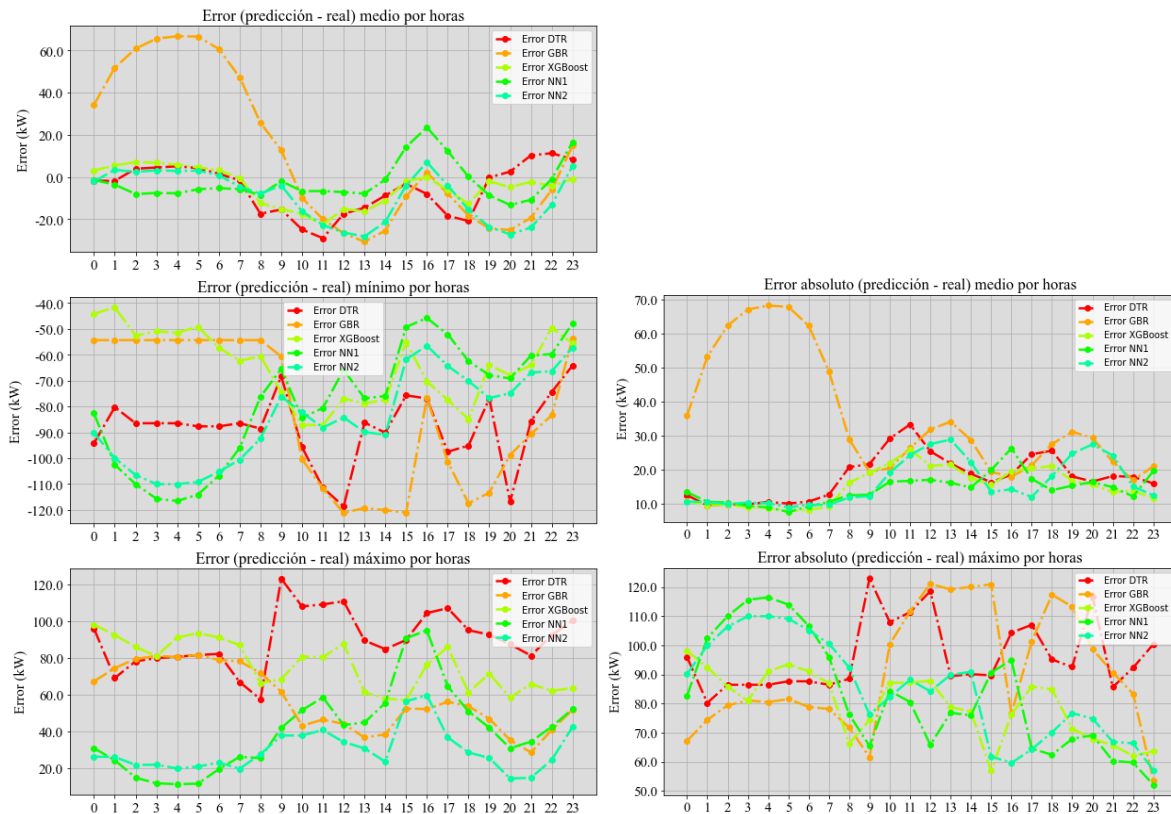


Figura 31 Evaluación de varias estadísticas del error de predicción horario

Centrándose en los errores medios real y absoluto se puede observar que el modelo RNP 1 tiene un comportamiento más estable que el resto de modelos, excepto en las horas pico de consumo en las que el error aumenta. También se puede comprobar como el modelo XGBoost, analizando el patrón de errores absoluto, tiene un comportamiento similar al DTR pero siempre un valor de error ligeramente inferior. Es interesante destacar que en las horas en las que el RNP 1 tiene unos errores elevados, estos dos últimos modelos presentan un comportamiento de predicción bueno, lo que revela la compatibilidad entre ellos.

Si se analizan, por otro lado, los valores máximos real y absoluto y mínimo, se puede verificar que no hay un modelo que se comporte significativamente peor que el resto en todas las horas, aunque en general el DTR tiene unos resultados peores que el resto de modelos (descartando GBR) y hay comportamientos dispares según se analice el error máximo o error mínimo.

En cuanto a estos últimos, el modelo XGBoost presenta menor error en la predicción de valores elevados de potencia (error mínimo) especialmente en las horas nocturnas mientras que, en el caso de errores máximos, presenta un comportamiento peor.

Por el contrario, los modelos basados en redes neuronales tienen un comportamiento opuesto al anterior, poseen el peor comportamiento en la predicción de valores elevados especialmente en las horas nocturnas mientras que en el caso de los errores máximos tienen el mejor comportamiento.

Estas diferencias entre algoritmos evidencian el hecho de que en las pruebas realizadas para este TFM, a pesar de predecir correctamente las tendencias y patrones, no se acerca a una predicción perfecta y no se ha encontrado el modelo que ajuste correctamente máximos y mínimos por igual. Esto puede deberse a diferentes causas:

- La cantidad de datos finalmente utilizada en esta prueba es escasa: los 10 transformadores que pertenecen al cluster 3, además del transformador objetivo de predicción.
- Los datos pueden tener errores asociados, especialmente los datos topológicos los cuales se han considerado como constantes para todo el periodo de datos disponibles. y que, finalmente, sólo se han usado datos de 10 transformadores del cluster.

Solventando estos problemas, haciendo un seguimiento más exhaustivo de los cambios topológicos y aumentando la cantidad de transformadores pertenecientes al mismo cluster, probablemente los resultados

mejorarían significativamente.

En definitiva y a la vista de los modelos aquí entrenados, se pone de manifiesto que, si bien tienen todos los modelos excepto el GBR tienen un buen comportamiento predictivo, la elección de un modelo u otro dependerá de la aplicación final de la predicción. Una de las opciones más ventajosas podría ser utilizar una combinación de varios de los modelos anteriores a través de un integrador.

4.3 Comparativa con otros métodos

Hasta ahora, en los dos apartados anteriores, se ha realizado una evaluación de la predicción de los modelos por medio de métricas comúnmente utilizadas y de representaciones gráficas que comparan el valor predicho con el valor real. No obstante, para ser más exhaustivos y evaluar esta metodología en el contexto de las metodologías actuales, en este apartado se han comparado también los resultados de los modelos anteriores con otras metodologías diferentes:

- a. Curvas de potencia activa normalizadas.
- b. Procedimiento publicado en BOE en términos de energía.

4.3.1 Curvas normalizadas de potencia activa normalizadas (PAN)

Este método consiste en estimar la potencia activa del transformador objetivo basándose únicamente en los históricos de potencia activa de los transformadores que pertenecen a su mismo cluster. En concreto, la metodología seguida se basa, en primer lugar, en normalizar el histórico de potencia activa de los transformadores del mismo clúster dividiendo la medida de carga real en cada instante por la potencia total contratada total de cada transformador y calcular el valor medio. Posteriormente, este valor medio de carga normalizada por potencia contratada se aplica al transformador objeto de predicción multiplicando el valor de potencia normalizada por la potencia total contratada del mismo.

La metodología anterior se ha puesto en práctica y se han calculado las mismas métricas para las estimaciones obtenidas. En la Tabla 16 está recogido el valor de las métricas junto con las obtenidas en la prueba con datos clusterizados, a modo de comparación.

Tabla 16 Comparativa de métricas entre datos clusterizados y método con potencia activa normalizada

Modelo	Evaluación con datos clusterizados	
	MAE (kW)	MSE (kW ²)
DTR	17.96	616.86
GBR	35.57	1826.57
XGBoost	15.28	411.29
RNP 1	14.31	393.53
RNP 2	16.23	490.96
PAN	38.06	1704.76

Se puede comprobar que haciendo un uso de modelos de inteligencia artificial, que tienen en cuenta no sólo la información histórica de medidas de potencia activa, sino todos los datos descritos en el apartado 2.2 se obtienen unas predicciones muchos más exactas y fiables que siguiendo métodos menos desarrollados. En la Figura 32 se muestra la misma representación para el método determinista con el que se han comparado los modelos evaluados. Viendo estas gráficas se confirma que los modelos basados en IA son más precisos.

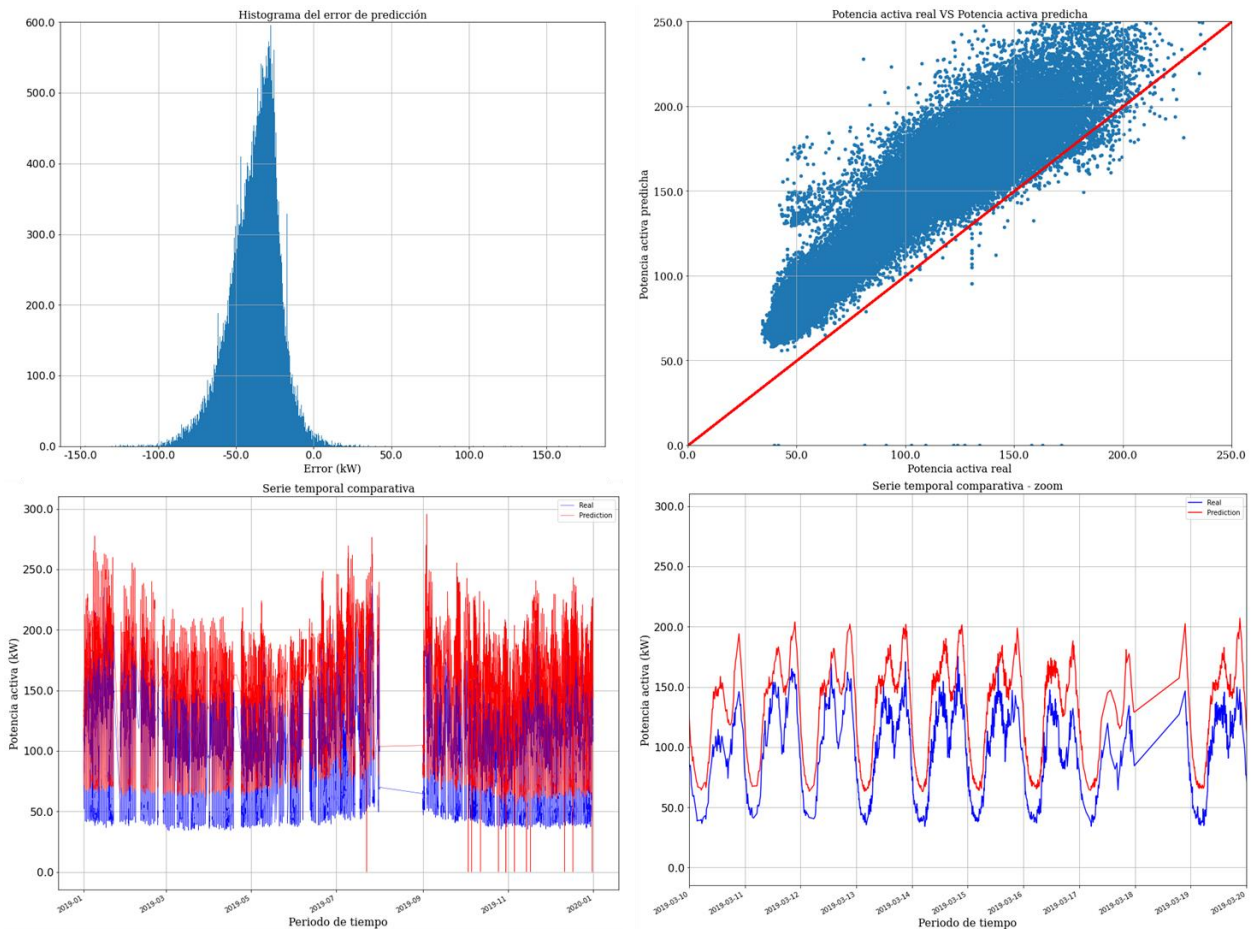


Figura 32 Resultados métricas para curvas de potencia de activa normalizadas

4.3.2 Procedimiento publicado en BOE en términos de energía

Con el objetivo de ofrecer una visión más amplia del rendimiento de la metodología propuesta, los resultados de predicción conseguidos con uno de los modelos, la Red Neuronal 1, se ha comparado con una herramienta que sea de uso generalizado por las distribuidoras y a la que se pueda tener acceso.

De esta manera, como se introdujo en el primer apartado de esta memoria, se han comparado los resultados conseguidos con la metodología aquí desarrollada con el método de cálculo de los perfiles de consumo para los puntos de medida tipo 4 y tipo 5 de consumidores que no dispongan de registro horario. Este es un procedimiento oficial publicado en el artículo 9 del Real Decreto 1435/2002, de 27 de diciembre, por el que se establece un método de cálculo del perfil de consumo de contadores a efectos de la liquidación de su consumo eléctrico por parte de las empresas comercializadoras. Para ello, todos los años se publica en el BOE una tabla con valores de referencia de la Demanda de Referencia y Perfiles Iniciales propuesto por la Comisión Nacional de los Mercados y Competencia.

Este método calcula los perfiles finales de energía horaria para un cliente concreto, siguiendo varias expresiones detalladas en el artículo publicado en el BOE y utilizando los valores de referencia que se calculan cada año. Como se dispone de la información del número de clientes y el tipo de tarifa de cada uno de ellos (2.0 o 3.0), se ha implementado dicha metodología para el año 2019 y todos los contadores del transformador objetivo de predicción, se han agregado para simular la predicción del consumo de energía y se ha incrementado un 3% para incluir una parte de consumo relativo a las pérdidas técnicas de la red.

Este perfil de consumo de energía horaria del transformador se ha comparado con las predicciones hechas con la metodología aquí propuesta en términos de energía activa horaria en lugar de potencia activa cinco-minutal. Este cálculo se ha hecho suponiendo que la potencia activa permanece constante durante los cinco minutos hasta que se registra la siguiente medida. Dicho cálculo se ha llevado tanto para la energía predicha como para la energía real, la primera a partir de las predicciones de potencia activa y la segunda a partir de las medidas reales

del transformador.

En la Tabla 17 se han recogida, al igual que en las pruebas y en la comparativa anterior, las métricas calculadas en esta comparativa para todos los modelos evaluados y para el procedimiento del BOE. En este caso, a diferencia de los anteriores, las métricas están calculadas en términos de energía.

Tabla 17 Comparativa de métricas en términos de energía entre modelos estudiados y procedimiento BOE

Modelo	Evaluación con datos clusterizados	
	MAE (kWh)	MSE (kWh ²)
DTR	136.65	45320.85
GBR	281.18	143047.47
XGBoost	118.57	30339.64
RNP 1	106.93	28117.58
RNP 2	125.12	36138.21
BOE	832.34	1054045.12

Como se puede ver, los resultados anteriores confirman que las predicciones efectuadas por los valores predichos por los modelos aquí evaluados mejoran el error asociado al procedimiento del BOE.

4.4 Otros alcances de predicción

Hasta ahora, el caso de uso que se ha abordado en este Trabajo Fin de Máster se ha centrado en entrenar modelos para predecir el histórico de potencia activa de un año completo del transformador objetivo (año 2019). Sin embargo, la manera en la que se han estructurado los datos de entrada a los modelos de predicción para su entrenamiento permitiría la aplicación de los mismos para hacer predicciones a futuro.

Siendo así, y para valorar la metodología aquí propuesta en un contexto diferente, se han evaluado los mismos modelos anteriores pero, en este caso, haciendo predicciones cinco-minutales a tres días vista (866 valores predichos) y se ha comparado con el valor real. En concreto, se han utilizado los de enero a noviembre, este último incluido, de los transformadores del clúster 3, para predecir la curva de potencia activa del transformador objetivo para los días 1, 2 y 3 diciembre. En la Figura 33 se presentan gráficamente los resultados obtenidos.

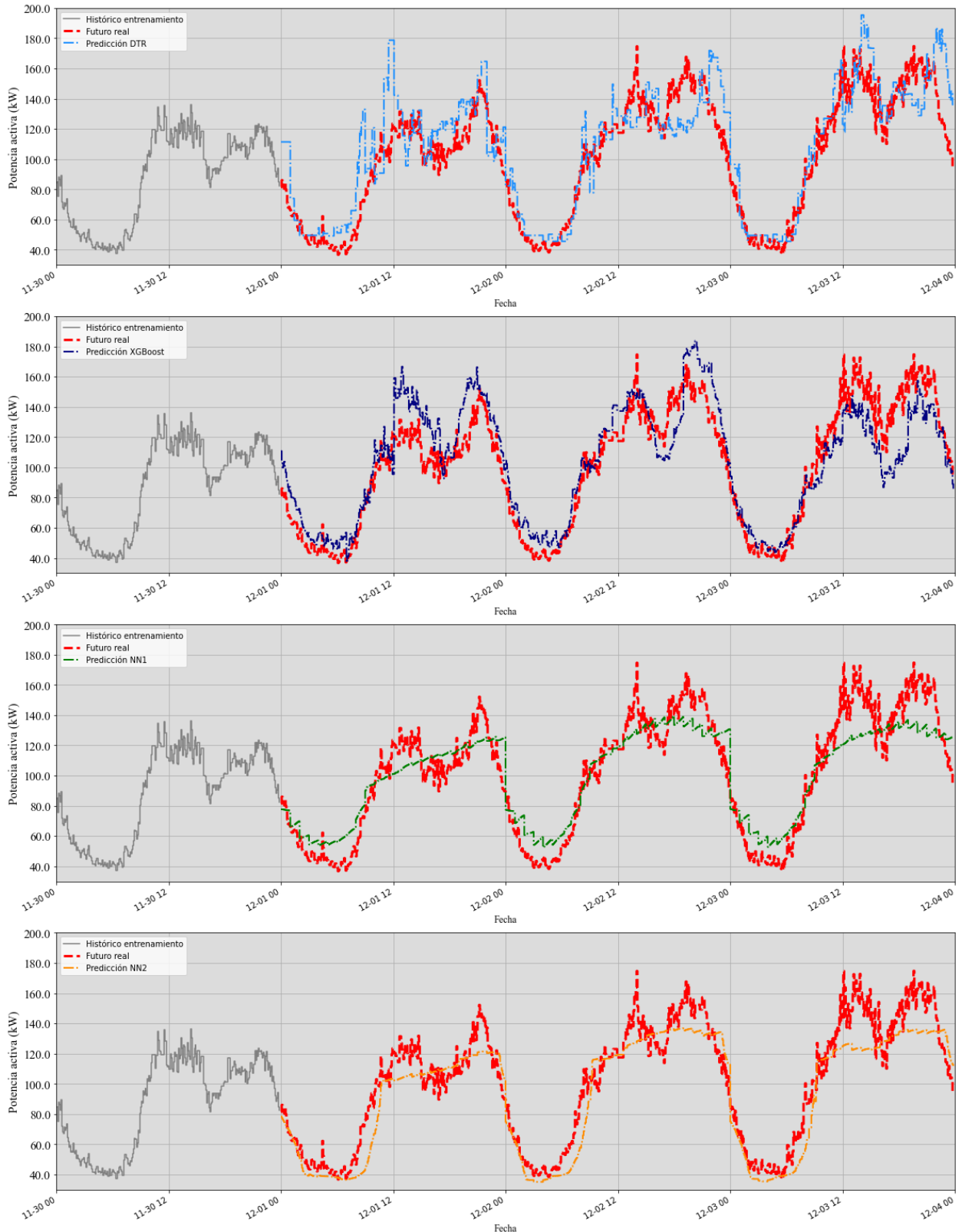


Figura 33 Comparativa de las predicciones a tres días vista con los modelos DTR, XGB, NN1 y NN2

Como se puede apreciar en las cuatro imágenes anteriores, de forma general, los cuatro modelos tienen un buen comportamiento en el caso de uso de predicción a tres días vista. No obstante, analizando más detenidamente cada una de ellas, destacan los modelos DTR y XGBoost por su predicción más precisa del patrón. En la Tabla 18 están recogidos, al igual que en el caso de uso principal del TFM, las métricas de la predicción donde se pone de manifiesto ese buen comportamiento.

Tabla 18 Comparativa de métricas en el caso de uso de predicción

Modelo	Evaluación predicción	
	MAE (kW)	MSE (kW²)
DTR	15.98	444.27
XGBoost	14.30	319.16
RNP 1	13.19	250.81
RNP 2	12.38	245.47

Por último, es importante destacar que, en los datos para hacer la predicción, se ha asumido conocida la temperatura de los tres días para los que se quiere predecir la potencia activa. Esta información puede encontrarse de forma gratuita en algunas APIs meteorológicas como la de AEMET.

5 CONCLUSIONES Y TRABAJO FUTURO

El primer requisito para el éxito es la capacidad de aplicar tus energías físicas y mentales a un problema, sin cesar y sin cansarte.

- Thomas A. Edison -

El objetivo principal de este Trabajo Fin de Máster ha sido establecer una metodología válida para predecir la potencia activa en transformadores MT/BT de los que no se dispone de un histórico de medidas de cualquier naturaleza eléctrica, transformadores no observables, y en concreto centrándose en la predicción de un transformador de tipo residencial.

Para ello se han aplicado y evaluado diferentes técnicas de analítica de datos e inteligencia artificial utilizando un conjunto de datos reales de potencia activa procedentes de sensores instalados en Centros de Transformación reales. Adicionalmente, se ha utilizado, junto a lo anterior, información topológica de la red suministrada por dichos transformadores pero limitando su uso para simular un contexto representativo del conocimiento de la red de una distribuidora. Con esto se pretende que los resultados aquí conseguidos sean perfectamente escalables a cualquier distribuidora. En resumen, se han aplicado y evaluado nuevas técnicas de inteligencia artificial para predecir la potencia activa de transformadores MT/BT en escenarios estrictamente realistas.

A lo largo de las distintas etapas de desarrollo de este trabajo, se ha puesto de manifiesto que el uso de datos reales para entrenar modelos de predicción, especialmente en el caso de uso que aquí se ha tratado, implica verdaderos desafíos:

- Disponer de una BB.DD. lo suficientemente grande y balanceada como para poder abordar el mayor espectro posible del parque de transformadores y permita, a los modelos de predicción, generalizar en su entrenamiento.
- Necesidad por parte de la distribuidora de tener un control exhaustivo y una actualización constante de la información recogida en sus BB.DD. Especialmente la información topológica, puesto que es una información dinámica que a menudo no se revisa pero que influye directamente sobre la potencia en transformadores y afecta, como consecuencia, a los modelos de predicción.

No obstante, se ha comprobado que, haciendo un tratamiento inteligente de los datos disponibles por medio de técnicas como el clustering, es posible obtener unos buenos resultados para los que, a la hora de decidir su aplicación, se deberá evaluar qué modelos o combinación de modelos se ajustan mejor a los objetivos esperados. Destacando, en el caso concreto de este trabajo, los modelos basados en árboles de decisión y los modelos basados en redes neuronales.

Además, es importante destacar que se ha constatado como la metodología aquí descrita mejora el comportamiento de otras técnicas más generales u oficialmente reconocidas que son ampliamente utilizadas.

Es por ello que la conclusión principal es que se confirma el potencial del uso de la infraestructura de datos que se está generando actualmente, en concreto, el potencial de la utilización de datos de transformadores de potencia MT/BT observables y técnicas de inteligencia artificial para predecir históricos de potencia activa en aquellos que no son observables es una opción válida e interesante.

Igualmente, además de la predicción de históricos, se han evaluado esos mismos modelos en el caso de uso de predicción de potencia activa a futuro, en concreto a tres días vista, y se consiguen unos resultados muy buenos. Estos resultados favorecen la aplicación de estos modelos en diferentes contextos.

En definitiva, los resultados aquí conseguidos invitan a valorar detenidamente las aplicaciones finales de uso, entre los que destacan:

- Predecir históricos de consumo en transformadores MT/BT remotos o no sensorizados que no disponen de una buena comunicación de las medidas de contadores.
- Garantizar la observabilidad en algoritmos de monitorización basados en Estimación de Estado, especialmente cuando, por imposibilidad física, las distribuidoras tienen Centros de Transformación o transformadores en poste sin equipos de medida. Un ejemplo de ello es el proyecto eCitySevilla, en la Isla de Cartuja, en donde el 2% del parque de transformadores no pudo sensorizarse.

Finalmente, se han destacado los siguientes trabajos futuros para mejorar el rendimiento de la metodología:

- Evaluar cómo de confiable y aplicable serían estas predicciones en aplicaciones finales. Un buen ejemplo sería caracterizar estas predicciones como pseudomedidas y utilizarla con herramientas de Estimación de Estado en escenarios controlados.
- Incluir en el conjunto de datos de entrenamiento variables relacionadas con la penetración de energía renovable o vehículo eléctrico en la red para ayudar a generalizar los modelos en estos contextos, que están empezando a ser cada vez más frecuentes.

REFERENCIAS

- [1] C. Rung-Fang, L. Rong-Ceng y L. Chan-Nan, «Distribution transformer load modelling using load research data,» *IEEE Transactions on Power Delivery*, vol. 7, no. 2, pp. 655-661, April 2002.
- [2] M. S. Giridhar y S. Sivanagaraju, «Distribution transformer load modelling with interval Type-2 Fuzzy Sets,» *Joint International Conference on Power Electrics, Drives and Energy Systems & Power*, vol. 2, nº 13, pp. 1-5, 2010.
- [3] C. Carmona, E. Romero-Ramos y J. Riquelme, «Fast and Reliable Distribution Load and State Estimator,» *Electric Power Systems Research*, pp. 110-124, 2013.
- [4] S. Kreutmayr, C. J. Steinhart, M. Finkel y C. Gutzmann, «Methodology for Annual Load Profile Estimation at the Outgoing Feeder of Distribution Transformers in Urban Areas,» *CIGRE Conference*, pp. 2032-9644, 2019.
- [5] H. M. Usman, R. El Shatshat y A. H. El-Hag, «Distribution Transformer kVA Load Estimation Using Smart Meter Data,» *IEEE Electrical Power and Energy Conference (EPEC)*, pp. 1-5, 2021.
- [6] J. Chen, D. Zhang y Y. Nanekaran, «Research of power load prediction based on boost clustering,» *Soft Comput*, pp. 6401-6413, 2021.
- [7] J. Jiménez Mares, L. Navarro, M. Quintero, G. C. y M. Pardo, «A Methodology for Energy Load Profile Forecasting Based on Intelligent Clustering and Smoothing Techniques,» *Energies*, p. 4040, 2020.
- [8] G. Rouwhorst, E. M. S. Duque, P. H. Nguyen y H. Slootweg, «Improving Clustering-Based Forecasting of Aggregated Distribution Transformer Loadings With Gradient Boosting and Feature Selection,» *IEEE Access*, vol. 10, pp. 443-455, 2022.
- [9] R. Fonteijn, T. Castelijn, M. Grond, P. H. Nguyen, J. Morren y H. Slootweg, «Short-term load forecasting on MV/LV transformer level,» *CIGRE*, pp. 2032-9644, 2019.
- [10] X. Haining, T. Yingjie, Z. Wei y H. Zhongyu, «Heavy overloaded forecasting of distribution transformers based on neural network,» *MATEC Web of Conferences*, 2020.
- [11] S. I. Kampezidou y S. Grijalva, «Distribution transformers short-term load forecasting models,» *IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1-5, 2016.
- [12] D. S. e. al., «Deep Learning-Based Short-Term Load Forecasting Approach in Smart Grid With Clustering and Consumption Pattern Recognition,» *IEEE Access*, vol. 9, pp. 54992-55008, 2021.
- [13] R. Carmona-Pardo, «Low Voltage Time Series Forecasting ML & DL Model Applied on Renewable Energy and Electric Vehicle Integration,» UMA. Master in Advanced Analytics on Big Data, 2019.
- [14] PASTORA, 2019. [En línea]. Available: <https://www.endesa.com/es/proyectos/todos-los-proyectos/transicion-energetica/redes-inteligentes/pastora-inteligencia-artificial-red-distribucion>.

- [15] M. L. Waskom, «Seaborn: statistical data visualization,» *The Open Journal*, 2021.
- [16] McKinsey. [En línea]. Available: <https://www.mckinsey.com/business-functions/quantumblack/our-insights/an-executives-guide-to-ai>.
- [17] Pedregosa y a. et, «Scikit-learn: Machine Learning in Python,» *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [18] J. Friedman, «Greedy Function Approximation: A Gradient Boosting Machine,» *The Annals of Statistics*, vol. 29, 2000.
- [19] c. Wikipedia, «Rectifier (neural networks),» 2022. [En línea]. Available: [https://en.wikipedia.org/w/index.php?title=Rectifier_\(neural_networks\)&oldid=1111203702](https://en.wikipedia.org/w/index.php?title=Rectifier_(neural_networks)&oldid=1111203702).

