



TRABAJO FIN DE MÁSTER

Aplicación de Machine Learning a la eficiencia energética en edificios. Caso práctico: edificio ETOPIA.

Implementation of Machine Learning to energy efficiency in buildings.
Case study: ETOPIA building.

Autor:

Elena Marcuello Van der Auwera

Director:

Carlos Monné Bailo

Máster en Ingeniería Industrial

Escuela de Ingeniería y Arquitectura

2023

RESUMEN

En un mundo donde la Inteligencia Artificial cobra cada vez más importancia, es fundamental su aplicación consciente, en beneficio de la humanidad y del medio ambiente. El Aprendizaje Automático es una de las disciplinas de este campo, por la cual se confiere a las máquinas la capacidad de identificar patrones en grandes conjuntos de datos y realizar predicciones. En este proyecto, se ha abordado un enfoque relacionado con la eficiencia energética. Para ello, se elige el edificio ETOPIA como caso de estudio, aprovechando los datos disponibles del mismo.

El aprendizaje en esta disciplina se materializa a través de un curso de la Universidad de Stanford. Posteriormente, se aplica este conocimiento al caso específico. Se recopilan y ordenan los datos relevantes del edificio y las variables climáticas pertinentes. Se lleva a cabo un cribado de datos, eliminando los que puedan afectar al rendimiento de los modelos. La implementación del modelo se divide en dos etapas: el comportamiento térmico y energético del edificio.

En primer lugar, se analiza el comportamiento térmico del edificio mediante modelos de regresión, que relacionan las variables climáticas externas con las temperaturas interiores. Se estudian diversos modelos y la elección final es un modelo de regresión lineal múltiple, gracias a su alto rendimiento y bajo coste computacional. No obstante, se identifican ciertos resultados que sugieren posibles mejoras en los sensores implantados y variables a considerar.

En la segunda etapa, se aborda el comportamiento energético del edificio, estableciendo modelos de regresión basados en las temperaturas interiores y el uso de los sistemas de climatización. Tras el análisis de varios modelos de regresión, estos demuestran que no se pueden establecer relaciones entre las variables. No obstante, la introducción de una nueva variable relacionada con la ocupación del edificio mejora de manera significativa los resultados.

Sin embargo, se encuentran limitaciones en ambos modelos propuestos. Respecto al comportamiento térmico, la falta de información climatológica específica resulta en la propuesta de implantación de una estación meteorológica en el lugar. En cuanto al comportamiento energético, la ausencia de variables para determinar el uso de los sistemas de climatización se resuelve con la introducción de una nueva variable de ocupación, que deberá ser incluida en la recopilación de datos en el edificio para estudiar la evolución de este modelo.

Las recomendaciones finales se basan en la implementación de estas mejoras, para poder validar posteriormente los algoritmos con datos actualizados. La fase final implicaría la aplicación de modelos mediante Machine Learning Operations, creando una aplicación que prediga variables objetivo basándose en las predicciones meteorológicas.

INDICE

1. INTRODUCCIÓN	6
1.1. CONTEXTO Y JUSTIFICACIÓN.....	6
1.2. OBJETIVOS DEL ESTUDIO	7
2. EFICIENCIA ENERGÉTICA Y MACHINE LEARNING.....	8
2.1. EFICIENCIA ENERGÉTICA EN EDIFICIOS: CONCEPTOS Y DESAFÍOS	8
2.2. APLICACIÓN DE MACHINE LEARNING EN LA EFICIENCIA ENERGÉTICA DE EDIFICIOS	9
3. METODOLOGÍA DE ESTUDIO.....	10
3.1. ENFOQUE DEL PROYECTO.....	10
3.2. SELECCIÓN DEL CASO DE ESTUDIO: EDIFICIO ETOPIA	11
3.3. RECOPILACIÓN DE DATOS Y FUENTES DE INFORMACIÓN	13
4. RECOPILACIÓN DE DATOS Y PREPROCESAMIENTO.....	16
4.1. DESCRIPCIÓN DE LOS DATOS RECOPILADOS	16
4.1.1 DATOS RECOGIDOS EN EL EDIFICIO ETOPIA.....	16
4.1.2. DATOS CLIMÁTICOS DE LA CIUDAD DE ZARAGOZA.....	19
4.2. ANÁLISIS Y PREPROCESAMIENTO DE DATOS	21
5. DESARROLLO DEL MODELO DE MACHINE LEARNING	24
5.1. PARÁMETROS EN LOS MODELOS DE MACHINE LEARNING	24
5.1.1. VARIABLES ATRIBUTO Y VARIABLES OBJETIVO.....	24
5.1.2. HIPERPARÁMETROS.....	24
5.1.3. AJUSTE DEL MODELO.....	24
5.1.4. SUBCONJUNTOS DE ENTRENAMIENTO.....	25
5.1.5. MÉTRICAS DE RENDIMIENTO.....	26
5.2. SELECCIÓN DE ALGORITMOS	27
5.3. IMPLEMENTACIÓN Y COMPARACIÓN DE LOS MODELOS TÉRMICOS.....	28
5.3.1. REGRESIÓN LINEAL MÚLTIPLE SOBRE EL CONJUNTO DE PRUEBA.....	30
5.4. IMPLEMENTACIÓN Y COMPARACIÓN DE LOS MODELOS ENERGÉTICOS	31
6. CONCLUSIONES Y RECOMENDACIONES	36
6.1 ANÁLISIS DE LOS PRINCIPALES RESULTADOS Y CONCLUSIONES.....	36
6.2. LIMITACIONES DEL ESTUDIO Y RECOMENDACIONES	37
LISTADO DE ACRÓNIMOS	39
REFERENCIAS BIBLIOGRÁFICAS.....	40

ANEXOS

ANEXO I. PREPROCESAMIENTO DE LOS DATOS DEL EDIFICIO ETOPÍA	42
ANEXO II. PREPROCESAMIENTO DE LOS DATOS CLIMÁTICOS DE LA CIUDAD DE ZARAGOZA .	51
ANEXO III. CÓDIGO DE PYTHON PARA EL PREPROCESAMIENTO DE LOS DATOS.....	54
ANEXO IV. MODELOS DE REGRESIÓN APLICADOS	57
ANEXO V. RESULTADOS DE LOS MODELOS TÉRMICOS DE REGRESIÓN	62
ANEXO VI. RESULTADOS DE LOS MODELOS ENERGÉTICOS DE REGRESIÓN	69
ANEXO VII. CÓDIGO DE PYTHON PARA LOS MODELOS TÉRMICOS	81
ANEXO VIII. CÓDIGO DE PYTHON PARA LOS MODELOS ENERGÉTICOS	89

1. INTRODUCCIÓN

1.1. CONTEXTO Y JUSTIFICACIÓN

En la actualidad, la Inteligencia Artificial (IA), en particular el campo de Data Science y Machine Learning, está ganando cada vez más relevancia en diversos sectores de la sociedad. Las máquinas tienen el potencial para analizar grandes volúmenes de datos y reconocer patrones. Esta capacidad, aplicada en beneficio de los humanos y del planeta, proporciona una oportunidad única para mejorar nuestra comprensión y toma de decisiones en múltiples áreas, desde la medicina hasta la ingeniería.

En un mundo en que nos enfrentamos a diversos desafíos ambientales y económicos es de gran importancia optimizar el uso de la energía. En este contexto hablamos de la eficiencia energética: el uso ineficiente de la energía no solo tiene impactos negativos en el medio ambiente, sino también en la economía y la calidad de vida de las personas. Los edificios juegan un papel fundamental en el consumo de energía, por lo que optimizar este consumo debe ser una prioridad.

La aplicación de Machine Learning en la mejora de la eficiencia energética en edificios, como el caso que aborda este estudio para el Edificio ETOPIA, es relevante por varias razones que se vinculan a los Objetivos de Desarrollo Sostenible (ODS) y sus Metas [1]:

- **Potencial de ahorro energético:** Mediante el uso de algoritmos, es posible identificar patrones y relaciones en los datos de comportamiento térmico y consumo energético, estos patrones pueden revelar áreas de ineficiencia o formas de optimizar el uso de la energía, y resultar en un ahorro de recursos energéticos.

Esto se relaciona con el Objetivo 7 en los ODS: Energía asequible y no contaminante.

- **Reducción de emisiones de carbono:** Al aplicar Machine Learning, se espera poder desarrollar modelos predictivos y estrategias de optimización que contribuyan a la mejora de la eficiencia energética del edificio. Esto tiene un impacto directo en la reducción de emisiones de carbono y otros gases de efecto invernadero, por lo que contribuye a una mayor sostenibilidad medioambiental.
- **Validación de la aplicabilidad del Machine Learning:** Al seleccionar un caso práctico como el edificio ETOPIA, donde se dispone de un gran volumen de datos, se puede validar la aplicabilidad y efectividad del Machine Learning en el contexto del

comportamiento térmico en edificios y el uso de equipos eléctricos, relacionados con la eficiencia energética.

Estos dos últimos puntos se relacionan con el Objetivo 9 de los ODS: Industria, innovación e infraestructuras.

En resumen, la aplicación de Machine Learning en el contexto de eficiencia energética se justifica debido a su potencial para generar ahorros energéticos, reducir las emisiones de carbono, validar la efectividad de estas tecnologías y ofrecer beneficios sociales y económicos.

1.2. OBJETIVOS DEL ESTUDIO

En este Trabajo de Fin de Máster (TFM) se aborda el caso práctico del edificio ETOPIA [2]. Este edificio es un Centro de Artes y Tecnología que fue inaugurado en el año 2013. Cuenta con una superficie construida de 16000 metros cuadrados, de los cuales 13000 son útiles. Se trata de un espacio que alberga actividades que promueven la cultura contemporánea, espacios de trabajo y formación, y exhibe expresiones artísticas vanguardistas, así como exposiciones relacionadas con la tecnología y la divulgación científica [3].

ETOPIA se divide en tres bloques diferentes, conectados entre sí para formar el edificio completo. Dentro de estas tres construcciones hay diversas salas de exposiciones, una residencia, laboratorios (fabricación digital, biología, sonido, imagen, etc.), oficinas, aulas y un auditorio, entre otros [4].

Este edificio ha sido seleccionado como caso práctico para este proyecto gracias a la amplia cantidad de datos que se han recogido a lo largo de los años. Desde 2018, se recolectan datos de las diferentes áreas del edificio, obteniendo así un gran número de variables referentes a temperaturas exteriores e interiores, porcentaje de uso de los sistemas de frío y calor, así como otros datos de las instalaciones. Estos sistemas y variables registradas se explicarán detalladamente en el apartado 3. *Metodología de estudio*.

Los objetivos detallados de este proyecto son:

- Investigar y aplicar técnicas de Machine Learning para el análisis de datos relacionados con las condiciones térmicas y el consumo energético del edificio ETOPIA.
- Identificar patrones y tendencias en los datos proporcionados, que permitan crear modelos de aprendizaje automático para ser aplicados en el futuro.
- Contribuir al conocimiento científico y técnico en el uso de Machine Learning en aplicaciones de eficiencia energética, mediante las conclusiones derivadas de este TFM.

2. EFICIENCIA ENERGÉTICA Y MACHINE LEARNING

2.1. EFICIENCIA ENERGÉTICA EN EDIFICIOS: CONCEPTO Y DESAFÍOS

La eficiencia energética en edificios hace referencia a la utilización racional de energía para abastecer sus necesidades con relación a la climatización, tanto para calefacción como para refrigeración, así como para agua caliente sanitaria, iluminación, etc. El conjunto de las soluciones empleadas, tanto en edificios nuevos como antiguos, con el objetivo de alcanzar un consumo de energía casi nulo, contribuyen a la mejora de su eficiencia energética [5].

La eficiencia energética en los edificios se relaciona con el concepto de Passivhaus: este es un estándar para la construcción de edificaciones, nacido en Alemania, con el objetivo de que estas tengan un consumo muy bajo de energía para su climatización. Esto se consigue contando con buenos aislamientos térmicos, controlando la calidad del aire interior, y aprovechando la energía del sol para una mejor climatización. Estas medidas, en conjunto, consiguen un menor gasto energético, además de reducir las pérdidas y la contaminación ambiental [6].

En los últimos años, se ha potenciado la mejora de la eficiencia energética en edificios en la Unión Europea, por ejemplo, mediante iniciativas como la subvención de proyectos sostenibles de diferente índole con los fondos Next Generation EU, tanto en el sector público como en el sector privado [7].

No obstante, la eficiencia energética en edificios presenta desafíos. Es necesario garantizar una utilización racional de la energía y, al mismo tiempo, garantizar el confort de los habitantes de ese espacio.

En España, concretamente, entre todos los retos en relación con la eficiencia energética, destaca el aislamiento de los edificios como área con mayor margen de mejora. Casi un 50% de los edificios es anterior a 1980, pero solo el 0.3% de los edificios existentes ha realizado intervenciones de rehabilitación energética [8].

Además, otro desafío radica en el correcto entendimiento del concepto de eficiencia energética, todavía ambiguo para muchos ciudadanos. A menudo se relaciona principalmente con la instalación de sistemas de energía renovable, dejando de lado otros aspectos que son tan o incluso más relevantes. Esto ha provocado que se hayan centrado muchos esfuerzos e inversiones en esta materia, en lugar de intentar reducir el gasto energético de los edificios.

Es fundamental promover los sistemas eficientes y de bajo consumo, sin olvidar que todos los ciudadanos deben disfrutar de estas mejoras, independientemente de su situación económica.

2.2. APPLICACIÓN DE MACHINE LEARNING EN LA EFICIENCIA ENERGÉTICA DE EDIFICIOS

Machine Learning se traduce como Aprendizaje Automático o Aprendizaje de Máquina. Estos términos, en inglés y español, se utilizan de manera indistinta a lo largo de este documento.

Se trata de conseguir que un programa informático aprenda a mejorar su rendimiento al realizar una tarea por sí mismo, a través de la experiencia. Se desarrollan algoritmos y modelos estadísticos, que requieren de datos estables para procesarlos y aprender de ellos.

Tanto el aprendizaje automático como la eficiencia energética son temas a la orden del día. El aprendizaje automático abre un abanico de posibilidades para gestionar y analizar grandes volúmenes de datos para hacer predicciones futuras. La eficiencia energética, en un mundo en el que el consumo energético crece de manera constante, es uno de los aspectos clave en la transición hacia un futuro más sostenible. Si se conectan ambas disciplinas, se puede aprender optimizar y reducir el consumo energético y, por tanto, contribuir a reducir el impacto medioambiental [9].

En un edificio como el de este caso de estudio, al poseer los datos de comportamiento térmico y uso de sistemas de climatización en períodos anteriores, se puede crear un modelo que averigüe las tendencias y pautas y que prediga los futuros modelos de consumo de energía.

Esto presenta numerosas ventajas de cara al futuro, por ejemplo:

- Ventajas prácticas, ya que facilita el entendimiento de cómo es el consumo y por qué es así, permitiendo cambiar los hábitos sin afectar la productividad o la calidad de vida.
- Ventajas económicas, al ser capaces de estimar los consumos y con ello la factura de energía, y tomar decisiones basadas en estos datos.

3. METODOLOGÍA DE ESTUDIO

3.1. ENFOQUE DEL PROYECTO

El presente proyecto se enfoca en la aplicación y validación de técnicas de Machine Learning, aprendidas mediante el curso “Machine Learning Specialization”, que se traduce como “Especialización en Aprendizaje Automático” de la Universidad de Stanford. Dentro de este curso, este proyecto se enfoca en el área de “Supervised Machine Learning: Regression and Classification”, que se traduce como “Aprendizaje automático supervisado: Regresión y clasificación”. El objetivo principal es aplicar la información obtenida mediante la realización de este curso al caso práctico del edificio ETOPIA [10] [11].

Se sigue un enfoque que se denomina supervisado, esto significa que se utilizan datos etiquetados para entrenar un modelo y hacer predicciones sobre nuevos datos. Se proporciona al algoritmo un conjunto de ejemplos de entrenamiento con las respuestas correctas, y el modelo aprende a encontrar patrones en los datos para realizar predicciones precisas en base a características similares.

En concreto, las técnicas de regresión juegan un papel fundamental en este proyecto. Mediante diversos algoritmos, se analiza la relación entre variables de entrada, como datos climáticos del exterior del edificio, y la variable de salida, como la temperatura interior del edificio o el porcentaje de uso de sistemas energéticos. A través del análisis de los datos históricos con los que se cuenta, el modelo puede aprender patrones y tendencias, y así realizar predicciones sobre el comportamiento térmico y el consumo energético futuro.

Se realiza un análisis y preprocesamiento de los datos disponibles, y se aplican algoritmos de Machine Learning para extraer información relevante, estimar los datos futuros y validar las técnicas utilizadas. Al aprovechar las capacidades del Machine Learning, es posible comprender mejor qué factores influyen en el comportamiento térmico y usos energéticos en el edificio, lo que puede resultar en estrategias más efectivas para mejorar su eficiencia energética.

Se espera obtener resultados significativos en términos de identificación de patrones y tendencias en los datos, así como en la predicción del comportamiento térmico y del uso de los sistemas de climatización del edificio ETOPIA. Estos resultados permiten proponer recomendaciones específicas y medidas concretas que puedan implementarse para continuar el estudio y mejorar la eficiencia energética del edificio.

3.2. SELECCIÓN DEL CASO DE ESTUDIO: EDIFICIO ETOPIA

El caso de estudio de este proyecto es el edificio ETOPIA de la ciudad de Zaragoza. Este edificio ha sido seleccionado debido a la gran cantidad de datos sobre variables relacionadas con su comportamiento térmico y energético, por ejemplo, las temperaturas de diferentes áreas del edificio, que han sido registrados desde el año 2018 [12].



Ilustración 1. Fotografía del edificio ETOPIA. Fuente: MSM Arquitectos.

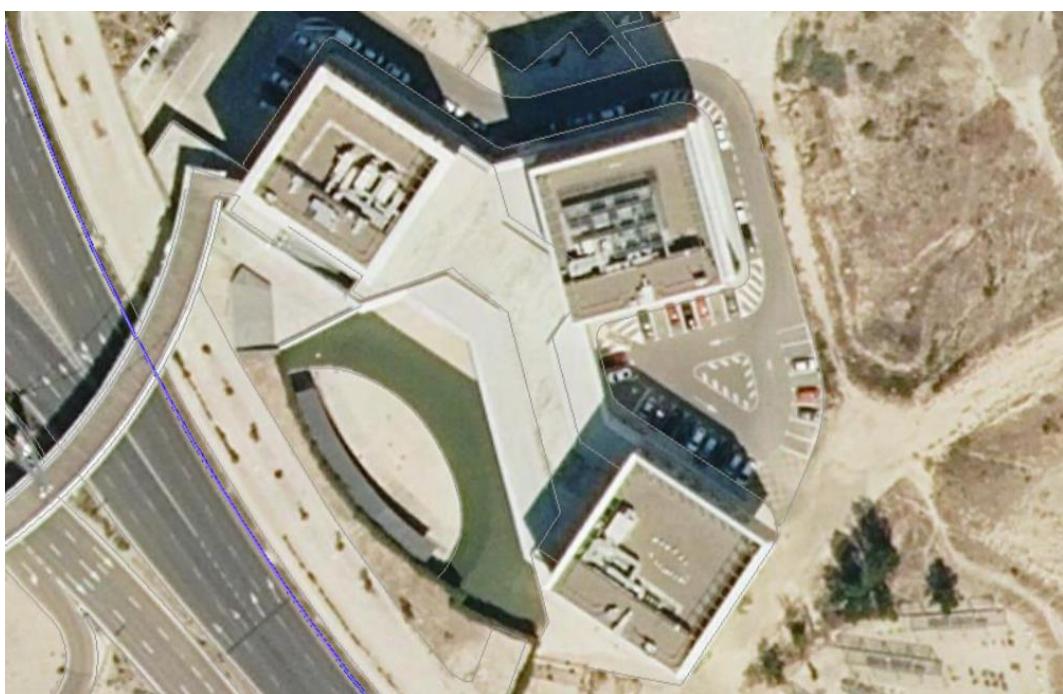


Ilustración 2. Fotografía aérea del edificio ETOPIA. Fuente: Catastro.

Como puede observarse en las imágenes superiores, este edificio se compone de tres bloques, conectados a través de la primera planta por la Galería Experimental, un corredor destinado a talleres individuales, encuentros y exposiciones temporales [13].

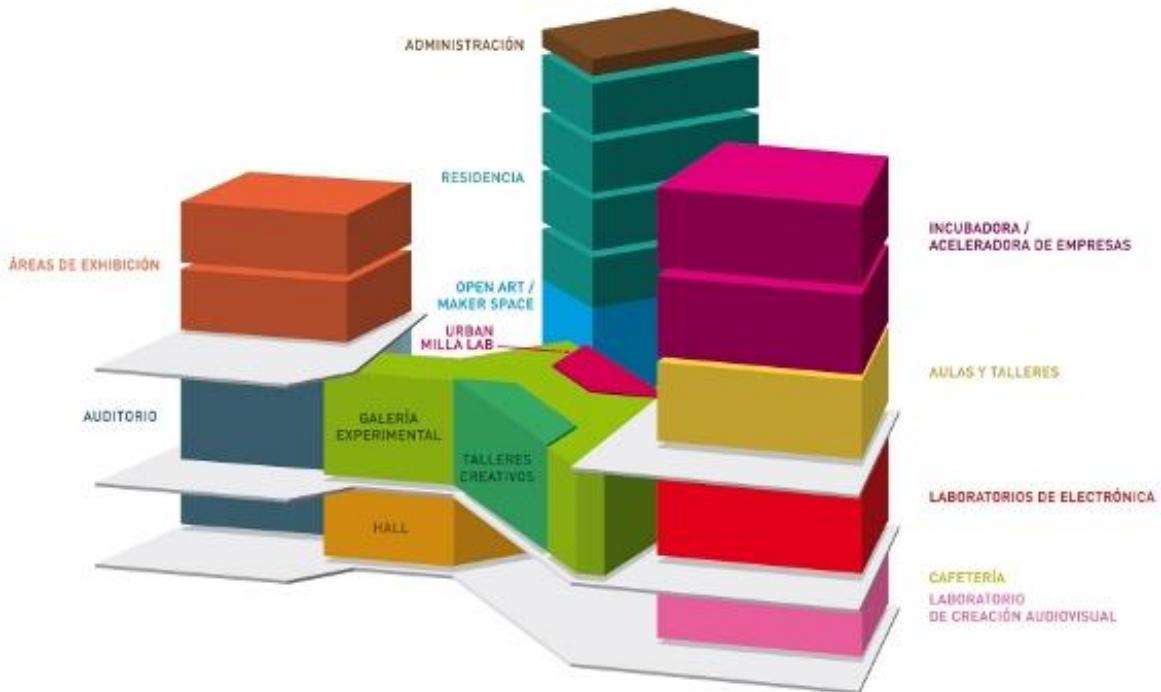


Ilustración 3. Esquema de bloques del edificio Etopía. Fuente: Impacto de las actividades artísticas de Etopía en la prensa digital aragonesa. Jéssica Blasco Quílez, 2015.

En este estudio, los bloques se han denominado como “edificios” y se explican a continuación, para entender en el transcurso de esta memoria acerca de qué zona se habla en cada caso:

- El bloque mostrado en la parte izquierda de la *Ilustración 3* se denomina **Edificio 1**.
- El bloque situado en el centro de la *Ilustración 3* se ha denominado **Edificio 2**.
- El bloque en la parte derecha de la *Ilustración 3* es el denominado **Edificio 3**.

A continuación, se describe cada uno de los bloques o edificios:

1. El **Edificio 1** está destinado a usos públicos. En este estudio se ha dividido en cuatro áreas separadas, cada una de las cuales cuenta con variables diferenciadas. Estas cuatro áreas son la entrada o hall, la planta baja que cuenta con el anfiteatro y el auditorio William J. Mitchell, la segunda y tercera planta que son áreas de exhibición de 400 m².
2. El **Edificio 2** aloja la residencia de artistas y la sección de administración. En este estudio se ha dividido en seis áreas separadas, cada una de las cuales también cuenta con variables diferenciadas. Estas seis áreas son la sala Open Art, la segunda planta, tercera,

cuarta y quinta plantas destinadas a la residencia y la sexta planta destinada a la administración.

3. El **Edificio 3** este edificio cuenta con oficinas, espacios de trabajo y laboratorios. Se ha dividido en cuatro áreas, cada una de las cuales se cuenta con sus variables. Estas cuatro áreas son la zona común, el primer aula, el segundo aula y el tercer aula.

Todas estas áreas diferenciadas cuentan con variables útiles para el estudio presente, tanto predictoras como objetivo, referentes a las temperaturas de las diferentes áreas y el uso de los equipos de climatización en cada una de estas zonas.

Las variables registradas se enumeran y explican en el apartado *4.1. Descripción de los datos recopilados*.

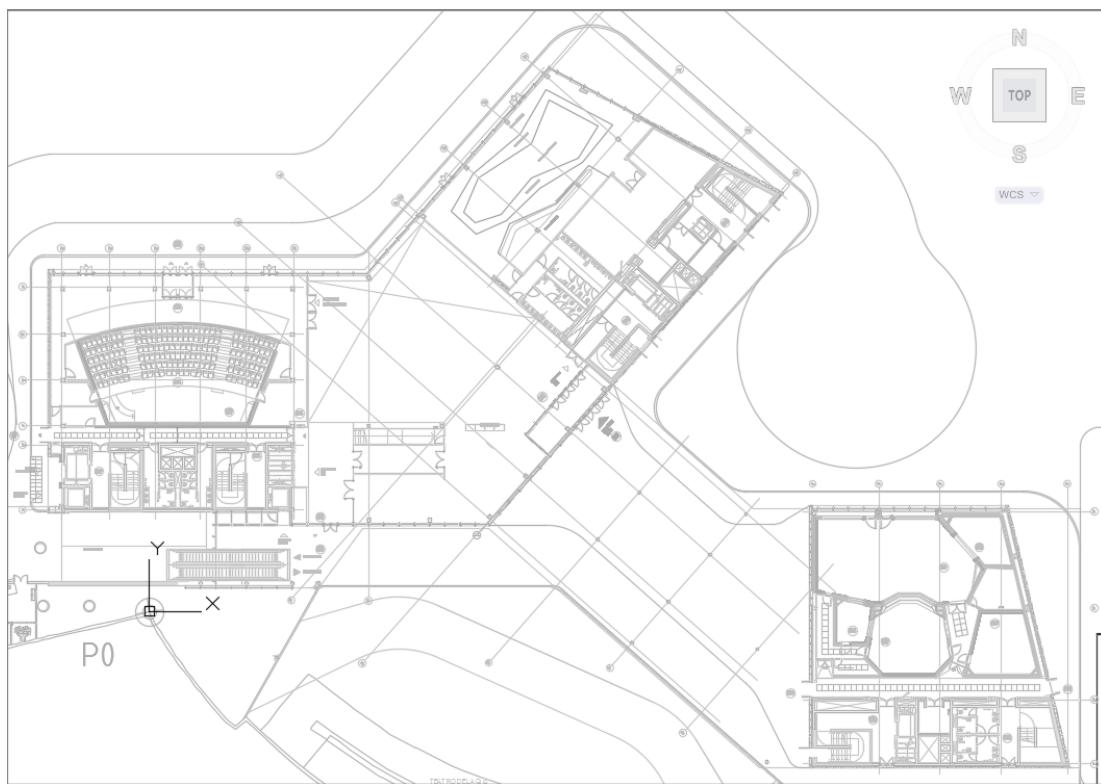


Ilustración 4. Plano Edificio ETOPIA. Fuente: Edificio ETOPIA.

3.3. RECOPILACIÓN DE DATOS Y FUENTES DE INFORMACIÓN

En un caso de estudio como el presente, resulta de gran importancia que las fuentes utilizadas para obtener información sean fiables y precisas, pues el algoritmo aprenderá de los datos recopilados, los cuales se consideran “respuestas correctas”.

En este caso, la principal fuente de información de datos es el propio edificio Etopía, ya que, tal como se ha mencionado, en este se dispone de un gran volumen de datos recopilado desde hace años para su posterior estudio [18].

No obstante, los datos climáticos exteriores son importantes en este estudio también. Los datos climáticos de la ciudad de Zaragoza tienen un papel fundamental en este estudio. Estos datos proporcionan información sobre las condiciones ambientales locales, como la temperatura, la humedad, la velocidad del viento, etc., que tienen un impacto directo en el comportamiento térmico del edificio y el uso de la climatización. Al relacionar estos datos con las variables registradas, se pueden identificar patrones y tendencias. Cuantos más datos y variables se tienen en consideración en el estudio, mejores estrategias se pueden implementar para optimizar la eficiencia energética y reducir el consumo, adaptando así las necesidades del edificio a las condiciones climáticas específicas de Zaragoza.

Sin embargo, en términos de datos climáticos de la ciudad de Zaragoza, en el edificio Etopía solo se recogen datos de la temperatura exterior, por lo que ha sido necesario recurrir a otras fuentes para obtener información sobre el clima en las mismas fechas en la ciudad de Zaragoza.

Es fundamental que estas fuentes sean oficiales y fiables para realizar un estudio preciso. Concretamente, se solicitaron estos datos al Departamento de Geografía y Ordenación del Territorio de la Universidad de Zaragoza. En esta facultad se dispone de una estación Davis, en la cual se miden a intervalos horarios variables relacionadas con la temperatura, el viento, las precipitaciones, la humedad, etcétera. Estos datos son de gran utilidad para su correlación con las variables de estudio del edificio Etopía [14].

Tal como se ha mencionado con anterioridad, los datos recogidos en el edificio Etopía presentan ciertos saltos temporales en los que no se tienen registro. A continuación, se muestra un gráfico ejemplo con algunas de las variables recogidas desde el año 2018 hasta la fecha para señalar estas discontinuidades.

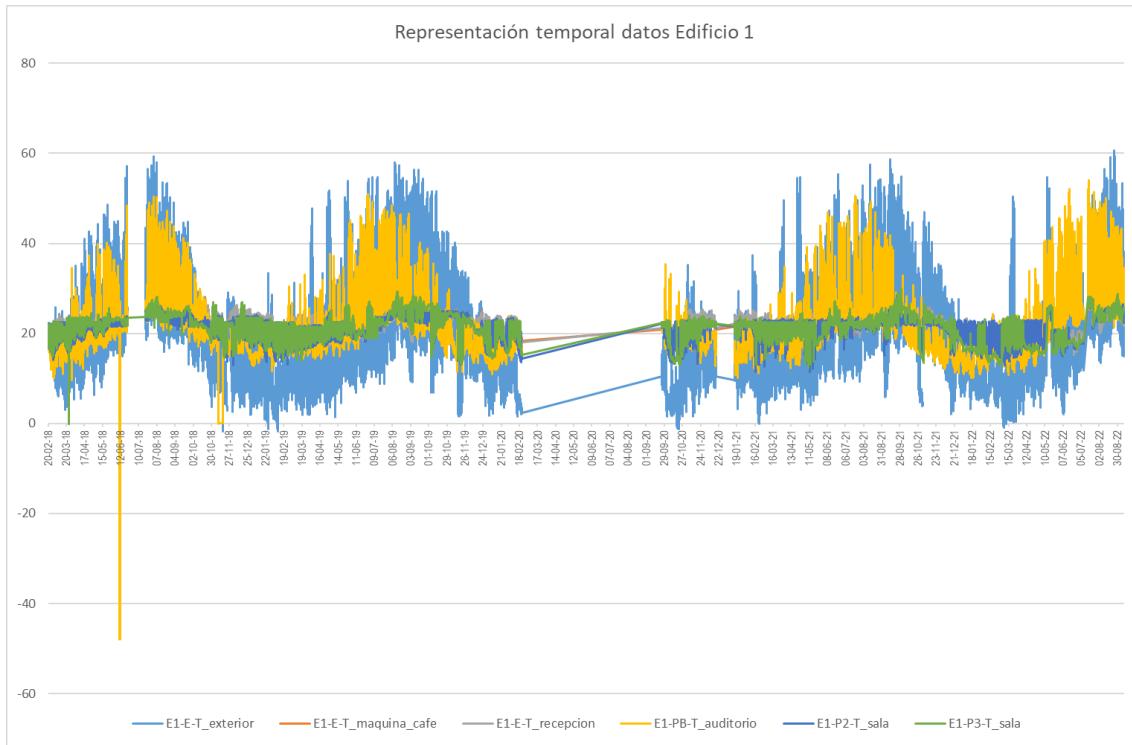


Ilustración 5. Gráfico de representación temporal de los datos del Edificio 1. Fuente: Elaboración propia.

La discontinuidad más evidente resulta en el año 2020, lo cual fue debido a la situación sobrevenida por la pandemia del COVID-19. No obstante, esta no es la única discontinuidad temporal existente, puede apreciarse otra en 2018 y en el inicio del año 2021.

Respecto a los datos recogidos en la facultad de Geografía de la Universidad de Zaragoza, no se dispone de estos desde la misma fecha que los del edificio Etopía, sino que existen registros desde enero de 2021.

Por tanto, debido la falta de datos meteorológicos y a los saltos temporales en la recolección de datos en el edificio Etopía, se ha decidido analizar los datos desde enero de 2021 hasta la fecha. Existe una gran cantidad de datos que se considera suficiente para que el algoritmo pueda trabajar correctamente, ya que se recogen en intervalos de quince minutos durante todo el periodo de tiempo considerado, casi tres años.

4. RECOPILACIÓN DE DATOS Y PREPROCESAMIENTO

4.1. DESCRIPCIÓN DE LOS DATOS RECOPILADOS

Como se ha explicado en los apartados previos, se han obtenido datos de dos fuentes diferentes. Por un lado, se tienen datos acerca del edificio Etopía, tanto de temperatura interior en las diferentes salas que lo conforman, como de utilización de los sistemas de calefacción y refrigeración del edificio. Por otro lado, se han obtenido los datos climáticos de la ciudad de Zaragoza por parte del Departamento de Geografía y Ordenación del Territorio de la Universidad de Zaragoza. Se explican a continuación las variables consideradas de cada una de las fuentes de información.

4.1.1 DATOS RECOGIDOS EN EL EDIFICIO ETOPIA

Como se explica en el punto 3.2. *Selección del caso de estudio: edificio ETOPIA*, el edificio consiste en tres bloques diferentes, que se han estudiado de manera independiente y dividido internamente por salas, de las cuales se detallan a continuación las variables obtenidas [18].

El **Edificio 1** se ha dividido en cuatro zonas separadas, con sus respectivas variables:

1. En la **ENTRADA** se han registrado:
 - a) Temperatura de la zona de la máquina de café, en grados centígrados [°C]
 - b) Temperatura de la zona de recepción [°C]
 - c) Temperatura de la zona del ascensor [°C]
 - d) Energía del suelo radiante, en porcentaje [%]: esta zona está climatizada mediante suelo radiante, una forma de calefacción basada en el transporte de agua por el interior de una serie de tubos embebidos en el suelo. Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema [15].



Ilustración 6. Fotografía de la entrada de Etopía. Fuente: MSM Arquitectos

2. En la **PLANTA BAJA** se han registrado:
- a) Temperaturas en el anfiteatro [°C]
 - b) Temperatura del auditorio [°C]
 - c) Energía de los sistemas de frío y calor del anfiteatro [%]: Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema.
 - d) Energía del sistema frío y calor del auditorio [%]: Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema.



Ilustración 7. Fotografía del auditorio William J. Mitchell. Fuente: Elaboración propia.

3. En la **SEGUNDA PLANTA** y en la **TERCERA PLANTA** se han registrado:
- a) Temperatura de la sala [°C]
 - b) Energía de los sistemas de frío y calor [%]: Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema.



Ilustración 8. Fotografía de una sala de exposiciones de Etopía. Fuente: Elaboración propia.

El **Edificio 2** se ha dividido en siete zonas separadas, con sus respectivas variables:

1. En la sala **OPEN ART** se han registrado:
 - a) Temperatura de la sala [°C]
 - b) Energía de los sistemas de frío y calor [%]: Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema.
2. En la **SEGUNDA PLANTA**, la **TERCERA PLANTA**, la **CUARTA PLANTA** y la **QUINTA PLANTA**, se han registrado:
 - a) Temperaturas en cada planta [°C]
 - b) Energía del suelo radiante, en porcentaje [%]: Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema.
 - c) Energía de tres sistemas Fancoil en cada planta [%]: El funcionamiento se basa en la transferencia de temperatura agua-aire, el aire que atraviesa el Fancoil absorbe energía en un intercambiador de calor mediante tuberías de agua. Los datos son en función de la apertura y cierre de la válvula que controla el sistema [16].



Ilustración 9. Fotografía residencia de artistas de Etopía. Fuente: Elaboración propia.

3. En la **SEXTA PLANTA** se han registrado:
 - a) Temperatura de la sala [°C]
 - b) Energía de los sistemas de frío y calor [%]: Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema.

El **Edificio 3** se ha dividido en cuatro áreas separadas, con las siguientes variables:

1. En la **ZONA COMÚN** las variables registradas han sido:
 - a) Temperaturas 1 y 2 de la sala [°C]

- b) Energía de los sistemas 1 y 2 de frío y calor [%]: Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema.



Ilustración 10. Fotografía laboratorio de Etopía. Fuente: Elaboración propia.

2. En las salas **AULA 1, AULA 2 y AULA 3** las variables registradas han sido:
- a) Temperatura de la sala [°C]
 - b) Energía de los sistemas de frío y calor [%]: Los datos recogidos son en función de la apertura y cierre de la válvula que controla el sistema.



Ilustración 11. Fotografía espacio de trabajo de Etopía. Fuente: Elaboración propia.

4.1.2 DATOS CLIMÁTICOS DE LA CIUDAD DE ZARAGOZA

En el apartado 3.3. *Recopilación de datos y fuentes de información*, se ha explicado que los datos climáticos de la ciudad de Zaragoza se obtuvieron gracias al Departamento de Geografía y Ordenación del Territorio de la Universidad de Zaragoza. En esta facultad, disponen de una

estación Davis que mide a intervalos horarios datos de temperatura, viento, precipitación, humedad, etcétera [14] [17].

Tanto los datos del edificio Etopía como los datos climáticos se recogen en intervalos de quince minutos. Las estaciones Davis como la presente pueden medir muchas variables meteorológicas, pero las presentadas a continuación son las que se han considerado de mayor relevancia de cara al estudio llevado a cabo en este proyecto.

1. **Temperatura:** Esta variable ha sido seleccionada ya que la temperatura exterior tiene un impacto significativo en la cantidad de energía necesaria para mantener en el interior de un edificio un nivel de confort adecuado. En temporadas frías, se requiere más energía para calentar el interior del edificio. En temporadas cálidas, se necesitará más energía para enfriarlo. Se han obtenido los datos de temperatura exterior en la ciudad en grados centígrados [°C]. Se tienen datos de temperatura máxima y mínima en el intervalo considerado, y de la temperatura media en el intervalo, que es la que se ha decidido considerar.
2. **Humedad:** La humedad relativa del aire también influye en el confort térmico, además de que puede afectar a la percepción de la temperatura. Se han obtenido los datos de humedad exterior relativa en la ciudad en porcentaje [%], es decir, de la cantidad de vapor de agua que hay en el aire respecto al máximo que podría haber cuando ese aire se satura.
3. **Viento:** Respecto a los datos obtenidos sobre el viento, la variable que considerar es la velocidad del viento en metros por segundo [m/s]. La velocidad del viento es un factor importante en la pérdida de calor y en el enfriamiento del edificio. En invierno, el viento puede aumentar la tasa de enfriamiento del edificio, lo que implica un uso mayor de los sistemas de calefacción para mantener una temperatura cómoda en el interior. Por el contrario, en verano, un viento suave puede ayudar a disipar el calor y reducir la carga de enfriamiento necesaria por parte del sistema de refrigeración.
4. **Precipitaciones:** Se han tomado los datos de cantidad de lluvia caída en el intervalo de tiempo considerado, en milímetros [mm]. Del mismo modo que sucede con la velocidad del viento, la cantidad de lluvia puede afectar la temperatura interior del edificio y el uso de los sistemas de climatización. En climas fríos y lluviosos, puede ser necesario un mayor uso de los sistemas de calefacción para mantener el confort, mientras que, en climas cálidos, la lluvia puede ayudar a reducir la necesidad de refrigeración.

4.2. ANÁLISIS Y PREPROCESAMIENTO DE DATOS

El preprocessamiento de los datos es una etapa inicial y fundamental para cualquier proyecto de Machine Learning. Para el presente estudio es de especial relevancia ya que, como se ha comentado en el punto anterior, los datos se han obtenido de fuentes diferentes y existen tanto picos incoherentes como saltos temporales. Esto hace que cobre gran importancia la fase de preparación y transformación de los datos “brutos”, para que sean adecuados y se puedan utilizar de manera efectiva.

Inicialmente, se realizó la recolección de datos del edificio ETOPIA, así como de los datos climáticos de la ciudad de Zaragoza, como se ha explicado anteriormente. Los datos recogidos se trasladaron a la herramienta de Excel, ya que el cribado inicial podía realizarse mediante este programa. Excel permite visualizar los datos mediante tablas y gráficos, lo que facilita la identificación inicial de patrones y tendencias, así como de comportamientos inusuales.

Los datos se ordenan cronológicamente desde 2018 hasta la fecha, en intervalos de quince minutos, lo cual permite tener una estructura temporal clara y facilitar el posterior análisis. Se realiza lo mismo con los datos meteorológicos obtenidos, desde 2021 hasta la fecha, los cuales también están recogidos en intervalos de quince minutos.

Tras analizar estos datos, se pueden observar discontinuidades y picos que corresponden a valores anómalos debido a errores de medición. Esto puede apreciarse en la *Ilustración 5. Gráfico de representación temporal de los datos del Edificio 1*. Además de encontrar anomalías en los datos del edificio Etopía, también se encuentran irregularidades en los datos climáticos recogidos, como se muestra en el siguiente ejemplo de la temperatura.

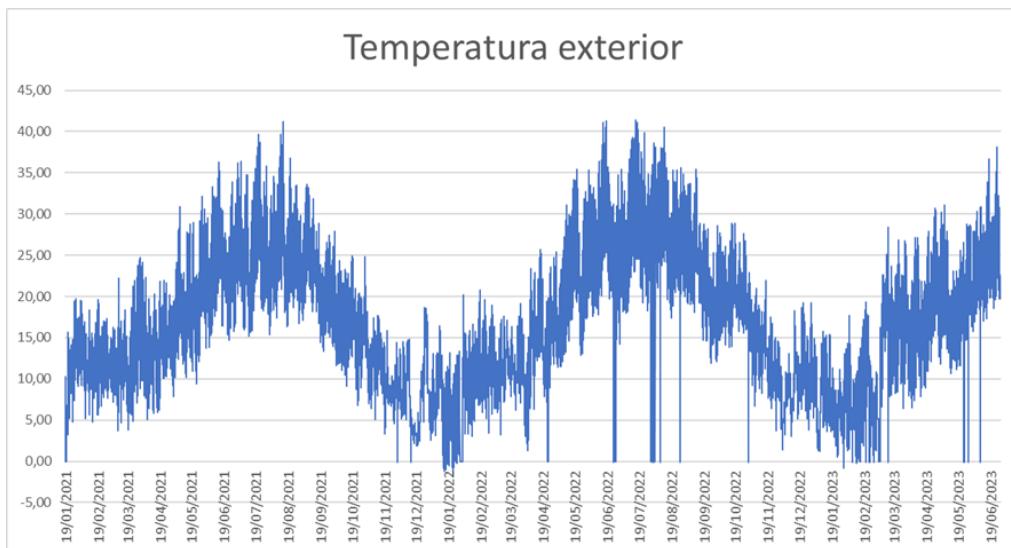


Ilustración 12. Gráfico temperatura exterior antes del preprocessamiento. Fuente: Elaboración propia.

Se toma la decisión de seleccionar los datos a partir del año 2021 para evitar aquellos intervalos con grandes discontinuidades, además de ajustarse a los datos climáticos disponibles. Esto asegura la coherencia temporal y reduce la presencia de datos que no resultan representativos.

Además, ha sido necesario limpiar los datos, lo cual implica eliminar valores inexistentes o erróneos, corrección de errores y normalización de los datos para asegurar que los datos se encuentrasen en una escala coherente y comparable.

Se identifican y eliminan aquellas filas que contuviesen valores nulos, para asegurar la integridad de los datos y evitar sesgos en el estudio.

Asimismo, se obtienen los valores máximos y mínimos para cada una de las variables, con el fin de revelar posibles picos atípicos. Estos valores extremos se revisan y, cuando sea necesario, se eliminan para asegurar la calidad y la fiabilidad de los datos.

A continuación, se muestra el ejemplo mostrado en la *Ilustración 5. Gráfico de representación temporal de los datos del Edificio 1*, en combinación con el gráfico de temperaturas exteriores expuesto anteriormente. En este caso se muestran esos datos una vez llevado a cabo el preprocesamiento, con lo que se ve un comportamiento normalizado de las variables.

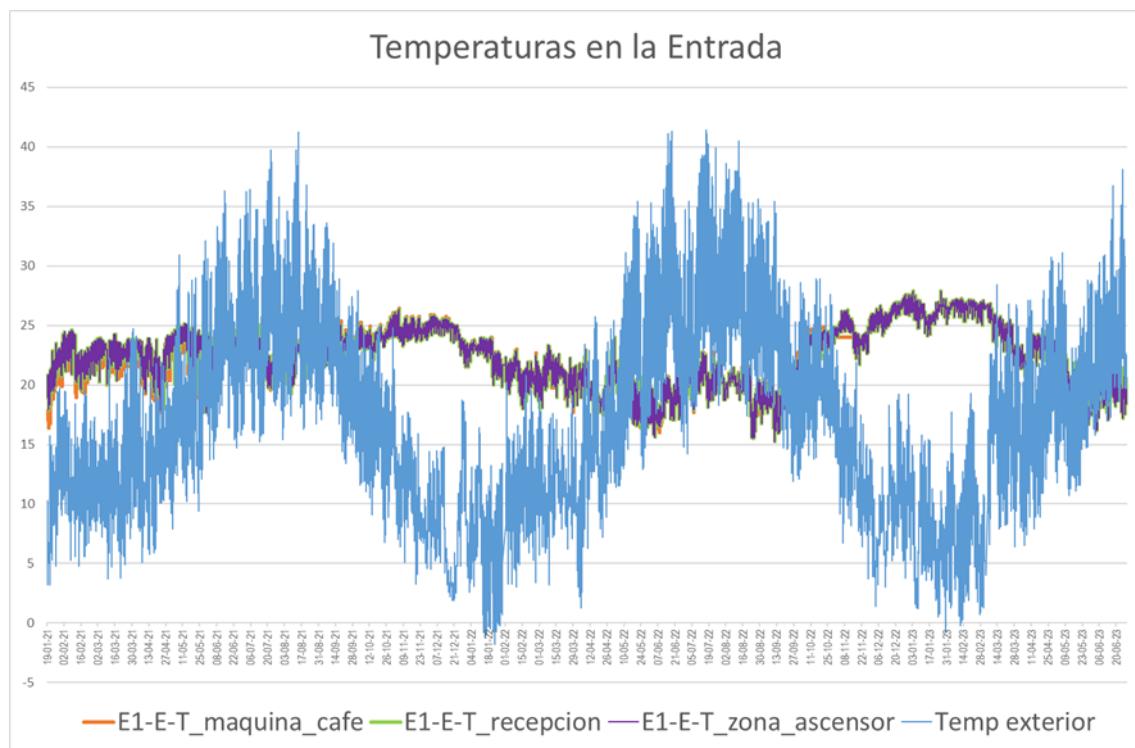


Ilustración 13. Gráfico temperaturas después del preprocesamiento. Fuente: Elaboración propia.

En los apartados *Anexo I. Preprocesamiento de los datos del edificio Etopía* y *Anexo II. Preprocesamiento de los datos climáticos de la ciudad de Zaragoza*, se explica con mayor detalle

el preprocessamiento llevado a cabo. Asimismo, se muestran gráficos y tablas explicativas de las variables del edificio que presentaban irregularidades antes y después de este preprocessamiento, así como de las variables climáticas consideradas en este estudio.

Por último, para poder poner en conjunto todos los datos climáticos junto con los datos térmicos y energéticos del edificio, se utilizó programación con Python mediante la herramienta Spyder. Este software y este lenguaje de programación son los utilizados también en la implementación del modelo de Machine Learning. En el *Anexo III. Código de Python para el preprocessamiento de los datos*, se encuentra el código utilizado para aunar los datos climáticos y del edificio.

5. DESARROLLO DEL MODELO DE MACHINE LEARNING

5.1. PARÁMETROS EN LOS MODELOS DE MACHINE LEARNING

A continuación, se procede a explicar las bases teóricas en modelos de aprendizaje automático, ya que estos términos se utilizan en el transcurso de este documento.

5.1.1. VARIABLES ATRIBUTO Y VARIABLES OBJETIVO

Las variables atributo o variables predictoras son, haciendo referencia a su nombre, aquellas que se utilizan para hacer predicciones. Estas variables son las de entrada del modelo y se utilizan para estimar o predecir el valor de otra variable. En este caso, las variables climáticas externas son las que van a definir el comportamiento del edificio, es decir, las variables de temperatura exterior, humedad exterior, viento o lluvia van a definir la temperatura que alcanza el edificio interiormente y por ende sus necesidades energéticas.

Por otro lado, se encuentran las variables objetivo, esto es, las variables que se intentan predecir o explicar con el modelo de aprendizaje automático. Son variables que se consideran dependientes o respuesta en relación con las variables atributo o predictoras, en este caso de estudio, las variables objetivo son las temperaturas interiores en el edificio y el uso de los sistemas de climatización [10] [11] [19].

5.1.2. HIPERPARÁMETROS

Los hiperparámetros se definen como aquellos parámetros que no se ajustan automáticamente durante el entrenamiento del modelo, a diferencia de los parámetros del modelo. Estos parámetros influyen en el comportamiento y rendimiento del modelo, se establecen antes del entrenamiento y no se modifican durante el entrenamiento. No obstante, en algunos modelos puede iterarse entre varios valores de hiperparámetros para encontrar el mejor [10] [11].

5.1.3. AJUSTE DEL MODELO

Una vez se obtiene entrena el modelo, hay que estudiarlo y prestar atención a los fenómenos de desajuste y sobreajuste.

El desajuste es un fenómeno que se produce cuando el modelo no se ajusta bien al conjunto de entrenamiento y hay un sesgo algo, esto es, el modelo es incapaz de adaptarse adecuadamente a los datos. Un sesgo alto implica que el modelo simplifica en exceso la relación entre las variables, lo que puede llevar a un mal ajuste de los datos de entrenamiento. Ambos términos

hacen referencia a que el modelo no captura de manera precisa la verdadera naturaleza de los datos.

Por otro lado, el sobreajuste en un modelo de aprendizaje automático ocurre cuando este se adapta extremadamente bien a los datos de entrenamiento, capturando incluso características insignificantes. Esta situación guarda una estrecha relación con una alta varianza, lo cual indica que el modelo es muy sensible a las fluctuaciones en los datos de entrenamiento. Esto puede resultar en errores en la generalización del modelo a nuevos datos que no hayan aparecido anteriormente, debido a su falta de capacidad para separar la señal del ruido.

La intención final es conseguir obtener modelos que se ajusten correctamente a los datos de entrenamiento pero que tengan, al mismo tiempo, capacidad de generalización. Es decir, que funcionen para nuevos datos que el algoritmo no haya visto antes [10] [11].

5.1.4. SUBCONJUNTOS DE ENTRENAMIENTO

A la hora de evaluar un modelo, no se pueden utilizar todos los datos disponibles en el conjunto de entrenamiento, sino que estos datos se dividen en tres subconjuntos para poder estudiar el comportamiento del modelo.

1. **Conjunto de entrenamiento (m_{train})**: se escogen aproximadamente 60% de los datos para entrenar los modelos sobre estos. Al calcular el rendimiento del modelo en este conjunto, se puede comenzar a interpretar qué tan bien se ajusta el modelo a los datos. Si el rendimiento es muy bajo en este conjunto, el modelo tampoco se ajustará bien a nuevos datos.
2. **Conjunto de validación cruzada (m_{cv})**: se escogen aproximadamente el 20% de los datos. Una vez se han entrenado los modelos de estudio con el conjunto de entrenamiento, se obtiene el rendimiento del modelo sobre el conjunto de validación cruzada. Gracias a la validación cruzada entre modelos, se selecciona el modelo que mejor se ajuste a los datos, es decir, el que tenga un rendimiento mayor. Este rendimiento da una idea de cómo de bien se ajusta el modelo a datos nunca vistos.
3. **Conjunto de prueba (m_{test})**: es en torno al 20% de los datos del conjunto. Una vez se ha seleccionado el modelo, se conoce su rendimiento de generalización gracias al rendimiento obtenido sobre este conjunto de prueba.

Una vez realizado este proceso, se selecciona el modelo más apropiado para el conjunto de los datos. El modelo seleccionado debe ser revisado y optimizado periódicamente [10] [11].

5.1.5. MÉTRICAS DE RENDIMIENTO

Se han decidido estudiar las siguientes métricas de rendimiento en este proyecto.

1. El **coeficiente de Determinación (R^2)**, es una medida que indica qué tan bien se ajusta el modelo de regresión a los datos. R^2 compara la variabilidad explicada por el modelo con la variabilidad total de los datos. El R^2 puede variar entre $-\infty$ y 1. Un valor de 1 significa un ajuste perfecto, mientras que 0 o valores por debajo, indicarían que el modelo no es mejor que una predicción constante. Se calcula de la siguiente manera:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad Ec. 1$$

Donde y_i son los valores reales de la respuesta, \hat{y}_i son los valores que predice el modelo y \bar{y} es la media de los valores reales.

En este estudio, se evalúa su valor para el conjunto de entrenamiento y para el conjunto de validación cruzada. Evaluarlo en el conjunto de entrenamiento proporciona una medida de cuánto se ajusta el modelo a los datos utilizados para el entrenamiento. Por otro lado, calcularlo en el conjunto de validación cruzada es crucial para evaluar la capacidad del modelo para generalizar a datos no vistos durante el entrenamiento, y comparar estos valores entre los diferentes modelos utilizados. Un valor alto en ambos conjuntos sugiere un buen rendimiento y una buena generalización del modelo.

El valor de R^2 también ofrece una estimación del sobreajuste o desajuste. Si es alto sobre el conjunto de entrenamiento, pero bajo sobre el conjunto de validación cruzada, es un indicio de sobreajuste. Por otro lado, si es bajo en ambos conjuntos, indica desajuste. Esto puede indicar que se deben ajustar los hiperparámetros del modelo.

2. **Error Absoluto Medio (MAE)**: este error mide el promedio absoluto de las diferencias entre las predicciones del modelo y los valores reales. Es una medida de cuánto se desvían, en promedio, las predicciones respecto a los valores reales. Se calcula de la siguiente manera:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad Ec. 2$$

Donde y_i son los valores reales de la respuesta, \hat{y}_i son los valores que predice el modelo y n es el número total de observaciones.

MAE se expresa en la misma unidad que los datos originales, lo que facilita su interpretación. Cuanto menor sea el valor de MAE, mejor será el rendimiento del modelo de observaciones.

3. El **NMAE (Normalized Mean Absolute Error)**, es otra métrica de rendimiento fácilmente interpretable. Se define como la media del error MAE sobre el conjunto de datos analizado.

Toma valores entre 0 y 1, interpretándose valores cercanos a 0 como una predicción casi perfecta de los datos, y cercanos a 1 como una mala predicción.

5.2. SELECCIÓN DE ALGORITMOS

En este caso, se ha considerado separar el modelo de aprendizaje en dos etapas:

1. **Comportamiento térmico del edificio:** Dado que las variables predictoras, es decir, las variables climáticas, afectan a cómo se comporta el edificio térmicamente, se deciden realizar en primera instancia modelos de regresión con estas variables como variables de entrada. En tal caso, las temperaturas internas en las diferentes zonas del edificio son las variables de salida u objetivo.

Para poder analizar el comportamiento térmico del edificio, se ha decidido realizar un cribado de los datos. Se eliminan para este caso los datos recogidos en aquellos momentos en que los sistemas de climatización estén en funcionamiento, ya que solo así puede garantizarse que el comportamiento térmico observado se deba únicamente a las condiciones ambientales.

Este cribado se ha realizado utilizando Python mediante la herramienta Spyder. En el *Anexo III. Código de Python para el preprocesamiento de los datos*, se encuentra el código utilizado para eliminar las filas no necesarias en este análisis.

2. **Comportamiento energético del edificio:** Una vez se conoce cómo se va a comportar el edificio térmicamente, se procede a establecer modelos de regresión que determinen como debe ser el funcionamiento de los sistemas de climatización en función de las temperaturas interiores alcanzadas en el edificio.

Para este caso, las variables predictoras o de entrada son las temperaturas interiores en las diferentes zonas del edificio. Por otro lado, como variables objetivo se incluyen los diferentes usos de climatización en el edificio.

Como se explica en el punto 3.1. *Enfoque del proyecto*, este trabajo se realiza bajo el enfoque del aprendizaje automático supervisado, ya que este tipo de modelos son aquellos que cuentan con datos correctos disponibles para entrenar un modelo, y a partir de estos se pueden hacer predicciones sobre nuevos datos. En concreto, la regresión es la base de este proyecto.

La regresión consiste en abordar la tarea de predecir valores numéricos continuos a partir de variables predictoras. Existen diversos modelos de regresión, con diversas características y aplicaciones.

Para el caso de estudio presente, se han seleccionado cuatro modelos de regresión altamente utilizados en aprendizaje automático. Estos son la regresión lineal múltiple, la regresión por máquina de vectores, la regresión por k vecinos más cercanos y la regresión por árboles aleatorios.

Estos modelos se han seleccionado por su aparición en proyectos de Machine Learning relacionados con el presente tema de estudio, como *Predicción de Consumo Energético de Oficinas mediante Modelos de Tipo Caja Negra* (Duarte Alexis Kazacos Winter País de Brito, 2017) [20] o *Aplicación de Machine Learning al consumo eléctrico de edificios inteligentes* (Mónica Meneses Díez, 2021) [21]. Estos proyectos validan la aplicación de los modelos sobre otros estudios de aprendizaje automático relacionados con este campo.

En el *Anexo IV. Modelos de regresión aplicados*, se realiza una introducción teórica a los modelos de regresión seleccionados para este caso de estudio.

5.3. IMPLEMENTACIÓN Y COMPARACIÓN DE LOS MODELOS TÉRMICOS

El primer paso es el entrenamiento de los modelos de Machine Learning, lo cual se realiza sobre el subconjunto de datos de entrenamiento (m_{train}). Una vez entrenados los algoritmos, se evalúa su rendimiento en otro subconjunto de datos denominado conjunto de validación cruzada (m_{cv}). Una vez se obtengan los rendimientos de cada uno de los modelos, gracias a la validación cruzada, se puede comparar el rendimiento de los modelos anteriormente explicados.

Para poder comparar los distintos modelos, es necesario que sean considerados los mismos conjuntos de entrenamiento y variables. Esto proporciona comparabilidad a los resultados. Por tanto, las variables de entrada serán para todos los casos las variables climáticas, y las de salida las diferentes temperaturas en el interior del edificio ETOPIA. Puede encontrarse su explicación detallada en el punto *4.1. Descripción de los datos recopilados*.

En el *Anexo VI. Resultados de los modelos térmicos de regresión*, pueden encontrarse los resultados detallados de los diferentes modelos considerados y en el *Anexo VII. Código de Python para los modelos térmicos*, la implementación en código de los modelos. En este apartado, se procede a realizar una comparativa entre esos resultados a fin de escoger el modelo óptimo.

A continuación, se observa una tabla comparativa sobre el rendimiento de los diferentes modelos sobre los datos de los conjuntos de validación cruzada.

Tabla 1. Comparativa de modelos de regresión para el comportamiento térmico. Fuente: elaboración propia

ZONA		VARIABLE SALIDA (T = Temperatura)	R ² MLR	R ² SVR	R ² KNN	R ² RFF
EDIFICIO 1	Entrada	T maquina café	0,5462	0,5335	0,4786	0,5262
		T recepción	0,3874	0,3668	0,3146	0,3589
		T zona ascensor	0,3874	0,3681	0,3146	0,3584
	Planta baja	T anfiteatro 1	0,8232	0,8288	0,7829	0,8209
		T anfiteatro 2	0,7864	0,794	0,7438	0,7801
		T auditorio	0,8307	0,7618	0,8535	0,8759
	Planta 2	T sala planta 2	0,6239	0,6191	0,5989	0,6304
	Planta 3	T sala planta 3	0,6828	0,6771	0,644	0,6586
	Open Art	Temperatura sala	0,7133	0,7126	0,708	0,7159
	Planta 2	Temperatura 1	0,4225	0,431	0,4266	0,4807
		Temperatura 2	0,5054	0,5119	0,4835	0,5363
EDIFICIO 2	Planta 3	Temperatura 1	0,4793	0,4817	0,4271	0,5197
		Temperatura 2	0,4909	0,5138	0,4509	0,5208
	Planta 4	Temperatura 1	0,49	0,5012	0,4132	0,4963
		Temperatura 2	0,5237	0,5323	0,4464	0,5359
	Planta 5	Temperatura 1	0,5468	0,548	0,489	0,5432
		Temperatura 2	0,5739	0,5672	0,5134	0,5722
	Planta 6	Temperatura sala	0,6462	0,6614	0,6197	0,6465
	Aula 1	Temperatura sala	0,6274	0,6386	0,6052	0,6272
	Aula 2	Temperatura sala	0,7287	0,7425	0,6816	0,7243
	Aula 3	Temperatura sala	0,8101	0,81	0,7729	0,8112
EDIFICIO 3	Zona común	Temperatura sala 1	0,8229	0,8237	0,7665	0,8256
		Temperatura sala 2	0,7725	0,7818	0,7459	0,7894

Al observar la tabla 1, se comparan los rendimientos de los diferentes modelos para todas las variables de salida estudiadas. En color naranja se aprecia para cada variable qué modelo otorga un mejor ajuste a los datos de validación cruzada. No obstante, puede observarse que los rendimientos son similares entre diferentes modelos para las mismas variables consideradas.

En Machine Learning, es de gran importancia optimizar los recursos con los que se trabaja, ya que se analizan grandes volúmenes de datos y esto puede requerir de un tiempo de cómputo muy alto. Por ello, se establece la lógica de escoger el modelo que, con rendimiento similar a otros con los que se compare, presente menor complejidad.

El modelo de regresión por árboles aleatorios es el que más veces otorga el mejor resultado, un total de 11 veces. Sin embargo, este modelo tiene un tiempo de cómputo mayor y mayor complejidad. Lo mismo ocurre con el modelo de regresión por máquina de vectores, que es

elegido 7 veces. Los modelos de regresión lineal múltiple y de k-vecinos más cercanos son los menos complejos, no obstante, el modelo de k-vecinos más próximos ha requerido de un valor alto en el hiperparámetro de número de vecinos (1000 vecinos) lo cual aumenta también su tiempo de cómputo y en ningún caso es escogido como modelo con mejor rendimiento.

Por tanto, en este caso, el modelo que se debe escoger es el **modelo de regresión lineal múltiple**, que otorga valores de rendimiento similares y tiene un coste computacional menor. En concreto, el modelo otorga valores de rendimiento de media un 3% menores a los modelos con mejores resultados.

5.3.1. REGRESIÓN LINEAL MÚLTIPLE SOBRE EL CONJUNTO DE PRUEBA

Una vez se ha determinado que el modelo escogido sea el modelo de regresión lineal múltiple, por su rendimiento similar a otros modelos y su sencillez computacional, se procede a evaluar su rendimiento sobre el conjunto de prueba. Este es el último paso para validar el modelo.

Tabla 2. Modelo de regresión lineal múltiple sobre conjunto de prueba. Fuente: elaboración propia

EDIFICIO	ZONA	VARIABLE SALIDA (T = Temperatura)	R ²	R ² prueba	MAE	NMAE
			entrenamiento			
EDIFICIO 1	Entrada	T maquina café	0,7248	0,4567	1,4502	0,0708
		T recepción	0,5832	0,1288	1,7626	0,0862
		T zona ascensor	0,5832	0,1288	1,7626	0,0862
	Planta baja	T anfiteatro 1	0,7984	0,8456	1,2602	0,0704
		T anfiteatro 2	0,7683	0,8278	1,1394	0,0606
		T auditorio	0,7809	0,7549	2,2842	0,1156
	Planta 2	T sala planta 2	0,6934	0,6104	1,7212	0,0892
	Planta 3	T sala planta 3	0,6599	0,6752	1,5564	0,0786
	Open Art	Temperatura sala	0,7082	0,5933	1,169	0,0572
		Temperatura 1	0,5309	0,4225	1,5041	0,0709
EDIFICIO 2	Planta 2	Temperatura 2	0,6581	0,3143	1,302	0,0621
		Temperatura 1	0,8751	0,512	1,0178	0,0471
	Planta 3	Temperatura 2	0,5232	0,4473	1,2038	0,0545
		Temperatura 1	0,5514	0,5475	1,1053	0,0498
	Planta 4	Temperatura 2	0,5753	0,5826	1,2079	0,0533
		Temperatura 1	0,6119	0,6719	0,9535	0,0411
	Planta 5	Temperatura 2	0,6119	0,6889	1,0024	0,0432
		Temperatura sala	0,6846	0,6747	0,9314	0,041
	Aula 1	Temperatura sala	0,6656	0,373	1,6079	0,0727
	Aula 2	Temperatura sala	0,6794	0,7475	0,9916	0,0459
EDIFICIO 3	Aula 3	Temperatura sala	0,7743	0,7314	1,2081	0,0559
	Zona común	Temperatura sala 1	0,854	0,8039	0,981	0,0461
		Temperatura sala 2	0,7887	0,7243	1,1307	0,0516

Se observa que el modelo sigue ajustándose a los nuevos datos con buenos rendimientos. Los rendimientos sobre el conjunto de prueba son, de media, un 28% peores que sobre el conjunto de entrenamiento y un 8% peores que sobre el conjunto de validación cruzada. Además, los valores se desvían de media 1,27°C del valor real.

Una excepción se da en el caso de la temperatura de la zona de recepción y del ascensor en la Entrada. Presenta un rendimiento un 78% peor que sobre el conjunto de entrenamiento, y la media de los errores en este caso es de 1,76 °C aproximadamente, un valor alto en comparación con los demás. Puede deberse a que los sensores en estas zonas no funcionen correctamente.

No obstante, si se observa el valor del NMAE, puede apreciarse que casi todos los valores se encuentran por debajo de 0,1. Esta métrica evalúa las predicciones del modelo, tomando valores entre 0 y 1, cuando es cercana a 0 significa que el modelo realiza predicciones casi perfectas. La media de estas métricas es de 0,063, lo que indica que el modelo realiza bien las predicciones.

Por último, cabe destacar que una variable muy interesante para tener en cuenta sería la irradiación. El edificio Etopía recibe mucha luz solar al encontrarse en un emplazamiento sin edificios colindantes. Además, varias zonas tienen grandes ventanales, por lo que la temperatura interior se ve afectada por este factor. La ocupación del edificio también podría ser una variable a tener en cuenta ya que afecta a las temperaturas en el interior. Se presentan las conclusiones detalladas en el apartado *6.1. Análisis de los principales resultados y conclusiones*.

5.4. IMPLEMENTACIÓN Y COMPARACIÓN DE LOS MODELOS ENERGÉTICOS

De igual modo que con los anteriores modelos, en primer lugar, se han de entrenar los modelos de Machine Learning sobre el subconjunto de datos de entrenamiento (m_{train}). Una vez entrenados los algoritmos, se evalúa su rendimiento en el conjunto de validación cruzada (m_{cv}), para poder realizar una comparativa.

Se consideran los mismos conjuntos de entrenamiento y las mismas variables para los diferentes modelos. En este caso, las variables de entrada serán las temperaturas interiores del edificio, y las de salida los usos de sistemas de calefacción o refrigeración en el edificio ETOPIA.

En el *Anexo VI. Resultados de los modelos energéticos de regresión*, pueden encontrarse los resultados detallados de los diferentes modelos considerados, y en el *Anexo VII. Código de Python para los modelos energéticos*, la implementación en código de los modelos. En este apartado, se realiza una comparativa entre los resultados a fin de escoger el modelo óptimo.

En este caso, los resultados emitidos inicialmente tras la implementación de los diferentes modelos de regresión fueron inesperados. Ninguno de los modelos estudiados podía establecer relaciones entre las variables de entrada, siendo estas las temperaturas internas del edificio, y las de salida, el uso de los sistemas de climatización. Tras analizar los datos con los que se contaba, podía observarse que los sistemas de climatización no se encendían siempre para las mismas temperaturas alcanzadas. Esto es debido, probablemente, al hecho de que no se activan dichos sistemas si la zona no va a ser utilizada. Se muestra a continuación una tabla comparativa de los resultados arrojados por los modelos en primera instancia.

Tabla 3. Comparativa inicial del rendimiento de los modelos energéticos de regresión. Fuente: elaboración propia

ZONA		VARIABLE SALIDA (E = Energía)	R ² MLR	R ² SVR	R ² KNN	R ² RFF
EDIFICIO 1	Entrada	E suelo radiante	0,2974	-0,2695	0,1755	0,0911
	Planta baja	E calor anfiteatro 1	-0,4252	-0,0325	-0,3348	-0,7946
		E frio anfiteatro 1	-0,2463	-0,0189	-0,1967	-9,5316
		E calor anfiteatro 2	-0,0803	-0,0491	-0,1213	-0,5675
		E frio anfiteatro 2	-0,8911	-0,0279	-0,5731	-6,9254
		E calor auditorio	0,0038	-0,0439	0,0698	-0,0811
	Planta 2	E frio auditorio	-0,0371	-0,0362	0,02023	-0,3309
		E calor planta 2	-0,02	-0,1228	0,1635	0,1489
	Planta 3	E frío planta 2	-0,4039	-0,0826	-0,4029	-5,0007
		E calor planta 3	-0,004	-0,06	0,0371	0,0312
		E frío planta 3	0,032	-0,0751	-0,055	-1,5114
		E calor Open Art	-0,0192	-0,1159	0,0418	0,037
EDIFICIO 2	Planta 2	E frío Open Art	0,0229	-0,0826	-0,0776	-0,0845
		E suelo radiante	0,0023	0,1377	0,0447	-0,1072
		E Fancoil 1	-0,3191	-0,0459	0,2575	0,223
		E Fancoil 2	-0,4514	-0,0417	0,2192	0,1491
		E Fancoil 3	-0,5477	-0,0639	0,2288	0,1925
	Planta 3	E suelo radiante	0,0104	-0,062	0,0824	-0,0736
		E Fancoil 1	-0,253	-0,0734	0,265	0,0519
		E Fancoil 2	-0,0432	-0,165	0,298	-0,1599
		E Fancoil 3	-0,3691	-0,1391	0,3466	-0,1548
	Planta 4	E suelo radiante	0,0125	-0,0442	0,06	-0,1653
		E Fancoil 1	-0,1084	-0,1034	0,1117	0,1293
		E Fancoil 2	-0,0417	-0,1049	0,1019	-0,2247
		E Fancoil 3	-0,1675	-0,151	0,1398	-0,0432
	Planta 5	E suelo radiante	-0,0022	-0,0205	-0,0024	-0,4885
		E Fancoil 1	-0,0016	-0,13	0,0119	0,0294
		E Fancoil 2	0,042	-0,2077	0,033	0,0717
		E Fancoil 3	0,1087	-0,2028	0,1146	0,1281
	Planta 6	E calor	-0,0021	-0,0463	0,0358	-0,0833
		E frio	-0,0091	-0,0981	0,0376	0,0236

EDIFICIO 3	Aula 1	E calor	0,0008	-0,0051	-0,0361	-0,0983
		E frio	-0,1817	-0,1901	-0,054	-0,0769
	Aula 2	E calor	-0,0006	-0,0011	-0,001	-0,0012
		E frio	-0,01993	-0,0907	0,0469	0,025
	Aula 3	E calor	-0,5402	-0,0268	-0,6133	-0,643
		E frio	-0,0093	-0,0697	0,0366	0,0163
	Zona común	E calor 1	-0,0633	-0,1475	-0,593	-0,1272
		E calor 2	0,0046	-0,0436	-0,126	-0,0926
		E frio 1	-0,7114	-0,0998	-0,03	-0,2335
		E frio 2	-0,2904	-0,1793	-0,103	-0,1918

En la tabla superior puede observarse que los valores de R^2 toman incluso valores negativos en muchos casos, lo cual indica que los modelos no pueden predecir correctamente los valores de las variables de salida.

En tal caso, ninguno de los modelos puede escogerse como válido para este proyecto. Cuando esto sucede en modelos de aprendizaje automático, es recomendable obtener otras variables predictoras que ayuden a explicar las variables objetivo. En este caso, es importante la puntualización de que las aulas no se utilizan constantemente. Por tanto, conocer la ocupación de los diferentes espacios resulta necesario en este estudio.

Al no disponer de tales datos, se decide crear una nueva variable de entrada denominada “ocupación”. Esta variable se establece con valor 0 si los sistemas de climatización no están en funcionamiento, y con valor 1 si lo están. Con esto se presupone que las salas solo han estado en uso cuando los sistemas de climatización han sido utilizados.

Este comportamiento no es real, ya que las salas han podido ser utilizadas sin necesidad de ser climatizadas. No obstante, sí que se cumple que las salas se climatizan únicamente cuando van a ser utilizadas.

La razón tras la creación de esta nueva variable es estudiar si ello pudiera mejorar el rendimiento de los modelos. Se muestran los resultados obtenidos en el *Anexo VI. Resultados de los modelos energéticos de regresión*, y a continuación una tabla comparativa de los diferentes modelos.

Tabla 4. Comparativa del rendimiento de los modelos de regresión con la variable de ocupación. Fuente:
elaboración propia

	ZONA	VARIABLE SALIDA (E = Energía)	R ² MLR	R ² SVR	R ² KNN	R ² RFF
EDIFICIO 1	Planta baja	E suelo radiante	0,992	0,9925	0,8213	0,9711
		E calor anfiteatro 1	0,1732	0,0725	0,1305	-0,2386
		E frio anfiteatro 1	0,1034	-0,0189	-0,0612	0,2975
		E calor anfiteatro 2	0,4456	-0,0489	0,203	0,4981
		E frio anfiteatro 2	0,1974	0,2979	0,1551	0,3653
		E calor auditorio	0,4635	0,3577	0,4539	0,6825
	Planta 2	E frio auditorio	0,2234	-0,0362	0,2695	0,2308
		E calor planta 2	0,4096	0,2168	0,5179	0,6581
	Planta 3	E frío planta 2	0,1469	-0,0826	-0,1256	-1,6471
		E calor planta 3	0,245	-0,0599	0,2101	0,0561
		E frío planta 3	0,3434	-0,0317	0,2095	0,429
EDIFICIO 2	Open Art	E calor Open Art	0,2997	0,1343	0,3379	0,3343
		E frío Open Art	0,2449	-0,0826	0,0313	-0,0848
	Planta 2	E suelo radiante	0,3219	-0,1377	0,6771	0,9747
		E Fancoil 1	-0,1412	-0,0049	0,5837	0,5897
		E Fancoil 2	-0,2942	0,1431	0,5697	-0,0082
		E Fancoil 3	-0,1577	-1,5536	0,5318	0,6884
	Planta 3	E suelo radiante	0,1451	-0,062	0,654	0,9441
		E Fancoil 1	0,2865	0,27	0,3757	0,4106
		E Fancoil 2	0,5534	0,3608	0,6115	0,8621
		E Fancoil 3	0,5473	0,2639	0,6708	0,9576
	Planta 4	E suelo radiante	0,0795	-0,0442	0,4879	0,9078
		E Fancoil 1	0,404	0,3355	0,406	0,3109
		E Fancoil 2	0,3277	-0,0949	0,3807	0,5725
		E Fancoil 3	0,5041	0,1998	0,4977	0,634
	Planta 5	E suelo radiante	0,0668	-0,0205	0,2592	0,5795
		E Fancoil 1	0,2938	0,1179	0,2346	0,2887
		E Fancoil 2	0,492	0,4014	0,6305	0,5843
		E Fancoil 3	0,5983	0,4505	0,4847	0,5828
	Planta 6	E calor	0,1626	-0,0462	0,2254	0,143
		E frio	0,3475	0,2055	0,3468	0,1948
EDIFICIO 3	Aula 1	E calor	0,0163	-0,0051	0,1146	-0,6422
		E frio	0,7899	0,7687	0,656	0,751
	Aula 2	E calor	0,0019	-0,0011	0,0006	0,0406
		E frio	0,7478	0,5748	0,5995	0,6652
	Aula 3	E calor	0,116	-0,015	-0,2972	0,1517
		E frio	0,4635	-0,0696	0,4289	0,5544
	Zona común	E calor 1	0,0902	-0,1475	0,1555	-0,0425
		E calor 2	0,012	-0,0436	0,0956	-0,3294
		E frio 1	-0,0028	-0,0314	0,2824	0,0748
		E frio 2	0,4075	-0,1062	0,2781	0,2855

Puede observarse que los valores de rendimiento sobre el conjunto de validación cruzada mejoran gracias a la implementación de esta nueva variable. Los modelos de regresión lineal múltiple mejoran de media un 42%, los de máquina de vectores mejoran en torno a un 21%, los de k-vecinos más próximos mejoran de media un 35% y los de regresión por árboles aleatorios un 60% aproximadamente. No obstante, sigue habiendo algunos valores negativos presentes y los rendimientos obtenidos son muy dispares entre sí.

Se observan valores cercanos a 1 para variables como el uso del suelo radiante, con métricas máximas como $R^2 = 0,9925$ en la zona de la entrada del Edificio 1, o $R^2 = 0,9747$ para el uso del suelo radiante en la planta 2 del Edificio 2.

Pero también se aprecia el caso opuesto, por ejemplo, para el uso de uno de los sistemas de calor de la Zona Común del Edificio 3, el mejor rendimiento es el que otorga el algoritmo de k-vecinos más cercanos, siendo este de $R^2 = 0,0956$.

El modelo que presenta mejor rendimiento en la mayoría de los casos es el de regresión mediante árboles aleatorios. Lo cual respalda las conclusiones obtenidas en [20].

El uso de esta nueva variable que determina la ocupación ayuda a obtener mejores modelos de regresión, pero no pueden ser considerados como válidos ya que esta variable es ficticia. Por tanto, para otorgar continuidad a este estudio debería poder recogerse el dato de ocupación real de las salas del edificio ETOPIA, para poder estudiar la validez real de los modelos presentados.

No obstante, gracias a la creación y uso de esta nueva variable, puede observarse que contar con los valores reales de ocupación de las aulas es muy probable que ayude a encontrar modelos más adecuados en el futuro.

6. CONCLUSIONES Y RECOMENDACIONES

6.1 ANÁLISIS DE LOS PRINCIPALES RESULTADOS Y CONCLUSIONES

En la sección 5.3. *Implementación y comparación de los modelos térmicos*, se ha demostrado que el modelo de regresión elegido para el comportamiento térmico del edificio ha proporcionado valores de rendimiento destacables en la mayoría de las variables. Estos resultados indican que el conjunto de variables de salida, compuesto por las temperaturas interiores del edificio Etopía, se puede predecir en su mayoría con exactitud utilizando este modelo.

Observando los valores del coeficiente de determinación del modelo elegido, el de Regresión Lineal Múltiple, obtenidos sobre un conjunto de datos nunca observado (el conjunto de prueba) se pueden obtener conclusiones de este modelo.

En primer lugar, ningún valor de R^2 es nulo o negativo, lo que indica que el modelo supera la predicción basada en el valor medio. No obstante, se identifican valores de R^2 que se consideran bajos, como sucede en las temperaturas de recepción y ascensor del Edificio 1 ($R^2 = 0,1288$), en la Planta 2 del Edificio 2 ($R^2 = 0,3143$), o en el Aula 1 del Edificio 3 ($R^2 = 0,373$). Es recomendable una revisión de los sensores, ya que los datos del conjunto de prueba corresponden a 2023, mientras los del conjunto de entrenamiento y validación cruzada pertenecen a 2021 y 2022, y otorgaban resultados más favorables ($R^2_{\text{recepción}} = 0,5462$; $R^2_{\text{planta 2}} = 0,4225$; $R^2_{\text{aula 1}} = 0,6274$).

Respecto a los valores de desviación en las predicciones realizadas por el modelo, el Error Absoluto Medio (MAE) promedio es de $1,277^\circ\text{C}$. La mayor desviación media se registra para la temperatura del auditorio, cuyo MAE tiene un valor de $2,284^\circ\text{C}$. A pesar de ello, el modelo presenta un rendimiento R^2 de 0,7549 para esta variable, lo que respalda la calidad de las predicciones realizadas. Además, los valores de NMAE inferiores a 0,1 sugieren que el modelo realiza predicciones casi perfectas en la mayoría de los casos.

No obstante, estos modelos tienen margen de mejora, las recomendaciones para la continuidad del estudio y su mejora se explicarán en el siguiente apartado 6.2. *Limitaciones del estudio y recomendaciones para futuras investigaciones*.

Por otro lado, se ha podido observar, tras los resultados expuestos en el punto 5.4. *Implementación y comparación de los modelos energéticos*, que estos modelos no han tenido el rendimiento esperado. Los modelos aplicados son modelos de regresión utilizados con gran frecuencia en proyectos de aprendizaje automático, que proporcionan predicciones buenas si

cuentan con la suficiente información. Al comparar modelos tan diversos, algunos de ellos utilizados no solo en aprendizaje supervisado sino también en aprendizaje no supervisado, y no obtener resultados positivos con ninguno de ellos, se concluyó que la problemática radica en la falta de información pertinente.

En un esfuerzo por abordar esta limitación, se propuso una nueva variable que reflejase la ocupación del edificio, considerando las zonas ocupadas solo cuando los sistemas de climatización estaban en funcionamiento. Esta hipótesis se fundamenta en parte en una realidad, ya que la climatización de las aulas no se enciende a no ser que se vaya a utilizar. Sin embargo, la variable creada no contempla situaciones en las que se utilizan las aulas sin necesidad de climatización.

A pesar de estas limitaciones, la introducción de esta variable pudo dar más información al algoritmo, y se observa que los resultados mejoran para todos los modelos. Por ello, se concluye que esta variable debería ser registrada para poder continuar con el proyecto en un futuro. En el siguiente apartado se explicará bajo qué recomendaciones se debe continuar el proyecto.

6.2. LIMITACIONES DEL ESTUDIO Y RECOMENDACIONES

Las limitaciones encontradas en el estudio respecto al comportamiento térmico del edificio fueron varias. Desde el inicio, se observó que la temperatura exterior registrada en el edificio Etopía registraba un gran conjunto de datos con valores atípicos, especialmente en los meses de verano. Al investigar la ubicación de este sensor en el edificio, se encontró desconocimiento por parte de los trabajadores sobre su localización. Como resultado, fue necesario buscar datos climáticos en fuentes externas.

Una recomendación clave para avanzar con este estudio sería considerar la instalación de una estación meteorológica Davis en el propio edificio. Los datos climatológicos con los que se ha realizado este estudio pertenecen a la Facultad de Geografía, lo que implica que los sensores están en una localización diferente al edificio. Además, no se pudo contar con los datos de irradiación, los cuales serían valiosos y podrían obtenerse mediante la propuesta de la estación Davis.

En cuanto al comportamiento energético del edificio, el objetivo del estudio era obtener información sobre el uso de los diferentes sistemas de climatización, expresado como el porcentaje de apertura de las válvulas. La principal limitación en el estudio radica en la falta de variables para su determinación. Al intentar estimar el uso de los sistemas de climatización únicamente con las variables de temperatura interior, los modelos utilizados no lograron

identificar relaciones significativas entre las mismas. Sin embargo, al introducir una variable nueva creada para el estudio, que hace referencia a la ocupación de las diversas zonas del edificio, los algoritmos pudieron identificar relaciones más sólidas entre las variables predictoras y objetivo.

Por ello, la principal recomendación radica en la implementación de esta nueva variable, que debe ser recogida en el edificio. Además de mejorar la precisión en la determinación del uso de los sistemas de climatización, esta variable de ocupación podría ser empleada para mejorar las predicciones del comportamiento térmico del edificio.

Una vez se disponga de las variables descritas, se deberían validar nuevamente los algoritmos empleados en este estudio. Los resultados emitidos indican que, con la adquisición de nuevos datos y variables, los modelos propuestos podrían mejorar significativamente.

Finalmente, una vez se hayan definido los modelos térmicos y energéticos, el último paso en cualquier proyecto de Machine Learning es la implementación del modelo, conocido como Machine Learning Operations (MLOps). Esto requiere ingeniería de software para crear un servidor que integre los modelos de aprendizaje automático y que sea monitoreado continuamente para mantener el sistema. El resultado final sería una aplicación en la que se ingresen los datos de entrada, es decir, los valores de las variables predictoras. Con los modelos adecuados integrados, esta aplicación realiza la predicción de las variables objetivo.

LISTADO DE ACRÓNIMOS

1. IA: Inteligencia Artificial
2. ML: Machine Learning
3. ODS: Objetivos de Desarrollo Sostenible
4. TFM: Trabajo de Fin de Máster
5. EU: European Union
6. COVID-19: Coronavirus Disease 2019
7. MAE: Median Absolute Error
8. MLR: Multiple Linear Regression
9. SVR: Support Vector Machine
10. KNN: K Nearest Neighbors
11. RFR: Random Forest Regressor
12. MLOps: Machine Learning Operations

BIBLIOGRAFÍA

1. Objetivos y metas de desarrollo sostenible. Naciones Unidas.
<https://www.un.org/> [Consulta: junio 2023]
2. Etopía Centro de Arte y Tecnología. Wikipedia.
<https://es.wikipedia.org/> [Consulta: mayo 2023]
3. ETOPIA. Página web del Centro.
<https://etopia.es/> [Consulta: mayo 2023]
4. Etopía Centro de Arte y Tecnología. Ayuntamiento de Zaragoza.
<https://www.zaragoza.es/> [Consulta: mayo 2023]
5. Artículo “*¿Qué es la eficiencia energética en edificios?*”. REHAU.
<https://www.rehau.com/> [Consulta: junio 2023]
6. Passivhaus. Wikipedia.
<https://es.wikipedia.org/> [Consulta: junio 2023]
7. Teresa Knoerr, 2023. Artículo “*Ciudades: retos y oportunidades para la eficiencia energética*”. La Ciudad Sostenible.
<https://www.ciudadssostenible.eu/> [Consulta: junio 2023]
8. Bruno Gutiérrez Cuevas, 2022. Artículo “*El reto de lograr edificios energéticamente eficientes*”. El País.
<https://elpais.com/> [Consulta: junio 2023]
9. Elodie Guillard, 2022. Artículo “*Predicción del Consumo de Energía mediante el Machine Learning y la IA*”. Dexma.
<https://www.dexma.com/es> [Consulta: junio 2023]
10. Andrew Ng. Curso “*Supervised Machine Learning: Regression and Classification*”. Universidad de Stanford. Coursera.
<https://www.coursera.org/> [Consulta: febrero 2023 – noviembre 2023]
11. Andrew Ng. Curso “*Advanced Learning Algorithms*”. Universidad de Stanford. Coursera.
<https://www.coursera.org/> [Consulta: febrero – noviembre 2023]
12. Proyecto centro de Arte y Tecnología Etopía. MSM Arquitectos.
<https://www.msmarquitectos.es/> [Consulta: julio 2023]
13. Jéssica Blasco Quílez, 2015. *Impacto de las actividades artísticas de Etopía en la prensa digital aragonesa*. Repositorio de la Universidad de Zaragoza. [Consulta: agosto 2023]
14. Datos climáticos de la ciudad de Zaragoza. Departamento de Geografía y Ordenación del Territorio de la Universidad de Zaragoza. [Consulta: junio – noviembre 2023]

15. Idoia Arnabat, 2020. Artículo “*Suelo Radiante Refrescante: Una solución integral de Climatización*”. Calor y frio.
<https://www.caloryfrio.com/> [Consulta: julio 2023]
16. Oscar Martín González, 2023. Artículo “*¿Qué es un Fancoil? cómo funciona, tipos y precio*”. Selectra.
<https://preciogas.com/> [Consulta: julio 2023]
17. Estación Davis Vantage Pro2. Davis Instruments.
<https://www.davisinstruments.com/> [Consulta: agosto 2023]
18. Datos recogidos en el edificio ETOPIA. Edificio ETOPIA. [Consulta: febrero – noviembre 2023]
19. Pablo Lucero, 2023. Artículo “*¿Qué es la variable objetivo, dentro del Machine learning?*”. Medium.
<https://pabloluceroai.medium.com/> [Consulta: julio 2023]
20. Duarte Alexis Kazacos Winter País de Brito, 2017. *Predicción de Consumo Energético de Oficinas mediante Modelos de Tipo Caja Negra*. Repositorio Escuela Técnica Superior de Ingenieros de Minas y Energía. Universidad Politécnica de Madrid. [Consulta: septiembre – noviembre 2023]
21. Mónica Meneses Díez, 2021. *Aplicación de Machine Learning al consumo eléctrico de edificios inteligentes*. Repositorio Universitat Oberta de Catalunya. [Consulta: agosto 2023]
22. Comprender la regresión del aprendizaje automático: Una guía completa. Mailchimp.
<https://mailchimp.com/> [Consulta: octubre 2023]
23. Joaquín Amat Rodrigo, 2023. Regresión lineal múltiple con Python. Ciencia de datos.
<https://cienciadedatos.net/> [Consulta: octubre 2023]
24. Joaquín Amat Rodrigo, 2020. Máquinas de Vector Soporte (SVM). Ciencia de datos.
<https://cienciadedatos.net/> [Consulta: octubre 2023]
25. Vídeo “*¿Cómo funciona SVM?*”. MindMachineTV. Youtube.
<https://www.youtube.com/> [Consulta: octubre 2023]
26. Máquina de Soporte Vectorial (SVM). Medium.
<https://medium.com/> [Consulta: octubre 2023]
27. Palak Airón, 2022. “*Una guía rápida para comprender un algoritmo KNN*”. Unite IA.
<https://www.unite.ai/> [Consulta: octubre 2023]
28. Artículo “*¿Qué es el algoritmo de k vecinos más cercanos?*”. IBM.
<https://www.ibm.com/> [Consulta: octubre 2023]
29. Joaquín Amat Rodrigo, 2020. Random Forest con Python. Ciencia de Datos.
<https://cienciadedatos.net/> [Consulta: octubre 2023]

ANEXO I. PREPROCESAMIENTO DE LOS DATOS DEL EDIFICIO ETOPIA

En el proceso de preprocesamiento de los datos del edificio Etopía se han llevado a cabo una serie de pasos, en primer lugar, utilizando la herramienta de Excel.

Primeramente, se procedió a trasladar los datos “brutos” a Excel, para poder trabajar con ellos adecuadamente. Se ordenaron por fecha y hora, desde los más antiguos hasta los actuales, y se procedió a obtener los máximos y mínimos de todas las variables consideradas importantes para poder detectar valores atípicos. A continuación, se detallan dichas variables y la inconsistencia en los datos obtenidos, separándolos en los tres bloques que conforman el edificio Etopía y las zonas explicadas en el punto 4.1.1. *DATOS RECOGIDOS EN EL EDIFICIO ETOPIA*.

A1.1. PREPROCESAMIENTO DE LOS DATOS DEL EDIFICIO 1

En primer lugar, se muestran los gráficos de las temperaturas durante el tiempo, de todos los datos recogidos desde 2018 hasta la fecha, para que puedan observarse tanto los picos como los intervalos de tiempo en los que no se registraron datos. Además, se muestran tablas de los máximos y mínimos de cada una de las variables para poder encontrar los picos incoherentes.

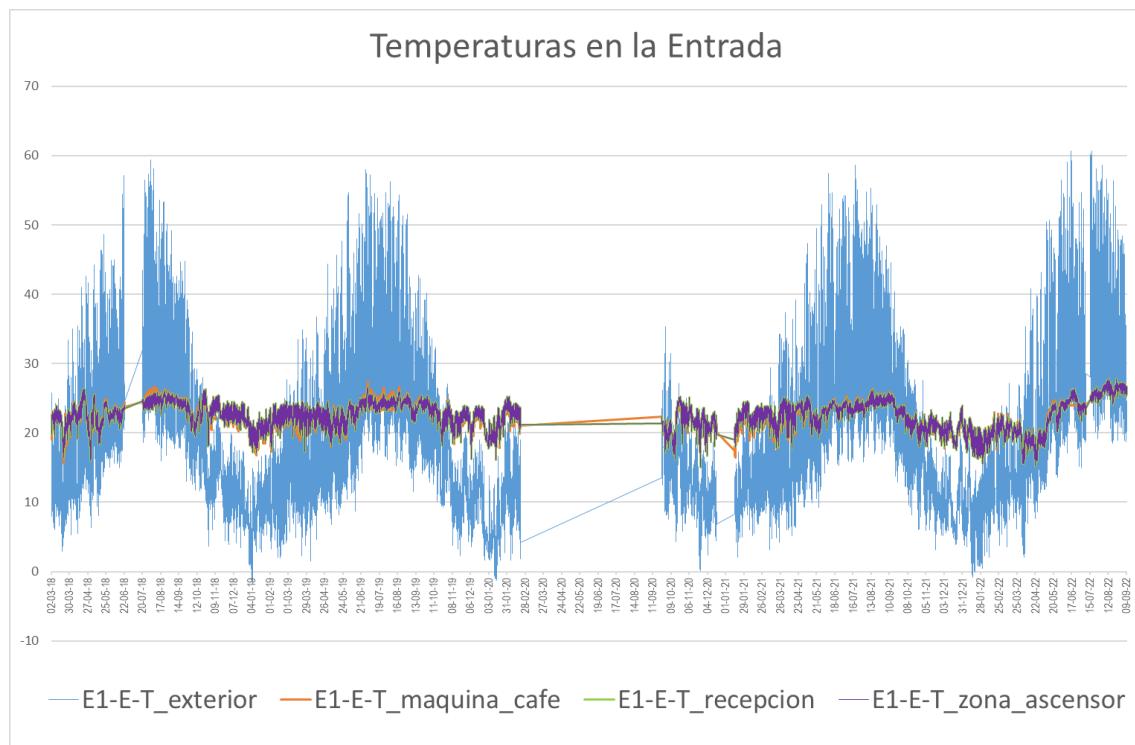


Ilustración 14. Gráfico inicial temperaturas en la entrada del Edificio 1. Fuente: elaboración propia

Tabla 5. Valores máximos y mínimos de los datos de la entrada del Edificio 1. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS		
VARIABLE	MÁXIMO	MÍNIMO
Temperatura exterior [°C]	60,71	-1,81
Temperatura máquina café [°C]	27,86	15,55
Temperatura recepción [°C]	27,86	14,99
Temperatura zona ascensor [°C]	27,86	14,99
Suelo radiante [%]	100	0

Puede observarse que los datos obtenidos respecto a la temperatura exterior presentan errores, dando valores de temperatura por encima de 60°C. Estos valores no son puntuales, sino que en la *Ilustración 12*, puede observarse que hay una gran cantidad de datos situados entre 40°C y 60°C, esto es debido a que el sensor se encuentra en el exterior del edificio y, probablemente, cuando incide el sol sobre el mismo se sobrecalienta dando lugar a datos erróneos.

Dado que se cuenta con los datos de temperatura exterior por parte del Departamento de Geografía y Ordenación del Territorio de la Universidad de Zaragoza, los cuales se muestran en el *Anexo II*, se decide desestimar el uso de los datos recogidos en el edificio para esta variable.

Los datos recogidos para las demás variables en la entrada se considera que son correctos.

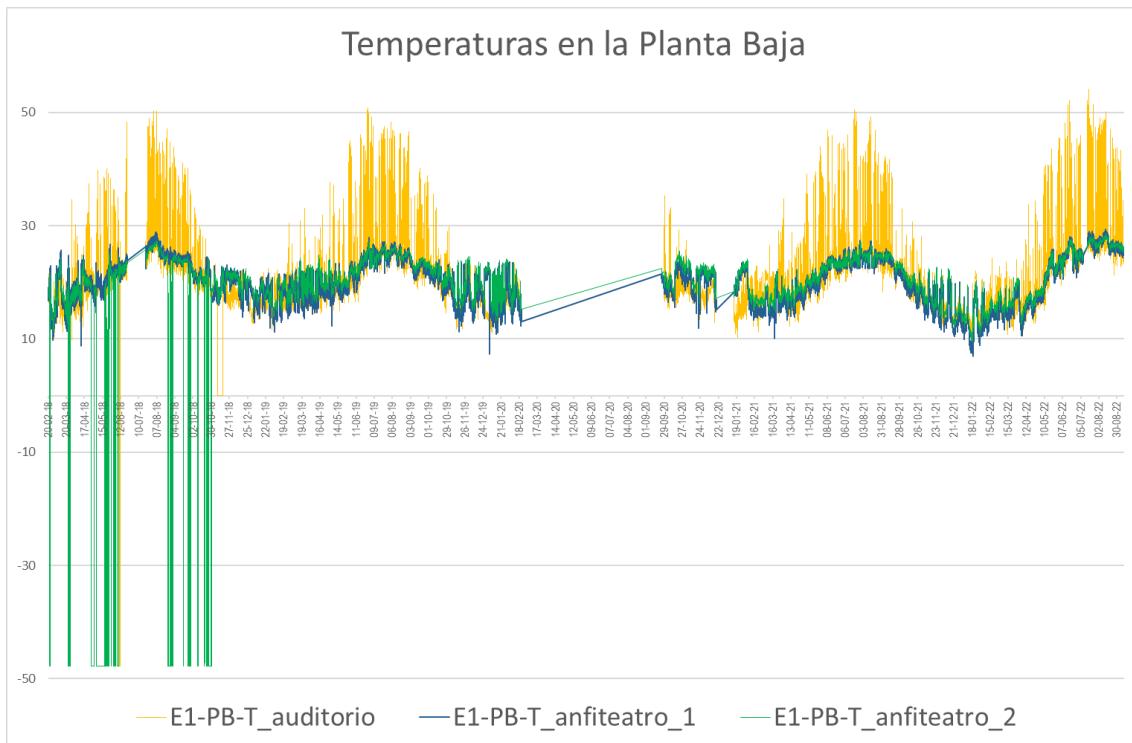


Ilustración 15. Gráfico inicial temperaturas en la planta baja del Edificio 1. Fuente: elaboración propia

Tabla 6. Valores máximos y mínimos de los datos de la planta baja del Edificio 1. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS		
VARIABLE	MÁXIMO	MÍNIMO
Temperatura anfiteatro 1 [°C]	29,3	7,05
Sistema calor anfiteatro 1 [%]	100,07	0
Sistema frío anfiteatro 1 [%]	100	0
Temperatura anfiteatro 2 [°C]	28,66	-47,83
Sistema calor anfiteatro 2 [%]	100,07	0
Sistema frío anfiteatro 2 [%]	100	0
Temperatura auditorio [°C]	54,07	-47,83
Sistema calor auditorio [%]	100,07	0
Sistema frío auditorio [%]	100,07	0

En este caso, se observan errores en los datos obtenidos para la planta baja del Edificio 1.

Respecto a la temperatura 2 del anfiteatro, esta presenta errores para sus valores mínimos, dando valores por debajo de -47°C. En este caso, estos valores son puntuales, como puede observarse en la *Ilustración 13*, lo cual puede deberse a fallos puntuales del sensor.

Esto mismo ocurre con los valores de temperatura en el auditorio, con valores inferiores a -47°C y superiores a 50°C. Tratándose de datos de temperatura interior de un edificio, no resulta lógico que superen los 30°C en tantos puntos como se ve en la *Ilustración 13*. Se ha decidido desestimar esta variable, ya que el sensor está probablemente dañado y estos datos alterarían los resultados coherentes.

Los datos recogidos para las demás variables se considera que son correctos. Respecto al porcentaje máximo de los sistemas de frío y calor, se ha considerado que el dato 100,07% corresponde en realidad al 100% de apertura de la válvula.

Tabla 7. Valores máximos y mínimos de los datos de la Planta 2 del Edificio 1. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS		
VARIABLE	MÁXIMO	MÍNIMO
Temperatura sala [°C]	28,37	11,55
Sistema calor [%]	100,07	0
Sistema frío [%]	100	0

Tabla 8. Valores máximos y mínimos de los datos de la Planta 3 del Edificio 1. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS		
VARIABLE	MÁXIMO	MÍNIMO
Temperatura sala [°C]	30,96	12,45
Sistema calor [%]	100,07	0
Sistema frío [%]	100,07	0

Tabla 9. Valores máximos y mínimos de los datos de la Cubierta del Edificio 1. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS		
VARIABLE	MÁXIMO	MÍNIMO
Sistema calor primario [%]	100,07	0
Sistema frío primario [%]	100,07	0

Tras la valoración conjunta de los datos climáticos junto a los datos obtenidos del edificio Etopía, se concretó realizar el estudio para aquellos datos recogidos desde el 2021, con objeto de asegurar la coherencia temporal y reducir la presencia de datos que no resultan representativos.

Se muestra a continuación cómo resultan los gráficos anteriormente mostrados, así como una tabla resumen de los valores máximos y mínimos de las diferentes variables, tras el preprocesamiento explicado.

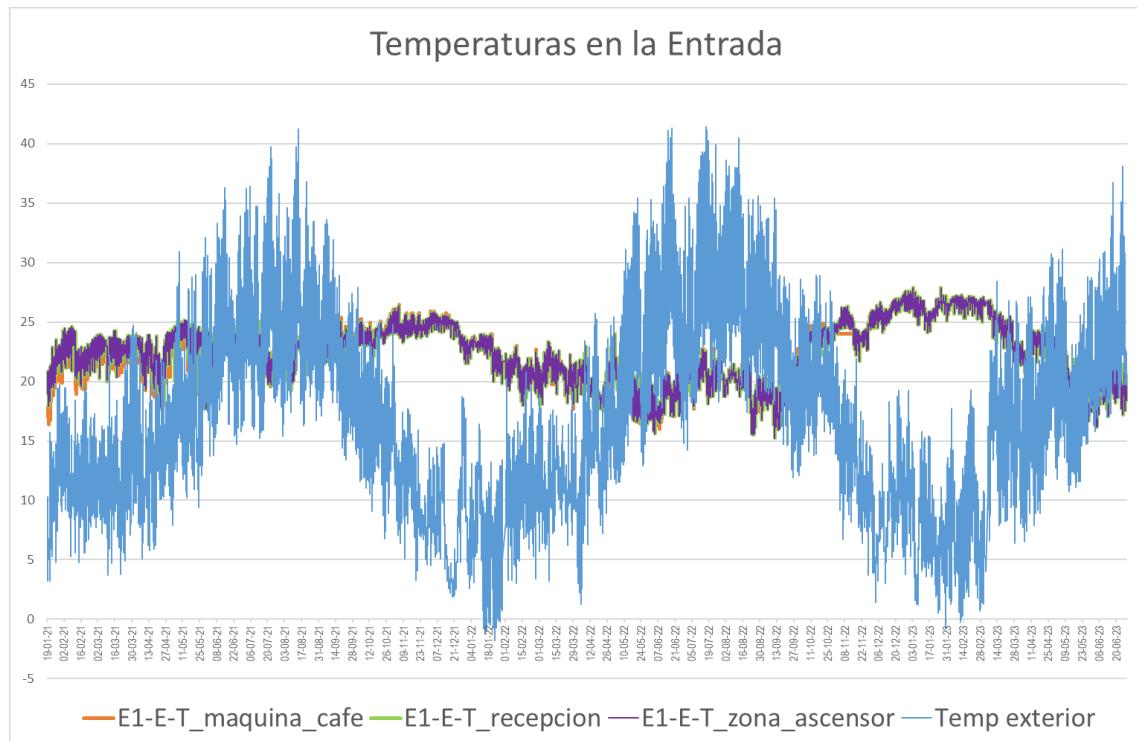


Ilustración 16. Gráfico temperaturas en la entrada del Edificio 1 después del preprocesamiento. Fuente: elaboración propia

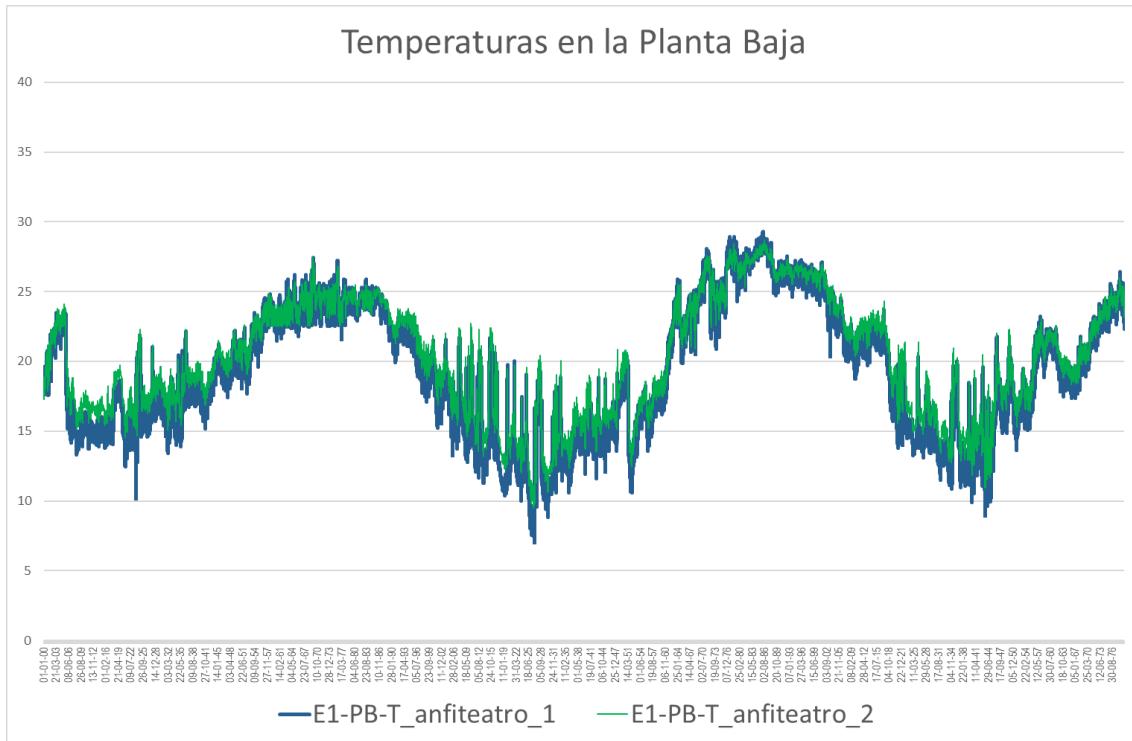


Ilustración 17. Gráfico temperaturas en la planta baja del Edificio 1 después del preprocesamiento. Fuente: elaboración propia

Tabla 10. Valores máximos y mínimos de los datos del Edificio 1 después del preprocesamiento. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS			
ZONA	VARIABLE	MÁXIMO	MÍNIMO
ENTRADA	Temperatura exterior [°C]	41,40	-1,80
	Temperatura máquina café [°C]	26,45	16,4
	Temperatura recepción [°C]	26,34	17,48
	Temperatura zona ascensor [°C]	26,34	17,48
	Suelo radiante [%]	100	0
PLANTA BAJA	Temperatura anfiteatro 1 [°C]	27,43	10,23
	Sistema calor anfiteatro 1 [%]	100	0
	Sistema frío anfiteatro 1 [%]	100	0
	Temperatura anfiteatro 2 [°C]	27,46	12,27
	Sistema calor anfiteatro 2 [%]	100	0
	Sistema frío anfiteatro 2 [%]	100	0
PLANTA 2	Temperatura sala [°C]	28,37	11,55
	Sistema calor [%]	100	0
	Sistema frío [%]	100	0
PLANTA 3	Temperatura sala [°C]	29,97	12,45
	Sistema calor [%]	100	0
	Sistema frío [%]	100	0
CUBIERTA	Sistema calor primario [%]	100	0
	Sistema frío primario [%]	100	0

En la tabla previa se muestran los nuevos valores entre los que se encuentran las temperaturas y el uso de los sistemas de climatización, tras el preprocesamiento aplicado a los datos. Puede observarse que estos datos se han normalizado. Las temperaturas dentro de los edificios no superan los 30°C ni llegan a bajar de 10°C en ningún caso. La temperatura en el exterior, por otro lado, tiene un valor máximo de 41,2°C y un mínimo de 1,9°C.

A1.2. PREPROCESAMIENTO DE LOS DATOS DEL EDIFICIO 2

En este punto se recoge el preprocesamiento de los datos del Edificio 2, acorde a lo explicado en el apartado anterior. A continuación, se muestran las tablas de los máximos y mínimos de cada una de las variables, de todos los datos recogidos desde 2018 hasta la fecha, para que puedan observarse los picos que puedan resultar incoherentes.

Tabla 11. Valores máximos y mínimos de los datos del Edificio 2 antes del preprocesamiento. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS			
ZONA	VARIABLE	MÁXIMO	MÍNIMO
OPEN ART	Temperatura sala [°C]	29,38	13,79
	Sistema calor [%]	100	0
	Sistema frío [%]	100,07	0
PLANTA 2	Temperatura 1 [°C]	31,78	14,34
	Temperatura 2 [°C]	32,67	13,61
	Suelo radiante [%]	100	0
	Fancoil 1 [%]	100	0
	Fancoil 2 [%]	100	0
	Fancoil 3 [%]	100	0
PLANTA 3	Temperatura 1 [°C]	32,47	14,63
	Temperatura 2 [°C]	33,3	15,14
	Suelo radiante [%]	100	0
	Fancoil 1 [%]	100	0
	Fancoil 2 [%]	100	0
	Fancoil 3 [%]	100	0
PLANTA 4	Temperatura 1 [°C]	33,42	15,94
	Temperatura 2 [°C]	34,47	15,62
	Suelo radiante [%]	100	0
	Fancoil 1 [%]	100	0
	Fancoil 2 [%]	100	0
	Fancoil 3 [%]	100	0
PLANTA 5	Temperatura 1 [°C]	34,19	19,38
	Temperatura 2 [°C]	34,8	17,07
	Suelo radiante [%]	100	0
	Fancoil 1 [%]	100	0
	Fancoil 2 [%]	100	0

	Fancoil 3 [%]	100	0
PLANTA 6	Temperatura sala [°C]	31,87	17,32
	Sistema calor [%]	100,07	0
	Sistema frío [%]	100,07	0
CUBIERTA	Sistema calor primario [%]	100,07	0
	Sistema frío primario [%]	100,07	0

Puede observarse que no existen datos de temperaturas o de uso de los sistemas de climatización que correspondan a picos o extremos ilógicos, sino que los datos recogidos están dentro de los valores esperados. Es por ello que no se ha considerado representar estos datos gráficamente en este apartado. Respecto al porcentaje máximo de los sistemas de frío y calor, se considera que el dato 100,07% corresponde en realidad al 100% de apertura de la válvula.

No obstante, cabe recalcar que estos datos corresponden a los recogidos en el edificio 2 desde 2018 y, para adecuarse al resto del estudio, se han modificado del mismo modo que los correspondientes al edificio 1, dejando únicamente aquellos datos desde 2021. Por tanto, se muestra a continuación la tabla resumen con los datos desde 2021 hasta la fecha.

Tabla 12. Valores máximos y mínimos de los datos del Edificio 2 después del preprocesamiento. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS			
ZONA	VARIABLE	MÁXIMO	MÍNIMO
OPEN ART	Temperatura sala [°C]	28,39	14,67
	Sistema calor [%]	100	0
	Sistema frío [%]	100	0
PLANTA 2	Temperatura 1 [°C]	31,78	14,34
	Temperatura 2 [°C]	32,67	13,61
	Suelo radiante [%]	100	0
	Fancoil 1 [%]	100	0
	Fancoil 2 [%]	100	0
	Fancoil 3 [%]	100	0
PLANTA 3	Temperatura 1 [°C]	32,47	14,63
	Temperatura 2 [°C]	33,3	15,14
	Suelo radiante [%]	100	0
	Fancoil 1 [%]	100	0
	Fancoil 2 [%]	100	0
	Fancoil 3 [%]	100	0
PLANTA 4	Temperatura 1 [°C]	33,42	15,94
	Temperatura 2 [°C]	34,47	15,62
	Suelo radiante [%]	100	0
	Fancoil 1 [%]	100	0
	Fancoil 2 [%]	100	0

	Fancoil 3 [%]	100	0
PLANTA 5	Temperatura 1 [°C]	34,19	19,38
	Temperatura 2 [°C]	34,8	17,07
	Suelo radiante [%]	100	0
	Fancoil 1 [%]	100	0
	Fancoil 2 [%]	100	0
	Fancoil 3 [%]	100	0
PLANTA 6	Temperatura sala [°C]	31,87	17,65
	Sistema calor [%]	100	0
	Sistema frío [%]	100	0
CUBIERTA	Sistema calor primario [%]	100	0
	Sistema frío primario [%]	100	0

A1.3. PREPROCESAMIENTO DE LOS DATOS DEL EDIFICIO 3

En este punto se recoge el preprocesamiento de los datos del Edificio 3, con concordancia a lo explicado en los apartados previos. A continuación, se muestran las tablas de los máximos y mínimos de cada una de las variables, de todos los datos recogidos desde 2018 hasta la fecha, para que puedan observarse los picos que puedan resultar incoherentes.

Tabla 13. Valores máximos y mínimos de los datos del Edificio 3 antes del preprocesamiento. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS			
ZONA	VARIABLE	MÁXIMO	MÍNIMO
AULA 1	Temperatura sala [°C]	33,04	13,6
	Sistema calor [%]	100,07	0
	Sistema frío [%]	100,07	0
AULA 2	Temperatura sala [°C]	30,71	15,53
	Sistema calor [%]	100,07	0
	Sistema frío [%]	100,07	0
AULA 3	Temperatura sala [°C]	32,96	14,09
	Sistema calor [%]	100,07	0
	Sistema frío [%]	100,07	0
ZONA COMÚN	Temperatura sala 1 [°C]	31,99	14,67
	Temperatura sala 2 [°C]	31,7	14,77
	Sistema calor 1 [%]	100,07	0
	Sistema calor 2[%]	100,07	0
	Sistema frío 1 [%]	100,07	0
	Sistema frío 2 [%]	100,07	0
CUBIERTA	Sistema calor primario [%]	100,07	0
	Sistema frío primario [%]	100,07	0

En este caso, tampoco existen datos de temperaturas o de uso de los sistemas de climatización que correspondan a picos incoherentes, sino que los datos recogidos también resultan dentro

de la normalidad. Es por ello que tampoco se han representado estos datos gráficamente en este apartado. Respecto al porcentaje máximo de los sistemas de frío y calor, se considera que el dato 100,07% corresponde en realidad al 100% de apertura de la válvula.

No obstante, estos datos corresponden a los recogidos desde 2018 y, para tener concordancia con los anteriores apartados, se han modificado del mismo modo que los de los edificios 1 y 2, dejando únicamente los datos desde 2021. Por tanto, se muestra a continuación la tabla resumen con los datos desde 2021 hasta la fecha.

Tabla 14. Valores máximos y mínimos de los datos del Edificio 3 después del preprocesamiento. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS			
ZONA	VARIABLE	MÁXIMO	MÍNIMO
AULA 1	Temperatura sala [°C]	32,37	14,82
	Sistema calor [%]	100	0
	Sistema frío [%]	100	0
AULA 2	Temperatura sala [°C]	30,71	16,05
	Sistema calor [%]	100	0
	Sistema frío [%]	100	0
AULA 3	Temperatura sala [°C]	32,96	14,94
	Sistema calor [%]	100	0
	Sistema frío [%]	100	0
ZONA COMÚN	Temperatura sala 1 [°C]	31,99	14,67
	Temperatura sala 2 [°C]	31,7	14,77
	Sistema calor 1 [%]	100	0
	Sistema calor 2[%]	100	0
	Sistema frío 1 [%]	100	0
	Sistema frío 2 [%]	100	0
CUBIERTA	Sistema calor primario [%]	100	0
	Sistema frío primario [%]	100	0

ANEXO II. PREPROCESAMIENTO DE LOS DATOS CLIMÁTICOS DE LA CIUDAD DE ZARAGOZA

En el preprocesamiento de los datos climáticos de la ciudad de Zaragoza se ha utilizado, en primera instancia, al igual que para los datos del edificio Etopía, la herramienta de Excel.

En primer lugar, los datos “brutos” se trasladaron a Excel, para poder trabajar con ellos convenientemente. Se ordenaron por fecha y hora, desde los más antiguos hasta los más recientes. Posteriormente, se obtuvieron los máximos y mínimos de todas las variables consideradas importantes para poder detectar valores atípicos, igual que se realizó con los datos energéticos del edificio Etopía.

A continuación, se muestran las tablas de valores máximos y mínimos, así como los gráficos de las variables de temperatura y humedad en el periodo del estudio. Gracias a ello, se pueden percibir los picos que pueden ser debidos a datos erróneos o nulos.

Tabla 15. Valores máximos y mínimos de los datos climáticos de Zaragoza. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS		
VARIABLE	MÁXIMO	MÍNIMO
Temperatura exterior [°C]	41,40	-1,80
Humedad exterior [%]	96	0
Velocidad del viento [m/s]	30,60	0,00
Cantidad de lluvia [mm]	10,20	0,00

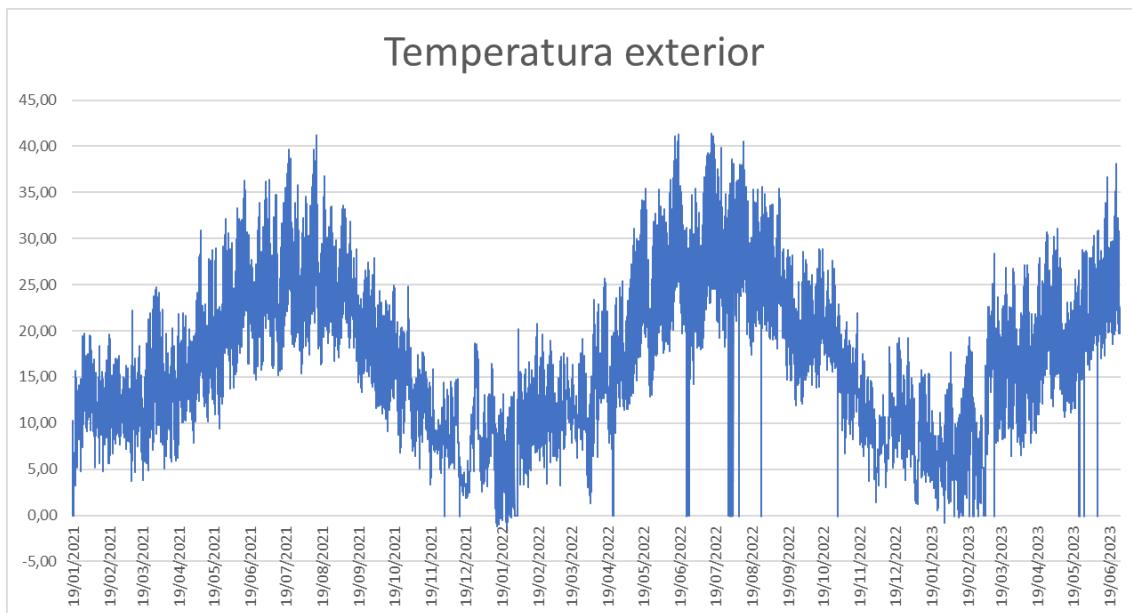


Ilustración 18. Gráfico temperatura exterior antes del preprocesamiento. Fuente: elaboración propia

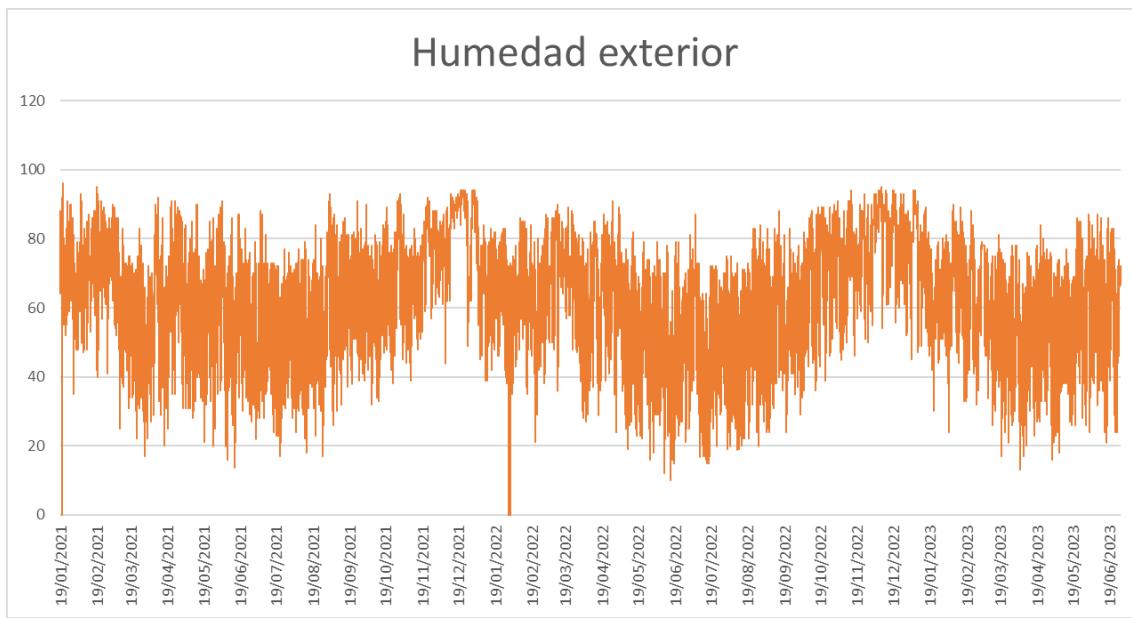


Ilustración 19. Gráfico humedad exterior antes del preprocesamiento. Fuente: elaboración propia

Como puede apreciarse mediante la observación de estos gráficos y la *tabla 11*, existen anomalías en los datos correspondientes a la temperatura y a la humedad de la ciudad. Estos picos correspondientes a un valor 0 para estas variables, se deben a datos nulos en dichos momentos. Es decir, momentos en los que los sensores tuvieron un funcionamiento incorrecto.

Por tanto, para el caso de los datos climáticos de la ciudad de Zaragoza, el cribado necesario en primer lugar fue el de estos datos nulos que otorgaron los sensores. Tras esta modificación, se enseña a continuación la tabla de máximos y mínimos, así como los gráficos de temperatura y humedad que presentaban irregularidades, en este caso sin tales anomalías.

Tabla 16. Valores máximos y mínimos de los datos climáticos de Zaragoza después del preprocesamiento. Fuente: elaboración propia

MÁXIMOS Y MÍNIMOS		
VARIABLE	MÁXIMO	MÍNIMO
Temperatura exterior [°C]	41,40	-1,80
Humedad exterior [%]	96	10
Velocidad del viento [m/s]	30,60	0,00
Cantidad de lluvia [mm]	10,20	0,00

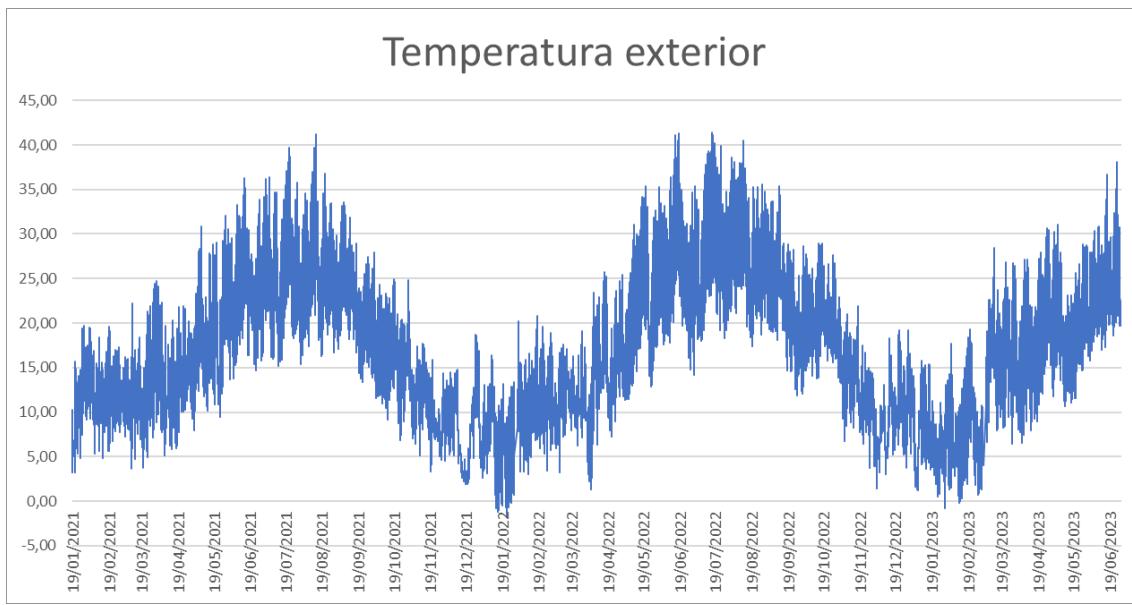


Ilustración 20. Gráfico temperatura exterior después del preprocesamiento. Fuente: elaboración propia

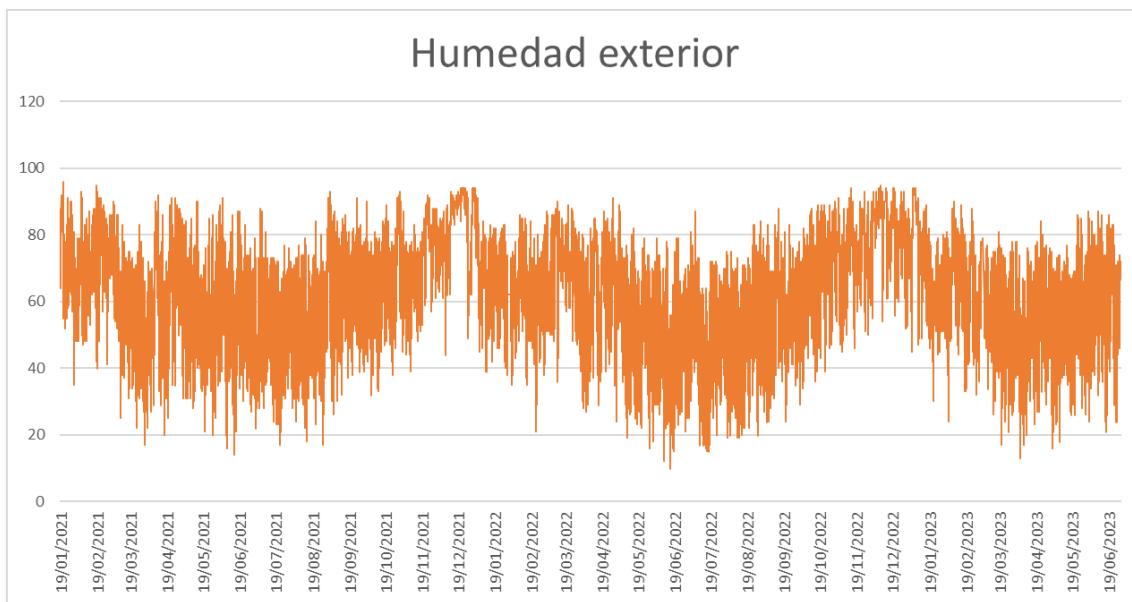


Ilustración 21. Gráfico humedad exterior después del preprocesamiento. Fuente: elaboración propia

ANEXO III. CÓDIGO DE PYTHON PARA EL PREPROCESAMIENTO DE LOS DATOS

A3.1. PREPROCESAMIENTO DE LOS DATOS CLIMÁTICOS Y ENERGÉTICOS

En este anexo del proyecto se recoge un ejemplo del código de Python utilizado para la unión de los datos climáticos de la ciudad de Zaragoza y de los datos térmicos y energéticos del edificio.

Para la realización del proyecto, los datos recogidos se dividieron en diferentes archivos de Excel, según el año al que pertenecen y al edificio al que se refieren. Dentro de los archivos Excel, los datos quedan divididos en diferentes hojas según la zona del edificio a la que pertenecen los datos. Las zonas del edificio se explican en el apartado 4. *RECOPILACIÓN DE DATOS Y PREPROCESAMIENTO*.

Se ha realizado de esta manera para agilizar el procesamiento a través del código escrito en Python. Además de optimizar el uso de la herramienta de Excel ya que, de lo contrario, esta se bloqueaba en diversas ocasiones al intentar trabajar con todo el volumen de datos del que se dispone.

El ejemplo en concreto hace referencia a los datos del Edificio 3 en el año 2023, concretamente de la zona denominada Aula 1. Para el resto de los edificios, zonas y años se cambian las respectivas variables en el código.

```
import pandas as pd

from openpyxl import load_workbook

#Rutas de los archivos Excel

archivo_clima = 'Datos_Climaticos.xlsx'
archivo_edificio = 'Edificio_3_2023.xlsx'
archivo_salida = 'coincidenciasEd3_2023.xlsx'

#Nombres de las hojas que contienen los datos en ambos archivos
hoja_clima = 'DATOS_2023'
hoja_edificio = 'Aula_1'

#Cargar los datos de los archivos Excel en DataFrames de pandas

df_clima = pd.read_excel(archivo_clima, sheet_name=hoja_clima)
```

```

df_edificio=pd.read_excel(archivo_edificio,sheet_name=hoja_edificio)

#Buscar coincidencias de fecha y hora en ambos DataFrames

coincidencias = pd.merge(df_clima, df_edificio, on=['Fecha',
'Hora'], how='inner')

#Eliminar filas duplicadas en caso de que existan

coincidencias = coincidencias.drop_duplicates()

#Cargar el archivo Excel de salida si ya existe

try:

    book = load_workbook(archivo_salida)

    writer = pd.ExcelWriter(archivo_salida, engine='openpyxl')

    writer.book = book

except FileNotFoundError:

    #Si el archivo no existe, crea uno nuevo

    writer = pd.ExcelWriter(archivo_salida, engine='xlsxwriter')

#Aregar el DataFrame 'coincidencias' a una hoja existente o nueva

coincidencias.to_excel(writer, sheet_name='ClimayA1', index=False)

#Guardar el archivo Excel con las coincidencias

writer.save()

```

A3.1. PREPROCESAMIENTO PARA EL MODELO TÉRMICO

En este anexo del proyecto se recoge un ejemplo del código de Python utilizado para el cribado de los datos para los que los sistemas de climatización del edificio estuvieron en funcionamiento. Se quiere analizar así, separadamente, el comportamiento térmico del edificio debido a las condiciones ambientales externas.

El ejemplo en concreto hace referencia a los datos del Edificio 3 en el año 2023, concretamente de la zona denominada Aula 1. Para el resto de los edificios, zonas y años se cambian las respectivas variables en el código.

```

import pandas as pd

from openpyxl import load_workbook

#Ruta del archivo Excel de origen

archivo_origen = 'ClimayEd3_2023.xlsx'

#Nombre de la hoja que deseas procesar

hoja_origen = 'ClimayA1'

#Leer el archivo Excel de origen en un DataFrame

df_origen = pd.read_excel(archivo_origen, sheet_name=hoja_origen)

#Definir las columnas en las que deseas evaluar los valores

columnas_evaluadas = ['E3-Aula_1-E_calor', 'E3-Aula_1-E_frio']

#Filtrar las filas donde al todas las columnas evaluadas sea igual a
#"0"

df_filtrado      = df_origen[df_origen[columnas_evaluadas].eq(0).all(
axis=1)]

#Ruta del nuevo archivo Excel de salida

archivo_salida = 'Ed3_2023.xlsx'

#Intentar cargar el archivo Excel de salida si ya existe

try:

    book = load_workbook(archivo_salida)

    writer = pd.ExcelWriter(archivo_salida, engine='openpyxl')

    writer.book = book

except FileNotFoundError:

    #Si el archivo no existe, crea uno nuevo

    writer = pd.ExcelWriter(archivo_salida, engine='openpyxl')

#Agregar el DataFrame filtrado a una hoja existente o nueva

df_filtrado.to_excel(writer,      sheet_name=f'Filtrado_{hoja_origen}', 
index=False)

#Guardar el archivo Excel con las filas filtradas

writer.save()

```

ANEXO IV. MODELOS DE REGRESIÓN APLICADOS

A4.1. REGRESIÓN LINEAR MÚLTIPLE

La regresión lineal múltiple (MLR por su término en inglés Multiple Linear Regression) es un método estadístico que intenta modelar la relación entre una variable continua y dos o más variables independientes mediante el ajuste de una ecuación lineal. [23]

La regresión lineal múltiple permite generar un modelo en el que el valor de la variable dependiente o respuesta (y) se determina a partir de un conjunto de variables independientes o predictoras ($x_1, x_2, x_3\dots$).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip} + \epsilon_i \quad Ec. 3$$

En este caso, se hace referencia al valor de y para una observación i concreta, mientras p es el número de predictores. La interpretación de los elementos del modelo es la siguiente:

- β_0 : es la ordenada en el origen, se corresponde con el valor promedio de la variable respuesta y cuando todos los predictores son cero.
- β_j : es el efecto promedio que tiene sobre la variable respuesta el incremento en una unidad de la variable predictora x_j , manteniéndose constantes el resto de las variables. Son los coeficientes parciales de regresión.
- ϵ : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo. Recoge el efecto de todas aquellas variables que influyen en y pero que no se incluyen en el modelo como predictoras.

En la gran mayoría de casos, los valores β_0 y β_j se desconocen, por lo que, a partir de una muestra, se obtienen sus estimaciones $\hat{\beta}_0$ y $\hat{\beta}_j$.

Ajustar el modelo consiste en estimar, con los datos disponibles, los valores de los coeficientes de regresión que dan lugar al modelo que con mayor probabilidad puede haber generado los datos observados.

El método más empleado es el ajuste por mínimos cuadrados ordinarios (OLS), que identifica como mejor modelo el plano que minimiza la suma de las desviaciones verticales entre cada dato de entrenamiento y la recta, elevadas al cuadrado. Expresado de forma matricial:

$$y = X^T \beta + \epsilon \quad Ec. 4$$

$$\hat{\beta} = \arg \min_{\beta} (y - X^T \beta)^2 \quad Ec. 5$$

Una vez estimados los coeficientes, se obtienen las estimaciones de cada observación, \hat{y}_j :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \cdots + \hat{\beta}_p x_{ip} \quad \text{Ec. 6}$$

A4.2. MÁQUINAS DE SOPORTE VECTORIAL PARA REGRESIÓN

Las Máquinas de Vectores de Soporte (SVM por su término en inglés Support Vector Machine) son una herramienta destacada en aprendizaje automático, aplicándose tanto en problemas de clasificación múltiple como en regresión. Este algoritmo se basa en el concepto de Maximal Margin Classifier, centrado en la idea de hiperplanos en espacios de múltiples dimensiones.

En un espacio de p dimensiones, un hiperplano se define como un subespacio plano y afín de dimensiones $p - 1$. Aunque visualizar hiperplanos en dimensiones superiores a tres no es intuitivo, el concepto de subespacio con $p-1$ dimensiones se mantiene.

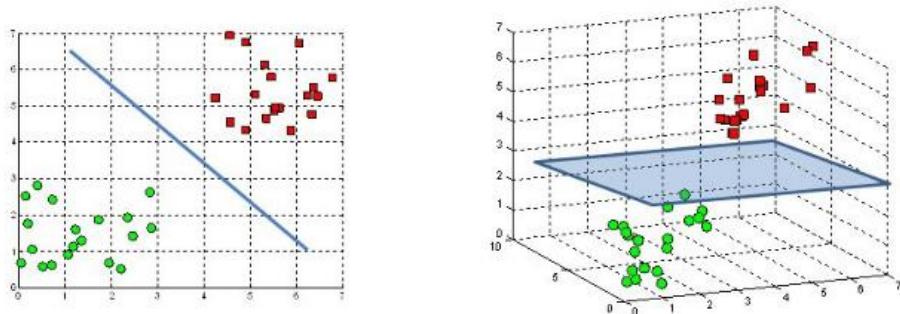


Ilustración 22. Hiperplanos en 2 dimensiones (izquierda) y 3 dimensiones (derecha). Fuente: Medium.

El hiperplano óptimo, conocido como Maximal Margin Hyperplane, es aquel que se encuentra más alejado de todas las observaciones de entrenamiento. La Ilustración 17 muestra el Maximal Margin Hyperplane, (línea negra continua) y su margen (las dos líneas discontinuas). Las tres observaciones equidistantes respecto al Maximal Margin Hyperplane que se encuentran a lo largo de las líneas discontinuas son los vectores soporte, que definen el maximal margin hyperplane.

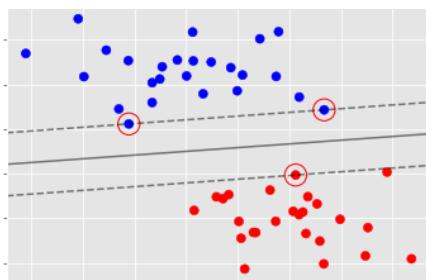


Ilustración 23. Ejemplo de Maximal Margin Hyperplane. Fuente: Ciencia de Datos

Sin embargo, el Maximal Margin Classifier tiene limitaciones prácticas, ya que raramente se encuentran casos donde las clases son perfecta y linealmente separables. Para abordar esto, surgen los Support Vector Classifiers, que permiten ciertas violaciones del margen o incluso del hiperplano.

El proceso de identificar el hiperplano correcto introduce el hiperparámetro C . Este controla el número y la severidad de las violaciones del margen que se toleran en el ajuste. Con $C = \infty$ no se permiten violaciones, equivaliendo al Maximal Margin Classifier. A medida que C se acerca a cero, se penalizan menos los errores, permitiendo más observaciones en el lado incorrecto del margen.

El Support Vector Classifier funciona bien cuando la separación entre clases es aproximadamente lineal. Sin embargo, su rendimiento decae en situaciones no lineales. Para abordar esto, se aumentan las dimensiones del espacio original. Aunque los grupos no sean linealmente separables en el espacio original, podrían serlo en un espacio de mayores dimensiones, como se observa en la *Ilustración 18*.

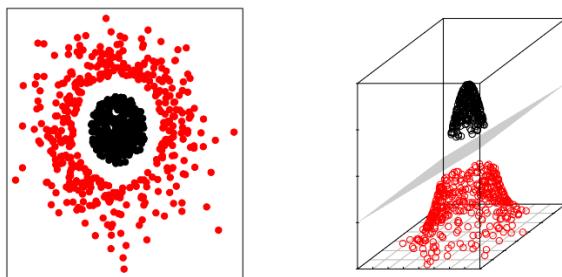


Ilustración 24. Ejemplo de datos con separación no lineal en dos dimensiones. Fuente: Ciencia de Datos

El método de Máquinas Vector Soporte (SVM) es una extensión del Support Vector Classifier obtenida al aumentar la dimensión de los datos. Los límites de separación lineales generados en el espacio aumentado se convierten en límites de separación no lineales al proyectarlos en el espacio original.

Para aumentar la dimensión, se recurre a funciones de transformación. Aquí es donde entra en juego el concepto de kernel, una función que devuelve el resultado del producto escalar entre dos vectores en un nuevo espacio dimensional. La elección del kernel es crucial, y uno común es el kernel lineal, el cual se utiliza en este proyecto, preferido en situaciones de gran volumen de datos debido a su menor gasto computacional en comparación con otros kernels. [24] [25]

A4.3. REGRESIÓN POR K-VECINOS MÁS CERCANOS

El algoritmo de K-vecinos más cercanos (o k-NN, K-Nearest Neighbors) es uno de los algoritmos de aprendizaje más utilizados por la sencillez que presenta. Es un algoritmo de aprendizaje supervisado que funciona bajo la premisa de que “las cosas que están cerca unas de otras son parecidas entre sí”.

Se trata de un método no paramétrico, esto significa que, a diferencia de algunos modelos paramétricos que asumen una forma específica para la relación entre las variables, k-NN clasifica o predice nuevos puntos de datos basándose en la similitud con los puntos de datos cercanos en el conjunto de entrenamiento.

Cuando se utiliza k-NN, se quieren encontrar los k vecinos más cercanos a un punto de consulta. Para ello, se calcula la distancia Euclídea entre el punto de consulta y todos los demás puntos en el conjunto de datos.

La distancia Euclídea es una medida de la "longitud" del segmento de línea que conecta dos puntos en un espacio Euclídeo. Para dos puntos p y q con coordenadas p_i y q_i en cada dimensión i , la fórmula de la distancia Euclídea en un espacio de m dimensiones es:

$$d(p, q) = \sqrt{\sum_{i=1}^m (q_i - p_i)^2} \quad Ec. 7$$

Luego, se seleccionan los k puntos con las distancias Euclídeas más pequeñas, lo que significa que son los vecinos más cercanos.

El valor que devuelve la regresión por k-NN es la media de los valores de las observaciones o k vecinos a los que apuntan estas k distancias más cercanas. Este enfoque pondera igualmente la contribución de todos los vecinos más cercanos.

Definir el hiperparámetro k requiere buscar el equilibrio, ya que diferentes valores pueden llevar a un sobreajuste o desajuste [27] [28].

A.4.4. REGRESIÓN POR BOSQUES ALEATORIOS

Un modelo Random Forest está formado por múltiples árboles de decisión individuales. Cada uno de estos árboles es entrenado con una muestra diferente de los datos de entrenamiento, generada mediante bootstrapping. El proceso de bootstrapping implica seleccionar

aleatoriamente observaciones del conjunto de datos original, permitiendo que una observación específica se elija más de una vez o no se elija en absoluto en una muestra específica.

Los modelos de Random Forest se fundamentan en la estrategia conocida como bagging. El bagging se basa en el ajuste de múltiples modelos, cada uno con un subconjunto distinto de los datos de entrenamiento. Para predecir, todos los modelos que forman el agregado participan aportando su predicción. Como valor final, se toma la media de todas las predicciones, es decir, se combinan las predicciones de todos los árboles que conforman el modelo [29].

Los hiperparámetros se pueden ajustar en un modelo de regresión por bosques aleatorios, en particular para evitar el sobreajuste del modelo, incluyen:

- Número de árboles: Representa la cantidad de árboles en el bosque. Aumentar el número de árboles puede mejorar el rendimiento, pero también aumenta el costo computacional.
- Profundidad máxima de los árboles: Controla la profundidad máxima de cada árbol en el bosque. Limitar la profundidad puede ayudar a prevenir el sobreajuste.
- Número mínimo de muestras para dividir un nodo: Especifica el número mínimo de muestras requeridas para dividir un nodo interno. Aumentar este valor puede ayudar a prevenir el sobreajuste.

ANEXO V. RESULTADOS DE LOS MODELOS TÉRMICOS DE REGRESIÓN

A5.1. RESULTADOS DEL MODELO DE REGRESIÓN LINEAL MÚLTIPLE

El modelo mediante regresión lineal múltiple se implementa con el código que puede encontrarse en el *Anexo VII. Códigos de Python para modelos de regresión térmica*. Pueden verse a continuación los datos obtenidos en la *Tabla 17. Resultados del modelo térmico de regresión lineal múltiple*.

Tabla 17. Resultados del modelo térmico de regresión lineal múltiple. Fuente: elaboración propia

REGRESIÓN LINEAL MÚLTIPLE					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	T maquina café	0,7248	0,5462	1,3793
		T recepción	0,5832	0,3874	1,6244
		T zona ascensor	0,5832	0,3874	1,6244
	Planta baja	T anfiteatro 1	0,7984	0,8232	1,8457
		T anfiteatro 2	0,7683	0,7864	1,8363
		T auditorio	0,7809	0,8307	2,3255
	Planta 2	Temperatura sala	0,6934	0,6239	2,2491
	Planta 3	Temperatura sala	0,6599	0,6828	2,1534
	Open Art	Temperatura sala	0,7082	0,7133	1,5556
EDIFICIO 2	Planta 2	Temperatura 1	0,5309	0,4225	1,5858
		Temperatura 2	0,6581	0,5054	1,6399
	Planta 3	Temperatura 1	0,5994	0,4793	1,6189
		Temperatura 2	0,5232	0,4909	1,7168
	Planta 4	Temperatura 1	0,5514	0,49	1,5872
		Temperatura 2	0,5753	0,5237	1,6525
	Planta 5	Temperatura 1	0,6119	0,5468	1,7534
		Temperatura 2	0,6119	0,5739	1,7659
	Planta 6	Temperatura sala	0,6846	0,6462	1,423
EDIFICIO 3	Aula 1	Temperatura sala	0,6656	0,6274	1,7753
	Aula 2	Temperatura sala	0,6794	0,7287	1,4361
	Aula 3	Temperatura sala	0,7743	0,8101	1,3809
	Zona común	Temperatura sala 1	0,854	0,8229	1,262
		Temperatura sala 2	0,7887	0,7725	1,4602

Se observa que, para este caso, el coeficiente de determinación R² en el entrenamiento da un valor mínimo de 0,5232 para una de las temperaturas en la Planta 3 del Edificio 2, y un valor máximo de 0,854 para la una de las temperaturas medidas en la Zona Común del Edificio 3.

Estos valores deben ser comparados con los obtenidos sobre el conjunto de validación cruzada, ya que R² calculado sobre este conjunto mide la capacidad del modelo para generalizar a datos

no vistos. En este caso, se observa un valor mínimo de 0,3874 para la temperatura en la zona del ascensor del Edificio 1, y uno máximo de 0,8307 para la temperatura en el auditorio.

Por último, los valores del error absoluto medio (MAE) se calculan también sobre el conjunto de validación cruzada. Este valor representa el tamaño promedio de los errores en la predicción, en este caso en grados centígrados. Se observa que las mayores desviaciones se producen en las temperaturas en el auditorio, alcanzando 2,32°C, mientras las mínimas se dan en la una de las temperaturas medidas en la Zona Común, siendo 1,262 °C.

A5.2. RESULTADOS DEL MODELO DE REGRESIÓN POR MÁQUINA DE VECTORES

El modelo de regresión por máquina de vectores se implementa con el código escrito en el Anexo VII. *Códigos de Python para modelos de regresión térmica*. Pueden verse a continuación los datos obtenidos en la *Tabla 18. Resultados del modelo térmico de regresión por máquina de vectores*.

Tabla 18. Resultados del modelo térmico de regresión por máquina de vectores. Fuente: elaboración propia

MÁQUINA DE VECTORES					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	T maquina café	0,7232	0,5335	1,4062
		T recepción	0,5812	0,3668	1,6553
		T zona ascensor	0,5811	0,3681	1,6535
	Planta baja	T anfiteatro 1	0,7972	0,8288	1,8054
		T anfiteatro 2	0,7673	0,794	1,7921
		T auditorio	0,7451	0,7618	2,4981
	Planta 2	Temperatura sala	0,6916	0,6191	2,2649
	Planta 3	Temperatura sala	0,6517	0,6771	2,1826
	EDIFICIO 2	Open Art	Temperatura sala	0,7067	0,7126
		Planta 2	Temperatura 1	0,5294	0,431
			Temperatura 2	0,6555	0,5119
		Planta 3	Temperatura 1	0,5983	0,4817
			Temperatura 2	0,5117	0,5138
		Planta 4	Temperatura 1	0,5452	0,5012
			Temperatura 2	0,5666	0,5323
	EDIFICIO 3	Planta 5	Temperatura 1	0,6086	0,548
			Temperatura 2	0,6086	0,5672
		Planta 6	Temperatura sala	0,6825	0,6614
	Aula 1	Temperatura sala	0,659	0,6386	1,7344
	Aula 2	Temperatura sala	0,6771	0,7425	1,3981
	Aula 3	Temperatura sala	0,7718	0,81	1,3773
	Zona común	Temperatura sala 1	0,8537	0,8237	1,2566
		Temperatura sala 2	0,7876	0,7818	1,4291

Se aprecia que, para este caso, el coeficiente de determinación R^2 en el entrenamiento otorga un valor mínimo de 0,5117 para una de las temperaturas en la Planta 3 del Edificio 2, el mismo comportamiento que en la regresión lineal múltiple. El valor máximo es de 0,8537 para la una de las temperaturas medidas en la Zona Común del Edificio 3, también la misma que para el modelo de regresión múltiple.

En el caso del conjunto de validación cruzada, se observa un valor mínimo de 0,3668 para la temperatura en la zona de recepción del Edificio 1, y uno máximo de 0,8288 para una de las temperaturas del anfiteatro.

Por último, respecto a los valores del error absoluto medio (MAE), se observa que las mayores desviaciones se producen en las temperaturas en el auditorio, alcanzando 2,4981°C, mientras las mínimas se dan en la una de las temperaturas medidas en la Zona Común, siendo 1,2566 °C.

A5.3. RESULTADOS DEL MODELO DE REGRESIÓN POR K VECINOS MÁS CERCANOS

El modelo mediante regresión por k-vecinos más próximos se implementa con el código que puede encontrarse en el *Anexo VII. Códigos de Python para modelos de regresión térmica*. Pueden verse a continuación los datos obtenidos en la *Tabla 19. Resultados del modelo térmico de regresión por k-vecinos más cercanos*.

Tabla 19. Resultados del modelo térmico de regresión por k-vecinos más cercanos. Fuente: elaboración propia

K-VECINOS MÁS CERCANOS					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	T maquina café	0,8717	0,4846	1,4638
		T recepción	0,804	0,3124	1,7248
		T zona ascensor	0,84	0,3124	1,7248
	Planta baja	T anfiteatro 1	0,9245	0,8014	1,9317
		T anfiteatro 2	0,9126	0,7584	1,9327
		T auditorio	0,92222	0,8575	2,2262
	Planta 2	Temperatura sala	0,8721	0,6052	2,2804
	Planta 3	Temperatura sala	0,857	0,6262	2,3171
	Open Art	Temperatura sala	0,8835	0,6867	1,5863
EDIFICIO 2	Planta 2	Temperatura 1	0,7909	0,4597	1,5279
		Temperatura 2	0,8412	0,52	1,5734
	Planta 3	Temperatura 1	0,808	0,491	1,5456
		Temperatura 2	0,7692	0,4924	1,6734
	Planta 4	Temperatura 1	0,7647	0,4755	1,5677
		Temperatura 2	0,7731	0,5159	1,6294
	Planta 5	Temperatura 1	0,7923	0,5226	1,7618
		Temperatura 2	0,7919	0,5449	1,8109

	Planta 6	Temperatura sala	0,8452	0,6203	1,4492
EDIFICIO 3	Aula 1	Temperatura sala	0,8406	0,5822	1,884
	Aula 2	Temperatura sala	0,8435	0,6893	1,5295
	Aula 3	Temperatura sala	0,8867	0,7824	1,5074
	Zona común	Temperatura sala 1	0,9285	0,809	1,3171
		Temperatura sala 2	0,8962	0,7635	1,4877

Se aprecia que, para este caso, el coeficiente de determinación R^2 en el entrenamiento, tiene valores altos. El valor mínimo es de 0,7647 para una de las temperaturas en la Planta 4 del Edificio 2. El valor máximo es de 0,9285 para la una de las temperaturas medidas en la Zona Común del Edificio 3.

Estos valores podrían apuntar a un buen comportamiento del modelo, no obstante, en el caso del conjunto de validación cruzada, los datos son menos inspiradores. La gran diferencia que se aprecia entre los valores de entrenamiento y de validación cruzada sugiere que el modelo presenta sobreajuste. En el caso de la regresión por k-vecinos más cercanos, se puede ajustar únicamente el hiperparámetro de “número de vecinos”. Al aumentar el número de vecinos, se aumenta el tiempo de cómputo del algoritmo. Se ha probado un aumento fuerte entre los 5 vecinos inicialmente a 1000 vecinos, para ver si esto podría mejorar significativamente los valores en el conjunto de validación cruzada. Se muestran los valores obtenidos a continuación.

Tabla 20. Resultados del modelo térmico modificado de regresión por k-vecinos más cercanos. Fuente: elaboración propia

K-VECINOS MÁS CERCANOS					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	T maquina café	0,7565	0,4786	1,4798
		T recepción	0,6422	0,3146	1,7355
		T zona ascensor	0,6422	0,3146	0,7355
	Planta baja	T anfiteatro 1	0,8326	0,7829	2,0788
		T anfiteatro 2	0,8095	0,7438	2,0405
		T auditorio	0,8466	0,8535	2,2552
	Planta 2	Temperatura sala	0,7484	0,5989	2,3461
	Planta 3	Temperatura sala	0,7303	0,644	2,277
	Open Art	Temperatura sala	0,7775	0,708	1,5424
EDIFICIO 2	Planta 2	Temperatura 1	0,6314	0,4266	1,5397
		Temperatura 2	0,7186	0,4835	1,5996
	Planta 3	Temperatura 1	0,652	0,4271	1,5905
		Temperatura 2	0,6002	0,4509	1,7047
	Planta 4	Temperatura 1	0,6	0,4132	1,6157
		Temperatura 2	0,6165	0,4464	1,6877
	Planta 5	Temperatura 1	0,6511	0,489	1,8063

		Temperatura 2	0,6519	0,5134	1,8412
	Planta 6	Temperatura sala	0,7234	0,6197	1,4497
EDIFICIO 3	Aula 1	Temperatura sala	0,7042	0,6052	1,8666
	Aula 2	Temperatura sala	0,7214	0,6816	1,5357
	Aula 3	Temperatura sala	0,7902	0,7729	1,5559
	Zona común	Temperatura sala 1	0,8444	0,7665	1,4427
		Temperatura sala 2	0,7966	0,7459	1,5439

Puede observarse que, gracias a la mejora en el hiperparámetro del número de vecinos, el modelo empeora su ajuste en el conjunto de entrenamiento y lo mejora ligeramente en el de validación cruzada. Estos valores se compararán más adelante junto a los demás modelos. Cabe destacar que el tiempo de cómputo del algoritmo es mayor cuantos más vecinos se decretan, y no debe aumentarse este número indefinidamente porque crea un modelo más complejo.

A5.4. RESULTADOS DEL MODELO DE REGRESIÓN POR BOSQUES ALEATORIOS

El modelo de regresión realizado por el método de bosques aleatorios se implementa con el código que puede encontrarse en el *Anexo VII. Códigos de Python para modelos de regresión térmica*. Pueden verse a continuación los datos obtenidos en la *Tabla 21. Resultados del modelo térmico de regresión por árboles aleatorios*.

Tabla 21. Resultados del modelo térmico de regresión por árboles aleatorios. Fuente: elaboración propia

ÁRBOLES ALEATORIOS					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	T maquina café	0,9477	0,5013	1,4369
		T recepción	0,9173	0,3253	1,7072
		T zona ascensor	0,9172	0,3244	1,7089
	Planta baja	T anfiteatro 1	0,9716	0,8033	1,9367
		T anfiteatro 2	0,9668	0,7571	1,9394
		T auditorio	0,973	0,8518	2,2601
	Planta 2	Temperatura sala	0,9522	0,6086	2,2611
	Planta 3	Temperatura sala	0,9416	0,6218	2,3275
	Open Art	Temperatura sala	0,9538	0,6795	1,6027
	Planta 2	Temperatura 1	0,9143	0,4602	1,5308
		Temperatura 2	0,9349	0,5222	1,5767
EDIFICIO 2	Planta 3	Temperatura 1	0,9283	0,502	1,5371
		Temperatura 2	0,9115	0,5008	1,6686
	Planta 4	Temperatura 1	0,9073	0,4771	1,5691
		Temperatura 2	0,9124	0,5161	1,6387
	Planta 5	Temperatura 1	0,918	0,5215	1,7715
		Temperatura 2	0,92	0,5473	1,8128

	Planta 6	Temperatura sala	0,9408	0,6129	1,4653
EDIFICIO 3	Aula 1	Temperatura sala	0,9413	0,5797	1,8859
	Aula 2	Temperatura sala	0,9397	0,6858	1,5385
	Aula 3	Temperatura sala	0,9562	0,7805	1,508
	Zona común	Temperatura sala 1	0,9739	0,8087	1,3197
		Temperatura sala 2	0,9611	0,7656	1,4811

Se observa que, en este caso, el ajuste a los datos de entrenamiento es muy cercano a 1 en todos los casos, tomando un valor máximo de 0,9739 para la temperatura 1 de la Zona Común, y un valor mínimo de 0,9073 para la temperatura 1 de la planta 4 en el Edificio 2. De modo similar a lo ocurrido con el modelo anterior de k-vecinos más próximos, estos valores podrían dar una idea de un comportamiento idóneo del modelo. No obstante, cuando se comparan con sus valores en el conjunto de validación cruzada, se observan valores muy dispares, lo que indica sobreajuste. Para estudiar si el modelo reduce su sobreajuste al ajustar sus hiperparámetros, los hiperparámetros que debemos modificar son la profundidad máxima, reduciéndola, y el número mínimo de muestras para dividir un nodo, reduciéndolo. Se modifican estos valores para observar si el modelo mejora en tal caso. Los resultados se muestran a continuación.

Tabla 22. Resultados del modelo térmico modificado de regresión por árboles aleatorios. Fuente: elaboración propia

ÁRBOLES ALEATORIOS					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	T maquina café	0,844	0,5262	1,3959
		T recepción	0,7547	0,3589	1,6672
		T zona ascensor	0,7538	0,3584	1,668
	Planta baja	T anfiteatro 1	0,9078	0,8209	1,8428
		T anfiteatro 2	0,8927	0,7801	1,8512
		T auditorio	0,9109	0,8759	2,0951
	Planta 2	Temperatura sala	0,8502	0,6304	2,2171
	Planta 3	Temperatura sala	0,8267	0,6586	2,2184
	Open Art	Temperatura sala	0,8589	0,7159	1,5038
	Planta 2	Temperatura 1	0,7343	0,4807	1,4831
		Temperatura 2	0,8098	0,5363	1,5347
EDIFICIO 2	Planta 3	Temperatura 1	0,7777	0,5197	1,4932
		Temperatura 2	0,7165	0,5208	1,6241
	Planta 4	Temperatura 1	0,7122	0,4963	1,5248
		Temperatura 2	0,7283	0,5359	1,5837
	Planta 5	Temperatura 1	0,7506	0,5432	1,7172
		Temperatura 2	0,7465	0,5722	1,7341
	Planta 6	Temperatura sala	0,8116	0,6465	1,3958
	Aula 1	Temperatura sala	0,8092	0,6272	1,7782
	Aula 2	Temperatura sala	0,807	0,7243	1,4407

	Aula 3	T sala	0,8609	0,8112	1,3963
	Zona común	T sala 1	0,9171	0,8256	1,2601
		T sala 2	0,8751	0,7894	1,4082

De manera similar a lo que ocurría con el modelo anterior, en este caso, al ajustar los hiperparámetros que causan el sobreajuste, el ajuste sobre los datos de entrenamiento se reduce y se mejora ligeramente el de los datos de validación cruzada.

ANEXO VI. RESULTADOS DE LOS MODELOS ENERGÉTICOS DE REGRESIÓN

A6.1. RESULTADOS DEL MODELO DE REGRESIÓN LINEAL MÚLTIPLE

El modelo mediante regresión lineal múltiple se implementa con el código que puede encontrarse en el *Anexo VIII. Códigos de Python para modelos de regresión energéticos*. Pueden verse a continuación los datos obtenidos en la *Tabla 23. Resultados iniciales del modelo energético de regresión por regresión lineal múltiple*.

Tabla 23. Resultados iniciales del modelo energético de regresión por regresión lineal múltiple. Fuente: elaboración propia

REGRESIÓN LINEAL MÚLTIPLE					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	E suelo radiante	0,1286	0,2974	26,0254
		E calor anfiteatro 1	0,1405	-0,4252	14,8088
	Planta baja	E frío anfiteatro 1	0,0987	-0,2463	7,9483
		E calor anfiteatro 2	0,0925	-0,0803	13,5888
		E frío anfiteatro 2	0,1484	-0,8911	14,7745
		E calor auditorio	0,022	0,0038	7,2394
		E frío auditorio	0,0286	-0,0371	4,5691
	Planta 2	E calor planta 2	0,0008	-0,02	18,992
		E frío planta 2	0,0504	-0,4039	12,3291
	Planta 3	E calor planta 3	0,001	-0,004	6,7605
		E frío planta 3	0,04	0,032	19,9996
EDIFICIO 2	Open Art	E calor Open Art	0,0009	-0,0192	12,2465
		E frío Open Art	0,0031	0,0229	8,2913
	Planta 2	E suelo radiante	0,0536	0,0023	17,5791
		E Fancoil 1	0,1121	-0,3191	15,0275
		E Fancoil 2	0,1073	-0,4514	15,4446
		E Fancoil 3	0,149	-0,5477	22,7725
	Planta 3	E suelo radiante	0,0875	0,0104	10,6494
		E Fancoil 1	0,1314	-0,253	18,356
		E Fancoil 2	0,1693	-0,0432	24,2111
		E Fancoil 3	0,192	-0,3691	27,9008
	Planta 4	E suelo radiante	0,0256	0,0125	6,886
		E Fancoil 1	0,054	-0,1084	19,5582
		E Fancoil 2	0,0959	-0,0417	18,5731
		E Fancoil 3	0,0902	-0,1675	25,1637
	Planta 5	E suelo radiante	0,0202	-0,0022	3,8185
		E Fancoil 1	0,022	-0,0016	14,9356
		E Fancoil 2	0,0779	0,042	20,4364
		E Fancoil 3	0,0918	0,1087	21,8179

EDIFICIO 3	Planta 6	E calor	0,0003	-0,0021	7,5576
		E frío	0,0024	-0,0091	13,5887
	Aula 1	E calor	0,0022	0,0008	1,2454
		E frío	0,052	-0,1817	32,6957
	Aula 2	E calor	0	-0,0006	0,097
		E frío	0,063	-0,01993	19,0103
	Aula 3	E calor	0,0286	-0,5402	11,7423
		E frío	0,0323	-0,0093	13,2892
	Zona común	E calor 1	0,0165	-0,0633	14,6583
		E calor 2	0,0339	0,0046	7,9923
		E frío 1	0,3157	-0,7114	26,432
		E frío 2	0,1065	-0,2904	31,6606

Es notorio que, para este caso, el coeficiente de determinación R^2 adopta valores muy bajos en comparación con los obtenidos para el modelo térmico. En el entrenamiento da un valor mínimo de 0 para el uso del sistema de calor del Aula 2 en el Edificio 3, y un valor máximo de 0,3157 para la uno de los sistemas de frío en la Zona Común del Edificio 3.

Estos valores son muy bajos, y el modelo se comporta peor sobre el conjunto de validación cruzada. En este caso, se observa un valor mínimo de -0,8911 para uno de los sistemas de refrigeración del anfiteatro, y uno máximo de 0,2974 para el uso del suelo radiante en la entrada.

Estos valores tan bajos, que adoptan incluso valores negativos, indican que el modelo no sabe interpretar correctamente las relaciones entre las variables de entrada y salida en el entrenamiento, por lo que las predice peor en el conjunto de validación cruzada.

Tal como se ha explicado en el apartado 5.4. *Implementación y comparación de los modelos energéticos*, se ha creado una variable ficticia de ocupación del edificio. A continuación, se muestra la misma tabla con los valores obtenidos añadiendo la ocupación del edificio.

Tabla 24. Resultados del modelo energético por regresión lineal múltiple considerando ocupación. Fuente: elaboración propia

REGRESIÓN LINEAL MÚLTIPLE					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R^2 entrenamiento	R^2 validación cruzada	MAE
EDIFICIO 1	Entrada	E suelo radiante	0,9782	0,992	0,9694
		E calor anfiteatro 1	0,4518	0,1732	9,2676
	Planta baja	E frío anfiteatro 1	0,1871	0,1034	5,9795
		E calor anfiteatro 2	0,3721	0,4456	9,7153
		E frío anfiteatro 2	0,4374	0,1974	8,0436
		E calor auditorio	0,4370	0,4635	3,9667
		E frío auditorio	0,1313	0,2234	2,9415

EDIFICIO 2	Planta 2	E calor planta 2	0,3186	0,4096	13,573
		E frío planta 2	0,1991	0,1469	9,4309
	Planta 3	E calor planta 3	0,1828	0,245	5,5058
		E frío planta 3	0,2804	0,3434	8,332
	Open Art	E calor Open Art	0,3058	0,2997	9,7682
		E frío Open Art	0,1951	0,2449	7,212
		E suelo radiante	0,2747	0,3219	17,2194
		E Fancoil 1	0,5062	-0,1412	12,1825
		E Fancoil 2	0,4948	-0,2942	12,9698
		E Fancoil 3	0,6367	-0,1577	17,7086
		E suelo radiante	0,243	0,1451	12,0669
		E Fancoil 1	0,4921	0,2865	12,4105
		E Fancoil 2	0,5447	0,5534	15,5154
		E Fancoil 3	0,6365	0,5473	15,8121
	Planta 4	E suelo radiante	0,0797	0,0795	7,7227
		E Fancoil 1	0,4206	0,404	11,7026
		E Fancoil 2	0,3076	0,3277	14,3825
		E Fancoil 3	0,4484	0,5041	15,6529
	Planta 5	E suelo radiante	0,0747	0,0668	4,5094
		E Fancoil 1	0,3513	0,2938	11,6262
		E Fancoil 2	0,5433	0,492	12,9206
		E Fancoil 3	0,526	0,5983	13,0603
	Planta 6	E calor	0,2033	0,1626	5,9828
		E frío	0,3542	0,3475	8,428
EDIFICIO 3	Aula 1	E calor	0,0293	0,0163	1,6562
		E frío	0,6638	0,7899	0,0472
	Aula 2	E calor	0,0214	0,0019	0,1025
		E frío	0,7413	0,7478	4,9932
	Aula 3	E calor	0,3477	0,116	4,2415
		E frío	0,3243	0,4635	5,6156
	Zona común	E calor 1	0,109	0,0902	15,6722
		E calor 2	0,1655	0,012	11,3966
		E frío 1	0,5809	-0,0028	19,408
		E frío 2	0,4857	0,4075	19,304

Se observa una mejora sustancial en el rendimiento del modelo, tanto en el conjunto de entrenamiento como en el de validación cruzada. El coeficiente de determinación R^2 en el entrenamiento da un valor mínimo de 0,0214 para el uso del sistema de calor del Aula 2 en el Edificio 3, y un valor máximo de 0,9782 para el uso del suelo radiante en la entrada.

En el caso del conjunto de validación cruzada, se observa un valor mínimo de -0,2942 para uno de los sistemas Fancoil de la planta 2 del Edificio 2, y uno máximo de 0,992 para el uso del suelo radiante en la entrada.

Pese a que aún se llegan a obtener valores negativos, esto es probablemente debido a que la variable creada no atiende a la realidad del edificio, por lo que tampoco pueden adoptarse como válidos estos modelos. El objetivo de añadir esta nueva variable predictora es mostrar que, de obtenerse los datos reales de ocupación del edificio, podrá continuarse este estudio y obtener modelos que puedan predecir el comportamiento futuro del edificio.

A6.2. RESULTADOS DEL MODELO DE REGRESIÓN POR MÁQUINA DE VECTORES

El modelo mediante regresión por máquina de vectores se implementa con el código que puede encontrarse en el *Anexo VIII. Códigos de Python para modelos de regresión energéticos*. Pueden verse a continuación los datos obtenidos en la *Tabla 1. Resultados del modelo térmico de regresión lineal múltiple*.

Tabla 25. Resultados iniciales del modelo energético por máquina de vectores. Fuente: elaboración propia

MÁQUINA DE VECTORES					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	E suelo radiante	-0,1901	-0,2695	21,1747
	Planta baja	E calor anfiteatro 1	-0,113	-0,0325	3,0311
		E frío anfiteatro 1	-0,0372	-0,0189	1,7703
		E calor anfiteatro 2	-0,0903	-0,0491	4,5733
		E frío anfiteatro 2	-0,1088	-0,0279	2,5535
		E calor auditorio	-0,063	-0,0439	3,4452
	Planta 2	E frío auditorio	-0,0185	-0,0362	2,4774
		E calor planta 2	-0,1838	-0,1228	9,9227
	Planta 3	E frío planta 2	-0,1116	-0,0826	4,1932
		E calor planta 3	-0,0507	-0,06	4,0589
		E frío planta 3	-0,0777	-0,0751	5,8722
EDIFICIO 2	Open Art	E calor Open Art	-0,0756	-0,1159	8,8899
		E frío Open Art	-0,0478	-0,0826	5,446
	Planta 2	E suelo radiante	-0,0418	0,1377	12,1052
		E Fancoil 1	-0,1156	-0,0459	3,8669
		E Fancoil 2	-0,1128	-0,0417	3,5769
		E Fancoil 3	-0,1427	-0,0639	5,9124
	Planta 3	E suelo radiante	-0,0288	-0,062	5,8423
		E Fancoil 1	-0,147	-0,0734	6,4698
		E Fancoil 2	-0,159	-0,165	14,1228
		E Fancoil 3	-0,1865	-0,1391	12,1745
	Planta 4	E suelo radiante	-0,0138	-0,0442	4,2382
		E Fancoil 1	-0,1483	-0,1034	8,8407
		E Fancoil 2	-0,0961	-0,1049	9,238
		E Fancoil 3	-0,1534	-0,151	12,9023
	Planta 5	E suelo radiante	-0,0114	-0,0205	2,0054

EDIFICIO 3		E Fancoil 1	-0,0883	-0,13	11,0223
		E Fancoil 2	-0,1319	-0,2077	16,4124
		E Fancoil 3	-0,1253	-0,2028	16,4558
	Planta 6	E calor	-0,0646	-0,0463	3,3733
		E frío	-0,1123	-0,0981	6,7042
	Aula 1	E calor	-0,0119	-0,0051	0,4252
		E frío	-0,3411	-0,1901	15,2386
	Aula 2	E calor	-0,0039	-0,0011	0,0803
		E frío	-0,1429	-0,0907	7,4308
	Aula 3	E calor	-0,0949	-0,0268	2,3588
		E frío	-0,0884	-0,0697	6,2588
	Zona común	E calor 1	-0,032	-0,1475	12,657
		E calor 2	-0,041	-0,0436	4,0715
		E frío 1	-0,283	-0,0998	8,2848
		E frío 2	-0,2693	-0,1793	14,8217

Del mismo modo que sucedía para el anterior caso, el coeficiente de determinación R^2 adopta valores muy bajos en comparación con los obtenidos para el modelo térmico. En el entrenamiento da un valor mínimo de -0,3411 para el uso del sistema de frío del Aula 1 en el Edificio 3, y un valor máximo de -0,0039 para la uno de los sistemas de calor en el Aula 2 del Edificio 3.

Estos valores son todos negativos, por lo que el modelo no sabe interpretar correctamente las relaciones entre las variables de entrada y salida en el entrenamiento, y lo transfiere al conjunto de validación cruzada. En este caso, se observa un valor mínimo de -0,2695 para el uso del suelo radiante en la entrada, y uno máximo de -0,0011 para el uso del sistema de calor en el Aula 2 del Edificio 3.

A continuación, se muestra la tabla con los valores obtenidos al añadir la ocupación del edificio.

Tabla 26. Resultados del modelo energético por máquina de vectores considerando ocupación. Fuente: elaboración propia

MÁQUINA DE VECTORES					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R^2 entrenamiento	R^2 validación cruzada	MAE
EDIFICIO 1	Entrada	E suelo radiante	0,977	0,9925	0,193
		E calor anfiteatro 1	-0,0008	0,0725	3,117
	Planta baja	E frío anfiteatro 1	-0,0372	-0,0189	1,7701
		E calor anfiteatro 2	-0,0902	-0,0489	4,5729
		E frío anfiteatro 2	0,3415	0,2979	3,2876
		E calor auditorio	0,3498	0,3577	3,2239
		E frío auditorio	-0,0185	-0,0362	2,4772
	Planta 2	E calor planta 2	0,2025	0,2168	9,6551

		E frío planta 2	-0,1116	-0,0826	4,193
EDIFICIO 2	Planta 3	E calor planta 3	-0,0507	-0,0599	4,0589
		E frío planta 3	-0,0318	-0,0317	5,8506
		E calor Open Art	0,1977	0,1343	8,7391
	Open Art	E frío Open Art	-0,0477	-0,0826	5,4461
		E suelo radiante	-0,0418	-0,1377	12,1054
	Planta 2	E Fancoil 1	0,4644	-0,0049	7,6782
		E Fancoil 2	0,3367	0,1431	6,1125
		E Fancoil 3	0,3708	-1,5536	14,1335
	Planta 3	E suelo radiante	-0,0288	-0,062	5,8424
		E Fancoil 1	0,4577	0,27	8,799
		E Fancoil 2	0,4257	0,3608	11,799
		E Fancoil 3	0,3295	0,2639	7,9155
	Planta 4	E suelo radiante	-0,0138	-0,0442	4,2386
		E Fancoil 1	0,3715	0,3355	8,9788
		E Fancoil 2	-0,0839	-0,0949	9,2262
		E Fancoil 3	0,2156	0,1998	11,7552
	Planta 5	E suelo radiante	-0,0114	-0,0205	2,0056
		E Fancoil 1	0,2358	0,1179	11,0573
		E Fancoil 2	0,5096	0,4014	12,7178
		E Fancoil 3	0,4551	0,4505	13,0782
	Planta 6	E calor	-0,0646	-0,0462	3,3735
		E frío	0,1549	0,2055	6,1029
	Aula 1	E calor	-0,0119	-0,0051	0,4254
		E frío	0,4246	0,7687	3,5299
	Aula 2	E calor	-0,0039	-0,0011	0,0803
		E frío	0,5635	0,5748	3,3814
	Aula 3	E calor	-0,0828	-0,015	2,3772
		E frío	-0,0883	-0,0696	6,2583
	Zona común	E calor 1	-0,032	-0,1475	12,657
		E calor 2	-0,041	-0,0436	4,0716
		E frío 1	0,4772	-0,0314	14,0259
		E frío 2	0,0245	-0,1062	15,4182

Se observa una mejora en el rendimiento del modelo, tanto en el conjunto de entrenamiento como en el de validación cruzada. El coeficiente de determinación R2 en el entrenamiento da un valor mínimo de -0,1116 para el uso del sistema de frío de la planta 2 en el Edificio 1, y un valor máximo de 0,977 para el uso del suelo radiante en la entrada.

En el caso del conjunto de validación cruzada, se observa un valor mínimo de -0,15536 para uno de los sistemas Fancoil de la planta 2 del Edificio 2, y uno máximo de 0,9925 para el uso del suelo radiante en la entrada.

Todavía se obtienen muchos valores negativos con este modelo, probablemente debido a que la variable creada no atiende a la realidad del edificio.

A6.3. RESULTADOS DEL MODELO DE REGRESIÓN POR K-VECINOS MÁS CERCANOS

El modelo mediante regresión por k-vecinos más cercanos se implementa con el código que puede encontrarse en el *Anexo VIII. Códigos de Python para modelos de regresión energéticos*. Pueden verse a continuación los datos obtenidos en la *Tabla 1. Resultados del modelo térmico de regresión lineal múltiple*.

Tabla 27. Resultados iniciales del modelo energético por k-vecinos más cercanos. Fuente: elaboración propia

K-VECINOS MÁS CERCANOS					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Entrada	E suelo radiante	0,2558	0,1755	21,5036
		E calor anfiteatro 1	0,4801	-0,3348	7,4212
	Planta baja	E frio anfiteatro 1	0,1475	-0,1967	5,7204
		E calor anfiteatro 2	0,382	-0,1213	8,0394
		E frio anfiteatro 2	0,3001	-0,5731	10,0727
		E calor auditorio	0,2664	0,0698	4,9153
		E frio auditorio	0,1028	0,02023	3,8286
	Planta 2	E calor planta 2	0,1155	0,1635	10,4061
		E frío planta 2	0,1105	-0,4029	10,0208
	Planta 3	E calor planta 3	0,0725	0,0371	4,4209
		E frío planta 3	0,0939	-0,055	11,1453
EDIFICIO 2	Open Art	E calor Open Art	0,0567	0,0418	9,567
		E frío Open Art	0,07	-0,0776	7,6322
	Planta 2	E suelo radiante	0,1095	0,0447	13,8298
		E Fancoil 1	0,5203	0,2575	4,9938
		E Fancoil 2	0,4381	0,2192	5,3165
		E Fancoil 3	0,6688	0,2288	6,5434
	Planta 3	E suelo radiante	0,1099	0,0824	7,9991
		E Fancoil 1	0,4582	0,265	9,8926
		E Fancoil 2	0,4771	0,298	15,4733
		E Fancoil 3	0,6652	0,3466	13,2953
	Planta 4	E suelo radiante	0,0649	0,06	5,5389
		E Fancoil 1	0,1761	0,1117	11,4955
		E Fancoil 2	0,2073	0,1019	11,5252
		E Fancoil 3	0,2542	0,1398	14,7811
	Planta 5	E suelo radiante	0,0881	-0,0024	2,5992
		E Fancoil 1	0,2164	0,0119	11,2723
		E Fancoil 2	0,3394	0,033	16,1717
		E Fancoil 3	0,3941	0,1146	15,4405
	Planta 6	E calor	0,0237	0,0358	5,6667

		E frio	0,0844	0,0376	9,8417
EDIFICIO 3	Aula 1	E calor	0,0294	-0,0361	1,203
		E frio	0,1886	-0,054	25,4779
	Aula 2	E calor	0,0384	-0,001	0,0894
		E frio	0,1992	0,0469	11,8542
	Aula 3	E calor	0,1355	-0,6133	8,5971
		E frio	0,1279	0,0366	9,9992
	Zona común	E calor 1	0,0823	-0,593	14,0481
		E calor 2	0,1423	-0,126	7,1586
		E frio 1	0,4826	-0,03	14,3751
		E frio 2	0,426	-0,103	19,8059

De igual modo que anteriormente, el coeficiente de determinación R^2 adopta valores bajos en comparación con los obtenidos para el modelo térmico. En el entrenamiento da un valor mínimo de 0,0237 para el uso del sistema de calor de la planta 6 del Edificio 2, y un valor máximo de 0,6688 para la uno de los sistemas Fancoil en la Planta 2 del Edificio 2.

Respecto al conjunto de validación cruzada. En este caso, se observa un valor mínimo de -0,6133 para el uso del sistema de calor en el aula 4 del Edificio 3, y uno máximo de 0,3466 para uno de los sistemas Fancoil en la Planta 3 del Edificio 2.

A continuación, se muestra la tabla con los valores obtenidos al añadir la ocupación del edificio.

Tabla 28. Resultados del modelo energético por k-vecinos más cercanos considerando ocupación. Fuente: elaboración propia

K-VECINOS MÁS CERCANOS					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R^2 entrenamiento	R^2 validación cruzada	MAE
EDIFICIO 1	Planta baja	E suelo radiante	0,9503	0,8213	3,9316
		E calor anfiteatro 1	0,7608	0,1305	3,7071
		E frio anfiteatro 1	0,3713	-0,0612	4,6546
		E calor anfiteatro 2	0,758	0,203	4,6295
		E frio anfiteatro 2	0,7707	0,1551	4,9687
		E calor auditorio	0,5908	0,4539	2,6653
		E frio auditorio	0,3965	0,2695	2,0579
	Planta 2	E calor planta 2	0,4527	0,5179	6,5534
		E frío planta 2	0,3586	-0,1256	7,2876
	Planta 3	E calor planta 3	0,2681	0,2101	3,6237
		E frío planta 3	0,3652	0,2095	8,4331
EDIFICIO 2	Open Art	E calor Open Art	0,3202	0,3379	7,5278
		E frío Open Art	0,3286	0,0313	6,7634
	Planta 2	E suelo radiante	0,7134	0,6771	6,3017
		E Fancoil 1	0,6862	0,5837	2,8807
		E Fancoil 2	0,6642	0,5697	2,7542

EDIFICIO 3		E Fancoil 3	0,8838	0,5318	3,0732
	Planta 3	E suelo radiante	0,6447	0,654	3,3572
		E Fancoil 1	0,6096	0,3757	8,1318
		E Fancoil 2	0,7749	0,6115	9,5766
		E Fancoil 3	0,8923	0,6708	7,3088
	Planta 4	E suelo radiante	0,5296	0,4879	3,0917
		E Fancoil 1	0,4713	0,406	8,016
		E Fancoil 2	0,4827	0,3807	8,4134
		E Fancoil 3	0,6214	0,4977	9,0608
	Planta 5	E suelo radiante	0,3554	0,2592	1,7261
		E Fancoil 1	0,4017	0,2346	9,5187
		E Fancoil 2	0,6299	0,6305	12,1932
		E Fancoil 3	0,6331	0,4847	10,3528
	Planta 6	E calor	0,2137	0,2254	4,4311
		E frio	0,3937	0,3468	6,6506
	Aula 1	E calor	0,1589	0,1146	0,6711
		E frio	0,6963	0,656	7,4981
	Aula 2	E calor	0,1933	0,0006	0,0862
		E frio	0,7868	0,5995	4,0587
	Aula 3	E calor	0,3838	-0,2972	4,2685
		E frio	0,3956	0,4289	4,5775
	Zona común	E calor 1	0,2614	0,1555	11,7584
		E calor 2	0,5948	0,0956	6,3976
		E frio 1	0,7309	0,2824	8,7927
		E frio 2	0,7164	0,2781	11,3141

Se observa que el rendimiento del modelo mejora, tanto en el conjunto de entrenamiento como en el de validación cruzada. El coeficiente de determinación R^2 en el entrenamiento da un valor mínimo de 0,1589 para el uso del sistema de calor del Aula 1 en el Edificio 3, y un valor máximo de 0,9503 para el uso del suelo radiante en la entrada.

En el caso del conjunto de validación cruzada, se observa un valor mínimo de -0,2972 para el sistema de calor del Aula 3 del Edificio 3, y uno máximo de 0,8213 para el uso del suelo radiante en la entrada.

Pese a la mejora, todavía se obtienen muchos valores negativos con este modelo, probablemente debido a que la variable creada no atiende a la realidad del edificio.

A6.4. RESULTADOS DEL MODELO DE REGRESIÓN POR BOSQUES ALEATORIOS

El modelo mediante regresión por bosques aleatorios se implementa con el código que puede encontrarse en el Anexo VIII. *Códigos de Python para modelos de regresión energéticos*. Pueden

verse a continuación los datos obtenidos en la *Tabla 1. Resultados del modelo térmico de regresión lineal múltiple*.

Tabla 29. Resultados iniciales del modelo energético por árboles aleatorios. Fuente: elaboración propia

BOSQUES ALEATORIOS					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R ² entrenamiento	R ² validación cruzada	MAE
EDIFICIO 1	Planta baja	Entrada	E suelo radiante	0,3974	0,0911
			E calor anfiteatro 1	0,7198	-0,7946
			E frio anfiteatro 1	0,7194	-9,5316
			E calor anfiteatro 2	0,6709	-0,5675
			E frio anfiteatro 2	0,7033	-6,9254
			E calor auditorio	0,6085	-0,0811
			E frio auditorio	0,4958	-0,3309
	Planta 2		E calor planta 2	0,1311	0,1489
			E frío planta 2	0,164	-5,0007
	Planta 3		E calor planta 3	0,0921	0,0312
			E frío planta 3	0,1604	-1,5114
EDIFICIO 2	Open Art		E calor Open Art	0,0722	0,037
			E frío Open Art	0,0885	-0,0845
	Planta 2		E suelo radiante	0,3346	-0,1072
			E Fancoil 1	0,7137	0,223
			E Fancoil 2	0,6095	0,1491
			E Fancoil 3	0,7923	0,1925
	Planta 3		E suelo radiante	0,3415	-0,0736
			E Fancoil 1	0,6242	0,0519
			E Fancoil 2	0,6087	-0,1599
			E Fancoil 3	0,7506	-0,1548
	Planta 4		E suelo radiante	0,3629	-0,1653
			E Fancoil 1	0,2952	0,1293
			E Fancoil 2	0,3529	-0,2247
			E Fancoil 3	0,36	-0,0432
	Planta 5		E suelo radiante	0,4866	-0,4885
			E Fancoil 1	0,393	0,0294
			E Fancoil 2	0,5297	0,0717
			E Fancoil 3	0,5288	0,1281
	Planta 6		E calor	0,0356	-0,0833
			E frio	0,097	0,0236
EDIFICIO 3	Aula 1		E calor	0,0572	-0,0983
			E frio	0,2055	-0,0769
	Aula 2		E calor	0,0596	-0,0012
			E frio	0,216	0,025
	Aula 3		E calor	0,1549	-0,643
			E frio	0,1577	0,0163
			E calor 1	0,3694	-0,1272

	Zona común	E calor 2	0,337	-0,0926	6,818
		E frio 1	0,6084	-0,2335	14,9274
		E frio 2	0,5305	-0,1918	19,6189

Igual que sucedía con el modelo anterior, el coeficiente de determinación R^2 adopta valores bajos en comparación con los obtenidos para el modelo térmico, pero mejores que con otros modelos. En el entrenamiento da un valor mínimo de 0,0356 para el uso del sistema de calor de la planta 6 del Edificio 2, y un valor máximo de 0,7923 para la uno de los sistemas Fancoil en la Planta 2 del Edificio 2.

Respecto al conjunto de validación cruzada. En este caso, se observa un valor mínimo de -9,5316 para el uso del sistema de frío en el Anfiteatro, y uno máximo de 0,223 para uno de los sistemas de frío de la Zona Común del Edificio 3.

A continuación, se muestra la tabla con los valores obtenidos al añadir la ocupación del edificio.

Tabla 30. Resultados del modelo energético por árboles aleatorios considerando ocupación. Fuente: elaboración propia

BOSQUES ALEATORIOS					
EDIFICIO	ZONA	VARIABLE SALIDA (T = temperatura)	R^2 entrenamiento	R^2 validación cruzada	MAE
EDIFICIO 1	Entrada	E suelo radiante	0,9889	0,9711	1,4737
	Planta baja	E calor anfiteatro 1	0,9054	-0,2386	3,9821
		E frio anfiteatro 1	0,8011	0,2975	1,421
		E calor anfiteatro 2	0,9263	0,4981	2,5383
		E frio anfiteatro 2	0,8776	0,3653	1,8011
		E calor auditorio	0,7134	0,6825	1,9211
	Planta 2	E frio auditorio	0,6671	0,2308	2,1491
		E calor planta 2	0,5216	0,6581	4,8827
		E frío planta 2	0,4598	-1,6471	8,4048
		E calor planta 3	0,4231	0,0561	4,2486
	EDIFICIO 2	E frío planta 3	0,5744	0,429	4,7555
	Planta 2	Open Art	0,4396	0,3343	8,5273
		E frío Open Art	0,4246	-0,0848	7,1985
		E suelo radiante	0,9957	0,9747	0,4115
		E Fancoil 1	0,8413	0,5897	2,957
		E Fancoil 2	0,7942	-0,0082	4,1614
	Planta 3	E Fancoil 3	0,9684	0,6884	1,9123
		E suelo radiante	0,9654	0,9441	0,4708
		E Fancoil 1	0,7396	0,4106	5,2195
		E Fancoil 2	0,8567	0,8621	3,0045
		E Fancoil 3	0,9442	0,9576	1,0629
	Planta 4	E suelo radiante	0,9792	0,9078	0,458
		E Fancoil 1	0,5944	0,3109	8,0525

		E Fancoil 2	0,6547	0,5725	4,9636
		E Fancoil 3	0,7188	0,634	5,6085
EDIFICIO 3	Planta 5	E suelo radiante	0,9445	0,5795	0,954
		E Fancoil 1	0,5664	0,2887	11,4221
		E Fancoil 2	0,8376	0,5843	9,7419
		E Fancoil 3	0,762	0,5828	9,0159
	Planta 6	E calor	0,2681	0,143	4,749
		E frio	0,4426	0,1948	7,7006
	Aula 1	E calor	0,3557	-0,6422	0,9756
		E frio	0,7245	0,751	5,2838
	Aula 2	E calor	0,5756	0,0406	0,1284
		E frio	0,8248	0,6652	3,3847
	Aula 3	E calor	0,4705	0,1517	3,1596
		E frio	0,4684	0,5544	3,6484
	Zona común	E calor 1	0,7735	-0,0425	13,3614
		E calor 2	0,913	-0,3294	6,4754
		E frio 1	0,8269	0,0748	8,9025
		E frio 2	0,793	0,2855	10,7164

Se observa una gran mejora en el rendimiento del modelo, tanto en el conjunto de entrenamiento como en el de validación cruzada. El coeficiente de determinación R^2 en el entrenamiento toma ahora un valor mínimo de 0,2681 para el uso del sistema de calor de la Planta 6 en el Edificio 2, y un valor máximo de 0,9957 para el uso del suelo radiante en la entrada.

En el caso del conjunto de validación cruzada, se observa un valor mínimo de -1,6471 para el sistema de frío de la planta 2 del Edificio 1, y uno máximo de 0,9747 para el uso del suelo radiante en la Planta 3 del Edificio 2.

Aunque aún se obtienen valores negativos para todos los modelos, esto se debe probablemente a que la variable creada no atiende a la realidad del edificio. Las diferentes estancias no son solo ocupadas si los sistemas de climatización se ponen en marcha.

Por ello, no pueden adoptarse como válidos estos modelos. No obstante, el interés en estudiar el comportamiento con esta nueva variable predictora radica en que se puede observar que, si se obtienen los datos reales de ocupación del edificio, podrá continuarse este estudio y obtener modelos que puedan predecir el comportamiento futuro del edificio.

ANEXO VII. CÓDIGOS DE PYTHON PARA MODELOS DE REGRESIÓN TÉRMICOS

A7.1. CÓDIGO DE PYTHON PARA EL MODELO TÉRMICO DE REGRESIÓN LINEAL MÚLTIPLE

El ejemplo en concreto hace referencia a los datos del Edificio 1. Para el resto de los edificios y sus respectivas zonas, se deben modificar los nombres de los archivos donde se alojan los datos y los nombres de las hojas de los archivos de Excel, así como las columnas de las variables.

```
import pandas as pd

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_squared_error, mean_absolute_error,
r2_score

from sklearn.preprocessing import StandardScaler

hojas = ['Filtrado_ClimayEntrada', 'Filtrado_ClimayPBaja',
'Filtrado_ClimayP2', 'Filtrado_ClimayP3']

for hoja in hojas:

    #Leer los datos de entrenamiento

    df_train = pd.read_excel('Ed1_2021filtrado.xlsx',
sheet_name=hoja)

    #Leer los datos de validación cruzada

    df_val = pd.read_excel('Ed1_2022filtrado.xlsx', sheet_name=hoja)

    #Variables de entrada (características)

    features = ['Temp_exterior', 'Humedad_exterior',
'Velocidad_viento', 'Cantidad_lluvia']

    if hoja == 'Filtrado_ClimayEntrada' or hoja ==
'Filtrado_ClimayPBaja':

        targets = df_train.columns[6:9] #columnas de las
temperaturas

    elif hoja == 'Filtrado_ClimayP2' or hoja == 'Filtrado_ClimayP3':

        targets = [df_train.columns[6]] #columnas de las
temperaturas

    for target in targets:

        #Variable de salida actual
```

```

y = df_train[target]

#Escalar las variables de entrada

scaler = StandardScaler()

X_train_scaled = scaler.fit_transform(df_train[features])

#Crear y entrenar el modelo de regresión lineal

model = LinearRegression()

model.fit(X_train_scaled, y)

#Realizar predicciones en el conjunto de entrenamiento

y_train_pred = model.predict(X_train_scaled)

#Calcular métricas de rendimiento en el entrenamiento

mse_train = mean_squared_error(y, y_train_pred)

r2_train = r2_score(y, y_train_pred)

print(f'\nMétricas para la variable de salida "{target}":')

print('Conjunto de entrenamiento:')

print(f'MSE: {mse_train:.4f}, R^2: {r2_train:.4f}')

#Escalar las variables de entrada del en validación cruzada

X_val_scaled = scaler.transform(df_val[features])

#Realizar predicciones en el conjunto de validación cruzada

y_val_pred = model.predict(X_val_scaled)

#Calcular métricas de rendimiento en el en validación cruzada

mse_val = mean_squared_error(df_val[target], y_val_pred)

mae_val = mean_absolute_error(df_val[target], y_val_pred)

r2_val = r2_score(df_val[target], y_val_pred)

nmae_val = mae_val / df_val[target].mean()

print(f'Conjunto de validación cruzada:')

print(f'MSE: {mse_val:.4f}, MAE: {mae_val:.4f}, NMAE: {nmae_val:.4f}, R^2: {r2_val:.4f}')

#Determinar los coeficientes del modelo

coeficientes = model.coef_

print(f'Coeficientes: {coeficientes}')

```

A7.2. CÓDIGO DE PYTHON PARA EL MODELO TÉRMICO DE REGRESIÓN POR MÁQUINA DE VECTORES

El ejemplo en concreto hace referencia a los datos del Edificio 1. Para el resto de los edificios y sus respectivas zonas, se deben modificar los nombres de los archivos donde se alojan los datos y los nombres de las hojas de los archivos de Excel, así como las columnas de las variables.

```
import pandas as pd

from sklearn.svm import LinearSVR

from sklearn.metrics import mean_squared_error, mean_absolute_error,
r2_score

from sklearn.preprocessing import StandardScaler

hojas = ['Filtrado_ClimayEntrada', 'Filtrado_ClimayPBaja',
'Filtrado_ClimayP2', 'Filtrado_ClimayP3']

for hoja in hojas:

    #Ler los datos de entrenamiento

    df_train = pd.read_excel('Ed1_2021filtrado.xlsx', sheet_name=hoja)

    #Ler los datos de validación cruzada

    df_val = pd.read_excel('Ed1_2022filtrado.xlsx', sheet_name=hoja)

    #Variables de entrada (características)

    features = ['Temp_exterior', 'Humedad_exterior',
'Velocidad_viento', 'Cantidad_lluvia']

    if hoja == 'Filtrado_ClimayEntrada' or hoja ==
'Filtrado_ClimayPBaja':

        targets = df_train.columns[6:9] #columnas de las temperaturas

    elif hoja == 'Filtrado_ClimayP2' or hoja == 'Filtrado_ClimayP3':

        targets = [df_train.columns[6]] #columnas de las temperaturas

    for target in targets:

        #Variable de salida actual

        y = df_train[target]

        #Escalar las variables de entrada

        scaler = StandardScaler()

        X_train_scaled = scaler.fit_transform(df_train[features])
```

```

#Crear el modelo SVR con kernel lineal

svr = LinearSVR(max_iter=10000)

#Realizar predicciones en el conjunto de entrenamiento

y_train_pred=svr.fit(X_train_scaled,y).predict(X_train_scaled)

#Calcular métricas de rendimiento en el entrenamiento

mse_train = mean_squared_error(y, y_train_pred)

r2_train = r2_score(y, y_train_pred)

print(f'\nMétricas para la variable de salida "{target}":')

print('Conjunto de entrenamiento:')

print(f'MSE: {mse_train:.4f}, R^2: {r2_train:.4f}')

#Escalar las variables de entrada del en validación cruzada

X_val_scaled = scaler.transform(df_val[features])

#Realizar predicciones en el conjunto de validación cruzada

y_val_pred = SVR.predict(X_val_scaled)

#Calcular métricas de rendimiento en el en validación cruzada

mse_val = mean_squared_error(df_val[target], y_val_pred)

mae_val = mean_absolute_error(df_val[target], y_val_pred)

r2_val = r2_score(df_val[target], y_val_pred)

nmae_val = mae_val / df_val[target].mean()

print(f'Conjunto de validación cruzada:')

print(f'MSE: {mse_val:.4f}, MAE: {mae_val:.4f}, NMAE: {nmae_val:.4f}, R^2: {r2_val:.4f}')

```

A7.3. CÓDIGO DE PYTHON PARA EL MODELO TÉRMICO DE REGRESIÓN POR K-VECINOS MÁS CERCANOS

El ejemplo en concreto hace referencia a los datos del Edificio 1. Para el resto de los edificios y sus respectivas zonas, se deben modificar los nombres de los archivos donde se alojan los datos y los nombres de las hojas de los archivos de Excel, así como las columnas de las variables.

```
import pandas as pd

from sklearn.neighbors import KNeighborsRegressor

from sklearn.metrics import mean_squared_error, mean_absolute_error,
r2_score

from sklearn.preprocessing import StandardScaler

hojas = ['Filtrado_ClimayEntrada', 'Filtrado_ClimayPBaja',
'Filtrado_ClimayP2', 'Filtrado_ClimayP3']

for hoja in hojas:

    #Leer los datos de entrenamiento

    df_train = pd.read_excel('Ed1_2021filtrado.xlsx', sheet_name=hoja)

    #Leer los datos de validación cruzada

    df_val = pd.read_excel('Ed1_2022filtrado.xlsx', sheet_name=hoja)

    #Variables de entrada (características)

    features = ['Temp_exterior', 'Humedad_exterior',
'Velocidad_viento', 'Cantidad_lluvia']

    if hoja == 'Filtrado_ClimayEntrada' or hoja ==
'Filtrado_ClimayPBaja':

        targets = df_train.columns[6:9] #columnas de las temperaturas

    elif hoja == 'Filtrado_ClimayP2' or hoja == 'Filtrado_ClimayP3':

        targets = [df_train.columns[6]] #columnas de las temperaturas

    for target in targets:

        #Variable de salida actual

        y = df_train[target]

        #Escalar las variables de entrada

        scaler = StandardScaler()

        X_train_scaled = scaler.fit_transform(df_train[features])
```

```

#Crear el modelo KNN

knn = KNeighborsRegressor(n_neighbors=1000)

#Entrenar el modelo

knn.fit(X_train_scaled, y_train)

#Realizar predicciones en el conjunto de entrenamiento

y_train_pred=knn.predict(X_train_scaled)

#Calcular métricas de rendimiento en m entrenamiento

mse_train = mean_squared_error(y, y_train_pred)

r2_train = r2_score(y, y_train_pred)

print(f'\nMétricas para la variable de salida "{target}":')

print('Conjunto de entrenamiento:')

print(f'MSE: {mse_train:.4f}, R^2: {r2_train:.4f}')

#Escalar las variables de entrada del m validación cruzada

X_val_scaled = scaler.transform(df_val[features])

#Realizar predicciones en el conjunto de validación cruzada

y_val_pred = knn.predict(X_val_scaled)

#Calcular métricas de rendimiento en el m validación cruzada

mse_val = mean_squared_error(df_val[target], y_val_pred)

mae_val = mean_absolute_error(df_val[target], y_val_pred)

r2_val = r2_score(df_val[target], y_val_pred)

nmae_val = mae_val / df_val[target].mean()

print(f'Conjunto de validación cruzada:')

print(f'MSE: {mse_val:.4f}, MAE: {mae_val:.4f}, NMAE: {nmae_val:.4f}, R^2: {r2_val:.4f}')

```

A7.4. CÓDIGO DE PYTHON PARA EL MODELO TÉRMICO DE REGRESIÓN POR ARBOLES ALEATORIOS

El ejemplo en concreto hace referencia a los datos del Edificio 1. Para el resto de los edificios y sus respectivas zonas, se deben modificar los nombres de los archivos donde se alojan los datos y los nombres de las hojas de los archivos de Excel, así como las columnas de las variables.

```
import pandas as pd

from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import mean_squared_error, mean_absolute_error,
r2_score

from sklearn.preprocessing import StandardScaler

hojas = ['Filtrado_ClimayEntrada', 'Filtrado_ClimayPBaja',
'Filtrado_ClimayP2', 'Filtrado_ClimayP3']

for hoja in hojas:

    #Ler los datos de entrenamiento

    df_train = pd.read_excel('Ed1_2021filtrado.xlsx', sheet_name=hoja)

    #Ler los datos de validación cruzada

    df_val = pd.read_excel('Ed1_2022filtrado.xlsx', sheet_name=hoja)

    #Variables de entrada (características)

    features = ['Temp_exterior', 'Humedad_exterior',
'Velocidad_viento', 'Cantidad_lluvia']

    if hoja == 'Filtrado_ClimayEntrada' or hoja ==
'Filtrado_ClimayPBaja':

        targets = df_train.columns[6:9] #columnas de las temperaturas

    elif hoja == 'Filtrado_ClimayP2' or hoja == 'Filtrado_ClimayP3':

        targets = [df_train.columns[6]] #columnas de las temperaturas

    for target in targets:

        #Variable de salida actual

        y = df_train[target]

        #Escalar las variables de entrada

        scaler = StandardScaler()

        X_train_scaled = scaler.fit_transform(df_train[features])
```

```

#Crear el modelo de bosques aleatorios

rf_model = RandomForestRegressor(n_estimators=100,
max_depth=10, min_samples_split=5)

#Entrenar el modelo

rf_model.fit(X_train_scaled, y_train)

#Realizar predicciones en el conjunto de entrenamiento

y_train_pred=rf_model.predict(X_train_scaled)

#Calcular métricas de rendimiento en el entrenamiento

mse_train = mean_squared_error(y, y_train_pred)

r2_train = r2_score(y, y_train_pred)

print(f'\nMétricas para la variable de salida "{target}":')

print('Conjunto de entrenamiento:')

print(f'MSE: {mse_train:.4f}, R^2: {r2_train:.4f}')

#Escalar las variables de entrada del en validación cruzada

X_val_scaled = scaler.transform(df_val[features])

#Realizar predicciones en el conjunto de validación cruzada

y_val_pred = knn.predict(X_val_scaled)

#Calcular métricas de rendimiento en el en validación cruzada

mse_val = mean_squared_error(df_val[target], y_val_pred)

mae_val = mean_absolute_error(df_val[target], y_val_pred)

r2_val = r2_score(df_val[target], y_val_pred)

nmae_val = mae_val / df_val[target].mean()

print('Conjunto de validación cruzada:')

print(f'MSE: {mse_val:.4f}, MAE: {mae_val:.4f}, NMAE: {nmae_val:.4f}, R^2: {r2_val:.4f}')

```

ANEXO VIII. CÓDIGOS DE PYTHON PARA MODELOS DE REGRESIÓN ENERGÉTICOS

A8.1. CÓDIGO DE PYTHON PARA EL MODELO ENERGÉTICO DE REGRESIÓN LINEAL MÚLTIPLE

El ejemplo en concreto hace referencia a los datos del Edificio 1 en la zona de la Entrada. Para el resto de los edificios y zonas, se cambian los nombres de los archivos donde se alojan los datos, los nombres de las hojas de los archivos de Excel y las columnas en las que se alojan las variables de entrada y salida.

```
import pandas as pd

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_absolute_error, r2_score

#hojas = ['ClimayEntrada', 'ClimayPBaja', 'ClimayP2', 'ClimayP3']

#Leer los datos de entrenamiento

df_train = pd.read_excel('ClimayEd1_2021.xlsx',
sheet_name='ClimayEntrada')

#Leer los datos de validación cruzada

df_val = pd.read_excel('ClimayEd1_2022.xlsx',
sheet_name='ClimayEntrada')

#Variables de entrada (características)

features = df_train.columns[6:9]

targets = df_train.columns[9:11] #Variables de salida (objetivo)

for target in targets:

    #Variable de salida actual

    y = df_train[target]

    X_train = (df_train[features])

    #Crear y entrenar el modelo de regresión lineal

    model = LinearRegression()

    model.fit(X_train, y)

    #Realizar predicciones en el conjunto de entrenamiento
```

```

y_train_pred = model.predict(X_train)

#Métricas de rendimiento en el conjunto de entrenamiento

r2_train = r2_score(y, y_train_pred)

print(f'\nMétricas para la variable de salida "{target}":')

print('Conjunto de entrenamiento:')

print(f'R^2: {r2_train:.4f}')

#Escalar las características del conjunto de validación cruzada

X_val = df_val[features]

#Realizar predicciones en el conjunto de validación cruzada

y_val_pred = model.predict(X_val)

#Métricas de rendimiento en el conjunto de validación cruzada

mae_val = mean_absolute_error(df_val[target], y_val_pred)

r2_val = r2_score(df_val[target], y_val_pred)

print('Conjunto de validación cruzada:')

print(f'MAE: {mae_val:.4f}, R^2: {r2_val:.4f}')

```

A8.2. CÓDIGO DE PYTHON PARA EL MODELO ENERGÉTICO DE REGRESIÓN POR MÁQUINA DE VECTORES

El ejemplo en concreto hace referencia a los datos del Edificio 1 en la zona de la Entrada. Para el resto de los edificios y zonas, se cambian los nombres de los archivos donde se alojan los datos, los nombres de las hojas de los archivos de Excel y las columnas en las que se alojan las variables de entrada y salida.

```

import pandas as pd

from sklearn.svm import LinearSVR

from sklearn.metrics import mean_absolute_error, r2_score

#hojas = ['ClimayEntrada', 'ClimayPBaja', 'ClimayP2', 'ClimayP3']

#Leer los datos de entrenamiento

df_train = pd.read_excel('ClimayEd1_2021.xlsx',
sheet_name='ClimayEntrada')

```

```

#Leer los datos de validación cruzada

df_val = pd.read_excel('ClimayEd1_2022.xlsx',
sheet_name='ClimayEntrada')

#Variables de entrada (características)

features = df_train.columns[6:9]

targets = df_train.columns[9:11] #Variables de salida (objetivo)

for target in targets:

    #Variable de salida actual

    y = df_train[target]

    X_train = (df_train[features])

    #Crear y entrenar el modelo de regresión lineal

    model = LinearSVR(max_iter=10000)

    model.fit(X_train, y)

    #Realizar predicciones en el conjunto de entrenamiento

    y_train_pred = model.predict(X_train)

    #Métricas de rendimiento en el conjunto de entrenamiento

    r2_train = r2_score(y, y_train_pred)

    print(f'\nMétricas para la variable de salida "{target}":')

    print('Conjunto de entrenamiento:')

    print(f'R^2: {r2_train:.4f}')

    #Escalar las características del conjunto de validación cruzada

    X_val = df_val[features]

    #Realizar predicciones en el conjunto de validación cruzada

    y_val_pred = model.predict(X_val)

    #Métricas de rendimiento en el conjunto de validación cruzada

    mae_val = mean_absolute_error(df_val[target], y_val_pred)

    r2_val = r2_score(df_val[target], y_val_pred)

    print(f'Conjunto de validación cruzada:')

    print(f'MAE: {mae_val:.4f}, R^2: {r2_val:.4f}')

```

A8.3. CÓDIGO DE PYTHON PARA EL MODELO ENERGÉTICO DE REGRESIÓN POR K-VECINOS MÁS CERCANOS

El ejemplo en concreto hace referencia a los datos del Edificio 1 en la zona de la Entrada. Para el resto de los edificios y zonas, se cambian los nombres de los archivos donde se alojan los datos, los nombres de las hojas de los archivos de Excel y las columnas en las que se alojan las variables de entrada y salida.

```
import pandas as pd

from sklearn.neighbors import KNeighborsRegressor

from sklearn.metrics import mean_absolute_error, r2_score

#hojas = ['ClimayEntrada', 'ClimayPBaja', 'ClimayP2', 'ClimayP3']

#Leer los datos de entrenamiento

df_train = pd.read_excel('ClimayEd1_2021.xlsx',
sheet_name='ClimayEntrada')

#Leer los datos de validación cruzada

df_val = pd.read_excel('ClimayEd1_2022.xlsx',
sheet_name='ClimayEntrada')

#Variables de entrada (características)

features = df_train.columns[6:9]

targets = df_train.columns[9:11] #Variables de salida (objetivo)

for target in targets:

    #Variable de salida actual

    y = df_train[target]

    X_train = (df_train[features])

    # Crear el modelo KNN

    model = KNeighborsRegressor(n_neighbors=1000)

    model.fit(X_train, y)

    #Realizar predicciones en el conjunto de entrenamiento

    y_train_pred = model.predict(X_train)

    #Métricas de rendimiento en el conjunto de entrenamiento

    r2_train = r2_score(y, y_train_pred)
```

```

print(f'\nMétricas para la variable de salida "{target}":')

print('Conjunto de entrenamiento:')

print(f'R^2: {r2_train:.4f}')

#Escalar las características del conjunto de validación cruzada

X_val = df_val[features]

#Realizar predicciones en el conjunto de validación cruzada

y_val_pred = model.predict(X_val)

#Métricas de rendimiento en el conjunto de validación cruzada

mae_val = mean_absolute_error(df_val[target], y_val_pred)

r2_val = r2_score(df_val[target], y_val_pred)

print(f'Conjunto de validación cruzada:')

print(f'MAE: {mae_val:.4f}, R^2: {r2_val:.4f}')

```

A8.4. CÓDIGO DE PYTHON PARA EL MODELO ENERGÉTICO DE REGRESIÓN POR ÁRBOLES ALEATORIOS

El ejemplo en concreto hace referencia a los datos del Edificio 1 en la zona de la Entrada. Para el resto de los edificios y zonas, se cambian los nombres de los archivos donde se alojan los datos, los nombres de las hojas de los archivos de Excel y las columnas en las que se alojan las variables de entrada y salida.

```

import pandas as pd

from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import mean_absolute_error, r2_score

#hojas = ['ClimayEntrada', 'ClimayPBaja', 'ClimayP2', 'ClimayP3']

#Leer los datos de entrenamiento

df_train = pd.read_excel('ClimayEd1_2021.xlsx',
sheet_name='ClimayEntrada')

#Leer los datos de validación cruzada

df_val = pd.read_excel('ClimayEd1_2022.xlsx',
sheet_name='ClimayEntrada')

#Variables de entrada (características)

```

```

features = df_train.columns[6:9]

targets = df_train.columns[9:11] #Variables de salida (objetivo)

for target in targets:

    #Variable de salida actual

    y = df_train[target]

    X_train = (df_train[features])

    # Crear y entrenar el modelo de regresión lineal

    model = RandomForestRegressor(n_estimators=100, max_depth=10,
min_samples_split=5)

    model.fit(X_train, y)

    #Realizar predicciones en el conjunto de entrenamiento

    y_train_pred = model.predict(X_train)

    #Métricas de rendimiento en el conjunto de entrenamiento

    r2_train = r2_score(y, y_train_pred)

    print(f'\nMétricas para la variable de salida "{target}":')

    print('Conjunto de entrenamiento:')

    print(f'R^2: {r2_train:.4f}')

    #Escalar las características del conjunto de validación cruzada

    X_val = df_val[features]

    #Realizar predicciones en el conjunto de validación cruzada

    y_val_pred = model.predict(X_val)

    #Métricas de rendimiento en el conjunto de validación cruzada

    mae_val = mean_absolute_error(df_val[target], y_val_pred)

    r2_val = r2_score(df_val[target], y_val_pred)

    print(f'Conjunto de validación cruzada:')

    print(f'MAE: {mae_val:.4f}, R^2: {r2_val:.4f}')

```