



GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

TRABAJO FIN DE GRADO PREDICCIÓN DE DEMANDA ELÉCTRICA

Autor: LUCAS OTERO FILGUEIRA

Director: MERCEDES VALLES RODRÍGUEZ

Co-Director: JAVIER RENESES

Madrid

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

PREDICCIÓN DE DEMANDA ELÉCTRICA

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2023/24 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.



Fdo.: LUCAS OTERO FILGUEIRA

Fecha: 22/07/2024

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Mercedes Vallés Rodríguez

Fecha: 22/07/2024

Agradecimientos

Quisiera expresar mi más sincero agradecimiento a todas las personas que han hecho posible la realización de este Trabajo de Fin de Grado.

En primer lugar, me gustaría agradecer a la empresa SimularT, donde tuve la oportunidad de realizar mi TFG. Mi agradecimiento especial a Mercedes y Javi, quienes, como directores de proyecto, me brindaron su apoyo, guía y conocimientos a lo largo de todo el proceso. Su paciencia, profesionalidad y dedicación fueron fundamentales para el desarrollo de este trabajo.

A mis padres, quienes también estudiaron en ICAI y me transmitieron el amor por la ingeniería. Gracias por transmitirme desde pequeño el deseo de ser ingeniero, por vuestro ejemplo y por vuestro apoyo incondicional en cada paso de mi formación.

A mis abuelos, por su constante apoyo y cariño, y por ser siempre una fuente de inspiración y motivación. Compartir con vosotros este logro es un orgullo y una alegría inmensa.

Finalmente, quiero agradecer a la ciudad de La Coruña, donde concluí y presenté mi TFG. Esta ciudad no solo me proporcionó un entorno propicio para el estudio y el trabajo, sino que también me acogió con calidez y me ofreció momentos inolvidables junto a mi familia.

A todos, gracias de corazón. Este logro es tanto mío como vuestro.

RESUMEN DEL PROYECTO

1. Introducción

Este proyecto se centra en el desarrollo de modelos predictivos para anticipar la demanda eléctrica, utilizando datos históricos de demanda y temperatura de España y Portugal. Los modelos implementados incluyen regresión lineal, Ridge, Lasso y el Tao Vanilla Benchmark (TVB), cada uno con distintas características y capacidades para ajustar los datos y prever la demanda.

La predicción de demanda eléctrica es crucial para la planificación y operación eficiente del sector eléctrico. Permite a las empresas eléctricas optimizar la generación y distribución de energía, reducir costos, y facilitar la integración de fuentes de energía renovables, promoviendo una red eléctrica más sostenible. La capacidad de prever la demanda también es vital para la programación del mantenimiento de infraestructuras y la gestión de la red en tiempo real, mejorando así la estabilidad y confiabilidad del suministro eléctrico.

El proyecto se desarrolla en varias fases: la recopilación y preprocesamiento de datos, el análisis y modelado predictivo, y la evaluación y comparación de los modelos. Los datos utilizados abarcan desde julio de 2020 hasta diciembre de 2023, excluyendo el período del confinamiento para evitar distorsiones. Se utilizan datos de demanda horaria y temperaturas diarias, con fuentes principales como Red Eléctrica para España y REN para Portugal.

El proyecto no solo busca mejorar la precisión de las predicciones de demanda a medio plazo, sino también contribuir a una red eléctrica más eficiente y sostenible. A través de este trabajo, se destacan los beneficios de la predicción de demanda para el sector eléctrico, alineándose con los Objetivos de Desarrollo Sostenible (ODS), especialmente en aspectos de energía asequible y no contaminante, ciudades y comunidades sostenibles, y acción por el clima. En resumen, este TFG aporta una herramienta estratégica para mejorar la gestión y eficiencia del sistema eléctrico, facilitando la transición hacia un modelo energético más sustentable.

2. Definición del Proyecto

El objetivo principal es el desarrollo y la implementación de modelos predictivos que permitan anticipar la demanda de electricidad. Este proyecto se fundamenta en la utilización de técnicas avanzadas de análisis de datos y modelado estadístico para ofrecer herramientas que mejoren la planificación y gestión del sistema eléctrico.

El proyecto emplea datos históricos de demanda y temperatura de España y Portugal, abarcando desde julio de 2020 hasta diciembre de 2023. Los datos incluyen la demanda horaria y las temperaturas diarias, obtenidos de fuentes oficiales como Red Eléctrica para España y REN para Portugal. Este conjunto de datos se utiliza para entrenar y validar varios modelos predictivos, entre los cuales se incluyen regresión lineal, Ridge, Lasso y el Tao Vanilla Benchmark (TVB).

3. Descripción del modelo/sistema/herramienta

El proyecto "Predicción de Demanda Eléctrica" se centra en el desarrollo de varios modelos predictivos para anticipar la demanda eléctrica en España y Portugal. Para ello, se emplean técnicas de regresión lineal, Ridge, Lasso y el modelo Tao Vanilla Benchmark (TVB), cada uno con sus particularidades y aplicaciones específicas.

Regresión Lineal: Es el modelo inicial utilizado debido a su simplicidad y facilidad de interpretación. La regresión lineal establece una relación directa entre la demanda eléctrica y variables predictoras como la temperatura y el tiempo. A pesar de su simplicidad, este modelo proporciona una visión básica y comprensible de cómo las variables influyen en la demanda eléctrica.

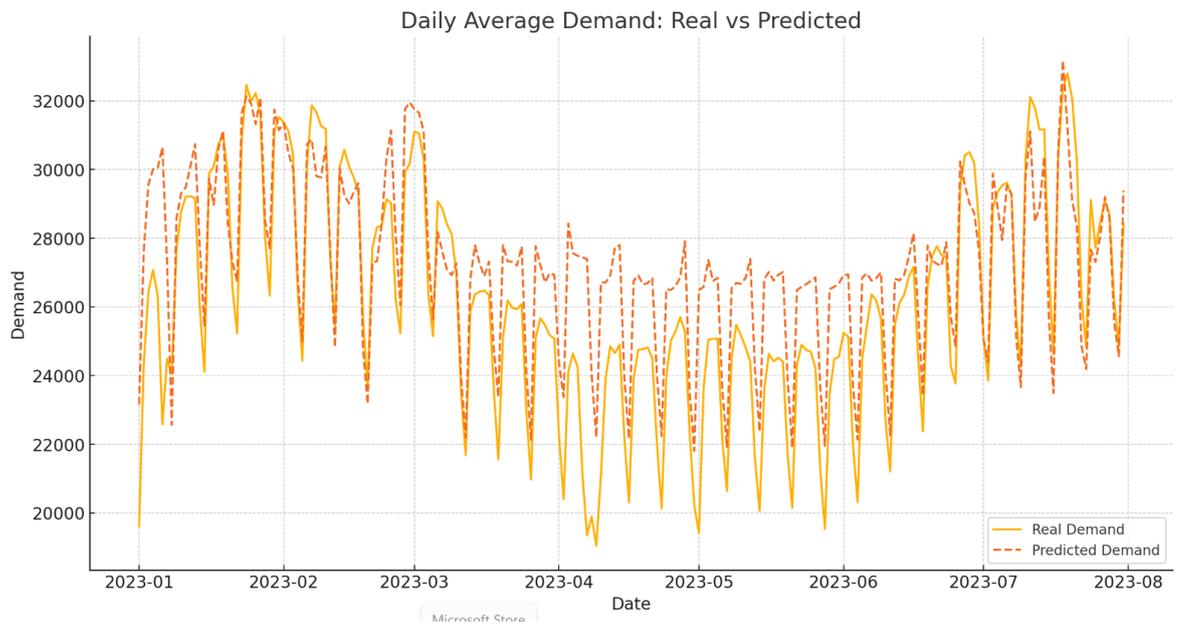
Regresión Ridge: Este modelo es una variante de la regresión lineal que incluye un término de penalización para evitar el sobreajuste y manejar la colinealidad entre variables predictoras. La regresión Ridge es útil para mejorar la capacidad de generalización del modelo y ofrecer predicciones más robustas cuando se enfrentan a datos con alta correlación entre variables.

Regresión Lasso: Similar al modelo Ridge, la regresión Lasso también añade un término de penalización, pero su principal ventaja es la capacidad de realizar selección de características. Esto significa que puede eliminar automáticamente las variables menos importantes, simplificando el modelo y mejorando su interpretabilidad.

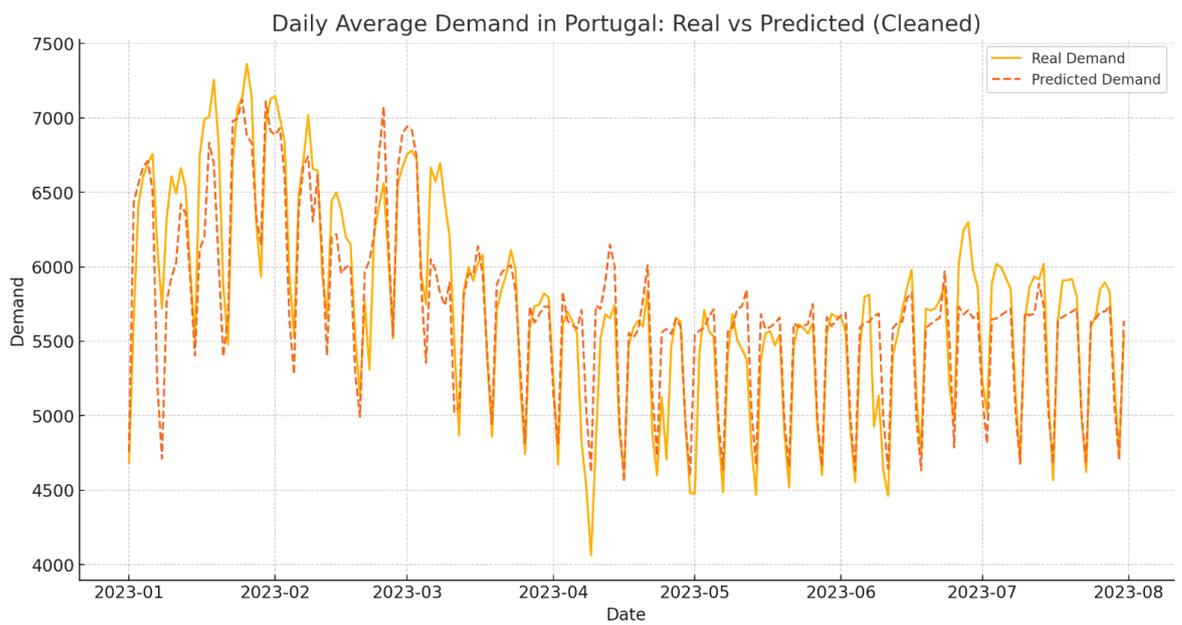
Modelo Tao Vanilla Benchmark (TVB): Este modelo, utilizado como referencia en competiciones de predicción de demanda, es conocido por su simplicidad y efectividad. Se basa en técnicas de regresión lineal múltiple y es útil para comparar la eficacia de otros modelos más complejos. El TVB es apreciado por su transparencia y buen rendimiento en la predicción de demanda a corto plazo.

Los modelos se entrenan y validan utilizando datos históricos de demanda y temperatura recopilados de fuentes oficiales como Red Eléctrica (España) y REN (Portugal). Los datos se preprocesan y se dividen en conjuntos de entrenamiento y validación para asegurar la precisión y la aplicabilidad de las predicciones. Este enfoque permite optimizar la planificación y gestión del sistema eléctrico, mejorando la eficiencia operativa y facilitando la integración de fuentes de energía renovable.

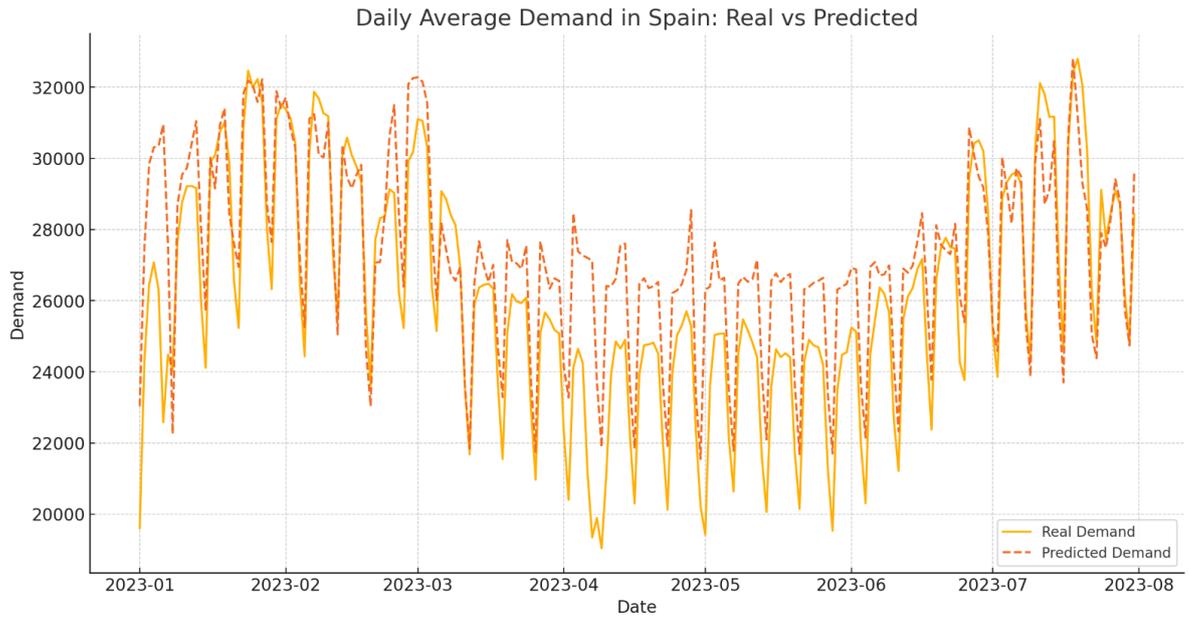
4. Resultados



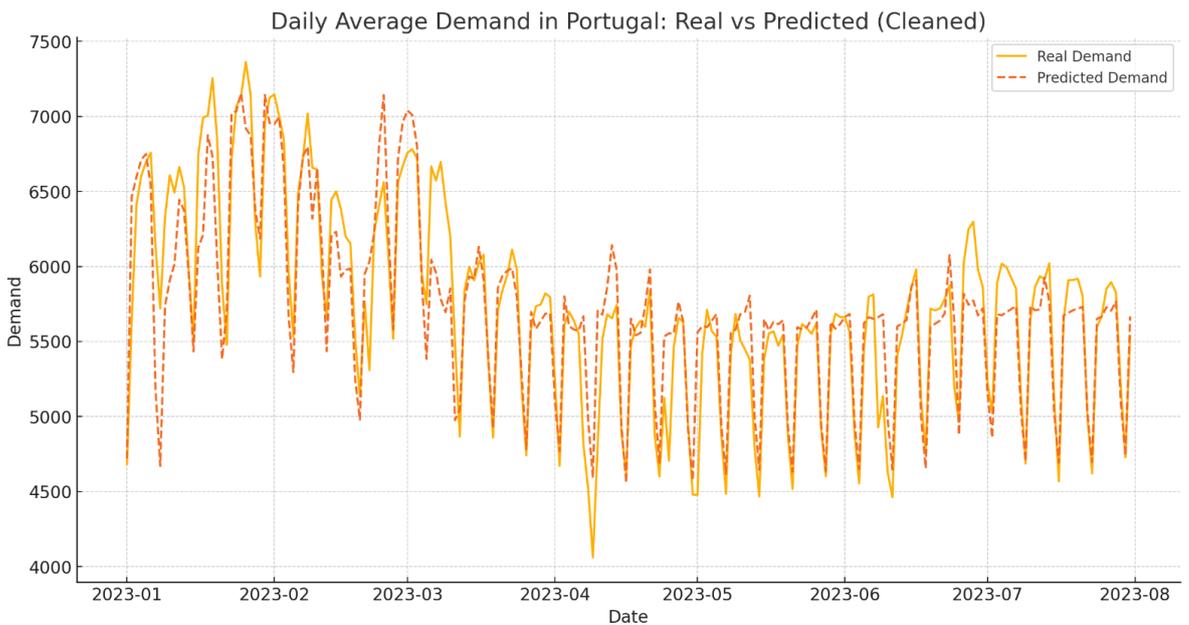
Predicción para España para (1-1-2023 hasta 31-07-2023) con Lasso



Predicción para Portugal para (1-1-2023 hasta 31-07-2023) con Lasso



Predicción para España para (1-1-2023 hasta 31-07-2023) con Ridge



Predicción para Portugal para (1-1-2023 hasta 31-07-2023) con Ridge

Model Error Metrics							
	Modelo	MAE España	MAPE España (%)	RMSE España	MAE Portugal	MAPE Portugal (%)	RMSE Portugal
1	TVB	1982.776	8.066074	2402.148	443.7327	8.296281	544.8969
2	Ridge	1904.888	7.807879	2315.515	419.8067	7.946575	510.5826
3	Lasso	1900.85	7.78368	2318.01	427.0236	7.9935	519.855

5. Conclusiones

En conclusión, el modelo Lasso se destaca como el mejor método de predicción de demanda eléctrica entre los modelos considerados. En España, Lasso muestra el menor MAE y MAPE, indicando una mayor precisión en términos absolutos y relativos. En Portugal, aunque Ridge tiene un RMSE ligeramente mejor, las diferencias en MAE y MAPE entre Ridge y Lasso son mínimas, con Lasso superando significativamente al modelo TVB. Esto sugiere que Lasso es casi tan eficaz como Ridge en manejar grandes errores, con el beneficio adicional de tener un mejor desempeño general en España.

Visualmente, el modelo Lasso también se ve ligeramente mejor en comparación con los otros modelos, mostrando predicciones más ajustadas a la demanda observada. Tanto Lasso como Ridge superan al modelo TVB utilizado como Benchmark, demostrando una mayor precisión y robustez en sus predicciones.

La simplicidad y capacidad de evitar el sobreajuste hacen que Lasso sea más robusto y generalizable. Al utilizar regularización L1, Lasso penaliza el valor absoluto de los coeficientes, llevando algunos a cero y seleccionando solo las características más relevantes. Esto resulta en modelos más simples y fáciles de interpretar. Por otro lado, Ridge utiliza regularización L2, que penaliza el cuadrado de los coeficientes, reduciéndolos sin llevarlos a cero, lo que es útil cuando todas las características son importantes.

En resumen, aunque Ridge y Lasso muestran un desempeño similar, Lasso se establece como el mejor modelo de predicción debido a su precisión, consistencia, simplicidad y capacidad de generalización. Lasso es ideal para aplicaciones prácticas donde se valora la simplicidad y la robustez frente al sobreajuste, y ambos modelos son superiores al TVB utilizado como referencia.

6. Referencias

- [1] Hyndman, R., Koehler, A.B., Ord, J.K., Snyder, R.D. (2008). "Forecasting with Exponential Smoothing: The State Space Approach". Springer.
- [2] Laptev, N., Yosinski, J., Li, L.E., Smyl, S. (2017). "Time-series extreme event forecasting with neural networks at Uber". International Conference on Machine Learning.
- [3] Li, C., Ding, Z., Yi, J., Lv, Y., Zhang, G. (2018). "Deep belief network based hybrid model for building energy consumption prediction". *Energies*, 11(1), 242.
- [4] Hoerl, A. E., & Kennard, R. W. (1970). "Ridge regression: Biased estimation for nonorthogonal problems". *Technometrics*, 12(1), 55-67.
- [5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Springer.

[6] Zou, H., & Hastie, T. (2005). "Regularization and variable selection via the elastic net". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320.

Índice de la memoria

Capítulo 1. Introducción	3
1.1 Estado de la cuestión o revisión de soluciones tecnológicas al problema analizado a otros similares	5
1.2 motivación o justificación de la realización del proyecto, para qué se hace.	6
1.3 Objetivos del proyecto	7
1.4 Alineación con los objetivos de desarrollo sostenible	7
1.5 fases del proyecto.....	9
1.6 Análisis de las herramientas y medios a emplear.....	10
Capítulo 2. Recopilación de datos.....	12
Capítulo 3. regresión lineal.....	20
3.1.1 Uso de los datos.....	23
3.2 Modelo tvb (tao vanilla benchmark)	23
3.2.1 Modelo Tao Vanilla Benchmark (TVB) en rstudio	27
3.2.2 Modelo rige.....	31
3.3 Regresión lasso	45
Capítulo 4. Comparación de los resultados según el modelo	65
Capítulo 5. Conclusiones del proyecto.....	69
Capítulo 6. Bibliografía.....	71

Capítulo 1. INTRODUCCIÓN

Este Trabajo Fin de Grado (TFG) trata de la predicción de la demanda. Dicha predicción desempeña un papel esencial en todos los ámbitos del sector eléctrico. Por una parte, permite a las empresas planificar de manera más eficaz la infraestructura de generación y distribución para satisfacer la demanda máxima esperada. Asimismo, esta anticipación posibilita ajustes en tiempo real, optimizando la eficiencia de las plantas generadoras, facilitando el casamiento de oferta y demanda en el mercado eléctrico. Además, facilita la integración de fuentes renovables intermitentes, contribuyendo a una red eléctrica más sostenible.

La capacidad de prever la demanda también es muy útil de cara a programar el mantenimiento de los activos de generación y la infraestructura eléctrica sin afectar el suministro en momentos de alta demanda. De la misma forma, la gestión de la red se beneficia al anticipar la demanda, permitiendo una operación más estable y confiable.

En términos económicos, la predicción consigue reducir costos al evitar inversiones innecesarias y minimizar el uso de fuentes de generación costosas durante picos imprevistos. En definitiva, la predicción de la demanda impulsa una planificación estratégica, operación eficiente y gestión efectiva de los recursos, promoviendo una red eléctrica sólida, confiable y económica.

En la predicción de demanda se trabaja con diferentes horizontes temporales, desde el corto al largo plazo. En cada uno de estos horizontes, toman relevancia distintos factores que afectan al consumo eléctrico.

En el corto plazo, elementos como las condiciones climáticas desempeñan un papel significativo, donde la temperatura y la humedad pueden afectar las preferencias y necesidades de consumo de las personas. Por otro lado, el calendario, tanto por las estaciones, los periodos especiales y los tipos de día, como por la hora del día influye en los patrones de demanda de manera notable.

En el medio y largo plazo, otros factores adquieren importancia. La actividad económica e industrial, junto con la evolución de la población, son determinantes en la configuración de la demanda a largo plazo. Las mejoras en la eficiencia energética también tienen un impacto significativo, ya que pueden modificar la cantidad y el tipo de recursos energéticos requeridos para satisfacer las necesidades de la sociedad.

Además, se observa una transformación en los patrones de consumo debido a la aparición de nuevas tecnologías y tendencias. La adopción de nuevos consumos, como los vehículos eléctricos y la electrificación de dispositivos que antes dependían de combustibles fósiles, como cocinas y sistemas de calefacción, está alterando el panorama energético y, por ende, la demanda asociada.

El cambio climático es otro factor importante que afecta la demanda a largo plazo. Los cambios en los patrones climáticos pueden tener impactos significativos en la disponibilidad de recursos naturales, la producción agrícola y otros aspectos de la economía, lo que a su vez influye en los patrones de consumo.

Otro aspecto relevante en la demanda eléctrica es el surgimiento del autoconsumo energético, especialmente a través de fuentes renovables como la energía solar. Esta tendencia puede evolucionar hacia modalidades más avanzadas, como el autoconsumo con almacenamiento mediante baterías o la utilización de vehículos eléctricos como sistemas de almacenamiento energético.

Finalmente, el comportamiento de los consumidores está siendo moldeado por avances tecnológicos que les permiten reaccionar de manera más ágil a las señales de precio y programar sus consumos de forma inteligente. Esta capacidad de adaptación a las condiciones del mercado puede tener un impacto significativo en la configuración de la demanda en el medio y largo plazo.

Es evidente la necesidad de incorporar de alguna manera la influencia de todos estos elementos en los modelos y técnicas de predicción de la demanda eléctrica en función del horizonte de predicción.

1.1 ESTADO DE LA CUESTIÓN O REVISIÓN DE SOLUCIONES TECNOLÓGICAS AL PROBLEMA ANALIZADO A OTROS SIMILARES

La previsión de la demanda en el sector eléctrico se aborda en distintos horizontes temporales. A corto plazo, que abarca desde las próximas horas a varios días, la predicción de demanda se suele utilizar para diseñar estrategias de participación en los mercados de energía, realizar ajustes inmediatos en la generación y operación de la red, para lo cual debe incorporar datos en tiempo real y considerar la variabilidad de las fuentes renovables.

En la predicción a medio plazo, que abarca desde semanas a meses, se emplea para anticipar la demanda mediante análisis de tendencias históricas, patrones estacionales y factores económicos. Facilita la planificación estratégica y la coordinación de recursos. Es utilizado para la programación del mantenimiento de generadores.

Por último, la previsión a largo plazo, con horizontes de años a décadas, proyecta la demanda teniendo en cuenta aspectos como el crecimiento económico, cambios demográficos y avances tecnológicos, guiando decisiones en la planificación de inversiones a gran escala. Estos enfoques temporales permiten una gestión integral y adaptativa del sistema eléctrico, muy importante para asegurar su eficiencia y sostenibilidad a lo largo del tiempo. Utilizando dichas predicciones para la construcción de nuevas líneas y subestaciones o para realizar modificaciones en el sistema existente.

Los métodos de predicción de la demanda varían según el horizonte temporal considerado. Para el corto plazo, se emplean métodos de series temporales como, modelización de regresión para incorporar eventos especiales y técnicas de aprendizaje automático como máquinas de soporte vectorial.

En el medio plazo, se suelen emplear análisis de regresión multivariable, métodos de suavización exponencial y modelos de descomposición que permiten abordar tendencias y estacionalidades.

Por último, en el largo plazo, se utilizan modelos de crecimiento exponencial, escenarios futuros y modelos económicos integrados que consideran factores económicos, demográficos y tecnológicos.

La elección de métodos depende de la disponibilidad de datos, la complejidad de los patrones y la precisión requerida, y a menudo se favorece la combinación de enfoques para obtener predicciones más robustas y precisas.

1.2 MOTIVACIÓN O JUSTIFICACIÓN DE LA REALIZACIÓN DEL PROYECTO, PARA QUÉ SE HACE.

La predicción de la demanda tiene un papel crucial dentro del sector eléctrico, siendo fundamental para una planificación eficaz de la generación y distribución de la energía eléctrica. Este TFG se enfoca en la predicción de la demanda eléctrica, con mayor énfasis en la predicción a medio plazo.

La anticipación de la demanda permite a las empresas eléctricas adaptarse con mayor eficacia a las fluctuaciones en el consumo. Esto no solo optimiza la eficiencia de las plantas generadoras, sino que también facilita el equilibrio entre la oferta y la demanda en el mercado eléctrico. La integración de fuentes renovables intermitentes se ve muy beneficiada con los métodos de predicción de demanda, promoviendo así una red eléctrica más sostenible.

La capacidad de prever la demanda impacta positivamente en la programación del mantenimiento, evitando interrupciones en el suministro durante momentos de alta demanda. De este modo la predicción de demanda evita gastos innecesarios, abaratando el proceso de producción. Además, la gestión anticipada de la red permite una operación más estable, beneficiando la planificación estratégica, la eficiencia operativa y la gestión efectiva de recursos.

Por último, una predicción fiable de la demanda de un sistema eléctrico puede ser utilizada como dato de entrada para otras aplicaciones de interés para el análisis y el estudio de los sistemas de energía eléctrica y los mercados. Un ejemplo de esto son los modelos de predicción, tanto fundamentales como estadísticos, de precios de los distintos mercados de energía eléctrica.

En resumen, en este TFG se busca mediante la predicción de la demanda conseguir mejorar la eficiencia operativa del sector eléctrico, contribuir a una transición hacia una red eléctrica más sostenible y con una mayor importancia de las fuentes renovables. De esta manera, se contribuye a mejorar la gestión del sistema eléctrico, asegurando su eficiencia y sostenibilidad tanto a corto como a largo plazo.

1.3 OBJETIVOS DEL PROYECTO

Este TFG se propone abordar los beneficios del uso de la predicción de la demanda en el sector eléctrico.

El objetivo principal es proponer una serie de modelos de predicción de la demanda eléctrica que se puedan aplicar a un conjunto grande de consumidores, tal como un sistema eléctrico o una cartera de clientes de una comercializadora, para un horizonte de medio plazo, es decir, de unos meses a un año. Otros objetivos se desglosan a continuación:

- Propuesta y ajuste de técnicas de modelado estadístico para predicción de la demanda eléctrica de fácil aplicación a sistemas reales.
 - Análisis y comparación de los modelos propuestos.
 - Aplicación de los modelos propuestos a sistemas reales. En particular, se prueban los modelos propuestos en los sistemas eléctricos de España y Portugal peninsulares.
-

1.4 ALINEACIÓN CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE

La predicción de la demanda eléctrica contribuye con los Objetivos de Desarrollo Sostenible, concretamente en Energía Asequible y No Contaminante (ODS 7), en Ciudades y Comunidades Sostenibles (ODS 11) y en Acción por el Clima (ODS 13).

La predicción de la demanda se relaciona de forma directa con el ODS 7, Energía Asequible y No Contaminante. Al buscar la optimización de recursos y costos en el sector eléctrico. Al prever la demanda con precisión, se podrían evitar inversiones innecesarias y minimizar el uso de fuentes de generación costosas durante picos imprevistos, lo que permitiría una mayor eficacia en la producción de energía. Esto se vería reflejado con un mayor peso de las renovables en el mix energético, promoviéndose de este modo una energía más limpia y sostenible.

También contribuye con el ODS 11, Ciudades y Comunidades Sostenibles. Al mejorar la planificación y eficiencia en la infraestructura eléctrica, beneficiando a comunidades y ciudades. La predicción precisa de la demanda asegura un suministro eléctrico confiable, evitando interrupciones en el servicio y promoviendo la estabilidad en el entorno urbano. Al optimizar la gestión de recursos, se contribuye a la construcción de comunidades más sostenibles.

Se podría relacionar con el ODS 13, Acción por el Clima. Ya que al anticipar con precisión la demanda, se optimiza la eficiencia de las plantas generadoras y se facilita la integración de fuentes renovables, reduciendo la dependencia de combustibles fósiles y disminuyendo las emisiones de gases de efecto invernadero. Fortaleciendo los esfuerzos para mitigar el cambio climático.

En resumen, la predicción de la demanda esta alineada con los Objetivos de Desarrollo Sostenible (ODS), ya que contribuye de forma activa a la acción por el clima, promoviendo ciudades y comunidades sostenibles, y trabajando hacia una energía no contaminante. La

predicción precisa de la demanda se convierte así en una herramienta estratégica para impulsar la sostenibilidad y eficiencia en el sector eléctrico.

1.5 FASES DEL PROYECTO

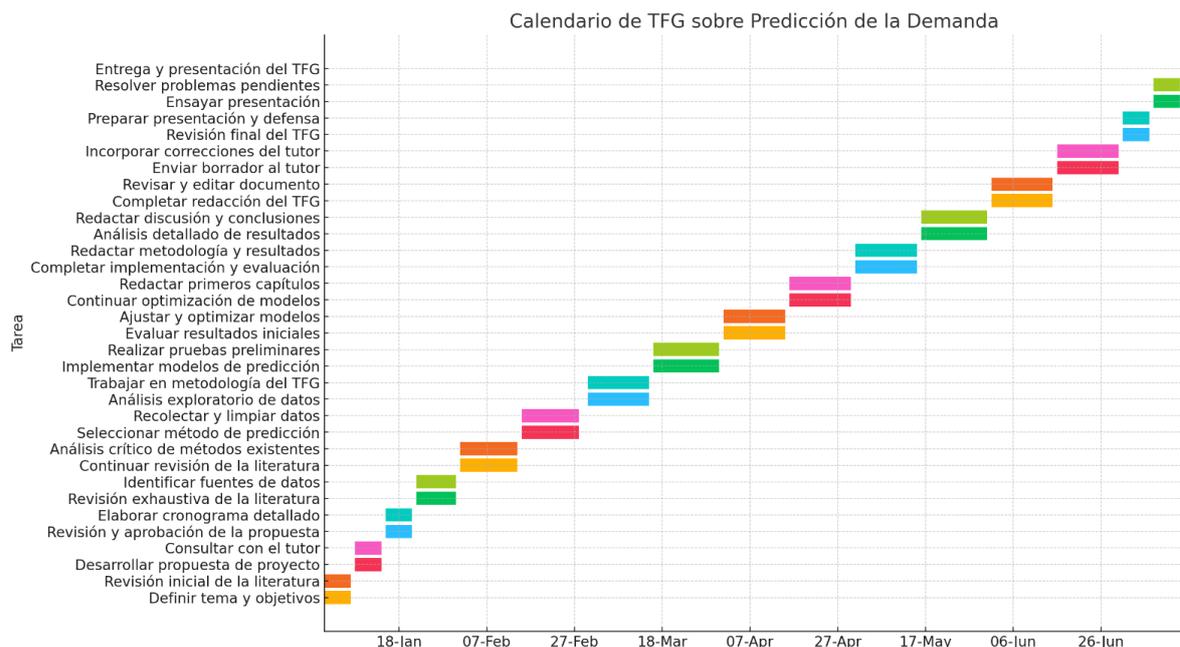


Ilustración 1

El gráfico de Gantt presenta una visión general del calendario para el TFG sobre la predicción de la demanda. Se distribuye desde enero hasta mediados de julio y está dividido en varias fases clave:

1. Inicio del Proyecto: definición del tema, objetivos y elaboración de la propuesta inicial.
2. Metodología y recolección de datos: selección del método de predicción adecuado y recopilación.
3. Implementación y análisis: implementación de los modelos de predicción, realización de pruebas preliminares y optimización de los modelos.

4. Programación del código de los modelos de predicción de la demanda eléctrica en Rstudio.
5. Revisión y correcciones: revisión del documento completo, incorporación de correcciones y feedback de los tutores.
6. Preparación final: ensayo de la presentación, resolución de problemas pendientes y preparación para la entrega y defensa del TFG.

1.6 ANÁLISIS DE LAS HERRAMIENTAS Y MEDIOS A EMPLEAR

En este apartado, se explorarán los recursos y herramientas necesarios para la implementación de técnicas de predicción de la demanda

Para llevar a cabo el análisis y la predicción de la demanda se ha utilizado principalmente el lenguaje R y el programa RStudio. Este lenguaje de programación ofrece una amplia gama de bibliotecas y paquetes especializados en el análisis de datos, la aplicación de técnicas estadísticas y el ajuste de modelos de machine learning.

El análisis descriptivo de datos y el ajuste de modelos de regresión lineal, así como otras variantes de estos modelos, presentan varias ventajas significativas en el ámbito de la predicción de la demanda de energía eléctrica. En primer lugar, estas técnicas permiten capturar de manera efectiva la estructura de los datos, incluyendo patrones estacionales, tendencias y ciclos. Esto proporciona una base sólida para el análisis preciso de la demanda en diferentes momentos del año.

Además, los modelos de regresión lineal y sus variantes son relativamente simples de entender e interpretar, lo que facilita su aplicación. Al utilizar métodos estadísticos y modelos matemáticos, es posible realizar análisis detallados de los datos y evaluar la precisión de las predicciones.

En resumen, la combinación de análisis descriptivo de datos, modelos de regresión lineal y herramientas de programación como R y Excel son eficaces para el análisis de la demanda de energía eléctrica. Mediante la aplicación de estas técnicas, se puede mejorar la planificación y gestión de los recursos energéticos, contribuyendo así a una mayor eficiencia y sostenibilidad en el suministro de energía eléctrica.

Capítulo 2. RECOPILACIÓN DE DATOS

En este capítulo se describen los datos utilizados en el proyecto, detallando su origen, estructura y el proceso de preparación necesario para su análisis. Se justifica su inclusión porque estos datos serán la base para desarrollar y comparar distintos modelos predictivos de demanda eléctrica. Además, se explica el rango de fechas considerado y cómo se han dividido los datos en conjuntos de entrenamiento y validación. En este proyecto se utilizan los datos desde el 1 de julio de 2020 hasta el 31 de diciembre de 2022 para el entrenamiento del modelo con el fin de evitar los datos anómalos obtenidos durante el confinamiento, que podrían distorsionar los resultados y reducir la eficacia del modelo. Los datos desde el 1 de enero de 2023 hasta el 31 de diciembre de 2023 se guardarán como datos de validación para comprobar y evaluar el rendimiento del modelo, asegurando así que se basa en un período más estable y representativo.

Para la realización del proyecto es necesario tener datos, los cuales son datos públicos de los operadores del sistema de cada uno de los sistemas. Red Eléctrica para España y REN para Portugal, dichos datos se encuentran en un archivo que contiene datos de demanda horaria de España como de Portugal. Los datos de Red Eléctrica para España y REN para Portugal los he obtenido de:

<https://www.ree.es/es/datos/publicaciones/boletines-mensuales>

<https://www.ren.pt>

Los datos de temperatura diaria de estaciones meteorológicas de referencia: Madrid (Retiro) para España y Oporto (aeropuerto) para Portugal. Son temperaturas de localidades concretas, pero se ha comprobado que tienen una alta representatividad de la temperatura media de sendos países. Esta información se ha utilizado para desarrollar una serie de modelos predictivos que permitan anticipar la demanda futura en función de estos parámetros. Este proceso incluye la carga y exploración de los datos, su preprocesamiento, un análisis para

identificar patrones y relaciones y la selección y entrenamiento de los modelos. Finalmente, se evalúan y comparan los modelos para asegurar su precisión y aplicabilidad en escenarios reales, con el objetivo de optimizar la predicción de demanda para la empresa.

Estos datos se han organizado en un archivo Excel que será trabajado utilizando RStudio con el objetivo de identificar posibles relaciones. La primera relación en la que trabajaremos será la relación que tenemos entre la temperatura ambiente y el consumo de electricidad a nivel nacional a lo largo del año, lo que permitirá comprender mejor cómo influyen las condiciones climáticas en el uso de energía eléctrica. Aquí lo que puedes comentar es que se observa estacionalidad anual, variabilidad entre unos años y otros, cierta tendencia decreciente en España en los últimos años.

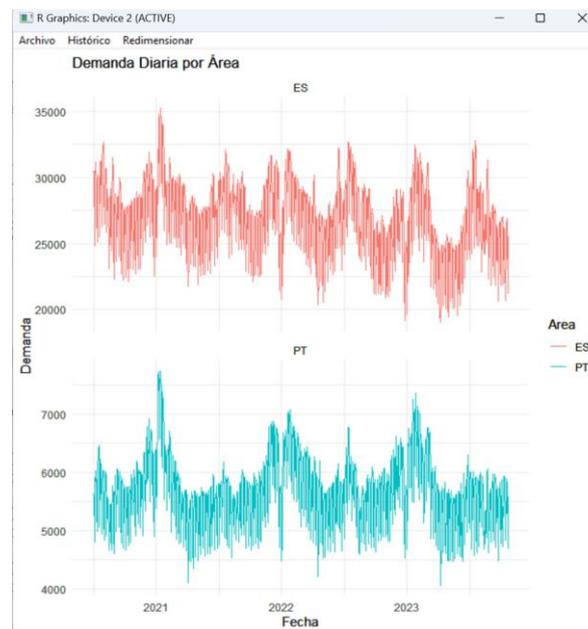


Ilustración 1

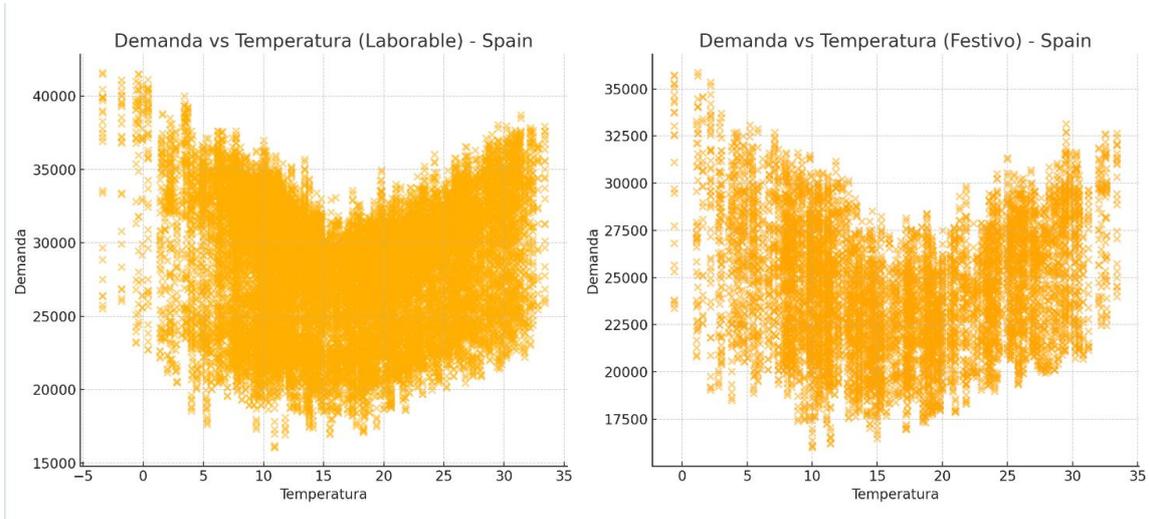


Ilustración 2

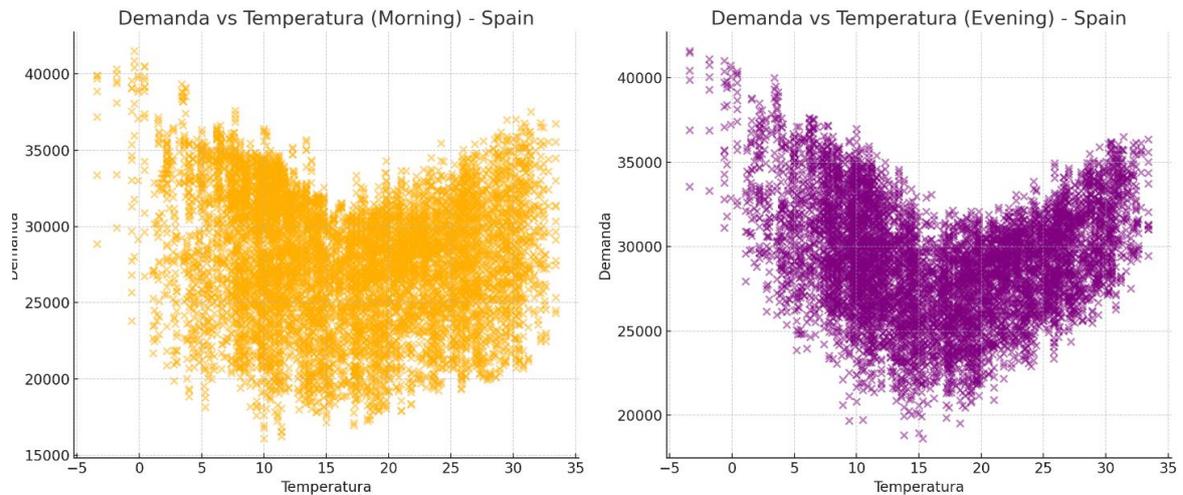


Ilustración 3

La forma de "U" en las gráficas de demanda versus temperatura generalmente indica dos picos de demanda en extremos opuestos de la escala de temperatura, con una disminución en la demanda a temperaturas moderadas. Esta forma puede ser explicada por varios factores:

A temperaturas bajas, la demanda de energía aumenta debido al uso intensivo de sistemas de calefacción. Hogares, oficinas y otros edificios utilizan más energía para mantener el confort térmico en invierno.

En temperaturas moderadas (15-20 grados), la necesidad de calefacción y refrigeración es mínima. Por lo tanto, la demanda de energía tiende a ser más baja y estable. Este rango suele ser confortable sin necesidad de intervención térmica significativa.

A temperaturas altas, la demanda de energía vuelve a aumentar debido al mayor uso de sistemas de aire acondicionado y refrigeración. Esto se observa particularmente durante los meses de verano cuando se busca mantener ambientes frescos.

Las pendientes en las gráficas reflejan la sensibilidad de la demanda de energía a los cambios de temperatura, y estas pendientes pueden variar por varias razones:

Durante las altas temperaturas, la pendiente de la curva de demanda es más pronunciada porque el uso de sistemas de refrigeración es mucho más intensivo. La energía requerida para enfriar espacios a niveles confortables puede ser significativamente alta, especialmente en zonas con climas cálidos y húmedos.

Similarmente, durante las bajas temperaturas, la pendiente es pronunciada debido al aumento del uso de sistemas de calefacción. Las regiones con inviernos severos experimentan un incremento en la demanda de energía para mantener temperaturas interiores adecuadas.

La forma plana observada en las gráficas en el rango de 15-20 grados puede ser atribuida a que este rango de temperatura es típicamente considerado confortable para muchas personas. Ni calefacción ni refrigeración adicionales son necesarios, resultando en una demanda de energía más baja y estable.

Durante estas temperaturas moderadas, la necesidad de utilizar sistemas de calefacción o aire acondicionado es mínima. Las actividades diarias pueden llevarse a cabo sin requerir ajustes significativos en el ambiente interior, lo que lleva a una menor variación en la demanda de energía.

En la ilustración 3 se observa, que las gráficas de demanda eléctrica versus temperatura para la mañana y la tarde en España, se pueden identificar varias diferencias y patrones significativos. En la gráfica de la mañana, la forma de "U" es menos pronunciada. La demanda es alta tanto en temperaturas bajas como altas, pero no disminuye tanto en temperaturas moderadas. En contraste, la gráfica de la tarde presenta una forma de "U" más clara y pronunciada, con una disminución significativa de la demanda en el rango de temperaturas moderadas, aproximadamente entre 15 y 20 grados.

Durante la mañana, la demanda es alta en temperaturas extremas, tanto bajas como altas, aunque el aumento en demanda a temperaturas muy altas (por encima de 25 grados) no es tan pronunciado como en la tarde. Por la tarde, la demanda eléctrica aumenta considerablemente en temperaturas altas, reflejando un uso intensivo de sistemas de aire acondicionado. Este comportamiento es más notable comparado con la mañana.

En temperaturas moderadas, la gráfica de la mañana muestra que la demanda no disminuye tanto como en la tarde. Este comportamiento sugiere que, durante la mañana, a pesar de las temperaturas moderadas, hay otros factores que mantienen la demanda relativamente alta, posiblemente relacionados con la actividad industrial y el inicio de operaciones diarias. Por otro lado, en la tarde, la demanda es considerablemente más baja en este rango, reflejando una menor necesidad de calefacción o refrigeración y quizás una reducción en actividades que requieran alta energía.

Finalmente, se observa una mayor dispersión de los datos en la mañana, lo que indica una variabilidad en la demanda para las mismas temperaturas. En la tarde, la dispersión de los datos es más concentrada, especialmente en el rango de temperaturas moderadas, indicando un comportamiento de demanda más predecible en este rango de temperatura.

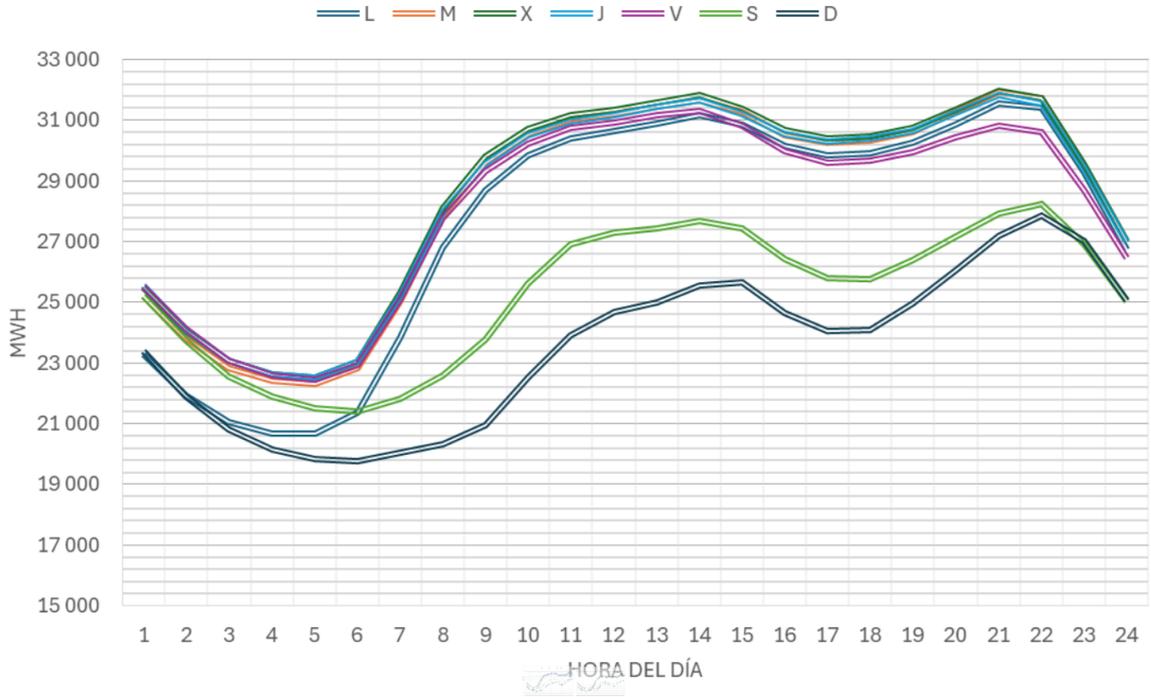


Ilustración 4

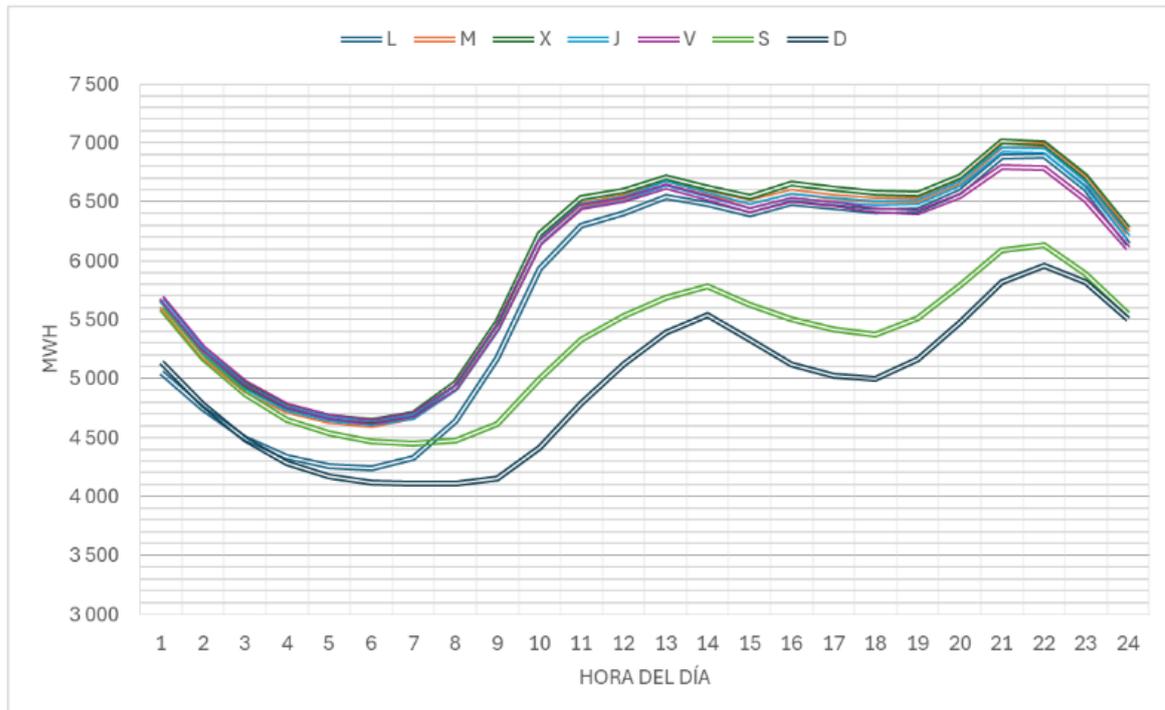


Ilustración 5

En la ilustración 4 (España) y 5 (Portugal) , se observa la diferencia en la demanda de energía eléctrica dependiendo de la hora del día y del día de la semana se debe a los patrones de actividad y comportamiento de las personas. Durante los días laborales (lunes a viernes), la demanda tiende a aumentar significativamente en las horas de la mañana (alrededor de las 8 a 9 AM) cuando la gente comienza su jornada laboral, y luego vuelve a subir en la tarde (alrededor de las 6 a 8 PM) cuando las personas regresan a sus hogares y utilizan más electricidad para actividades domésticas.

En los fines de semana (sábado y domingo), la demanda general es más baja comparada con los días laborales. Esto se debe a la disminución de actividades industriales y comerciales y al hecho de que muchas personas tienden a tener horarios más flexibles, lo que reduce los picos de demanda en las horas específicas del día. Especialmente el domingo, se observa una demanda considerablemente más baja, reflejando una menor actividad en general.

En resumen, la demanda de energía eléctrica varía a lo largo del día y de la semana debido a los ciclos de trabajo, actividades diarias y el uso de electrodomésticos y sistemas de climatización en los hogares, que son más intensivos durante los días laborales y las horas pico de la mañana y la tarde.

Capítulo 3. REGRESIÓN LINEAL

Para poder realizar una primera predicción de demanda se ha desarrollado un código en RStudio que se explicará a continuación. El objetivo del código es analizar y predecir la demanda eléctrica en España y Portugal utilizando datos históricos de demanda y temperatura. El propósito de un modelo como los propuestos aquí es obtener un perfil de demanda que se asemeje al real en el corto plazo, y también que pueda ser utilizado para proyectar una demanda base a largo plazo. A esta proyección se debería incorporar posteriormente el impacto de factores de largo plazo utilizando una metodología diferente. Esto representa un aspecto de mejora futuro para continuar este proyecto. Para realizarlo se han seguido los siguientes pasos:

1. Carga y preprocesamiento de datos: Lectura de los datos previamente recopilados de demanda eléctrica y temperatura. En estos datos se ha evitado coger datos de la época del confinamiento.

	A	B	C	D
1	Fecha	Hora	Demanda	Area
2	01/07/2020	0	27618.2	ES
3	01/07/2020	1	25650	ES
4	01/07/2020	2	24574.9	ES
5	01/07/2020	3	23948.2	ES
6	01/07/2020	4	23982.8	ES
7	01/07/2020	5	24005	ES
8	01/07/2020	6	25408.5	ES

Ilustración 6

	A	B	C
1	Fecha	Temp	Area
2	01/07/2020	28	ES
3	02/07/2020	26.9	ES
4	03/07/2020	23.9	ES
5	04/07/2020	26	ES
6	05/07/2020	27.9	ES
7	06/07/2020	29.6	ES
8	07/07/2020	30.1	ES
9	08/07/2020	30.6	ES

Ilustración 7

2. Análisis de datos: Estructuración de los datos.
3. Modelado predictivo: Ajuste de un modelo lineal para predecir la demanda eléctrica en función de la temperatura y el tipo de día.
4. Visualización de resultados: Creación de gráficos para comparar la demanda observada con la demanda predicha.

En el programa Rstudio he utilizado las librerías ``readr``, ``dplyr``, ``lubridate`` y ``ggplot2`` para la carga, manipulación, análisis y visualización de los datos. Los datos fueron leídos desde archivos CSV (Excel) mostrados en las ilustraciones 5 y 6. Dichos archivos contienen la demanda eléctrica y la temperatura diaria, por día y hora la demanda, y una única temperatura por día. La manipulación inicial incluyó la conversión de fechas a un formato reconocible por R y la creación de nuevas columnas que clasifican los días de la semana y el tipo de día (laboral, sábado, domingo).

Para el modelo de predicción se ajustó un modelo lineal simple para predecir la demanda eléctrica (``Demanda``) en función de la temperatura (``Temp``) y el tipo de día (``Tipo_Dia``).

Con este modelo he intentado capturar la relación entre estas variables y la demanda eléctrica. La fórmula del modelo es:

$$\text{Demanda} = \beta_0 + \beta_1 \times \text{Tipo_Dia} + \beta_2 \times \text{Temp} + \epsilon$$

En la siguiente fórmula intervienen los siguientes términos:

- β_0 es la intercepción, esto es la ordenada en el origen, esto es el valor que tiene la demanda cuando todas las variables independientes tienen el valor 0.
- β_1 y β_2 son los coeficientes de regresión para el tipo de día y la temperatura, respectivamente.
- ϵ es el término de error que representa cómo varía la demanda que no puede explicarse a través de las variables independientes.

Para incluir esto en mi código de Rstudio he utilizado la función de Rstudio, `lm()`, para ajustar el modelo de regresión lineal.

- Coeficientes de Regresión (β_0 , β_1 , β_2): Estos coeficientes indican el impacto de cada variable independiente en la demanda.
- R^2 : Una medida estadística que indica el porcentaje de la varianza en la variable dependiente que es explicada por las variables independientes en un modelo de regresión. Representa qué tan bien se ajustan los datos al modelo.
- Error Estándar: Indica la precisión de los coeficientes estimados.
- Valores p: Muestran la significancia estadística de cada variable independiente en el modelo.

¿Por qué he utilizado un modelo lineal?

Un modelo lineal es una buena opción inicial para entender las relaciones básicas entre las variables. Es simple de interpretar y puede proporcionar una visión general de cómo factores como la temperatura y el tipo de día afectan la demanda eléctrica. Aunque puede no capturar

todas las complejidades de la demanda eléctrica, ofrece una base desde la cual se pueden construir modelos más complejos.

3.1.1 USO DE LOS DATOS

En el contexto de un modelo estadístico o de aprendizaje automático, la división de datos en conjuntos de entrenamiento y prueba es una práctica estándar y esencial para garantizar la validez y la eficacia del modelo.

Datos de entrenamiento: El conjunto de entrenamiento se utiliza para ajustar el modelo. Esto implica encontrar los parámetros óptimos que minimicen el error en este conjunto de datos.

Datos de prueba: El conjunto de prueba se utiliza para evaluar el rendimiento del modelo. Dado que el modelo no ha visto estos datos durante el entrenamiento, proporciona una estimación imparcial de su capacidad para generalizar a datos nuevos y no vistos.

Voy a usar los datos desde el 1 de julio de 2020 hasta el 31 de diciembre de 2022 para el entrenamiento del modelo con el fin de evitar los datos anómalos obtenidos durante el confinamiento, que podrían distorsionar los resultados y reducir la eficacia del modelo. Los datos desde el 1 de enero de 2023 hasta el 31 de diciembre de 2023 se guardarán como datos de validación para comprobar y evaluar el rendimiento del modelo, asegurando así que se basa en un período más estable y representativo.

3.2 *MODELO TVB (TAO VANILLA BENCHMARK)*

El Modelo TVB (Tao's Vanilla Benchmark), Hong, T., Pinson, P., & et al. (2014). *Tao's Vanilla Benchmark (TVB) model.*, es una técnica de predicción utilizada ampliamente en el

contexto de la predicción de la demanda a corto plazo de energía (STLF, por sus siglas en inglés). Este modelo se basa en técnicas de regresión lineal múltiple y es conocido por su simplicidad, transparencia y buen rendimiento, características esenciales para un modelo de referencia.

El modelo Tao Vanilla Benchmark (TVB) es una herramienta ampliamente reconocida en el campo de la predicción de demanda eléctrica. Desarrollado para proporcionar un estándar de comparación en competiciones de predicción de energía, como el Global Energy Forecasting Competition (GEFCom), el TVB se destaca por su simplicidad y efectividad.

Principales ventajas del modelo TVB

1. Simplicidad: La especificación del modelo es directa, lo que facilita su implementación y entendimiento.
2. Benchmarking: Sirve como referencia para comparar la eficacia de modelos más complejos.
3. Competencias: Ha sido utilizado como base en diversas competiciones de predicción de energía, demostrando su robustez y fiabilidad en escenarios reales.

La definición del modelo es la siguiente:

$$y = \beta_0 + \beta_1 \text{Trend}_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t)$$

con:

$$f(T_t) = \beta_6 T_t + \beta_7 T_t^2 + \beta_8 T_t^3 + \beta_9 T_t M_t + \beta_{10} T_t^2 M_t + \beta_{11} T_t^3 M_t + \beta_{12} T_t H_t + \beta_{13} T_t^2 H_t + \beta_{14} T_t^3 H_t$$

- y Variable dependiente o respuesta que se está modelando, es la demanda eléctrica en el tiempo t.
- β_0 es la intercepción. Representa el valor esperado de y cuando todas las variables independientes son cero

- β_1 Trend captura la tendencia a largo plazo. Trend: Variable que captura la tendencia temporal en los datos, puede ser una variable continua que aumenta linealmente con el tiempo (por ejemplo, los números de semanas o meses).
- β_2 M_t representa el efecto mensual. M: Variable categórica que representa el mes del año. Entra como dummies (0/1) donde cada mes es una categoría diferente (por ejemplo, M_1 para enero, M_2 para febrero, etc.).
- β_3 W_t efectos del día de la semana. W: Variable categórica que representa la semana del año o la semana dentro de un mes. También entra como dummies (0/1).
- β_4 H_t son los efectos de la hora del día. H: Variable categórica que representa la hora del día. También entra como dummies (0/1).
- β_5 $W_t H_t$ es una interacción entre el día de la semana y la hora del día. W:H: Variable de interacción que captura el efecto combinado de la semana y la hora. Representa la interacción entre W y H.
- $f(T_t)$ es una función polinómica de la temperatura que incluye términos hasta el tercer grado e interacciones con otras variables. T_t : Podría representar cualquier otra variable o conjunto de variables que afectan y de manera no lineal o compleja.

Mejoras sobre la predicción anterior:

1. **Variables de Interacción:** El modelo TVB añade interacciones entre la temperatura y otras variables como el mes y la hora, lo cual permite capturar mejor las variaciones en la demanda.
2. **Términos Polinomiales:** Al incluir términos polinomiales de la temperatura, se mejora la capacidad del modelo para capturar relaciones no lineales entre la temperatura y la demanda.
3. **Enfoque Estructurado:** Utiliza un enfoque estructurado que combina variables dependiendo del día en el que nos encontremos y de las condiciones meteorológicas, lo cual es más efectivo para la predicción de la demanda

Para realizar esta predicción realicé:

1. Unión por Fecha y Hora: Se realizó la unión de los datos por las columnas Fecha y Hora para asegurar una relación uno-a-uno.
2. Conversión de Variables a Numéricas: Variables como Mes, Dia_Semana, y Hora se convirtieron a numéricas antes de realizar operaciones aritméticas.
3. Términos Polinomiales y de Interacción: Se crearon términos polinomiales para la temperatura y se calcularon interacciones entre temperatura y otras variables.

Esto permite una predicción más precisa de la demanda eléctrica, aprovechando tanto las variables meteorológicas como las del día en el que nos encontramos, mejorando la capacidad del modelo para capturar patrones complejos en los datos.

3.2.1 MODELO TAO VANILLA BENCHMARK (TVB) EN RSTUDIO

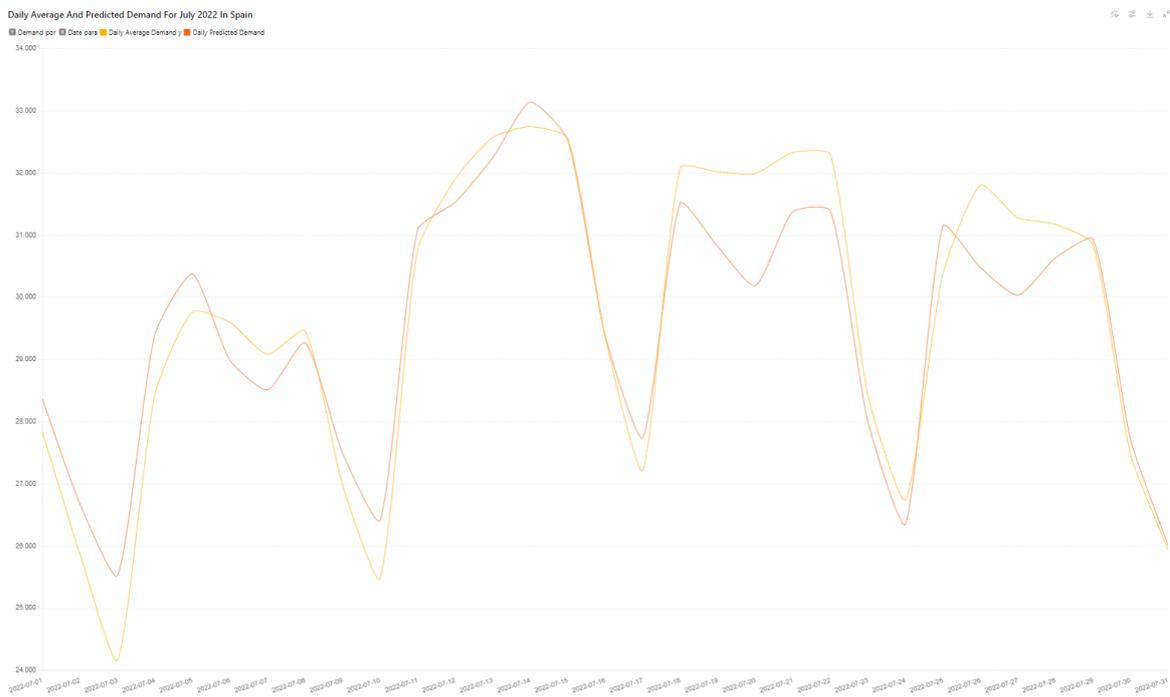


Ilustración 8

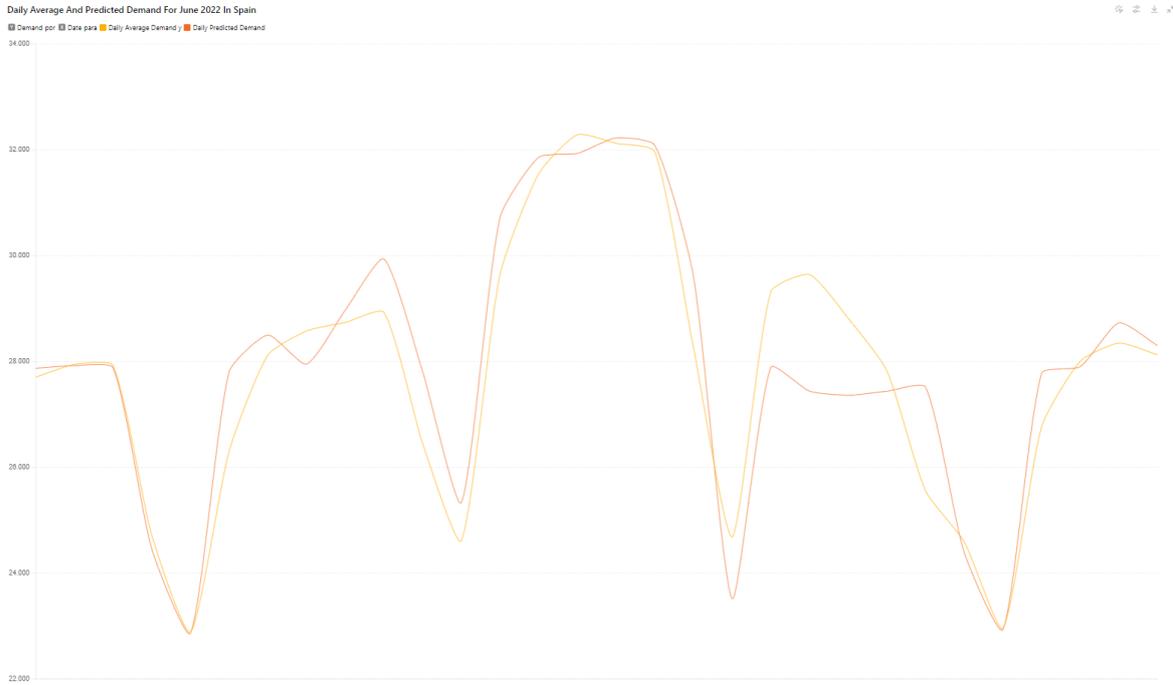


Ilustración 9



Ilustración 10



Ilustración 11

Las gráficas muestran las demandas observadas y predichas en junio y julio en España y Portugal, con los datos de entrenamiento, por tanto, solo aparecen datos hasta diciembre de 2022.

En España, la línea amarilla muestra la demanda observada en los meses de junio y julio a lo largo de los años.

La línea roja muestra la demanda predicha por el modelo TAO Vanilla Benchmark.

Se observa una demanda con picos muy marcados en los meses de junio y julio. Las predicciones del modelo se desvían ligeramente de la demanda observada, especialmente en los picos.

La demanda observada en Portugal es similar a España, la línea amarilla muestra la demanda observada. La línea roja muestra la demanda predicha.

También podemos observar que en Portugal, la demanda tiene picos marcados en junio y julio. El modelo muestra una desviación leve respecto a la demanda observada. Las predicciones fluctúan y esto se ve con los picos observados en la demanda real.

El modelo parece tener dificultades para capturar la estacionalidad y los picos de demanda observados en los datos reales. Al tratarse de un modelo Simplista, El TAO Vanilla Benchmark puede ser demasiado básico para capturar la complejidad de la demanda observada.

Para mejorar el modelo se podría, probar modelos más complejos o ajustados que puedan capturar mejor la estacionalidad y los patrones observados en los datos. También podríamos aumentar el horizonte de datos, es decir, utilizar más datos históricos podría ayudar a capturar mejor las tendencias. Otra opción sería usar Feature Engineering, incluir variables adicionales que puedan influir en la demanda para mejorar las predicciones.

Las gráficas muestran que el modelo TAO Vanilla Benchmark está prediciendo con precisión, salvo algunas imprecisiones, la demanda observada en junio y julio en España y Portugal. Las predicciones se desvían levemente de los datos reales, especialmente en los picos de demanda. Esto indica que el modelo podría mejorar con algunos ajustes, posiblemente incluyendo el uso de un modelo más avanzado y una mejor ingeniería de características para capturar la complejidad de los datos de demanda.

3.2.2 MODELO RIDGE

Después de realizar una investigación he encontrado 2 modelos, empezaré explicando el modelo de Ridge y posteriormente hablaré del modelo de Lasso.

Para justificar la utilización de esta técnica, se puede analizar la correlación entre variables independientes (x). Es cierto que el modelo TAO tiene muchas y seguramente haya un problema de este tipo.

La regresión Ridge es una técnica ampliamente utilizada en el campo del aprendizaje automático y la estadística para manejar problemas de predicción, particularmente cuando hay colinealidad entre las variables predictoras o cuando se quiere evitar el sobreajuste del modelo. A continuación, se explica cómo funciona esta técnica y por qué es útil para la predicción de la demanda.

¿Qué es la Regresión Ridge?

La regresión Ridge es una forma de regularización que modifica la fórmula estándar de la regresión lineal añadiendo un término de penalización a la suma de los coeficientes al cuadrado. Esto se hace para reducir la varianza del modelo y mejorar su capacidad de generalización a nuevos datos. La fórmula general de la regresión Ridge es:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2N} \sum_{i=1}^N (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

Aplicando el teorema a mis variables, tenemos:

$$y_t = \beta_0 + \beta_1 \text{Trend}_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t)$$

donde:

$$f(T_t) = \beta_6 T_t + \beta_7 T_t^2 + \beta_8 T_t^3 + \beta_9 T_t M_t + \beta_{10} T_t^2 M_t + \beta_{11} T_t^3 M_t + \beta_{12} T_t H_t + \beta_{13} T_t^2 H_t + \beta_{14} T_t^3 H_t$$

- y_t : Demanda en el tiempo t .
- $Trend_t$: Tendencia en el tiempo t .
- M_t : Mes en el tiempo t .
- W_t : Día de la semana en el tiempo t .
- H_t : Hora en el tiempo t .
- T_t : Temperatura en el tiempo t .
- T_t^2, T_t^3 : Potencias de la temperatura.
- $T_t M_t, T_t^2 M_t, T_t^3 M_t$: Interacciones de temperatura con el mes.
- $T_t H_t, T_t^2 H_t, T_t^3 H_t$: Interacciones de temperatura con la hora.

Las gráficas proporcionadas muestran la relación entre la demanda observada y la demanda predicha utilizando un modelo de regresión lineal Ridge para España y Portugal. A continuación, se presentan los comentarios sobre los resultados observados:

En España podemos observar que existe, correlación general, esto quiere decir que, hay una correlación positiva visible entre la demanda observada y la demanda predicha, lo que indica que el modelo captura los patrones de los datos. Se observa dispersión de datos. Existe una considerable dispersión alrededor de la línea de identidad (donde la demanda observada sería igual a la predicha). Esto sugiere que, aunque el modelo tiene una tendencia general correcta.

En Portugal la correlación general es Similar a la observada previamente en España, se observa una correlación positiva entre la demanda observada y la predicha.

Comparado con el modelo TAO Vanilla Benchmark, el modelo Ridge parece ofrecer una mejor correlación entre la demanda observada y la predicha, indicando que ha capturado más información relevante de los datos de entrenamiento.

A pesar de mejorar significativamente al modelo TVB este modelo tiene algunas limitaciones, predicciones a cero. En el caso de Portugal, la presencia de muchas predicciones a cero es un indicativo de que el modelo puede estar subestimando la demanda en ciertos casos, lo que podría deberse a datos insuficientes, características mal definidas o posibles errores en el procesamiento de los datos.

Para mejorar este modelo algunas de los posibles cambios para mejorar este modelo serían, mejorar la ingeniería de características, esto quiere decir, incluir más variables que puedan influir en la demanda, como factores climáticos, eventos especiales o cambios económicos.

Otra opción sería incluir modelos más complejos, para ello podríamos considerar el uso de modelos más complejos que puedan capturar mejor la no linealidad y la estacionalidad en los datos, como los modelos de árboles de decisión o redes neuronales.

El modelo de regresión lineal Ridge ha mejorado la predicción de la demanda en comparación con el modelo TAO Vanilla Benchmark, mostrando una correlación positiva entre la demanda observada y la predicha. En el caso de Portugal, las predicciones a cero indican posibles problemas que necesitan ser abordados. Para mejorar las predicciones, sería recomendable revisar y enriquecer las características utilizadas y considerar modelos más complejos.

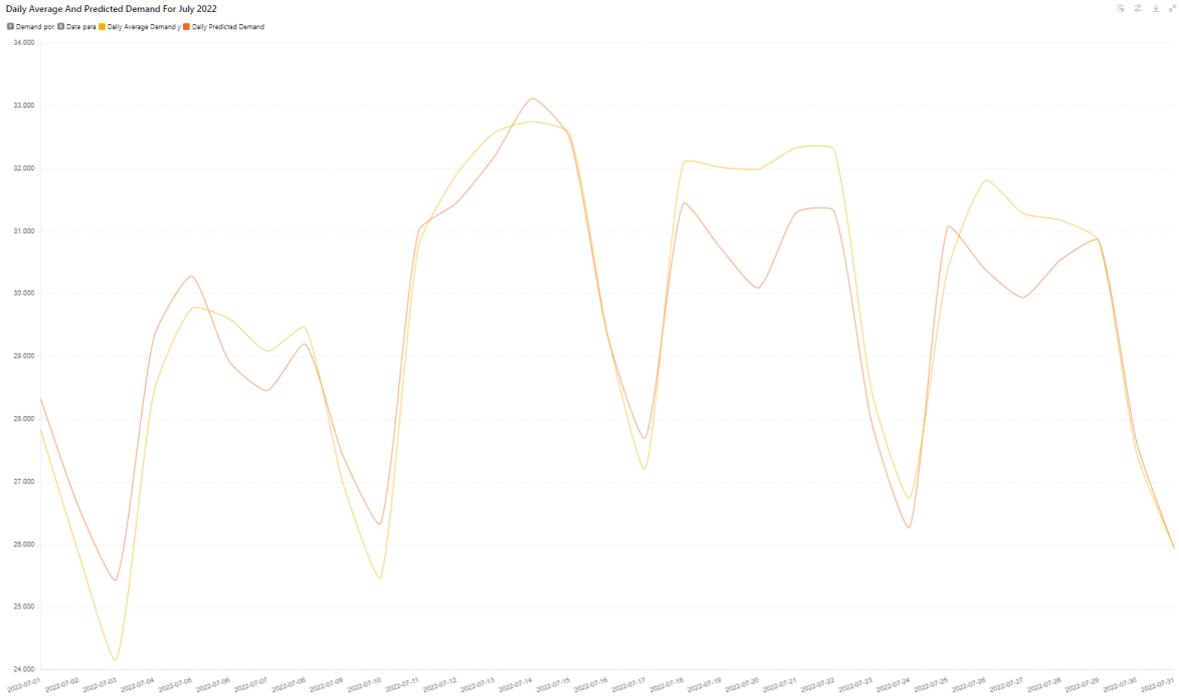


Ilustración 12



Ilustración 13



Ilustración 14

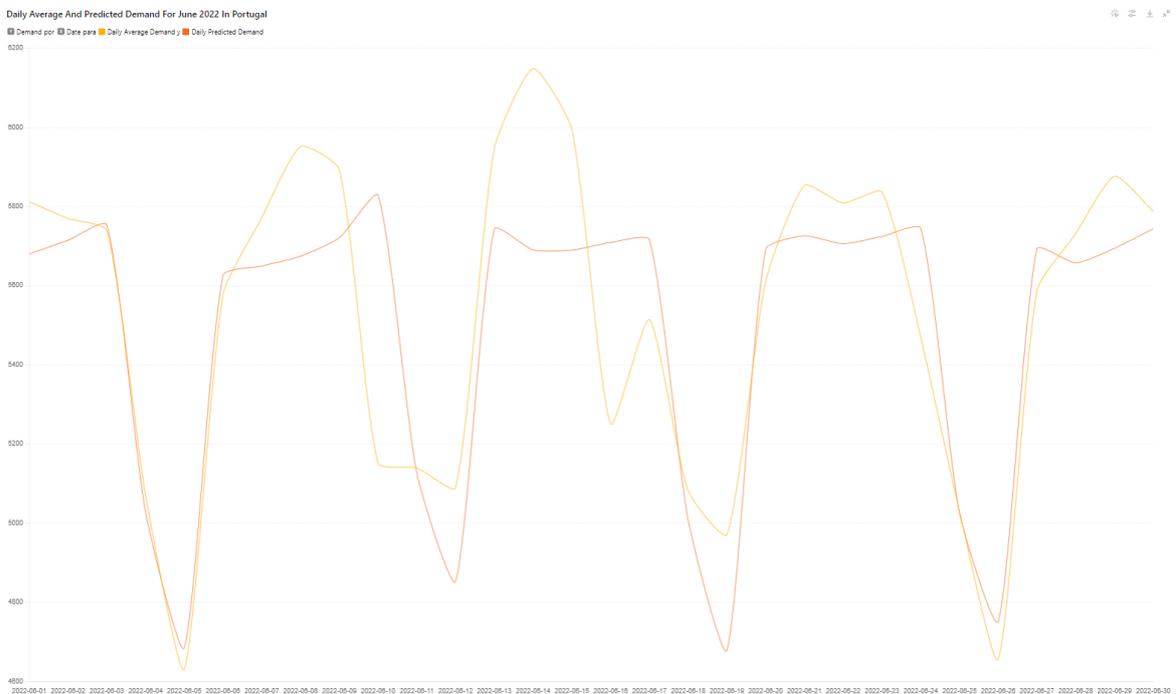


Ilustración 15

En las ilustraciones se observa el modelo de Ridge con los datos de entrenamiento.

En la gráfica de España, la demanda observada (amarilla) y la demanda predicha (rojo) presentan leves diferencias, especialmente en los valores extremos (picos y valles).

En general, el modelo parece no tener muchas dificultades para capturar la variabilidad diaria de la demanda.

En 2020 y 2021, la demanda observada muestra una tendencia decreciente mientras que la predicción sigue una trayectoria más suave y menos variable.

En 2022, la demanda observada presenta un aumento, pero nuevamente, la predicción parece subestimar este cambio y sigue una tendencia más lineal.

La estacionalidad de la demanda no parece ser capturada adecuadamente por el modelo. Las variaciones diarias y semanales observadas en la demanda real no se reflejan bien en las predicciones.

En Portugal, la demanda observada muestra variaciones más pronunciadas, mientras que la predicción tiende a ser más lineal y suavizada.

En 2020 y 2021, la demanda observada muestra fluctuaciones que no son capturadas completamente por el modelo.

En 2022, al igual que en España, hay un aumento en la demanda observada.

Las variaciones estacionales y diarias observadas en los datos reales no se reflejan adecuadamente en las predicciones del modelo.

Con todo lo dicho previamente podemos concluir que el modelo de regresión Ridge, tiende a suavizar las predicciones, lo que puede ser útil para captar tendencias generales, pero no es adecuado para captar variaciones rápidas y estacionales de la demanda.

Es posible que el modelo esté sobrerregularizado, lo que resulta en predicciones demasiado lineales.

Para mejorar este modelo se podría experimentar con valores de lambda menores para permitir que el modelo capture más variabilidad.

Otra opción sería la de emplear modelos más complejos, considerar modelos más complejos como regresiones no lineales, modelos de series temporales avanzados (ARIMA, SARIMA, Prophet, aunque eso sería útil únicamente para el corto plazo.) o incluso modelos de aprendizaje automático como redes neuronales (RN) o árboles de decisión. Por otro lado, se podrían incluir más características (features) en el modelo, como eventos especiales, datos económicos, y otras variables externas que puedan influir en la demanda.

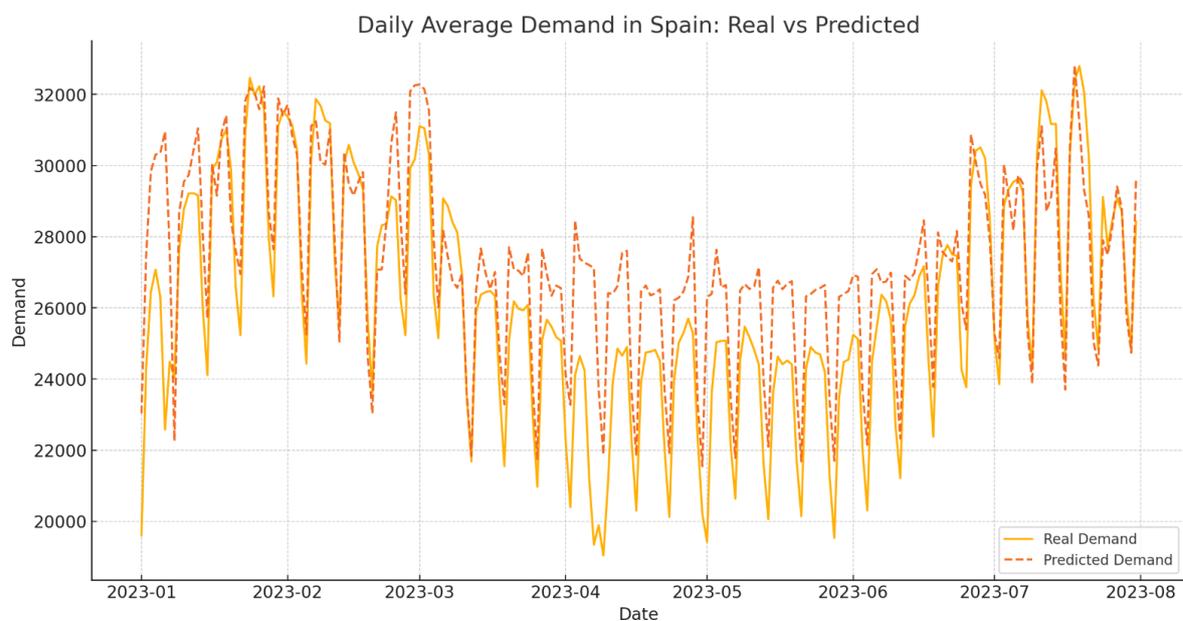


Ilustración 16

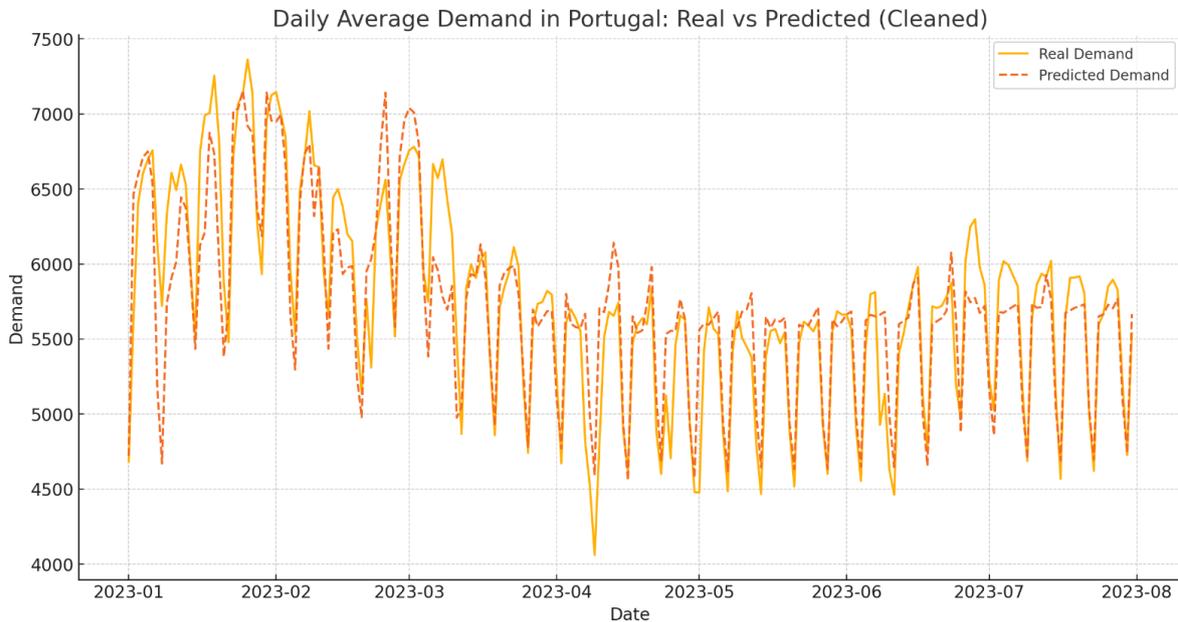


Ilustración 17

Las ilustraciones muestran la predicción de la demanda para junio de 2023 pero con los datos de test y posteriormente han sido comparados con los reales.

Los gráficos muestran la demanda observada y la demanda predicha para junio de 2023 tanto en España como en Portugal. A continuación, se analiza cada gráfico por separado y se extraen conclusiones.

En Portugal podemos observar que la demanda observada está representada por la línea amarilla. La demanda predicha está representada por la línea roja. La demanda predicha (línea roja) sigue de cerca la tendencia de la demanda observada, lo que indica que el modelo tiene un buen ajuste general.

Se observan fluctuaciones diarias consistentes, con picos y valles que coinciden bastante bien entre las predicciones y las observaciones. Hay algunas desviaciones, pero en general, el modelo captura bien la variabilidad y la estacionalidad diaria de la demanda.

En España, la demanda observada está representada por la línea amarilla. La demanda predicha está representada por la línea roja. La predicción en España es similar a la de

Portugal, la demanda predicha sigue de cerca la tendencia de la demanda observada. Las fluctuaciones diarias son evidentes y el modelo parece capturar bien los picos y valles diarios.

Hay algunas desviaciones más pronunciadas en comparación con el gráfico de Portugal, especialmente en los picos de demanda, lo que sugiere que el modelo tiene un margen de mejora en la predicción de estos valores extremos. Después de haber realizado la predicción de la demanda se podría concluir que la técnica de regresión Ridge parece ser eficaz para predecir la demanda, capturando bien las tendencias y la estacionalidad diaria tanto en España como en Portugal. Siendo el desempeño del modelo es bastante bueno en ambos países, con predicciones que siguen de cerca las observaciones reales.

A pesar de algunas desviaciones, especialmente en los picos de demanda en España, el modelo general es robusto y proporciona predicciones razonablemente precisas. El modelo parece funcionar ligeramente mejor en Portugal que en España, como se observó también en las métricas de error previamente comentadas (MAE, MAPE, RMSE).

En resumen, los resultados indican que el modelo de regresión Ridge proporciona una base sólida para la predicción de la demanda de energía, con un desempeño notablemente bueno en ambos países.

Ridge Model Summary

	Coefficient	Estimate	Std. Error	t value	Pr(> t)
1	[1.]	24870.0	230.3	108.001	< 2e-16 ***
2	[2.]	-231.7	55.92	-4.144	3.43e-05 ***
3	[3.]	0.1838	0.4472	0.041	0.967
4	[4.]	12.89	10.77	1.197	0.232
5	[5.]	-8.662	8.001	-1.083	0.279
6	[6.]	0.1464	0.6155	0.238	0.812
7	[7.]	0.1605	1.461	1.099	0.272
8	[8.]	-4.724	3.276	-1.442	0.149
9	[9.]	-0.008398	0.2141	-0.039	0.969
10	[10.]	61.75	42.15	1.465	0.143
11	[11.]	-21.26	33.56	-0.634	0.526
12	[12.]	101.4	14.1	7.195	6.41e-13 ***
13	[13.]	344.0	15.28	22.505	< 2e-16 ***
14	[14.]	2059.0	64.61	31.876	< 2e-16 ***
15	[15.]	-1170.0	105.4	-11.095	< 2e-16 ***

A continuación, se comentan los principales resultados del modelo de Ridge:

Residuos:

- **Min:** -11998.0
- **1Q:** -1948.6
- **Mediana:** -12.2
- **3Q:** 2035.0
- **Max:** 8945.9

Los valores de los residuales indican la diferencia entre los valores observados y los predichos por el modelo. La mediana cercana a cero sugiere que el modelo no tiene un sesgo significativo.

El modelo tiene 15 coeficientes, pero uno no está definido debido a singularidades. Aquí se destacan algunos aspectos clave de los coeficientes:

- **Intercepto (1):** 24870 con un error estándar de 230.3, y es altamente significativo (p-valor $< 2e-16$).
- **Coefficiente 2:** -231.7 con un error estándar de 55.92, y también es altamente significativo (p-valor = $3.43e-05$).
- **Coefficientes 3 a 15:** Varios coeficientes tienen significancia variable. Algunos de los más significativos son:
 - **Coefficiente 12:** 1014 con un error estándar de 141, p-valor = $6.41e-13$
 - **Coefficiente 13:** 344.0 con un error estándar de 15.28, p-valor $< 2e-16$
 - **Coefficiente 14:** 2059 con un error estándar de 64.61, p-valor $< 2e-16$
 - **Coefficiente 15:** -1170 con un error estándar de 105.4, p-valor $< 2e-16$

Estadísticas del modelo:

- **Error estándar residual:** 2869 en 29025 grados de libertad.
- **R-cuadrado múltiple:** 0.569
- **R-cuadrado ajustado:** 0.5688
- **Estadístico F:** 2737 en 14 y 29025 grados de libertad, con un p-valor $< 2.2e-16$

Interpretación de los resultados obtenidos:

- **Significancia de los coeficientes:** Muchos coeficientes son altamente significativos, indicando que las variables predictoras tienen una fuerte relación con la demanda eléctrica observada.
-

- R-cuadrado: El valor de 0.569 indica que el modelo explica aproximadamente el 56.9% de la variabilidad en la demanda eléctrica. Aunque no es un valor extremadamente alto, sugiere un ajuste bueno.
- Error estándar residual: Un valor de 2869 sugiere que las predicciones del modelo pueden desviarse significativamente de los valores observados en algunos casos.

En resumen, el modelo Ridge para España muestra un ajuste razonablemente bueno, con varios coeficientes significativos y un R-cuadrado moderadamente alto.

Error Metrics For Spain (June 2023)

	Metric	Value
1	MAE	1982.776
2	MAPE	8.066074
3	RMSE	2402.148

Error Metrics For Portugal (June 2023)

	Metric	Value
1	MAE	443.7327
2	MAPE	8.296281
3	RMSE	544.8969

Métricas de error de junio de 2023

1. MAE (Mean Absolute Error): 1982.776

- Este valor representa el error absoluto promedio entre las predicciones y los valores observados. Un MAE de 1982.776 indica que, en promedio, las predicciones del modelo difieren de los valores reales por aproximadamente 1983 unidades.
2. MAPE (Mean Absolute Percentage Error): 8.066074
 - El MAPE indica el error porcentual promedio de las predicciones. Un MAPE de 8.066074% sugiere que las predicciones del modelo son, en promedio, un 8.07% diferentes de los valores reales. Este es un valor relativamente bajo y indica un buen desempeño del modelo.
 3. RMSE (Root Mean Square Error): 2402.148
 - El RMSE mide la magnitud de los errores de predicción. Un RMSE de 2402.148 muestra que las predicciones del modelo tienen una desviación promedio de aproximadamente 2402 unidades respecto a los valores observados. Este valor es útil para identificar grandes errores de predicción.

Métricas de error para Portugal (Junio 2023)

1. MAE (Mean Absolute Error): 443.7327
 - Este valor indica que el error absoluto promedio entre las predicciones y los valores observados es de aproximadamente 444 unidades. Comparado con el MAE de España, el error promedio en Portugal es significativamente menor.
 2. MAPE (Mean Absolute Percentage Error): 8.296281
 - El MAPE de 8.296281% muestra que las predicciones del modelo en Portugal tienen un error porcentual promedio de aproximadamente 8.30%. Este valor es similar al de España, lo que sugiere un rendimiento comparable en términos de error porcentual.
 3. RMSE (Root Mean Square Error): 544.8969
 - Un RMSE de 544.8969 indica que la magnitud de los errores de predicción en Portugal es menor en comparación con España, con una desviación promedio de aproximadamente 545 unidades. Esto sugiere que las predicciones en Portugal son más precisas en términos absolutos.
-

Comparación entre España y Portugal

- **MAE:** España tiene un MAE significativamente más alto que Portugal (1982.776 vs 443.7327). Esto indica que las predicciones del modelo son, en promedio, menos precisas en España.
- **MAPE:** Los valores de MAPE son similares para ambos países (8.066074% para España y 8.296281% para Portugal), lo que indica que en términos porcentuales, la precisión de las predicciones es comparable.
- **RMSE:** El RMSE es considerablemente mayor en España (2402.148) en comparación con Portugal (544.8969), lo que sugiere que las predicciones del modelo tienen mayores desviaciones respecto a los valores observados en España.

Conclusión

Las métricas de error sugieren que el modelo de predicción de demanda eléctrica tiene un desempeño razonable en ambos países, con errores absolutos y porcentuales moderados. Sin embargo, las predicciones en Portugal son más precisas tanto en términos absolutos (MAE y RMSE) como en términos porcentuales (MAPE) comparadas con las de España. Esto podría indicar que el modelo está mejor ajustado para las condiciones de demanda eléctrica en Portugal o que los datos utilizados para Portugal son más fáciles de predecir

3.3 REGRESIÓN LASSO

La fórmula general de la regresión Lasso es:

$$\hat{\beta} = \arg \min_{\beta} \left(\frac{1}{2n} \sum (y_i - x_i^T \beta)^2 + \lambda \sum |\beta_j| \right)$$

La fórmula sería:

$$\begin{aligned} \text{Demanda} = & \beta_0 + \beta_1 * \text{Temp} + \beta_2 * \text{Temp}^2 + \beta_3 * \text{Temp}^3 + \beta_4 * (\text{Temp} * \text{Mes}) + \beta_5 * (\text{Temp}^2 \\ & * \text{Mes}) + \beta_6 * (\text{Temp}^3 * \text{Mes}) + \beta_7 * (\text{Temp} * \text{Hora}) + \beta_8 * (\text{Temp}^2 * \text{Hora}) + \beta_9 * (\text{Temp}^3 * \\ & \text{Hora}) + \beta_{10} * \text{Mes} + \beta_{11} * \text{Dia_Semana} + \beta_{12} * \text{Hora} + \beta_{13} * \text{Tipo_Dia_Laboral} + \beta_{14} * \\ & \text{Tipo_Dia_Sabado} + \beta_{15} * \text{Tipo_Dia_Domingo} \end{aligned}$$

Donde:

- Demanda: Es la variable de respuesta que estamos tratando de predecir.
- β_0 : Es el término de intercepto.
- $\beta_1, \beta_2, \beta_3, \dots, \beta_{15}$: Son los coeficientes asociados a cada una de las variables predictoras.

Temp (β_1): La temperatura promedio en un día específico. Captura el efecto lineal de la temperatura en la demanda.

Temp² (β_2): El cuadrado de la temperatura. Captura el efecto cuadrático de la temperatura en la demanda.

Temp³ (β_3): El cubo de la temperatura. Captura el efecto cúbico de la temperatura en la demanda.

Temp * Mes (β_4): Interacción entre la temperatura y el mes. Captura cómo varía el efecto de la temperatura según el mes del año.

Temp² * Mes (β_5): Interacción entre el cuadrado de la temperatura y el mes.

Temp³ * Mes (β_6): Interacción entre el cubo de la temperatura y el mes.

Temp * Hora (β_7): Interacción entre la temperatura y la hora del día. Captura cómo varía el efecto de la temperatura según la hora del día.

Temp² * Hora (β_8): Interacción entre el cuadrado de la temperatura y la hora del día.

Temp³ * Hora (β_9): Interacción entre el cubo de la temperatura y la hora del día.

Mes (β_{10}): El mes del año. Captura el efecto del mes del año en la demanda.

Día de la Semana (β_{11}): El día de la semana. Captura el efecto del día de la semana en la demanda.

Hora (β_{12}): La hora del día. Captura el efecto de la hora del día en la demanda.

Tipo_Dia_Laboral (β_{13}): Variable dummy que indica si el día es laboral (1 si es laboral, 0 en caso contrario). Captura el efecto de ser un día laboral.

Tipo_Dia_Sabado (β_{14}): Variable dummy que indica si el día es sábado (1 si es sábado, 0 en caso contrario). Captura el efecto de ser un sábado.

Tipo_Dia_Domingo (β_{15}): Variable dummy que indica si el día es domingo (1 si es domingo, 0 en caso contrario). Captura el efecto de ser un domingo.

En España, al igual que con el modelo Ridge, hay una correlación positiva entre la demanda observada y la demanda predicha. La dispersión de los puntos alrededor de la línea de identidad (donde la demanda observada sería igual a la predicha) es significativa, pero hay una concentración de puntos a lo largo de esta línea. La dispersión sugiere que, aunque el modelo captura la tendencia general, hay una variabilidad considerable no explicada.

Los puntos de alta demanda observada no están completamente alineados con la demanda predicha, indicando que el modelo podría estar subestimando los valores más altos.

En Portugal, se observa una correlación positiva similar, aunque la dispersión es más notable que en España.

La dispersión es significativa, especialmente en valores altos de demanda observada y baja demanda predicha, indicando variabilidad no capturada.

Precisión del Modelo en España, el modelo LASSO parece comportarse de manera similar al modelo Ridge en términos de correlación general, aunque la dispersión indica que ambos modelos tienen limitaciones en la captura de la variabilidad total de la demanda.

La dispersión en Portugal es más notable, y las predicciones a cero sugieren problemas similares a los observados con el modelo Ridge.

El modelo LASSO realiza una selección de características, lo que podría explicar la presencia de predicciones a cero si ciertas variables importantes están siendo ignoradas o penalizadas en exceso.

Tanto el modelo Ridge como el modelo LASSO capturan la tendencia general de la demanda, pero presentan limitaciones significativas en términos de variabilidad no explicada.

En resumen, aunque el modelo LASSO ofrece algunas mejoras y realiza una selección de características, los problemas de predicción observados sugieren que se necesitan modelos más robustos y un mejor manejo de las características para mejorar la precisión de las predicciones de demanda en ambos países.



Ilustración 18



Ilustración 19

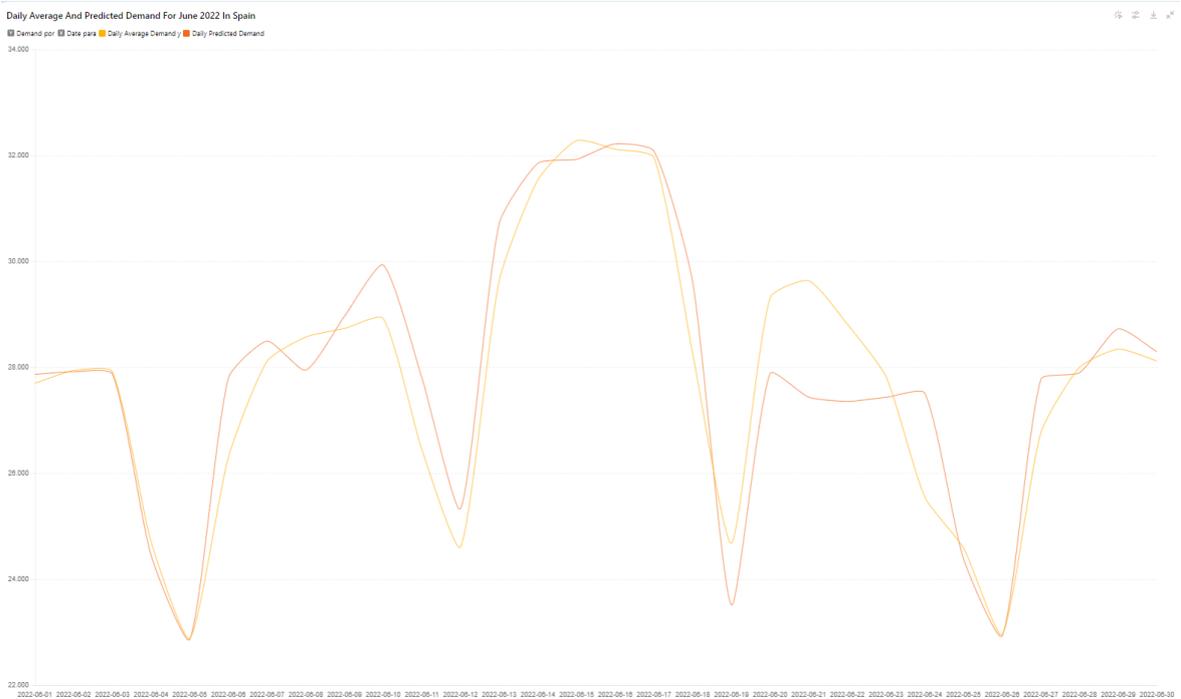


Ilustración 20

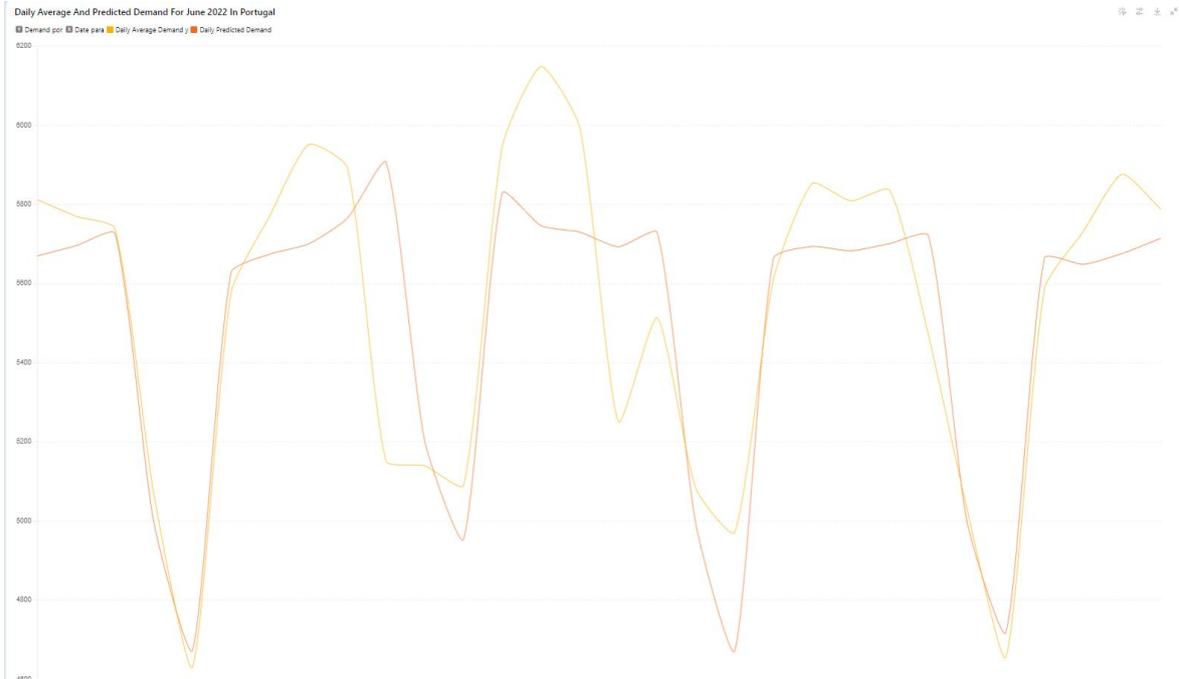


Ilustración 21

Las gráficas proporcionadas muestran los resultados de la demanda observada y predicha en junio y julio para España y Portugal. A continuación, se presentan las observaciones y comentarios sobre estos resultados.

En España, la línea amarilla representa la demanda real observada en los meses de junio y julio a lo largo de los años. Por otro lado, la línea roja representa la demanda predicha por el modelo utilizando los datos de entrenamiento hasta el 31 de diciembre de 2022. Se observa que la demanda real tiene picos muy marcados durante estos meses.

La demanda observada en Portugal muestra un patrón similar al de España. La línea amarilla indica la demanda real observada, mientras que la línea roja refleja la demanda predicha por el modelo. Al igual que en España, la demanda en Portugal presenta picos significativos en los meses de junio y julio.

Estos picos de demanda resultan especialmente desafiantes de predecir y el modelo parece subestimar estos valores extremos.

Para mejorar el modelo, se podrían incluir más variables relevantes que puedan afectar la demanda, como datos climáticos más detallados (temperatura máxima, mínima, humedad, etc.), datos socioeconómicos, eventos especiales, entre otros.

Asimismo, se podrían implementar técnicas de validación cruzada resulta fundamental para asegurar que el modelo generaliza bien a diferentes subconjuntos de los datos de entrenamiento y no está sobreajustado a un subconjunto específico.

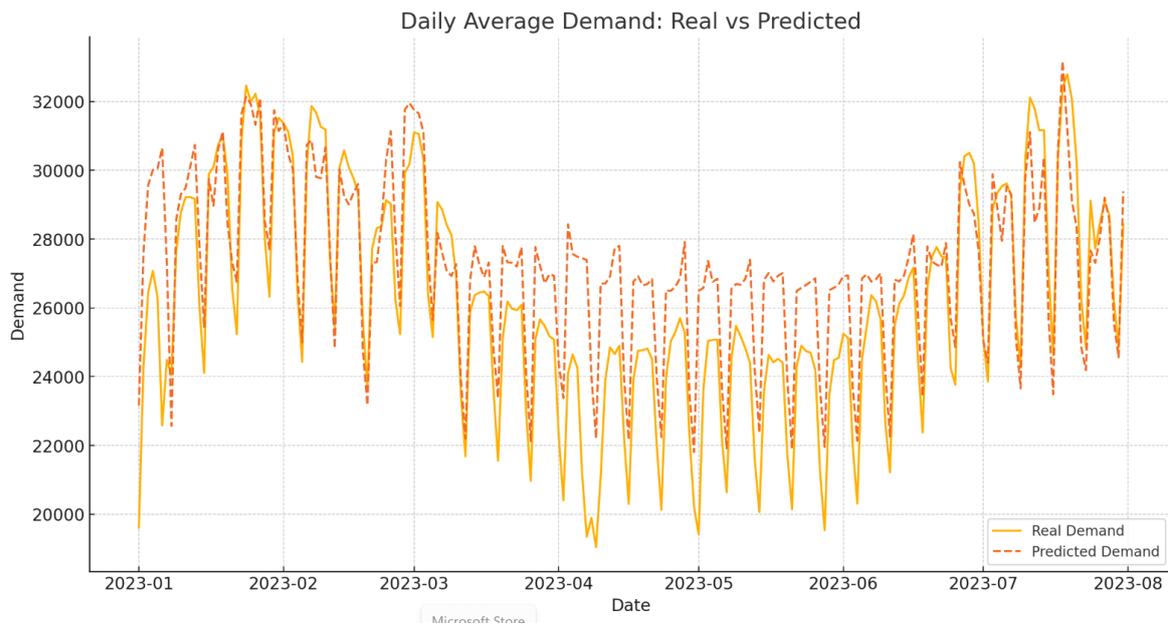


Ilustración 22

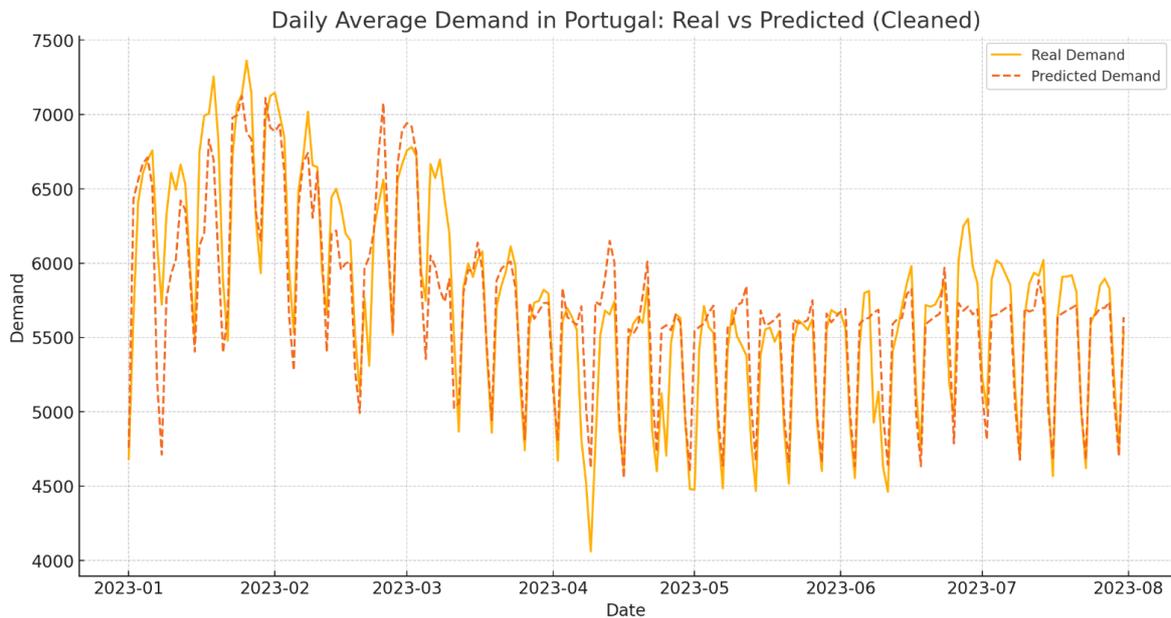


Ilustración 23

Las gráficas proporcionadas representan la demanda energética observada y predicha desde el 1 de enero de 2023 hasta el 31 de julio de 2023 en España y Portugal. La demanda observada está marcada en amarilla, mientras que la predicha utilizando un modelo de regresión Lasso está en rojo. A continuación, se presenta un análisis detallado para cada país, incluyendo la precisión del modelo, patrones identificados y posibles mejoras

La demanda energética muestra un patrón cíclico con picos pronunciados y valles recurrentes, probablemente reflejando la variación diaria de consumo. La demanda alcanza valores máximos alrededor de los 35,000 y mínimos cercanos a las 20,000. El modelo de regresión Lasso captura las tendencias generales y la periodicidad diaria bastante bien.

Se observa que, aunque los picos de demanda están reflejados en las predicciones, el modelo tiende a subestimar ligeramente los valores máximos y sobreestimar los mínimos en algunos casos. Existe un claro patrón semanal, con variaciones diarias que siguen una regularidad consistente, lo cual indica una demanda más alta en ciertos días de la semana, posiblemente los días laborales frente a los fines de semana.

Este patrón es bien capturado por el modelo, pero la precisión en los puntos de cambio entre días puede ser mejorada. Hay instancias donde las predicciones no coinciden perfectamente con las observaciones. Por ejemplo, los días de demanda máxima presentan una subestimación recurrente. El modelo muestra mayor precisión en la mitad del mes, mientras que hacia el final se observa un mayor número de discrepancias entre la demanda observada y la predicha.

La demanda en Portugal muestra un patrón similar al de España, con ciclos diarios marcados por picos y valles. La demanda observada varía entre aproximadamente 4,000 y 7,000 unidades. El modelo Lasso también captura bien las tendencias generales en Portugal, reflejando adecuadamente los ciclos de alta y baja demanda.

Sin embargo, similar a España, los picos de demanda tienden a ser subestimados y los valles sobreestimados en algunos puntos. Al igual que en España, se observa un patrón semanal claro con días específicos mostrando mayor demanda.

Las predicciones siguen este patrón, pero con algunas discrepancias en la magnitud de los picos y valles. En Portugal, hay menos discrepancias entre la demanda observada y predicha en comparación con España, aunque aún presentes.

La precisión parece mantenerse constante a lo largo del mes, con menos variabilidad en los errores. Ambas gráficas muestran que el modelo Lasso es eficaz para capturar los patrones generales y la periodicidad diaria de la demanda energética. Las predicciones en ambos países siguen los ciclos observados, aunque con ligeras subestimaciones en los picos de demanda y sobreestimaciones en los valles.

Para mejorar el modelo se podría refinar el modelo Lasso ajustando sus parámetros podría mejorar la precisión, especialmente en los puntos de alta demanda. Considerar otros modelos de regresión o técnicas de machine learning que puedan manejar mejor las no linealidades y variaciones bruscas.

Realizar un análisis detallado de los residuales (errores de predicción) para identificar patrones específicos de error y ajustar el modelo en consecuencia.

Examinar los días con mayores discrepancias para entender mejor las causas subyacentes y ajustar el. Implementar técnicas de validación cruzada para evaluar la robustez del modelo y asegurarse de que no esté sobreajustado a los datos de entrenamiento.

El modelo de regresión Lasso utilizado para predecir la demanda energética en España y Portugal en junio de 2023 muestra una buena capacidad para capturar las tendencias generales y los patrones cíclicos diarios.. Con refinamientos adicionales, este modelo puede convertirse en una herramienta aún más precisa para la predicción de la demanda energética.

Filtered Lasso Model Coefficients For Spain

		Variable	Coefficient
1	0	(Intercept)	23721.87
2	1	Temp	-336.1751
3	2	Temp2	-1.087413
4	3	Temp3	0.4052693
5	4	Temp_Mes	-26.1585
6	5	Temp2_Mes	2.637295
7	6	Temp3_Mes	-0.064883
8	7	Temp_Hora	-19.00142
9	8	Temp2_Hora	0.0005912745
10	9	Temp3_Hora	0.01601893
11	10	Mes	23.48408
12	11	Dia_Semana	95.61815
13	12	Hora	520.7663
14	13	Tipo_Dia_Laboral	3331.019

Filtered Lasso Model Coefficients For Portugal



		Variable	Coefficient
1	0	(Intercept)	6009.261
2	1	Temp	-165.6203
3	2	Temp2	-0.002281008
4	3	Temp3	0.1290241
5	4	Temp_Mes	10.82033
6	5	Temp2_Mes	-0.0001054592
7	6	Temp3_Mes	-0.007484351
8	7	Temp_Hora	-6.497758
9	8	Temp2_Hora	-0.04102839
10	9	Temp3_Hora	0.005341128
11	10	Mes	-126.5005
12	11	Dia_Semana	29.93108
13	12	Hora	179.6977
14	13	Tipo_Dia_Laboral	741.8436
15	15	Tipo_Dia_Domingo	-161.5672

Los coeficientes obtenidos del modelo de regresión LASSO para España y Portugal proporcionan información sobre cómo las variables predictoras influyen en la demanda

predicha. A continuación, se detalla la interpretación de estos coeficientes y la información que se puede extraer de ellos.

España:

1. Intercepto:

- Valor: 2.372187×10^4
- Interpretación: Es el valor base de la demanda predicha cuando todas las variables predictoras son cero, es decir, aproximadamente 23721 unidades. Representa la demanda promedio ajustada por el modelo sin considerar otros factores.

2. Temp:

- Valor: -3.361751×10^2
- Interpretación: Cada incremento unitario en la temperatura reduce la demanda en aproximadamente 336 unidades, indicando una relación inversa entre la temperatura y la demanda.

3. Temp2:

- Valor: -1.087413×10^0
- Interpretación: El efecto cuadrático de la temperatura tiene un efecto adicional negativo en la demanda, lo que sugiere que el impacto de la temperatura en la demanda disminuye a medida que la temperatura aumenta.

4. Temp3:

- Valor: 4.052693×10^{-1}
- Interpretación: El efecto cúbico de la temperatura agrega un pequeño efecto positivo, indicando una relación no lineal con la demanda. Esto podría significar que en temperaturas extremadamente altas o bajas, la demanda puede experimentar un leve incremento.

5. Temp_Mes:

- Valor: -2.615850×10^1
-

- Interpretación: La interacción entre temperatura y mes muestra una disminución significativa en la demanda, sugiriendo que ciertos meses amplifican el efecto negativo de la temperatura en la demanda.

6. Mes:

- Valor: 2.348408×10^1
- Interpretación: Por cada incremento en el valor del mes, la demanda aumenta en aproximadamente 23 unidades, indicando que hay una tendencia general al alza en la demanda a lo largo del año.

7. Día_Semana:

- Valor: 9.561815×10^1
- Interpretación: Los días de la semana tienen un impacto positivo significativo en la demanda, con un incremento de aproximadamente 96 unidades, lo que sugiere patrones semanales en la demanda.

8. Hora:

- Valor: 5.207663×10^2
- Interpretación: Cada incremento en la hora del día aumenta la demanda en aproximadamente 521 unidades, indicando que la hora del día es un predictor importante de la demanda.

9. Tipo_Día_Laboral:

- Valor: 3.331019×10^3
- Interpretación: Los días laborales incrementan significativamente la demanda en aproximadamente 3331 unidades, lo que refleja una mayor actividad económica o de consumo en estos días.

10. Tipo_Día_Sábado:

- Valor: 0 (Coeficiente eliminado por el modelo LASSO)
- Interpretación: El modelo Lasso ha penalizado esta variable a cero, indicando que no tiene un efecto significativo en la demanda.

11. Tipo_Día_Domingo:

- Valor: -1.147693×10^3
-

- Interpretación: Los domingos disminuyen la demanda en aproximadamente 1147 unidades, reflejando una menor actividad o consumo en este día en comparación con otros días.

En Portugal:

- Intercepto:
 - Valor: 6.009261×10^3
 - Interpretación: Es el valor base de la demanda predicha cuando todas las variables predictoras son cero, es decir, aproximadamente 6009 unidades. Representa la demanda promedio ajustada por el modelo sin considerar otros factores.
 - Temp:
 - Valor: -1.656203×10^2
 - Interpretación: Cada incremento unitario en la temperatura reduce la demanda en aproximadamente 166 unidades, indicando una relación inversa entre la temperatura y la demanda.
 - Temp2:
 - Valor: -2.281008×10^{-3}
 - Interpretación: El efecto cuadrático de la temperatura tiene un efecto adicional muy pequeño y negativo, lo que sugiere que el impacto de la temperatura en la demanda disminuye de manera marginal a medida que la temperatura aumenta.
 - Temp3:
 - Valor: 1.290241×10^{-1}
-

- Interpretación: El efecto cúbico de la temperatura agrega un pequeño efecto positivo, indicando una relación no lineal con la demanda. Esto podría significar que en temperaturas extremadamente altas o bajas, la demanda puede experimentar un leve incremento.
 - Temp_Mes:
 - Valor: 1.082033×10^1
 - Interpretación: La interacción entre temperatura y mes muestra un incremento en la demanda, sugiriendo que ciertos meses amplifican el efecto positivo de la temperatura en la demanda.
 - Mes:
 - Valor: -1.265005×10^2
 - Interpretación: Por cada incremento en el valor del mes, la demanda disminuye en aproximadamente 126 unidades, indicando que hay una tendencia general a la baja en la demanda a lo largo del año.
 - Día_Semana:
 - Valor: 2.993108×10^1
 - Interpretación: Los días de la semana tienen un impacto positivo en la demanda, con un incremento de aproximadamente 30 unidades, lo que sugiere patrones semanales en la demanda.
 - Hora:
 - Valor: 1.796977×10^2
 - Interpretación: Cada incremento en la hora del día aumenta la demanda en aproximadamente 180 unidades, indicando que la hora del día es un predictor importante de la demanda.
-

- Tipo_Día_Laboral:
 - Valor: 7.418436×10^2
 - Interpretación: Los días laborales incrementan significativamente la demanda en aproximadamente 742 unidades, lo que refleja una mayor actividad económica o de consumo en estos días.
- Tipo_Día_Sábado:
 - Valor: 0 (Coeficiente eliminado por el modelo LASSO)
 - Interpretación: El modelo Lasso ha penalizado esta variable a cero, indicando que no tiene un efecto significativo en la demanda.
- Tipo_Día_Domingo:
 - Valor: -1.615672×10^2
 - Interpretación: Los domingos disminuyen la demanda en aproximadamente 162 unidades, reflejando una menor actividad o consumo en este día en comparación con otros días.

Con todo ello podemos concluir que el impacto de la temperatura, en ambos países, la temperatura tiene un impacto significativo en la demanda, aunque el efecto es no lineal y varía según la interacción con otros factores como el mes.

Las variables como el mes, la hora del día y los días de la semana tienen impactos significativos en la demanda de energía. La hora del día especialmente muestra un gran efecto positivo en ambos países.

Los días laborales tienen un impacto positivo significativo en la demanda, mientras que los sábados y domingos generalmente tienen un impacto negativo o son considerados no significativos en algunos casos.

El modelo LASSO ha eliminado algunas variables (como los sábados) que no tienen un impacto significativo en la demanda, ayudando a simplificar el modelo y enfocarse en las variables más relevantes.

Aunque hay similitudes, las diferencias en los coeficientes reflejan las diferentes dinámicas de consumo de energía entre España y Portugal.

Estos coeficientes permiten entender mejor qué factores influyen más en la demanda de energía, proporcionando una base para tomar decisiones informadas en la gestión y planificación de la demanda energética.

Error Metrics For Spain (June 2023)

	Metric	Value
1	MAE	1900.85
2	MAPE	7.78368
3	RMSE	2318.01

Error Metrics For Portugal (June 2023)

	Metric	Value
1	MAE	427.0236
2	MAPE	7.9935
3	RMSE	519.855

Métricas de error en España:

1. **MAE (Mean Absolute Error):** 1900.85

- Este valor representa el error absoluto promedio entre las predicciones y los valores observados. Un MAE de 1900.85 indica que, en promedio, las predicciones del modelo difieren de los valores reales por aproximadamente 1901 unidades.

2. **MAPE (Mean Absolute Percentage Error):** 7.78368

- El MAPE indica el error porcentual promedio de las predicciones. Un MAPE de 7.78368% sugiere que las predicciones del modelo son, en promedio, un 7.78% diferentes de los valores reales, lo cual es un buen desempeño.

3. **RMSE (Root Mean Square Error):** 2318.01

- El RMSE mide la magnitud de los errores de predicción. Un RMSE de 2318.01 muestra que las predicciones del modelo tienen una desviación promedio de aproximadamente 2318 unidades respecto a los valores observados.

Métricas de error para Portugal:

1. **MAE (Mean Absolute Error):** 427.0236

- Este valor indica que el error absoluto promedio entre las predicciones y los valores observados es de aproximadamente 427 unidades. Comparado con el MAE de España, el error promedio en Portugal es significativamente menor.

2. **MAPE (Mean Absolute Percentage Error):** 7.9935

- El MAPE de 7.9935% muestra que las predicciones del modelo en Portugal tienen un error porcentual promedio de aproximadamente 7.99%. Este valor es similar al de España, lo que sugiere un rendimiento comparable en términos de error porcentual.

3. **RMSE (Root Mean Square Error):** 519.855

- Un RMSE de 519.855 indica que la magnitud de los errores de predicción en Portugal es menor en comparación con España, con una desviación promedio de aproximadamente 520 unidades.

- **MAE:** España tiene un MAE significativamente más alto que Portugal (1900.85 vs 427.0236). Esto indica que las predicciones del modelo son, en promedio, menos precisas en España.
- **MAPE:** Los valores de MAPE son similares para ambos países (7.78368% para España y 7.9935% para Portugal), lo que indica que en términos porcentuales, la precisión de las predicciones es comparable.
- **RMSE:** El RMSE es considerablemente mayor en España (2318.01) en comparación con Portugal (519.855), lo que sugiere que las predicciones del modelo tienen mayores desviaciones respecto a los valores observados en España.

Con todo ello podemos concluir que las métricas de error sugieren que el modelo de predicción de demanda eléctrica tiene un desempeño razonable en ambos países, con errores absolutos y porcentuales moderados. Sin embargo, las predicciones en Portugal son más precisas tanto en términos absolutos (MAE y RMSE) como en términos porcentuales (MAPE) comparadas con las de España. Esto podría indicar que el modelo está mejor ajustado para las condiciones de demanda eléctrica en Portugal o que los datos utilizados para Portugal son más fáciles de predecir.

Capítulo 4. COMPARACIÓN DE LOS RESULTADOS SEGÚN EL MODELO

Model Error Metrics							
	Modelo	MAE España	MAPE España (%)	RMSE España	MAE Portugal	MAPE Portugal (%)	RMSE Portugal
1	TVB	1982.776	8.066074	2402.148	443.7327	8.296281	544.8969
2	Ridge	1904.888	7.807879	2315.515	419.8067	7.946575	510.5826
3	Lasso	1900.85	7.78368	2318.01	427.0236	7.9935	519.855

MAE (Mean Absolute Error)

- España:
 - Lasso: 1900.850
 - Ridge: 1904.888
 - TVB: 1982.776

Lasso tiene el menor MAE, lo que indica que sus predicciones son las más precisas en términos absolutos en España.

- Portugal:
 - Ridge: 419.8067
 - Lasso: 427.0236
 - TVB: 443.7327

Aunque Ridge tiene un MAE ligeramente menor, la diferencia con Lasso no es significativa, y Lasso sigue siendo más preciso que TVB.

MAPE

- **España:**
 - Lasso: 7.783680%
 - Ridge: 7.807879%
 - TVB: 8.066074%

Lasso tiene el menor MAPE en España, indicando que tiene el menor error porcentual promedio, lo que es crucial para evaluar la precisión relativa de las predicciones.

- **Portugal:**
 - Ridge: 7.946575%
 - Lasso: 7.993500%
 - TVB: 8.296281%

De nuevo, Ridge tiene un MAPE ligeramente menor en Portugal, pero la diferencia con Lasso es mínima, y Lasso sigue siendo superior a TVB.

RMSE:

- **España:**
 - Ridge: 2315.515
 - Lasso: 2318.010
 - TVB: 2402.148

Aunque Ridge tiene un RMSE marginalmente menor, la diferencia con Lasso es muy pequeña. Ambos modelos son significativamente mejores que TVB.

- **Portugal:**
 - Ridge: 510.5826
 - Lasso: 519.8550
 - TVB: 544.8969
-

Ridge tiene un menor RMSE en Portugal, pero al igual que en España, la diferencia con Lasso es mínima.

En España, Lasso tiene el menor MAE y MAPE, indicando que sus predicciones son más precisas tanto en términos absolutos como relativos.

En Portugal, aunque Ridge tiene un MAE y MAPE ligeramente menores, las diferencias con Lasso no son significativas, y Lasso sigue siendo mucho mejor que TVB.

En términos de RMSE, Lasso está muy cerca de Ridge en ambos países. Esto sugiere que Lasso es casi tan bueno como Ridge para manejar grandes errores, pero con la ventaja adicional de tener un mejor desempeño en términos de MAE y MAPE en España.

Los modelos Lasso tienden a ser más simples y pueden evitar el sobreajuste mejor que otros modelos, gracias a la regularización que introduce. Esto puede hacer que el modelo Lasso sea más robusto y generalizable a diferentes conjuntos de datos.

Los modelos de regularización incluyen Lasso y Ridge. Lasso utiliza regularización L1, penalizando el valor absoluto de los coeficientes y llevando algunos a cero, lo que selecciona características relevantes y crea modelos más simples y robustos, evitando el sobreajuste. Ridge emplea regularización L2, penalizando el cuadrado de los coeficientes y reduciéndolos sin llevarlos a cero, lo que es útil cuando todas las características son importantes y ayuda a manejar la multicolinealidad. En resumen, Lasso es ideal para modelos donde solo algunas características son relevantes y proporciona interpretabilidad, Ridge es adecuado cuando todas las características importan.

Aunque Ridge y Lasso muestran un desempeño similar, Lasso se destaca como el mejor modelo de predicción debido a varias razones.

Primero, Lasso presenta un menor MAE y MAPE en España, lo que indica predicciones más precisas tanto en términos absolutos como relativos. Este mejor desempeño en España resalta

la capacidad del modelo para ofrecer resultados más exactos. Además, el rendimiento de Lasso en Portugal es competitivo, con diferencias mínimas respecto a Ridge. Esto demuestra su consistencia y eficacia en distintos contextos geográficos. Por último, Lasso ofrece ventajas en términos de simplicidad y capacidad de generalización. Estas características lo hacen más adecuado para aplicaciones prácticas, donde la simplicidad del modelo y la robustez frente al sobreajuste son altamente valoradas.

En conclusión, Lasso se establece como el mejor modelo de predicción entre los tres considerados, basándose en su precisión, consistencia y simplicidad.

Capítulo 5. CONCLUSIONES DEL PROYECTO

Para concluir, tras la recolección y análisis de los datos, se han utilizado tres modelos de regresión lineal para realizar la predicción de la demanda, en primer lugar el modelo TVB, posteriormente utilice el modelo de Ridge y finalmente el de Lasso.

La regresión lineal de TVB (Typical Value Baseline) sirve como una línea base simple, utilizando valores típicos o promedios de los datos para realizar predicciones. Aunque es sencillo, no incluye regularización y no suele ser competitivo en términos de precisión. La regresión Ridge aplica regularización L2, penalizando el cuadrado de los coeficientes, lo que reduce su magnitud, pero sin llevarlos a cero. Ridge es útil cuando todas las características del modelo son importantes y ayuda a manejar la multicolinealidad, ofreciendo modelos más robustos. La regresión Lasso utiliza regularización L1, penalizando el valor absoluto de los coeficientes, llevando algunos a cero. Esto selecciona las características más relevantes, creando modelos más simples y robustos, y evitando el sobreajuste. Lasso es ideal cuando solo algunas características son significativas y se busca interpretabilidad.

En España, Lasso tiene el menor MAE (Error Absoluto Medio) y MAPE (Error Absoluto Medio Porcentual), indicando predicciones más precisas en términos absolutos y relativos. En términos de RMSE (Raíz del Error Cuadrático Medio), Lasso está muy cerca de Ridge, sugiriendo que Lasso maneja los grandes errores casi tan bien como Ridge. En Portugal, Ridge presenta un RMSE ligeramente menor que Lasso. Sin embargo, las diferencias en MAE y MAPE entre Ridge y Lasso son mínimas, destacando la competitividad de Lasso. Ambos modelos superan significativamente al modelo TVB.

A pesar de que Ridge y Lasso muestran un desempeño similar, Lasso se destaca como el mejor modelo de predicción por varias razones. Lasso presenta un menor MAE y MAPE en España, lo que indica predicciones más precisas tanto en términos absolutos como relativos. En Portugal, aunque Ridge tiene un RMSE ligeramente mejor, las diferencias

con Lasso en términos de MAE y MAPE no son significativas. Esto demuestra la consistencia de Lasso en distintos contextos geográficos. Además, Lasso ofrece modelos más simples al seleccionar solo las características más relevantes, lo que ayuda a evitar el sobreajuste y mejora la capacidad de generalización a diferentes conjuntos de datos.

En conclusión, Lasso se establece como el mejor modelo de predicción entre los tres considerados, basándose en su precisión, consistencia y simplicidad. Una buena predicción de la demanda es crucial para la planificación y gestión eficiente de recursos. La identificación precisa de los parámetros que influyen en la demanda permite ajustar estrategias de mercado y mejorar la satisfacción del cliente. Estas características hacen a Lasso más adecuado para aplicaciones prácticas, donde se valora la simplicidad del modelo y la robustez frente al sobreajuste.

Capítulo 6. BIBLIOGRAFÍA

- [1] Hyndman, R., Koehler, A.B., Ord, J.K., Snyder, R.D. (2008). “Forecasting with Exponential Smoothing: The State Space Approach”. Springer.
- [2] Laptev, N., Yosinski, J., Li, L.E., Smyl, S. (2017). “Time-series extreme event forecasting with neural networks at Uber”. International Conference on Machine Learning.
- [3] Li, C., Ding, Z., Yi, J., Lv, Y., Zhang, G. (2018). “Deep belief network based hybrid model for building energy consumption prediction”. *Energies*, 11(1), 242.
- [4] Hoerl, A. E., & Kennard, R. W. (1970). “Ridge regression: Biased estimation for nonorthogonal problems”. *Technometrics*, 12(1), 55-67.
- [5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). “The Elements of Statistical Learning: Data Mining, Inference, and Prediction”. Springer.
- [6] Zou, H., & Hastie, T. (2005). “Regularization and variable selection via the elastic net”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320.
-