

ANEXO I. PREPROCESAMIENTO DE LOS DATOS DEL EDIFICIO ETOPIÁ

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.

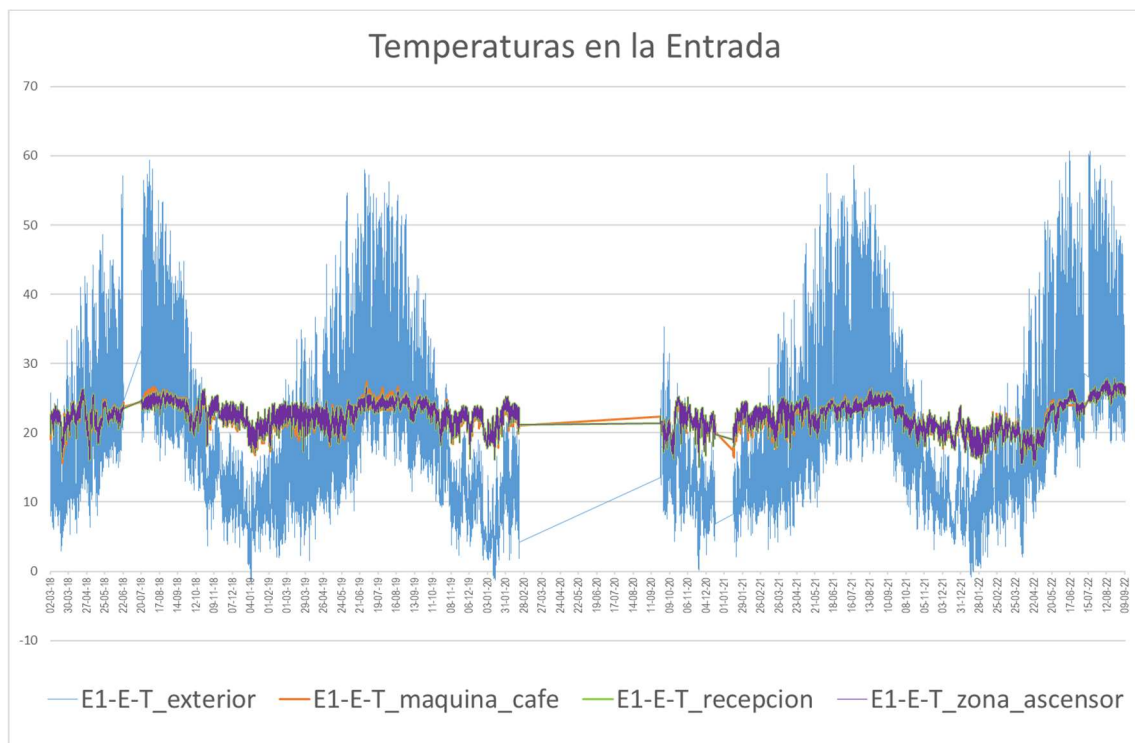


Ilustración 14. Gráfico inicial temperaturas en la entrada del Edificio 1. Fuente: elaboración propia

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.

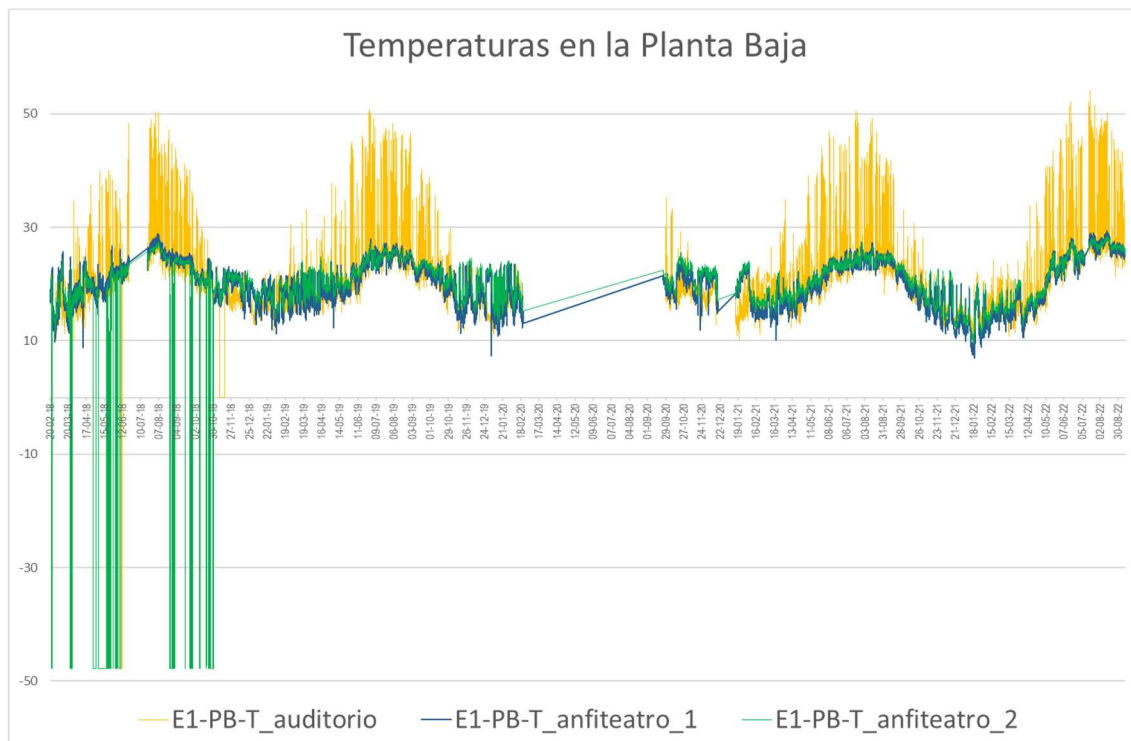


Ilustración 15. Gráfico inicial temperaturas en la planta baja del Edificio 1. Fuente: elaboración propia

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 Etópía, 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.

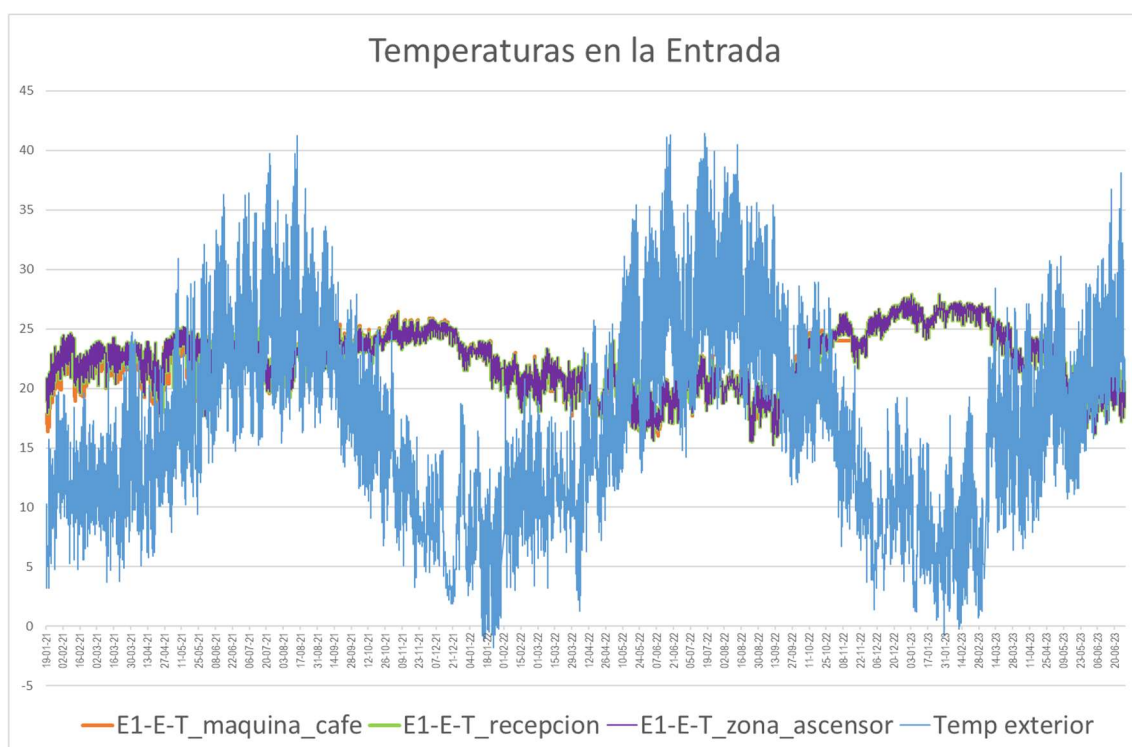


Ilustración 16. Gráfico temperaturas en la entrada del Edificio 1 después del preprocesamiento. Fuente: elaboración propia

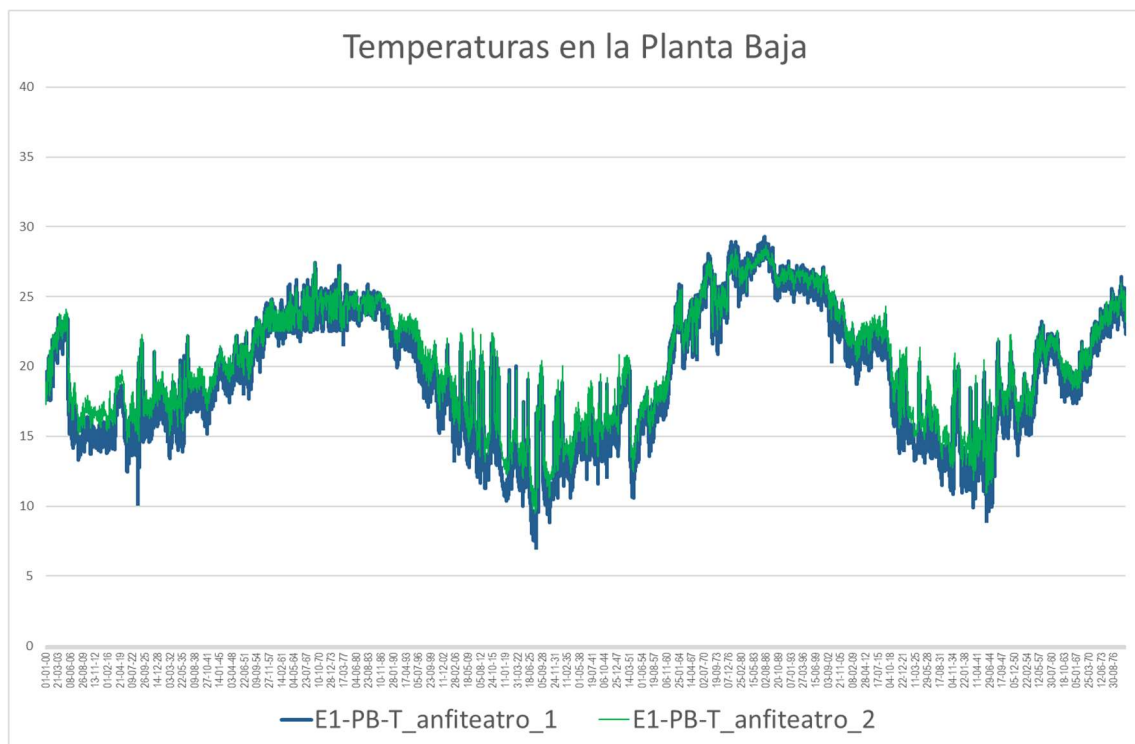


Ilustración 17. Gráfico temperaturas en la planta baja del Edificio 1 después del preprocesamiento. Fuente: elaboración propia

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

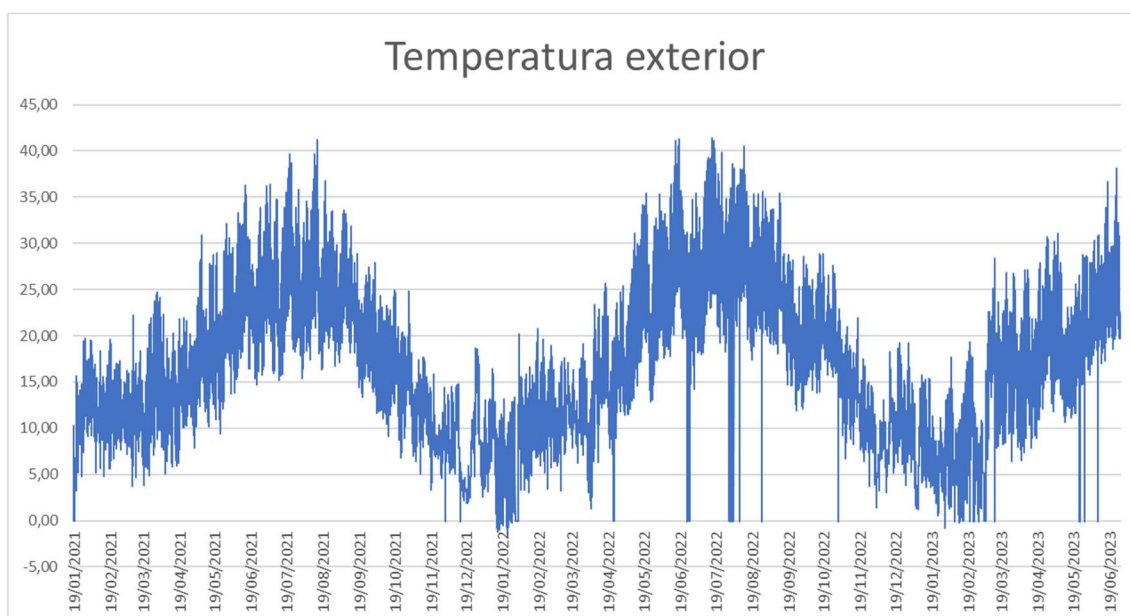


Ilustración 18. Gráfico temperatura exterior antes del preprocesamiento. Fuente: elaboración propia

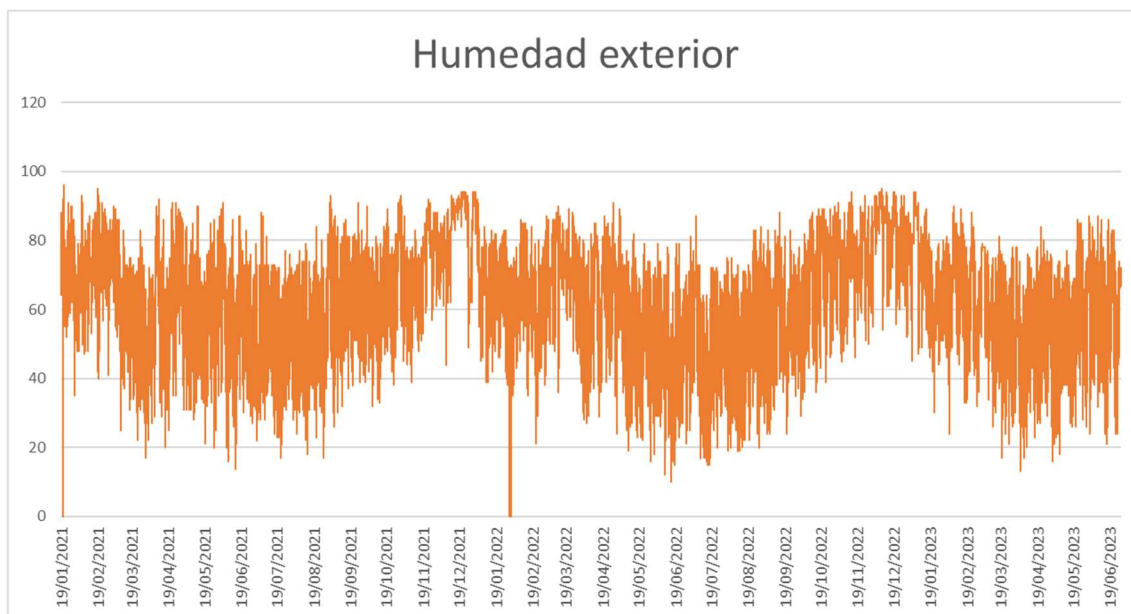


Ilustración 19. Gráfico humedad exterior antes del preprocesamiento. Fuente: elaboración propia

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

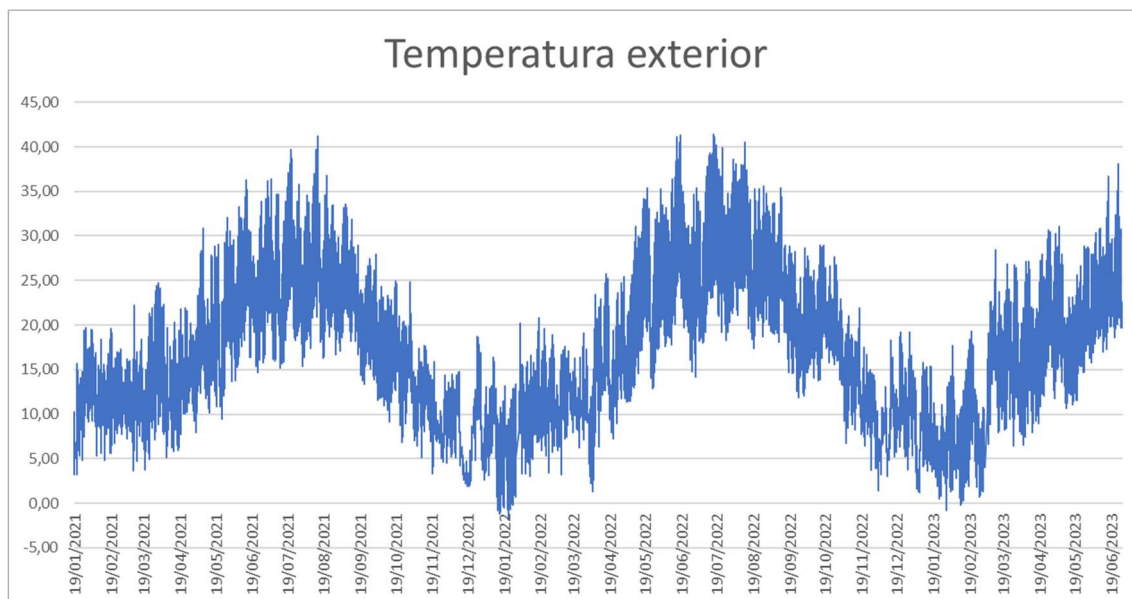


Ilustración 20. Gráfico temperatura exterior después del preprocesamiento. Fuente: elaboración propia

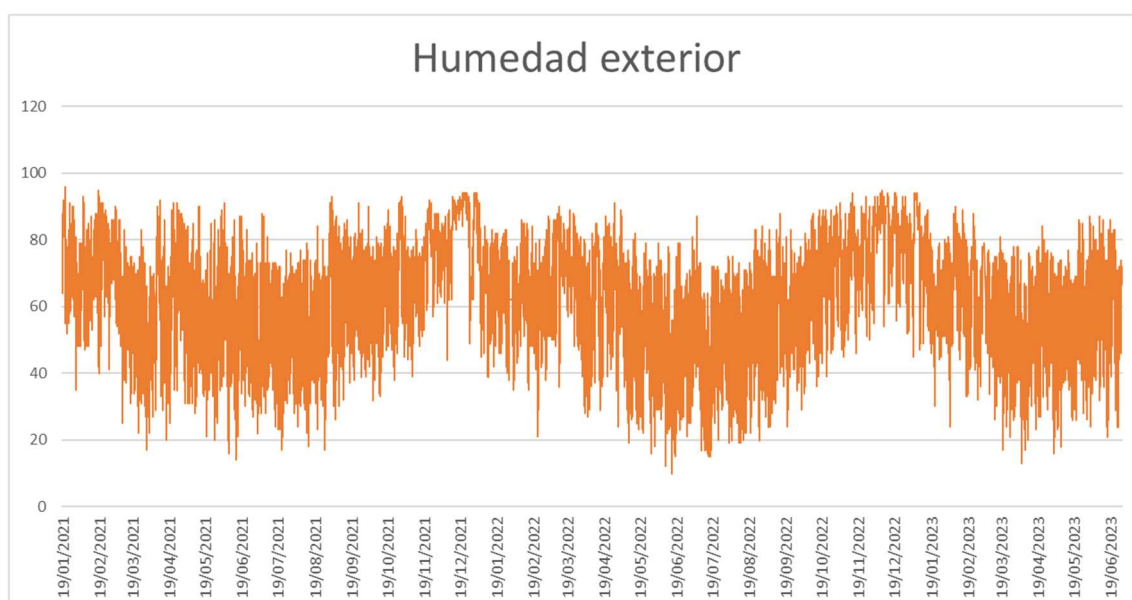


Ilustración 21. Gráfico humedad exterior después del preprocesamiento. Fuente: elaboración propia

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. RECOPIACIÓN 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')

#Agregar 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 ($x_1, x_2, x_3 \dots$).

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

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:

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

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

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

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

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.

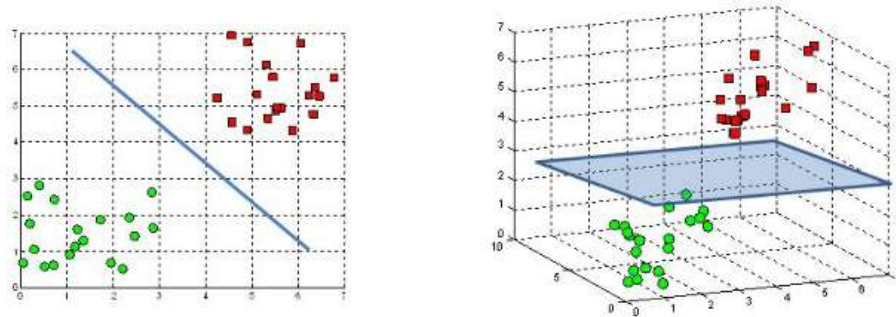


Ilustración 22. Hiperplanos en 2 dimensiones (izquierda) y 3 dimensiones (derecha). Fuente: Medium.

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.

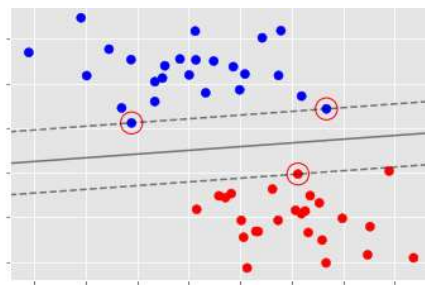


Ilustración 23. Ejemplo de Maximal Margin Hyperplane. Fuente: Ciencia de Datos

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*.

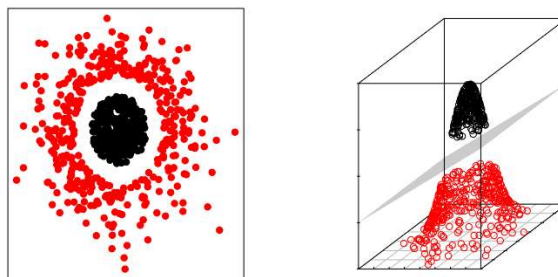


Ilustración 24. Ejemplo de datos con separación no lineal en dos dimensiones. Fuente: Ciencia de Datos

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 Euclidiana entre el punto de consulta y todos los demás puntos en el conjunto de datos.

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

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

Luego, se seleccionan los k puntos con las distancias Euclidianas 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
EDIFICIO 2	Open Art	Temperatura sala	0,7082	0,7133	1,5556
	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	1,5583
	Planta 2	Temperatura 1	0,5294	0,431	1,5747
		Temperatura 2	0,6555	0,5119	1,633
	Planta 3	Temperatura 1	0,5983	0,4817	1,618
		Temperatura 2	0,5117	0,5138	1,6748
	Planta 4	Temperatura 1	0,5452	0,5012	1,5598
		Temperatura 2	0,5666	0,5323	1,6259
	Planta 5	Temperatura 1	0,6086	0,548	1,7315
		Temperatura 2	0,6086	0,5672	1,771
	Planta 6	Temperatura sala	0,6825	0,6614	1,3844
EDIFICIO 3	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
EDIFICIO 2	Open Art	Temperatura sala	0,8835	0,6867	1,5863
	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
EDIFICIO 2	Open Art	Temperatura sala	0,7775	0,708	1,5424
	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

EDIFICIO 3		Temperatura 2	0,6519	0,5134	1,8412
	Planta 6	Temperatura sala	0,7234	0,6197	1,4497
	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
EDIFICIO 2	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
	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
EDIFICIO 2	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
	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
EDIFICIO 3	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
	Planta baja	E calor anfiteatro 1	0,1405	-0,4252	14,8088
		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
	Planta baja	E calor anfiteatro 1	0,4518	0,1732	9,2676
		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

	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
EDIFICIO 2	Open Art	E calor Open Art	0,3058	0,2997	9,7682
		E frío Open Art	0,1951	0,2449	7,212
	Planta 2	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
	Planta 3	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
		E frío auditorio	-0,0185	-0,0362	2,4774
	Planta 2	E calor planta 2	-0,1838	-0,1228	9,9227
		E frío planta 2	-0,1116	-0,0826	4,1932
	Planta 3	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	Planta 6	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
	Aula 1	E calor	-0,0646	-0,0463	3,3733
		E frío	-0,1123	-0,0981	6,7042
	Aula 2	E calor	-0,0119	-0,0051	0,4252
		E frío	-0,3411	-0,1901	15,2386
	Aula 3	E calor	-0,0039	-0,0011	0,0803
		E frío	-0,1429	-0,0907	7,4308
EDIFICIO 3	Zona común	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
		E calor planta 3	-0,0507	-0,0599	4,0589
EDIFICIO 2	Planta 3	E frío planta 3	-0,0318	-0,0317	5,8506
	Open Art	E calor Open Art	0,1977	0,1343	8,7391
		E frío Open Art	-0,0477	-0,0826	5,4461
	Planta 2	E suelo radiante	-0,0418	-0,1377	12,1054
		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
EDIFICIO 3	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
	Planta baja	E calor anfiteatro 1	0,4801	-0,3348	7,4212
		E frío anfiteatro 1	0,1475	-0,1967	5,7204
		E calor anfiteatro 2	0,382	-0,1213	8,0394
		E frío anfiteatro 2	0,3001	-0,5731	10,0727
		E calor auditorio	0,2664	0,0698	4,9153
		E frío 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	Entrada	E suelo radiante	0,9503	0,8213	3,9316
	Planta baja	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

	Planta 3	E Fancoil 3	0,8838	0,5318	3,0732
		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
EDIFICIO 3	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	Entrada	E suelo radiante	0,3974	0,0911	20,5954
	Planta baja	E calor anfiteatro 1	0,7198	-0,7946	9,2042
		E frío anfiteatro 1	0,7194	-9,5316	19,6979
		E calor anfiteatro 2	0,6709	-0,5675	10,5628
		E frío anfiteatro 2	0,7033	-6,9254	21,7801
		E calor auditorio	0,6085	-0,0811	5,5123
		E frío auditorio	0,4958	-0,3309	4,4161
	Planta 2	E calor planta 2	0,1311	0,1489	10,5616
		E frío planta 2	0,164	-5,0007	18,8831
	Planta 3	E calor planta 3	0,0921	0,0312	4,4313
		E frío planta 3	0,1604	-1,5114	17,3573
EDIFICIO 2	Open Art	E calor Open Art	0,0722	0,037	9,6258
		E frío Open Art	0,0885	-0,0845	7,67
	Planta 2	E suelo radiante	0,3346	-0,1072	15,4408
		E Fancoil 1	0,7137	0,223	4,7681
		E Fancoil 2	0,6095	0,1491	5,0866
		E Fancoil 3	0,7923	0,1925	6,0498
	Planta 3	E suelo radiante	0,3415	-0,0736	8,8439
		E Fancoil 1	0,6242	0,0519	11,6849
		E Fancoil 2	0,6087	-0,1599	20,329
		E Fancoil 3	0,7506	-0,1548	17,9596
	Planta 4	E suelo radiante	0,3629	-0,1653	7,0794
		E Fancoil 1	0,2952	0,1293	12,8067
		E Fancoil 2	0,3529	-0,2247	16,3517
		E Fancoil 3	0,36	-0,0432	18,8271
	Planta 5	E suelo radiante	0,4866	-0,4885	3,9826
		E Fancoil 1	0,393	0,0294	11,432
		E Fancoil 2	0,5297	0,0717	15,9205
		E Fancoil 3	0,5288	0,1281	15,1686
	Planta 6	E calor	0,0356	-0,0833	6,2568
		E frío	0,097	0,0236	9,787
EDIFICIO 3	Aula 1	E calor	0,0572	-0,0983	1,204
		E frío	0,2055	-0,0769	25,1027
	Aula 2	E calor	0,0596	-0,0012	0,089
		E frío	0,216	0,025	11,6944
	Aula 3	E calor	0,1549	-0,643	8,5646
		E frío	0,1577	0,0163	9,8755
		E calor 1	0,3694	-0,1272	14,7653

	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
		E frio auditorio	0,6671	0,2308	2,1491
	Planta 2	E calor planta 2	0,5216	0,6581	4,8827
		E frío planta 2	0,4598	-1,6471	8,4048
	Planta 3	E calor planta 3	0,4231	0,0561	4,2486
		E frío planta 3	0,5744	0,429	4,7555
EDIFICIO 2	Open Art	E calor Open Art	0,4396	0,3343	8,5273
		E frío Open Art	0,4246	-0,0848	7,1985
	Planta 2	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
		E Fancoil 3	0,9684	0,6884	1,9123
		E suelo radiante	0,9654	0,9441	0,4708
	Planta 3	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
		E suelo radiante	0,9792	0,9078	0,458
	Planta 4	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
	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
EDIFICIO 3	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 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 = model.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}')

#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:

    #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 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 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 = SVR.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}')

```

A7.3. 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:

    #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 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 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}')

```

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(f'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}')

```