



# Centro Universitario de la Defensa en la Escuela Naval Militar

## TRABAJO FIN DE GRADO

*Modelado de la inmisión del gas radón en espacios cerrados*

## Grado en Ingeniería Mecánica

**ALUMNO:** Ángela Cristina Merello Pérez

**DIRECTORES:** Arturo González Gil

Ana Ogando Martínez

**CURSO ACADÉMICO:** 2024-2025

Universida<sub>de</sub>Vigo





# Centro Universitario de la Defensa en la Escuela Naval Militar

## TRABAJO FIN DE GRADO

*Modelado de la inmisión del gas radón en espacios cerrados*

**Grado en Ingeniería Mecánica**  
Intensificación en Tecnología Naval  
Cuerpo General

UniversidadeVigo



## **RESUMEN**

La presencia de gas radón en espacios cerrados constituye un riesgo para la salud pública, por su relación con el cáncer de pulmón tras exposiciones prolongadas. Como parte de una de las líneas de investigación del CUD-ENM centradas en la monitorización y control del gas radón en espacios cerrados, este trabajo pretende desarrollar un modelo predictivo capaz de estimar la tasa exhalación de gas radón a través del suelo.

Para ello, se construyó una base de datos a partir de mediciones de la concentración de radón en un entorno controlado, complementadas con el registro de diversas variables tomadas en el mismo intervalo temporal. Esto permitió establecer relaciones entre las condiciones del entorno y los niveles de radón observados y estimar la tasa de exhalación de radón.

A partir de esta base de datos, se entrenaron y optimizaron distintos algoritmos de inteligencia artificial. Tras un proceso iterativo de mejora, el modelo Perceptrón Multicapa (MLP) fue el que proporcionó mejores resultados empleando únicamente cuatro variables para predecir la tasa de exhalación. Este trabajo contribuye así al desarrollo de una herramienta capaz de anticipar concentraciones elevadas de radón y facilitar la implementación de medidas preventivas o de mitigación.

## **PALABRAS CLAVE**

Gas radón, Tasa de exhalación, Inteligencia artificial, Modelos predictivos.



# AGRADECIMIENTOS

Este trabajo pone fin a una etapa intensa y llena de experiencias en la ENM. Han sido cinco años de mucho esfuerzo, aprendizaje y momentos que me llevo conmigo para siempre. Ahora empieza una nueva etapa, con ganas de seguir creciendo y cumpliendo metas.

Empiezo dando las gracias a mis tutores, Arturo González Gil y Ana Ogando Martínez, por su implicación, su paciencia y su apoyo durante todos estos meses.

A mi familia, gracias de corazón. En especial a mis padres y a mis dos hermanos, por estar ahí en cada paso, por confiar en mí incluso cuando ni yo lo hacía, y por ser siempre ese apoyo que nunca falla. Todo esto habría sido mucho más difícil sin vosotros.

A mi pareja, gracias por el apoyo que me ha dado durante esta etapa, por ayudarme a ver luz cuando todo se complicaba y por alegrarte conmigo de cada pequeño logro. Ha sido un apoyo constante que ha hecho más llevadero el camino.

Y cómo no, a mi promoción, la 425. Gracias por todo lo vivido juntos. Han sido años llenos de momentos buenos, otros más duros, pero siempre compartidos. Me quedo con las risas, los nervios, las complicidades... y con la certeza de que esta etapa no habría sido igual sin vosotros. A mis compañeras de estudio, gracias por sacar una sonrisa incluso en los peores días y por hacer equipo cuando más hacía falta.

Gracias a todos los que habéis formado parte de este camino.





## CONTENIDO

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| Contenido .....                                                                      | 6  |
| Índice de Figuras .....                                                              | 8  |
| Índice de Tablas.....                                                                | 10 |
| 1 Introducción y objetivos .....                                                     | 11 |
| 1.1 Introducción de la problemática del gas radón .....                              | 11 |
| 1.2 Problemática del gas radón en la Escuela Naval Militar .....                     | 13 |
| 1.3 Objetivos .....                                                                  | 15 |
| 1.4 Estructura .....                                                                 | 15 |
| 2 Estado del arte .....                                                              | 16 |
| 2.1 Generalidades del gas radón .....                                                | 16 |
| 2.1.1 Origen del gas radón .....                                                     | 16 |
| 2.1.2 Comportamiento del gas radón en interiores .....                               | 17 |
| 2.2 Factores que afectan a la inmisión del gas radón en edificios .....              | 17 |
| 2.2.1 Factores geológicos .....                                                      | 17 |
| 2.2.2 Factores constructivos .....                                                   | 18 |
| 2.2.3 Factores meteorológicos .....                                                  | 19 |
| 2.2.4 Factores cíclicos: mareas .....                                                | 20 |
| 2.3 Marco normativo y medidas de protección .....                                    | 21 |
| 2.4 Medición de los niveles de radón .....                                           | 24 |
| 2.4.1 Técnicas para la medición de la concentración de radón en aire .....           | 24 |
| 2.4.2 Medición de exhalación de radón desde superficies .....                        | 24 |
| 2.5 Uso de la IA para predecir la concentración del gas radón .....                  | 25 |
| 2.5.1 Algoritmos de predicción .....                                                 | 26 |
| 2.5.2 Revisión bibliográfica del uso de la IA para la predicción de radón .....      | 31 |
| 3 Materiales y métodos .....                                                         | 32 |
| 3.1 Ubicación y características del lugar de estudio .....                           | 32 |
| 3.2 Toma de datos experimentales .....                                               | 33 |
| 3.2.1 Tasa de exhalación de radón .....                                              | 34 |
| 3.2.2 Variables externas e internas relacionadas con la concentración de radón ..... | 38 |
| 3.3 Tratamiento de datos experimentales .....                                        | 41 |
| 3.3.1 Tasa de exhalación .....                                                       | 41 |
| 3.3.2 Variables externas e internas relacionadas con la concentración de radón ..... | 41 |
| 3.4 Selección de variables .....                                                     | 43 |
| 3.5 Entrenamiento de los modelos predictivos .....                                   | 43 |

|                                                                      |    |
|----------------------------------------------------------------------|----|
| 4 Resultados .....                                                   | 47 |
| 4.1 Filtrado y análisis de datos.....                                | 47 |
| 4.2 Selección de variables.....                                      | 48 |
| 5 Conclusiones .....                                                 | 58 |
| 5.1.1 Aportaciones del estudio.....                                  | 58 |
| 5.1.2 Resumen de los hallazgos .....                                 | 58 |
| 5.1.3 Implicaciones del estudio .....                                | 59 |
| 5.1.4 Implicaciones y líneas futuras .....                           | 59 |
| 6 Bibliografía.....                                                  | 60 |
| Anexo I: implicaciones sociales, y/o ambientales y/o económicas..... | 64 |
| Anexo II: reflexiones éticas y sociales .....                        | 65 |
| Anexo III: códigos de Python.....                                    | 66 |

## ÍNDICE DE FIGURAS

|                                                                                                         |    |
|---------------------------------------------------------------------------------------------------------|----|
| Figura 1-1 Mapa de Galicia de las casas con concentraciones superiores a 300 Bq/m <sup>3</sup> [9]..... | 12 |
| Figura 1-2 Visor cartográfico IGME en la ENM [14].....                                                  | 13 |
| Figura 1-3 Mapa de concentración de radón en la ENM [13].....                                           | 14 |
| Figura 1-4 Representación en 3D de la tasa de exhalación de radón [16] .....                            | 14 |
| Figura 2-1 Cadena de desintegración del uranio[7] .....                                                 | 16 |
| Figura 2-2 Influencia de la proporción envolvente en contacto con el edificio[8] .....                  | 17 |
| Figura 2-3 Influencia del grado de saturación del agua del terreno [22].....                            | 18 |
| Figura 2-4 Relación entre las fases lunares y las concentraciones de radón [24] .....                   | 21 |
| Figura 2-5 Clasificación de las zonas según geografía [26].....                                         | 22 |
| Figura 2-6 Barrera de protección frente al radón [25].....                                              | 22 |
| Figura 2-7 Cámara de aire ventilada [25].....                                                           | 23 |
| Figura 2-8 Despresurización del terreno [25].....                                                       | 23 |
| Figura 3-1 Emplazamiento del CUD dentro de la Escuela Naval militar .....                               | 32 |
| Figura 3-2 Plano de la planta baja del CUD con la ubicación del lugar de estudio .....                  | 33 |
| Figura 3-3 Funcionamiento general del sistema de ventilación en los despachos [19] .....                | 33 |
| Figura 3-4 RandonEye RD200 [40] .....                                                                   | 34 |
| Figura 3-5 Cámara de acumulación [16].....                                                              | 34 |
| Figura 3-6 Soporte [16] .....                                                                           | 35 |
| Figura 3-7 Ejemplo de la concentración de radón en la cámara de acumulación [31].....                   | 36 |
| Figura 3-8 Relación de la concentración entre los diferentes puntos [16] .....                          | 36 |
| Figura 3-9 Chat de Telegram para controlar la apertura y cierre de cubo.....                            | 37 |
| Figura 3-10 Montaje del dispositivo de ventilación.....                                                 | 37 |
| Figura 3-11 Representación en Grafana de las concentraciones de radón .....                             | 38 |
| Figura 3-12 Datos obtenidos en Meteo Galicia [42] .....                                                 | 39 |
| Figura 3-13 Calendario académico curso 2023-2024 [44] .....                                             | 40 |
| Figura 3-14 Registro del control del fancoil.....                                                       | 40 |
| Figura 3-15 Ciclo de apertura/cierre.....                                                               | 41 |
| Figura 3-16 Valor medio de las variables durante los distintos periodos.....                            | 42 |
| Figura 3-17 Moda de las variables cualitativas durante los diferentes periodos.....                     | 42 |
| Figura 3-18 Valor de las variables por cada periodo de cálculo de la tasa de exhalación .....           | 43 |
| Figura 3-19 Parámetros del modelo Random Forest.....                                                    | 44 |
| Figura 3-20 Parámetros del modelo SVR.....                                                              | 44 |
| Figura 3-21 Parámetros del modelo XGBoost .....                                                         | 44 |
| Figura 3-22 Parámetros del modelo GBM .....                                                             | 44 |
| Figura 3-23 Parámetros del modelo MLP .....                                                             | 44 |

|                                                                                       |    |
|---------------------------------------------------------------------------------------|----|
| Figura 4-1 Valores obtenidos manualmente frente a los obtenidos automáticamente ..... | 48 |
| Figura 4-2 Matriz de correlaciones.....                                               | 48 |
| Figura 4-3 Correlación con la tasa de exhalación .....                                | 49 |
| Figura 4-4 Variables utilizadas en cada una de las pruebas .....                      | 50 |
| Figura 4-5 Representación de los valores de <b>2</b> en cada una de las pruebas ..... | 51 |
| Figura 4-6 Representación de los valores MAE en cada una de las pruebas .....         | 51 |
| Figura 4-7 Representación de los valores RMSE en cada una de las pruebas .....        | 51 |
| Figura 4-8 Representación de los valores predichos frente a los valores reales .....  | 52 |
| Figura 4-9 Residuos modelo MLP .....                                                  | 53 |
| Figura 4-10 Residuos modelo GBM .....                                                 | 53 |
| Figura 4-11 Residuos modelo regresión lineal.....                                     | 54 |
| Figura 4-12 Residuos modelo SVR.....                                                  | 54 |
| Figura 4-13 Curva de aprendizaje MLP .....                                            | 55 |
| Figura 4-14 Curva de aprendizaje GBM .....                                            | 55 |
| Figura 4-15 Curva aprendizaje regresión lineal .....                                  | 56 |
| Figura 4-16 Curva de aprendizaje SVR .....                                            | 56 |

## ÍNDICE DE TABLAS

|                                                                                         |    |
|-----------------------------------------------------------------------------------------|----|
| Tabla 1-1 Concentración media radón en diferentes comunidades autónomas [7].....        | 12 |
| Tabla 2-1 Concentración de radón en función de los materiales de construcción [23]..... | 19 |
| Tabla 3-1 Dimensiones cámara acumulación gas [16].....                                  | 35 |

# 1 INTRODUCCIÓN Y OBJETIVOS

## 1.1 Introducción de la problemática del gas radón

El radón es una de las principales fuentes de radiación ionizante natural a la que se encuentra expuesta la población general, lo que lo convierte en un factor de riesgo significativo para la salud pública [1] [2]. Se trata de un gas noble de origen natural que se genera a partir de la desintegración del uranio presente en el suelo, las rocas y el agua, estando ampliamente distribuido en diversos entornos geológicos a nivel mundial. A pesar de ser inodoro, incoloro e insípido, su presencia en espacios cerrados puede tener consecuencias graves para la salud humana.

En los últimos años, el interés científico y político por el radón ha aumentado debido a la creciente evidencia que lo vincula con el cáncer de pulmón. Diversos estudios lo han destacado como la principal causa de esta enfermedad en no fumadores y la segunda en fumadores, después del tabaco [3]. La Organización Mundial de la Salud (OMS) estima que entre un 3% y un 14% de los casos de cáncer de pulmón a nivel mundial están relacionados con la exposición al radón en el hogar [4]. En Europa, se calcula que este gas es responsable del 9% de las muertes por esta enfermedad, mientras que en España representa aproximadamente el 4%, con alrededor de 1.500 muertes anuales [5].

Su peligrosidad radica en su capacidad de emitir partículas radiactivas durante su desintegración. Cuando estas partículas son inhaladas, pueden quedar depositadas en los tejidos pulmonares, lo que incrementa la probabilidad de mutaciones celulares y, en consecuencia, el desarrollo de patologías graves. Este riesgo se agrava en aquellas personas que permanecen expuestas a concentraciones elevadas de radón en espacios cerrados durante periodos prolongados [6].

El radón supone un desafío adicional porque no puede detectarse sin el uso de equipos especializados. Su gran capacidad de movilidad le permite filtrarse en los edificios a través de grietas, y sistemas de drenaje. Si la ventilación no es adecuada, su acumulación en espacios cerrados puede aumentar significativamente los riesgos para la salud.

La actividad del radón, fundamental para evaluar los riesgos, se mide en Becquerel (Bq), una unidad que representa el número de desintegraciones radiactivas por segundo. En espacios interiores, su concentración se indica en Becquerel por metro cúbico de aire (Bq/m<sup>3</sup>). Según el Consejo de Seguridad Nuclear (CSN), las viviendas en España tienen una concentración media anual de 43 Bq/m<sup>3</sup>.

En 1991, el CSN lanzó el proyecto MARNA con el objetivo de evaluar la radiación gamma existente en España y detectar áreas con mayor riesgo de acumulación de radón. La Tabla 1-1 muestra las concentraciones medias de radón detectadas en viviendas en diferentes comunidades autónomas. Estas diferencias se deben principalmente a las características geológicas de cada región.

| Comunidad autónoma | Concentración media anual de radón (Bq/m <sup>3</sup> ) |
|--------------------|---------------------------------------------------------|
| Galicia            | 118                                                     |
| Madrid             | 95                                                      |
| Extremadura        | 90                                                      |
| Castilla y León    | 68                                                      |
| Canarias           | 64                                                      |

**Tabla 1-1 Concentración media radón en diferentes comunidades autónomas [7]**

En Galicia, la geología granítica facilita la emisión y acumulación del gas radón. Estudios recientes indican que el 70% de la población gallega reside en áreas donde la concentración de radón supera los 300 Bq/m<sup>3</sup> [8], un nivel que triplica el umbral recomendado por la Organización Mundial de la Salud (OMS), fijado en 100 Bq/m<sup>3</sup> de media anual. En Figura 1-1 se muestran en color marrón aquellas poblaciones en las que el 70% de las viviendas triplican dicho valor.

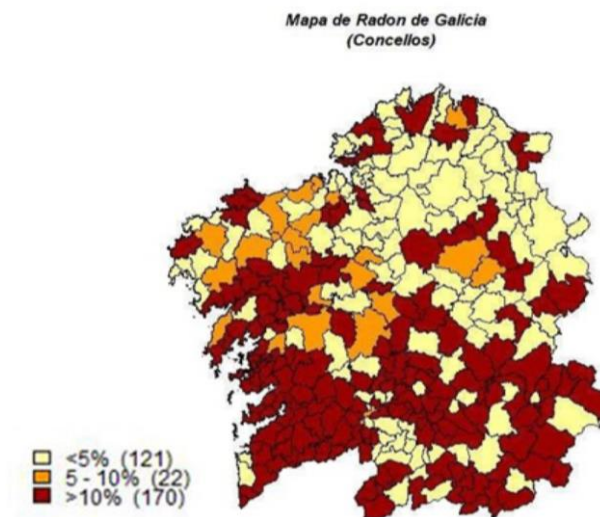


**Figura 1-1 Mapa de Galicia de las casas con concentraciones superiores a 300 Bq/m<sup>3</sup> [9]**

Con el objetivo de abordar esta problemática, España sigue la Directiva 2013/59/Euratom de la Unión Europea [10], estableciendo en el Código Técnico de la Edificación (CTE) un nivel de referencia promedio anual de 300 Bq/m<sup>3</sup> [11]. Esta normativa busca reducir la exposición al radón mediante medidas como la mejora de la ventilación en nuevas construcciones y la implementación de soluciones correctivas en edificios existentes. Sin embargo, su cumplimiento efectivo sigue siendo un desafío, especialmente en zonas con alta incidencia, donde las características geológicas demandan esfuerzos adicionales para garantizar la seguridad de los habitantes.

Para mitigar los riesgos asociados al radón, se han identificado tres enfoques principales que abordan construcciones nuevas como ya existentes. En edificaciones de nueva construcción, se recomienda la integración de medidas preventivas desde la fase de diseño, como barreras antirradiación o sistemas de ventilación. En edificios ya construidos, se pueden adoptar reformas estructurales, como el sellado de grietas y la mejora del aislamiento. Sin embargo, estas soluciones pueden no siempre ser viables debido a limitaciones constructivas o económicas [12].

En estos casos, la ventilación se presenta como la alternativa más efectiva. La ventilación natural, mediante la apertura de ventanas, permite renovar el aire y dispersar el gas, aunque su eficacia depende de factores externos, lo que limita su aplicación. Por ello, los sistemas de ventilación mecánica representan una opción más eficiente, ya que, al no depender de condiciones externas, permiten regular la calidad del aire de forma continua y controlada, logrando así una reducción sostenida de la concentración de radón [7].

No obstante, para controlar eficazmente los niveles de radón en espacios interiores, es esencial anticipar su comportamiento. La concentración de este gas varía en función de múltiples factores internos y externos, por lo que prever dichas fluctuaciones permite aplicar medidas preventivas de forma más eficiente.

En este contexto, la inteligencia artificial se presenta como una herramienta clave. A través de modelos que procesan información sobre las características del edificio, el entorno y las condiciones de uso, es posible predecir las variaciones en la concentración de radón. Esta información permite ajustar la ventilación de forma automatizada y precisa, evitando acumulaciones peligrosas y optimizando la eficiencia energética.

## 1.2 Problemática del gas radón en la Escuela Naval Militar

La Escuela Naval Militar (ENM), ubicada en Pontevedra, se enfrenta a un desafío debido a las elevadas concentraciones de radón en varios de sus edificios, lo que podría representar un riesgo para la salud de su personal y estudiantes. Este fenómeno está relacionado con la geología del terreno, que favorece la emisión de radón y da lugar a niveles superiores a los establecidos por la normativa europea. Por ello, desde 2019 se llevan a cabo diversos estudios para comprender mejor la distribución del radón en la ENM y evaluar sus posibles efectos.

Estudios realizados por la AN Camacho Martín y el AN Lázaro Benavides han demostrado que dentro de la Escuela Naval Militar existen dos tipos de suelos claramente diferenciados [13] [14]. Estas diferencias pueden observarse en el visor cartográfico del IGME (ver Figura 1-2), que muestra las áreas con terrenos graníticos, representados en rosa, y las zonas compuestas por acumulaciones de grava, arcillas y limos, marcadas en gris. Los terrenos graníticos, debido a su alta concentración de uranio y permeabilidad, facilitan la emisión del gas radón, mientras que las áreas con depósitos de grava y arcillas limitan su exhalación. Esta diferencia en la composición del suelo explica, en parte, las variaciones significativas en las concentraciones de radón observadas entre los edificios de la ENM.

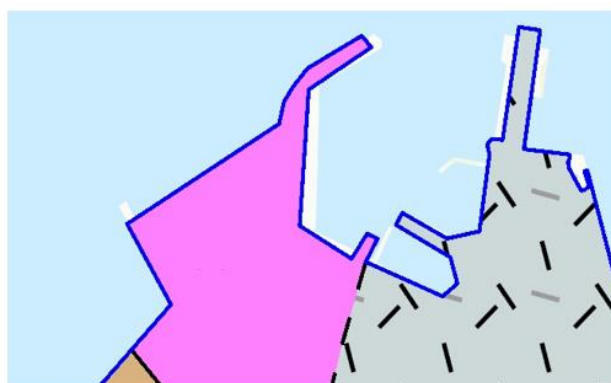


Figura 1-2 Visor cartográfico IGME en la ENM [14]

Se llevaron a cabo numerosas mediciones en distintos puntos de la ENM para evaluar la concentración de radón. Como se observa en la Figura 1-3, varias áreas registraron niveles elevados de este gas, destacando especialmente el Edificio de Investigación y la planta baja del Cuartel Marqués de la Victoria antes de realizarse la reforma, donde se superaron los límites recomendados por la normativa española. Además, en algunos despachos del Centro Universitario de la Defensa (CUD), las mediciones

revelaron que, en ausencia de ventilación, las concentraciones de radón pueden triplicar los valores permitidos por la normativa, lo que subraya la necesidad de una atención prioritaria [13] [15].

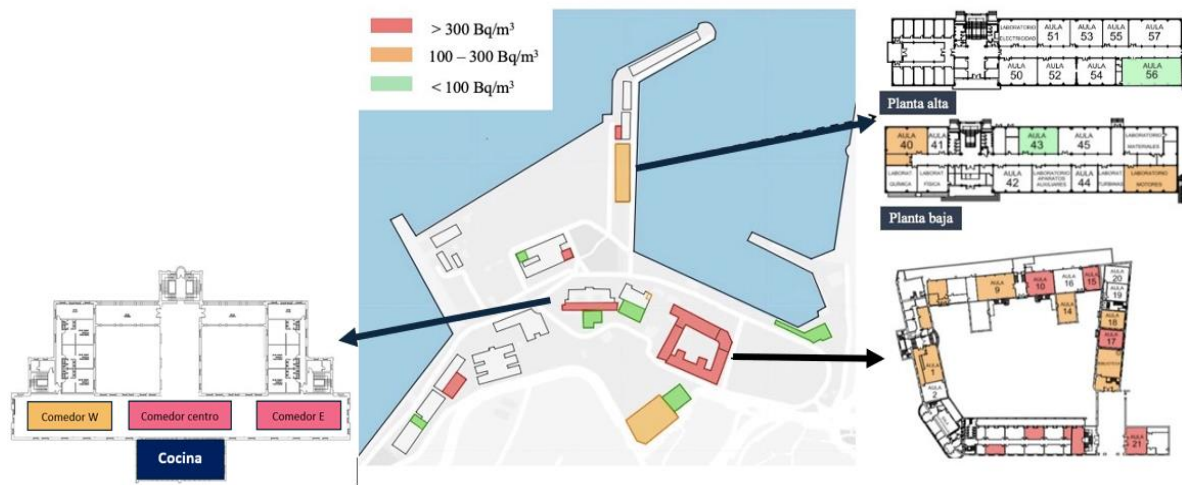


**Figura 1-3 Mapa de concentración de radón en la ENM [13]**

En este contexto, una de las líneas de investigación del CUD es el estudio de la entrada de radón en edificios, así como su comportamiento y distribución dentro de los mismos. Entre los estudios llevados a cabo, destaca el realizado por la AN Marín Alonso [16], en el que se caracterizó la inmisión del gas radón en un despacho del CUD a través del suelo. El estudio ha demostrado que la entrada de radón no ocurre de manera homogénea, sino que se concentra en zonas específicas, como se observa en la Figura 1-4. Esta variabilidad está influenciada por las características del suelo y la presencia de grietas en el subsuelo, que facilitan su infiltración. Estas zonas coinciden con la presencia de grietas en el subsuelo, que actúan como vías preferentes para la infiltración del radón. Gracias al mapa de inmisión generado en este estudio, es posible estimar la cantidad de radón en el despacho a partir de la medición de un único punto, aprovechando la relación entre las zonas de mayor y menor concentración.

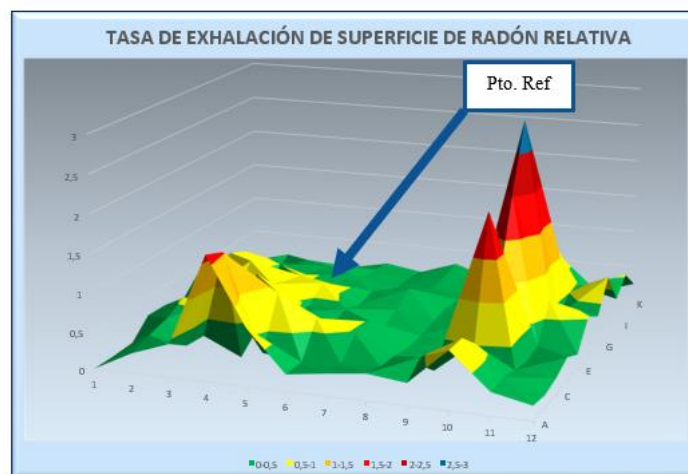


**Figura 1-4 Representación en 3D de la tasa de exhalación de radón [16]**

A partir de estos hallazgos, el presente Trabajo de Fin de Grado pretende ampliar la investigación mediante el desarrollo de un modelo predictivo capaz de estimar la tasa de inmisión del gas radón. El modelo se centrará en predecir la cantidad total de radón que entra al despacho 114 a través de un punto específico del suelo, utilizando como punto de referencia el mostrado en el mapa de la Figura 1-4. Para ello, se tendrán en cuenta diversas variables que describen las condiciones internas y externas del despacho en el momento de las mediciones.

### 1.3 Objetivos

La meta final del presente trabajo es desarrollar un modelo predictivo capaz de estimar la tasa exhalación de gas radón en el despacho 114 del CUD. Este modelo se construye a partir del análisis de variables internas y externas, con el objetivo de, en un futuro, poder anticiparse a niveles elevados de radón y facilitar la implementación de medidas de mitigación de forma más eficaz. No se busca innovar en el campo de los modelos predictivos, sino evidenciar su aplicabilidad en la gestión preventiva de la calidad del aire interior. Para ello, se plantean los siguientes objetivos:

- Identificar las variables que más influyen en la entrada y acumulación de radón, teniendo en cuenta factores como la ventilación, las condiciones meteorológicas, el uso del espacio y la actividad humana.
- Analizar la relación entre estas variables y la tasa de exhalación de radón, aplicando herramientas de correlación para seleccionar las más significativas.
- Desarrollar y comparar diferentes modelos de predicción utilizando algoritmos de aprendizaje automático, evaluando su capacidad para estimar la tasa de exhalación de radón.
- Validar y optimizar el rendimiento de los modelos desarrollados, garantizando su fiabilidad y precisión en diferentes condiciones.

### 1.4 Estructura

En este apartado se expone de forma general la estructura seguida a lo largo del Trabajo de Fin de Grado.

Tras una introducción inicial, el Capítulo 2 presenta un análisis del estado del arte relacionado con el gas radón, abordando su origen, comportamiento en interiores, factores que influyen en su inmisión, la normativa vigente, técnicas de medición, así como la posible aplicación de la inteligencia artificial en su predicción.

En el Capítulo 3 se expone el desarrollo práctico del trabajo, detallando el lugar de estudio, los equipos utilizados, la metodología de toma de datos, así como el proceso seguido para la estandarización y análisis de las mediciones experimentales.

El Capítulo 4 recoge los principales resultados obtenidos, incluyendo el rendimiento de los modelos predictivos implementados, el análisis de residuos y la evaluación de las curvas de aprendizaje.

Finalmente, el Capítulo 5 presenta las conclusiones extraídas del estudio, además de proponer futuras líneas de investigación enfocadas a la mejora del modelado y control del gas radón en espacios cerrados.

Además, se incluyen varios anexos complementarios: el código fuente empleado en el desarrollo del modelo predictivo, una reflexión sobre las implicaciones éticas y sociales del trabajo, así como un análisis específico de sus posibles repercusiones sociales, económicas y ambientales.

## 2 ESTADO DEL ARTE

### 2.1 Generalidades del gas radón

#### 2.1.1 Origen del gas radón

El radón es un gas noble radiactivo de origen natural que se genera en la cadena de desintegración del uranio (ver Figura 2-1), un elemento presente en la corteza terrestre. Existen tres isótopos principales de este gas: el torón (Rn-220), el actinón (Rn-219) y el radón (Rn-222). Entre ellos, el radón (Rn-222) destaca por ser el isótopo con mayor impacto radiológico debido a su vida media de 3,824 días, significativamente superior a la del torón y el actinón, cuyos periodos de desintegración son mucho más breves [6] [7]. Su gran movilidad le permite acumularse en espacios cerrados, donde puede alcanzar concentraciones peligrosas para la salud. Además de su impacto radiológico, el radón (Rn-222) es el isótopo más abundante entre los tres. Por estas razones, y para simplificar su referencia en adelante, este isótopo será denominado simplemente como radón.

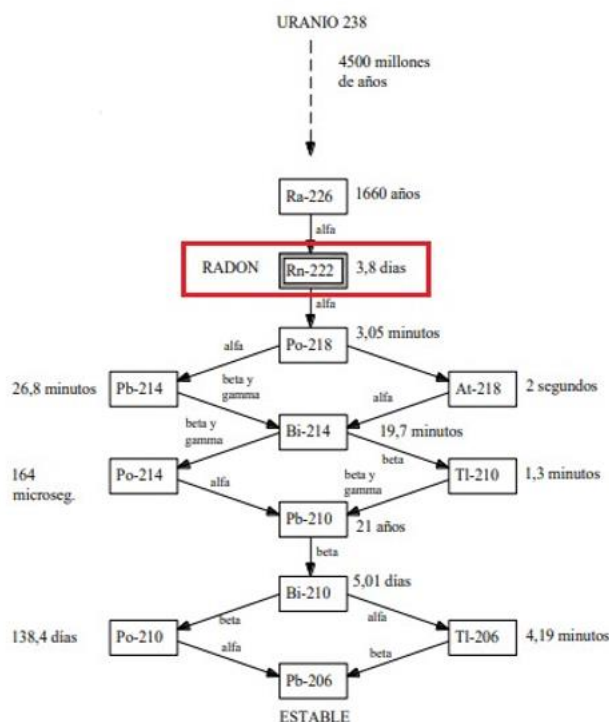


Figura 2-1 Cadena de desintegración del uranio[7]

### 2.1.2 Comportamiento del gas radón en interiores

El radón que se encuentra en los edificios proviene principalmente de tres fuentes: el suelo, los materiales de construcción y el agua subterránea, siendo el suelo la fuente predominante. Este gas se infiltra en los edificios a través de grietas, fisuras y juntas en los cimientos, así como mediante la porosidad de los materiales en contacto con el terreno [17]. Su entrada está influenciada por las características del terreno, como por ejemplo la permeabilidad y la presión diferencial entre el interior y el exterior del edificio, que facilita su movimiento hacia los espacios cerrados.

La cantidad de radón que puede ingresar desde el suelo depende directamente de la superficie del edificio en contacto con este. Como se muestra en Figura 2-2, a mayor superficie, mayor será la cantidad de radón que pueda infiltrarse, especialmente si hay grietas o fisuras en los cimientos. Sin embargo, el fenómeno disminuye significativamente cuando se emplean barreras anti-radón o sistemas de ventilación en cámaras de aire, que dificultan el paso del gas hacia el interior del edificio y reducen su acumulación en las zonas habitables [18].

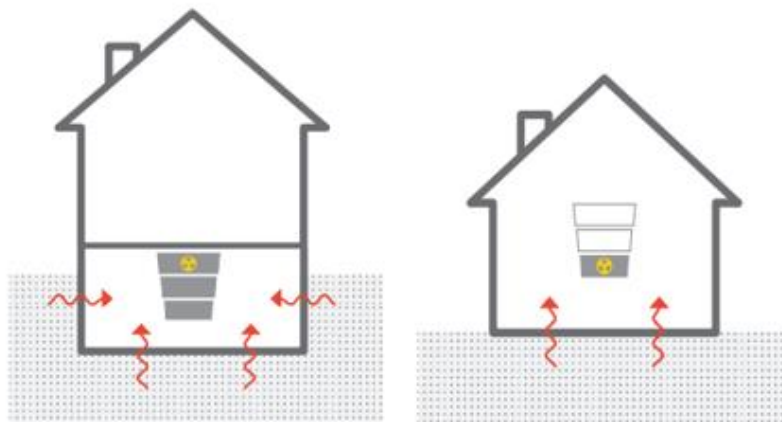


**Figura 2-2 Influencia de la proporción envolvente en contacto con el edificio[8]**

Una vez dentro del edificio, el radón no se distribuye de manera uniforme debido a su mayor densidad en comparación con el aire. Tiende a acumularse en las plantas bajas y sótanos, donde la ventilación es más limitada, dificultando su dispersión. Por el contrario, en las plantas superiores, la mayor circulación de aire facilita su dilución, reduciendo así su concentración [19].

La distribución del radón también está influenciada por la comunicación entre los distintos espacios del edificio. La ausencia de un aislamiento adecuado entre los sótanos y las zonas habitables permite que el gas se desplace hacia áreas ocupadas, aumentando el riesgo de exposición. Asimismo, los sistemas de ventilación juegan un papel clave en la concentración de radón en interiores. Un sistema bien diseñado favorece la dilución del gas, reduciendo su impacto. En cambio, si la ventilación provoca depresiones internas, puede generar un efecto de succión que facilita la entrada del radón desde el subsuelo, incrementando su presencia en el ambiente.

Además de los factores internos, las concentraciones de radón en interiores están condicionadas por elementos externos, como son los factores geológicos, los constructivos y los meteorológicos.

## 2.2 Factores que afectan a la inmición del gas radón en edificios

### 2.2.1 Factores geológicos

La inmición de gas radón en los edificios está estrechamente relacionada con las características geológicas del terreno donde se ubican. Uno de los factores más influyentes es la composición del subsuelo, particularmente el contenido de uranio presente en las rocas. Los suelos graníticos, debido a su alto contenido de uranio, generan mayores cantidades de radón. Por el contrario, los suelos arcillosos, aunque también ricos en uranio, liberan menos gas debido a su baja permeabilidad. Por último, las rocas

calcáreas y otros materiales sedimentarios, aunque contienen menos uranio, destacan por favorecer la liberación del radón gracias a su estructura porosa y fracturada [20].

La permeabilidad del terreno es un factor fundamental que influye en el desplazamiento del radón hacia la superficie. En general, los suelos con una alta conectividad entre poros permiten que el gas se mueva con mayor facilidad, mientras que los terrenos compactos o saturados de agua dificultan su avance [21]. Además, la presencia de grietas, fracturas o discontinuidades en el subsuelo puede aumentar localmente la permeabilidad, generando caminos preferentes por los que el radón asciende más rápidamente. Por ello, incluso en zonas donde el suelo parece poco permeable, como los suelos arcillosos, pueden detectarse altas concentraciones de radón si existen estas irregularidades estructurales.

El grado de saturación de agua en el terreno también influye en la movilidad del radón. Como se observa en Figura 2-3, en suelos secos, el gas asciende con mayor facilidad hacia los edificios, mientras que en terrenos saturados de agua, su desplazamiento se ve obstaculizado debido a que el agua tiene un coeficiente de difusión mucho menor que el aire [21]. Este efecto se acentúa si el terreno circundante está húmedo, generando un gradiente que facilita la migración del gas hacia zonas secas, como los interiores de los edificios.

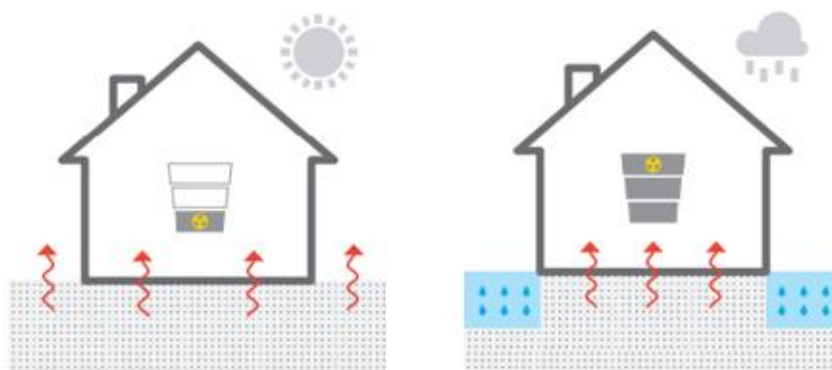


Figura 2-3 Influencia del grado de saturación del agua del terreno [22]

### 2.2.2 Factores constructivos

Los factores constructivos juegan un papel esencial en la inmisión del gas radón en los espacios interiores. La composición de los materiales utilizados en la construcción de una vivienda puede ser una fuente importante de exhalación de este gas. Diversos estudios han demostrado que algunos materiales pueden contribuir con niveles adicionales de radón, que oscilan entre 5 y 20 Bq/m<sup>3</sup> en espacios cerrados. Esta variación depende del tipo de material, su permeabilidad y la presencia de elementos radiactivos en su composición.

Un estudio realizado por el “*National Radiation Protection Institute*” identificó diferencias significativas en la emisión de radón entre materiales como el hormigón, los ladrillos cerámicos y el granito [23]. Estas diferencias se deben principalmente a la presencia de uranio y otros elementos radiactivos en su estructura.

La Tabla 2.1 permite visualizar y comparar los niveles de emisión de estos materiales, facilitando la toma de decisiones en el diseño constructivo para minimizar la presencia de radón en los espacios interiores. Por esta razón, es recomendable optar por materiales de baja permeabilidad al radón y aplicar revestimientos impermeabilizantes que limiten su liberación.

Como se observa en dicha tabla, materiales como las cenizas volantes, la arcilla y el hormigón presentan los valores más altos de concentración de radón, superando en algunos casos los 190 Bq/kg. Esto evidencia la importancia de seleccionar cuidadosamente los componentes utilizados en la construcción, especialmente en zonas con elevada susceptibilidad geológica. La elección de materiales con baja radiactividad natural puede contribuir significativamente a reducir la exposición prolongada al radón en interiores.

| <b>Materiales de construcción</b> | <b>Mínimo (Bq/kg)</b> | <b>Máximo (Bq/kg)</b> |
|-----------------------------------|-----------------------|-----------------------|
| Ladrillos                         | 45,2                  | 143                   |
| Hormigón                          | 21,1                  | 192                   |
| Arena                             | 13,3                  | 41                    |
| Arcilla                           | 40,9                  | 199                   |
| Cemento                           | 36,5                  | 88                    |
| Cenizas volantes                  | 75,5                  | 363                   |
| Limos                             | 12,5                  | 94                    |
| Yesos                             | 12,1                  | 86                    |

**Tabla 2-1 Concentración de radón en función de los materiales de construcción [23]**

Además de la cuidadosa selección de materiales, otro factor determinante en la inmisión del radón en espacios habitables es la ventilación del edificio. Los sistemas de ventilación pueden generar fenómenos de sobrepresión y depresión que influyen directamente en la forma en que el radón se infiltra desde el subsuelo o los elementos constructivos. En condiciones de depresión, el aire interior puede actuar como una fuerza de succión, atrayendo radón desde el terreno y a través de materiales porosos como ciertos tipos de hormigón y ladrillo, lo que puede derivar en un aumento significativo de su concentración. En contraste, una ventilación controlada y bien diseñada que mantenga una ligera sobrepresión en el interior puede reducir eficazmente la entrada de radón, favoreciendo así un ambiente más saludable.

### *2.2.3 Factores meteorológicos*

La inmisión de gas radón en espacios interiores está sujeta a una amplia gama de factores meteorológicos, que modifican tanto su liberación desde el suelo como su comportamiento dentro de los edificios. Los principales factores meteorológicos que impactan la inmisión del gas radón son los que a continuación se explican:

#### **Presión atmosférica**

La presión atmosférica destaca como uno de los factores más influyentes en la inmisión del gas radón en los espacios interiores debido a su capacidad de modificar los gradientes de presión entre el terreno y los edificios [7]. Estos gradientes facilitan o restringen el movimiento del gas, siendo particularmente significativos en terrenos permeables y en edificaciones con fisuras que actúan como vías de acceso.

En el caso de una disminución de la presión atmosférica, los gradientes resultantes tienen un efecto directo sobre la entrada de radón. Se crea una diferencia entre la presión del terreno y la del aire exterior [18]. Este desajuste genera un efecto de succión que intensifica el flujo del radón desde el suelo hacia el interior de los edificios. En estas condiciones, los niveles de radón pueden aumentar significativamente en pocas horas.

Por el contrario, durante periodos de alta presión atmosférica, los gradientes cambian de dirección, dificultando la entrada del radón en los edificios. En estas circunstancias, el gas tiende a acumularse en el subsuelo, reduciendo su concentración en los espacios interiores. Este comportamiento se observa cuando las condiciones exteriores son más estables y las variaciones son mínimas.

Además de estos cambios bruscos, existen variaciones más graduales como las mareas barométricas, que también afectan las concentraciones de radón. Estas variaciones diurnas y semi-diurnas, provocadas

por el calentamiento y enfriamiento de la atmósfera, generan flujos continuos de radón hacia el interior de los edificios, incluso en ausencia de grandes gradientes de presión.

### **Temperatura**

La temperatura está directamente relacionada con la presión atmosférica, y su influencia en la concentración del gas radón se manifiesta principalmente a través de los gradientes de presión que se generan entre el subsuelo y los espacios interiores. Aunque el impacto directo de la temperatura es reducido, dado que los cambios de presión que genera son mínimos, en torno a 1 Pascal, [7] [18]. A pesar de su efecto limitado, este factor puede amplificar la movilidad del radón al interactuar con factores como la ventilación o la permeabilidad del suelo.

Un fenómeno asociado a la temperatura es el conocido como "efecto chimenea", que refleja cómo las diferencias térmicas entre el interior y el exterior de un edificio generan corrientes de aire ascendentes. Estas corrientes internas alteran la distribución de presión en distintos niveles de la estructura, especialmente en la base, donde se crea un gradiente negativo que favorece la succión del gas desde el subsuelo. Este fenómeno es especialmente relevante en épocas frías, cuando la diferencia de temperatura es mayor y la presión interior tiende a ser inferior a la del exterior.

### **Viento**

El viento influye en la concentración del gas radón de manera significativa, pudiendo tanto reducirla como incrementarla según las condiciones del entorno y las características de las edificaciones. Este impacto está condicionado por tres factores principales: la velocidad del viento, su dirección y la configuración arquitectónica de los edificios, incluyendo la orientación, la estanqueidad y la presencia de aberturas o grietas en la estructura[23].

La velocidad del viento genera diferencias de presión entre distintas zonas del edificio, especialmente en las fachadas, techos y otras superficies expuestas al flujo de aire. Estas diferencias de presión, conocidas como "efecto de succión" o "efecto Venturi", facilitan la entrada de radón desde el subsuelo hacia los espacios habitables. A medida que aumenta la velocidad del viento, estas diferencias se intensifican, provocando un mayor gradiente de presión que incrementa el flujo de aire y, por consiguiente, de radón hacia el interior.

### **Precipitaciones**

La permeabilidad del suelo se ve directamente afectada por la cantidad de agua que contiene, especialmente durante los periodos de lluvia. El agua que se infiltra en el terreno satura los poros, disminuyendo su permeabilidad y dificultando el transporte del gas radón hacia la superficie. Esta saturación obliga al radón a desplazarse hacia las áreas secas que se encuentran bajo los edificios, donde el terreno, protegido por la estructura, permanece más permeable y actúa como una vía preferente para su emisión [7]. Por eso, cuando llueve, se espera más radón en los edificios.

Las lluvias, además, suelen coincidir con periodos de baja presión atmosférica. Estas condiciones intensifican el movimiento del radón desde el subsuelo hacia los espacios interiores, lo que genera un aumento temporal de las concentraciones en los edificios. Cuando las precipitaciones se prolongan, el terreno se satura de forma constante, lo que reduce la emisión de radón hacia el exterior y disminuye su acumulación en el interior, especialmente si las edificaciones cuentan con sistemas de ventilación adecuados [23].

#### **2.2.4 Factores cíclicos: mareas**

Las mareas, tanto terrestres como oceánicas, influyen en la inmisión de gas radón al modificar las condiciones del subsuelo y favorecer su exhalación hacia la superficie. Las mareas terrestres, originadas por las fuerzas gravitacionales de la Luna y el Sol, generan ciclos periódicos de compresión y dilatación en la corteza terrestre. Este movimiento altera las presiones internas del terreno, creando gradientes que facilitan el desplazamiento del radón atrapado en las capas subterráneas hacia la atmósfera. Este

fenómeno resulta especialmente relevante en terrenos porosos o fracturados, donde la estructura geológica permite un flujo más eficiente del gas hacia el exterior.

Un aspecto destacable de las mareas terrestres es su relación con las fases de la Luna. Durante las fases de Luna llena y nueva, las fuerzas gravitacionales alcanzan su máxima intensidad, amplificando los efectos de las mareas sobre la corteza terrestre. Sin embargo, el incremento en la exhalación de radón no es inmediato. Estudios, como el denominado *Tidal Synchrony of Built Environment Radon Levels in the UK*, han identificado un retraso de 3 a 4 días en este proceso [24]. Este retardo depende principalmente de la profundidad de las capas rocosas afectadas; a mayor profundidad, más tiempo tardan las presiones en propagarse y generar un aumento en la exhalación del gas.

La Figura 2-4, muestra la relación entre la concentración de radón y las fases lunares, evidenciando un patrón cíclico con picos que coinciden con los periodos de mayor atracción gravitacional. Esto ilustra la complejidad de la interacción entre las fuerzas gravitacionales y las propiedades del terreno.

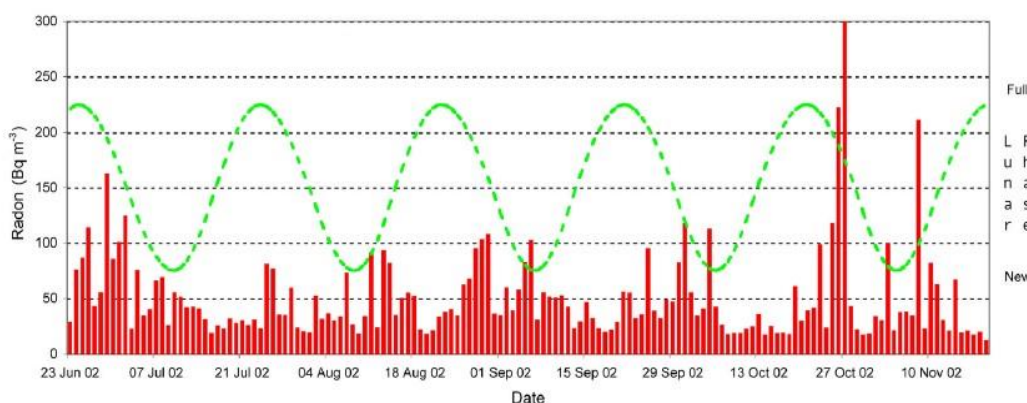


**Figura 2-4 Relación entre las fases lunares y las concentraciones de radón [24]**

En las regiones costeras, las mareas oceánicas afectan la liberación de radón de manera localizada al alterar periódicamente el nivel freático y las presiones en el subsuelo. Este efecto puede aumentar o reducir la exhalación del gas, dependiendo del tipo de terreno. Los suelos altamente permeables, como los arenosos y gravosos, son especialmente sensibles a estos cambios.

## 2.3 Marco normativo y medidas de protección

En España, la exposición al radón en edificios está regulada por el CTE. Este marco normativo establece medidas específicas para garantizar que las concentraciones de este gas no superen el límite anual promedio de 300 Bq/m³ [25]. Estas regulaciones están en línea con la Directiva 2013/59/EURATOM de la Unión Europea [10].

El CTE establece medidas específicas según el tipo de edificio, diferenciando entre construcciones nuevas y edificaciones ya existentes. En los edificios de nueva construcción, se pueden aplicar soluciones constructivas más efectivas desde la fase de diseño, permitiendo una mayor integración de barreras preventivas contra la entrada del gas radón. En cambio, en los edificios existentes, las restricciones técnicas o presupuestarias pueden limitar la implementación de medidas, por lo que es necesario adaptar las soluciones a cada caso [11] [22].

Además del tipo de edificación, la ubicación geográfica determina la clasificación del edificio en una de las tres zonas de riesgo como se puede ver en Figura 2-5. Esta clasificación influye en la necesidad y el tipo de medidas a aplicar para mitigar la exposición al radón.

En dicho mapa, Las zonas en color rojo corresponden a áreas de alta exposición, donde el riesgo de superar los niveles recomendados es significativamente mayor. Las zonas en amarillo indican una exposición media o baja, mientras que las áreas en verde representan una baja probabilidad de acumulación de radón.



**Figura 2-5 Clasificación de las zonas según geografía [26]**

Las zonas de riesgo bajo, con concentraciones de radón inferiores a  $150 \text{ Bq/m}^3$ , no requieren la implementación de medidas adicionales de mitigación. Las zonas de riesgo medio, donde las concentraciones de radón oscilan entre  $150$  y  $300 \text{ Bq/m}^3$ , requieren la aplicación de medidas de protección básica, asegurando el cumplimiento de al menos una de las siguientes estrategias [11] [27]:

- **Barreras de protección:** Consisten en membranas impermeables instaladas bajo la cimentación para evitar que el gas radón se infiltre desde el suelo al interior del edificio. Como se muestra en la Figura 2-6, las barreras se colocan estratégicamente para bloquear el ascenso del gas, funcionando en conjunto con la solera del sótano y otros elementos constructivos. Están fabricadas con láminas de polietileno de alta densidad u otros materiales con recubrimientos especiales que aseguren su durabilidad y resistencia al deterioro. Su instalación debe ser precisa, ya que cualquier fisura o discontinuidad puede comprometer su funcionamiento. Para asegurar un aislamiento total y una mayor efectividad, suelen combinarse con juntas selladas [7]

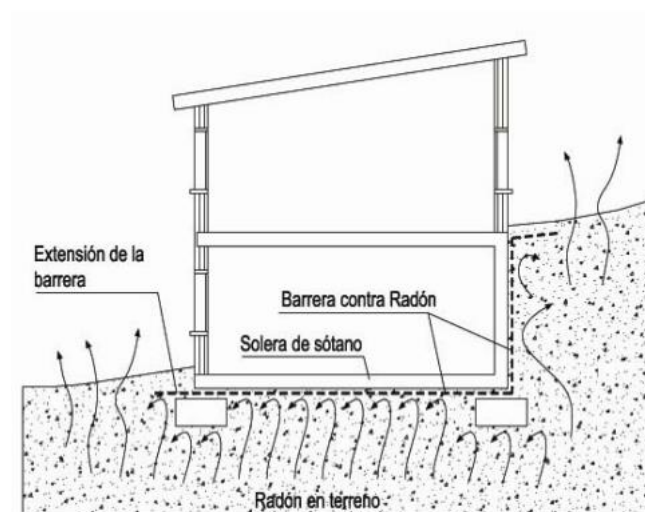


**Figura 2-6 Barrera de protección frente al radón [25]**

- **Cámaras de aire ventiladas:** Son espacios creados entre la cimentación y el terreno que permiten la circulación de aire y la evacuación del radón antes de que se acumule en el interior del edificio. Como se muestra en la Figura 2-7, las cámaras de aire ventiladas funcionan mediante

la creación de una capa intermedia de aire que permite que el gas se disipe antes de ingresar en las zonas habitadas. Estas cámaras pueden incluir conductos de ventilación natural o sistemas mecánicos de extracción para mejorar la eficiencia de la eliminación del gas. Además, se pueden diseñar con entradas y salidas de aire estratégicamente ubicadas para garantizar un flujo continuo de ventilación [11].

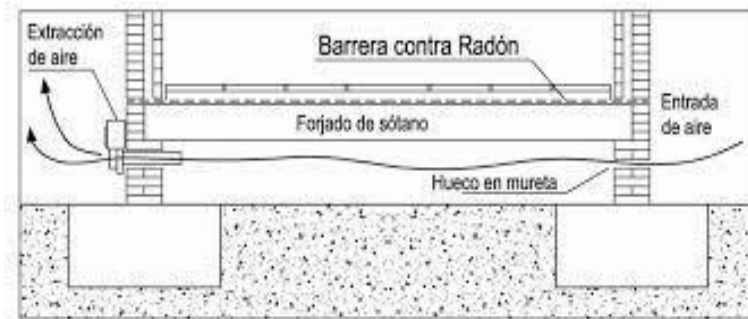


**Figura 2-7 Cámara de aire ventilada [25]**

Ambas soluciones tienen como objetivo evitar la entrada del radón, pero operan de manera diferente. Mientras que las barreras de protección crean un bloqueo físico que impide la infiltración del gas, las cámaras de aire ventiladas facilitan su evacuación antes de que pueda penetrar en el edificio. En muchas ocasiones, la combinación de ambas medidas maximiza la eficacia de la protección, especialmente en zonas con niveles de riesgo medio y alto [27].

Por último, en las zonas de riesgo alto, con concentraciones superiores a 300 Bq/m<sup>3</sup>, es imprescindible aplicar medidas de protección elevada. En estos casos, se implementan sistemas de despresurización del terreno, diseñados para generar una presión negativa bajo la estructura mediante tubos y extractores. Tal y como se muestra en la Figura 2-8, estos sistemas extraen el gas del subsuelo a través de un punto de extracción ubicado estratégicamente bajo la cimentación. El gas es conducido a través de un tubo de extracción hasta un conducto de evacuación, que lo libera al exterior mediante un extractor. Este mecanismo evita la acumulación de radón en el interior del edificio y puede ser potenciado con un extractor adicional si es necesario. Su uso es especialmente eficaz en áreas donde la emanación de radón es elevada, logrando una reducción significativa de su concentración en el interior [28].

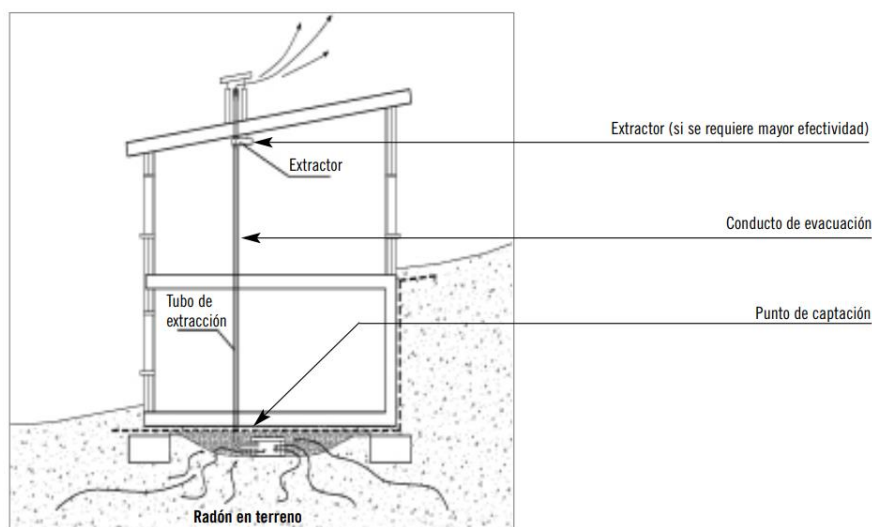


**Figura 2-8 Despresurización del terreno [25]**

En situaciones donde no es posible aplicar medidas constructivas debido a limitaciones técnicas o económicas en edificaciones existentes, el CTE permite la adopción de soluciones alternativas para reducir la concentración de radón [25]. Estas estrategias complementarias buscan minimizar la presencia del gas sin intervenir en la estructura del edificio. Entre ellas, la ventilación destaca como una medida fundamental para mejorar la calidad del aire interior y disminuir la concentración de radón.

## 2.4 Medición de los niveles de radón

### 2.4.1 Técnicas para la medición de la concentración de radón en aire

Existen distintas técnicas para medir la concentración de radón en interiores, las cuales pueden clasificarse en mediciones pasivas y activas. Los métodos pasivos requieren un análisis posterior en laboratorio y utilizan mediciones integradas para obtener un valor medio de concentración del gas en un periodo determinado. Por el contrario, los métodos activos proporcionan mediciones en tiempo real mediante sensores que permiten evaluar variaciones en la concentración del radón de manera continua.

#### Mediciones pasivas mediante integración

Dentro de los métodos pasivos, las mediciones integradas son particularmente útiles para estudios de exposición prolongada, ya que permiten identificar patrones de acumulación del gas en el tiempo. Entre los principales métodos se encuentran:

- **Colectores de carbón activo:** Contienen partículas de carbón con alta capacidad de absorción. Estos dispositivos se colocan en el área de estudio durante un periodo específico para que el radón se adhiera a la superficie del carbón. Posteriormente, la muestra es analizada en laboratorio mediante espectrometría gamma para medir la concentración del gas absorbido [29].
- **Detectores de trazas nucleares:** Consisten en láminas de polímero sensibles a la radiación alfa emitida por los productos de descomposición del radón. Las partículas de radón interactúan con la superficie del polímero, dejando pequeñas marcas o trazas que se analizan en laboratorio con microscopía óptica o electrónica [30].
- **Electretes:** Son dispositivos compuestos por un disco de teflón cargado eléctricamente que mide la ionización producida por la radiación alfa del radón y sus descendientes. La pérdida de carga eléctrica en el electrete es proporcional a la concentración de radón en el ambiente y permite obtener mediciones precisas con costos relativamente bajos.

#### Mediciones activas en continuo

Los métodos activos, por su parte, permiten realizar mediciones en tiempo real, lo que facilita la detección de variaciones en la concentración del radón y la evaluación inmediata de las medidas de mitigación. Entre estos métodos destacan los sensores de radón en continuo, dispositivos que registran datos de forma constante y pueden conectarse a sistemas de monitoreo remoto para un análisis más detallado y prolongado.

### 2.4.2 Medición de exhalación de radón desde superficies

La medición de la exhalación de radón desde superficies se realiza comúnmente mediante el método de acumulación, establecido en la norma UNE-EN ISO 11665-7 [31]. Este método consiste en colocar una cámara hermética sobre la superficie de interés, donde el gas se acumula durante un tiempo determinado. A partir de los datos recogidos, se calcula la tasa de exhalación mediante dos enfoques: la medición en continuo y la medición puntual.

#### Medición en continuo

Dentro de la medición en continuo de la tasa de exhalación, se pueden emplear dos técnicas principales: la circulación de aire y la difusión.

La circulación de aire utiliza una bomba integrada para generar un flujo constante entre la cámara de acumulación y el detector, asegurando mediciones homogéneas y reduciendo acumulaciones localizadas de gas. Sin embargo, su costo elevado y la necesidad de calibraciones frecuentes pueden limitar su aplicación en estudios de bajo presupuesto. Este método es similar a los sensores activos de radón en continuo, ya que ambos permiten un monitoreo constante y la detección precisa de variaciones dinámicas en la concentración del gas.

Por el contrario, la medición por difusión no emplea una bomba, sino que permite que el radón se desplace de manera natural hasta el detector. Esta técnica es más sencilla y menos invasiva, lo que la hace adecuada para estudios donde la reducción de costos y la simplicidad del equipo son prioritarias. No obstante, su precisión depende de factores ambientales, lo que exige una evaluación cuidadosa en cada caso. Su funcionamiento se asemeja al de las cámaras de acumulación pasivas, pues ambas técnicas dependen únicamente de la difusión del gas en un entorno controlado. Aunque más accesible, este método puede presentar limitaciones en estudios que requieren alta precisión.

### **Medición puntual**

En la medición puntual, se recogen muestras de aire en momentos específicos del periodo de acumulación, generalmente al inicio y al final, con el fin de determinar la concentración de radón liberada durante ese intervalo de tiempo. Estas muestras se analizan posteriormente en laboratorio, lo que permite calcular la tasa de exhalación a partir de la diferencia de concentraciones y del volumen de la cámara. Esta técnica proporciona resultados precisos sin necesidad de monitoreo continuo, por lo que resulta adecuada para estudios puntuales, de bajo costo o en entornos con recursos limitados.

No obstante, al tratarse de mediciones aisladas, no permite detectar variaciones dinámicas en la exhalación del gas, lo que puede ser una limitación en estudios donde se requiere un análisis detallado de la evolución temporal. Aun así, su rapidez, sencillez de implementación y bajo requerimiento tecnológico la convierten en una herramienta útil para estudios preliminares o como complemento de otras técnicas más avanzadas.

## **2.5 Uso de la IA para predecir la concentración del gas radón**

La predicción de la exhalación de radón es un desafío debido a la interacción de múltiples factores ambientales y geológicos. Variables como las precipitaciones, la temperatura, la presión atmosférica y la velocidad del viento influyen significativamente en la variabilidad del radón. Asimismo, las características del suelo, como su porosidad y composición mineralógica, determinan la cantidad de radón que puede liberarse al ambiente. Debido a esta compleja interacción de factores, es necesario contar con herramientas avanzadas para modelar con precisión la exhalación del gas. En este contexto, la inteligencia artificial (IA) ha surgido como una solución eficaz, mejorando la precisión de los modelos predictivos y superando las limitaciones de los métodos tradicionales. Estos modelos son capaces de reconocer patrones no lineales en los datos y adaptarse a distintas condiciones ambientales, lo que los convierte en herramientas especialmente valiosas en estudios de predicción ambiental.

La recopilación y calidad de los datos son esenciales para garantizar la fiabilidad de estos modelos. La IA predictiva requiere información detallada y confiable para generar estimaciones precisas. Para ello, el proceso de manejo de datos comienza con la recopilación desde diversas fuentes. Posteriormente, se lleva a cabo un preprocesamiento en el que los datos se limpian y estructuran para eliminar inconsistencias, valores atípicos o registros incompletos que puedan afectar la precisión de los modelos. Una vez organizados, se normalizan y escalan para garantizar su compatibilidad con los algoritmos de IA, optimizando así su capacidad de aprendizaje y predicción.

Además de contar con datos de calidad, es fundamental identificar qué variables tienen mayor influencia en la exhalación de radón para que los modelos trabajen con información relevante y representativa. Este proceso mejora la precisión de las predicciones y contribuye a la optimización del desempeño de los modelos, asegurando resultados más confiables y robustos.

### 2.5.1 Algoritmos de predicción

Para estimar concentraciones de radón con precisión, es necesario seleccionar un algoritmo de IA adecuado según el tipo y cantidad de datos disponibles. Factores como la cantidad de datos históricos, la complejidad de las relaciones entre las variables, la necesidad de interpretabilidad y el tiempo de procesamiento influyen en la elección del modelo. Los modelos de predicción del radón pueden clasificarse en dos enfoques principales: modelos de clasificación, que agrupan los niveles de radón en distintas categorías de riesgo, y modelos de regresión, que generan valores numéricos exactos para analizar su variabilidad.

Los modelos de clasificación permiten categorizar los niveles de radón en rangos de riesgo, como "bajo", "moderado" o "alto", en función de las variables meteorológicas y geológicas disponibles. Esta metodología facilita la interpretación de los resultados y la toma de decisiones en estrategias de mitigación.

Por otro lado, los modelos de regresión estiman con precisión la concentración de radón en valores numéricos, permitiendo un análisis detallado de su variabilidad y tendencias a lo largo del tiempo. Este enfoque es útil para estudios que requieren predicciones cuantitativas específicas. Entre los enfoques más utilizados se encuentran:

#### Regresión lineal

Es uno de los modelos más simples y fundamentales en el análisis de datos. Su objetivo es establecer una relación lineal entre una variable dependiente y una o más variables independientes, permitiendo hacer predicciones y analizar la influencia de cada variable en el resultado [32].

La ecuación general de la regresión lineal es:

$$f(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad \text{Ec. 1}$$

donde:

- $f(x)$ : es la predicción de la variable dependiente.
- $\beta_0$ : es el intercepto o termino independiente.
- $\beta_1, \dots, \beta_p$  : los coeficientes que indican la influencia de cada variable independiente.
- $x_1, \dots, x_p$  : es la variable dependiente.

Para que el modelo aprenda, se deben encontrar valores óptimos de los coeficientes  $\beta$ . Para ello, hay que minimizar el error entre los valores predichos y los reales usando una función llamada error cuadrático medio:

$$J(\beta) = \frac{1}{N} \sum_{i=1}^N \left( y_i - \left( \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \right) \right)^2 \quad \text{Ec. 2}$$

donde:

- $J(\beta)$ : es la función de pérdida que mide el error.
- $y_i$ : es el valor real de la variable dependiente.
- $N$ : es el número de observaciones.

El valor óptimo de los coeficientes se puede calcular de forma exacta resolviendo un sistema de ecuaciones lineales, conocido como la ecuación normal:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad \text{Ec. 3}$$

donde:

- $X$ : es la matriz de diseño que contiene las variables independientes.
- $y$ : es el vector de valores observados de la variable dependiente.

Una vez determinados los coeficientes  $\hat{\beta}$ , el modelo puede utilizarse para hacer predicciones sobre nuevos datos.

### Support Vector Regression (SVR)

Un modelo basado en máquinas de soporte vectorial que busca encontrar una función capaz de predecir valores dentro de un margen de tolerancia  $\varepsilon$ . A diferencia de otros métodos de regresión, SVR introduce flexibilidad en la predicción al no penalizar errores pequeños, enfocándose únicamente en aquellos que superan dicho margen [33].

Su ecuación general es:

$$f(x) = \langle w, x \rangle + b \quad \text{Ec. 4}$$

donde:

- $w$ : representa los pesos asociados a cada variable.
- $x$ : es el vector de entrada.
- $b$ : es un término de ajuste.

El objetivo de SVR es encontrar la mejor función  $f(x)$  minimizando la complejidad del modelo y, al mismo tiempo, controlando los errores que exceden el margen  $\varepsilon$ . Para ello, se introducen las variables de holgura  $\varepsilon_i$  y  $\varepsilon_i^*$ , que permiten manejar errores por encima y por debajo del margen, respectivamente.

La optimización del modelo se plantea como:

$$\min_{w, b, \varepsilon, \varepsilon^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\varepsilon + \varepsilon_i^*) \quad \text{Ec. 5}$$

El parámetro  $C$  desempeña un papel crucial en el modelo, ya que regula el compromiso entre la simplicidad del modelo y la penalización de los errores que superan el margen  $\varepsilon$ . Un valor alto de  $C$  implica una penalización mayor a los errores fuera del margen, lo que puede llevar a un mejor ajuste, pero también a una menor capacidad de generalización.

Para modelar relaciones no lineales, SVR puede utilizar funciones núcleo (*kernels*) que transforman los datos a un espacio de mayor dimensión sin necesidad de calcular explícitamente dicha transformación. Esto permite que el modelo capture patrones más complejos sin perder eficiencia computacional. En este caso, la función se expresa como:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad \text{Ec. 6}$$

donde:

- $K(x_i, x)$ : es la función núcleo elegida, que mide la similitud entre los puntos en el espacio transformado.
- $\alpha_i, \alpha_i^*$ : son los multiplicadores de Lagrange, obtenidos en la solución del problema dual.

### Random Forest

Es una técnica de ensemble que combina múltiples árboles de decisión para mejorar la precisión y robustez del modelo, tanto en problemas de clasificación como de regresión.

El algoritmo genera varios árboles de decisión, entrenando cada uno con una muestra *bootstrap* (con reemplazo) de los datos originales. Además, en cada nodo, se selecciona aleatoriamente un subconjunto

de variables para determinar la mejor partición. Esto reduce la correlación entre árboles y mejora la capacidad de generalización del modelo.

En clasificación, la predicción final se obtiene mediante votación mayoritaria de todos los árboles:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_B(x)\} \quad \text{Ec. 7}$$

donde

- $\hat{y}$  : es la predicción final del modelo.
- $T_B(x)$  : representa la predicción del árbol  $b$  para la instancia  $x$ .
- $B$  : es el número total de árboles en el ensemble.

*Mode* denota la clase más frecuente entre todas las predicciones de los árboles.

En regresión, la predicción final se obtiene promediando las salidas de todos los árboles:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad \text{Ec. 8}$$

### Gradient Boosting Machines (GBM)

Es una técnica de ensemble que combina secuencialmente árboles de decisión débiles para mejorar la precisión. Su principio fundamental es corregir iterativamente los errores de predicción de los modelos anteriores [34].

El proceso comienza con un modelo base que minimiza la función de pérdida:

$$f_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^N L(y_i, \gamma) \quad \text{Ec. 9}$$

donde :

- $L(y_i, \gamma)$  : es la función de pérdida que mide el error.
- $f_0(x)$  : es la mejor aproximación constante de la variable objetivo.

En cada iteración, se ajusta el modelo calculando el gradiente negativo de la función de pérdida, que representa los errores a corregir:

$$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x_i)=f_{m-1}(x_i)} \quad \text{Ec. 10}$$

donde:

- $r_{im}$  : es el residuo o error en la observación  $i$  en la iteración  $m$ .
- $f(x_i)$  : es la predicción del modelo en la observación  $x_i$ .
- $\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}$  : es la derivada de la función de pérdida con respecto a la predicción del modelo.

Estos residuos indican la dirección y la magnitud del ajuste necesario. Luego, se entrena un nuevo modelo débil para predecir estos residuos y se actualiza el modelo global:

$$f_m(x) = f_{m-1}(x) + \nu h_m(x) \quad \text{Ec. 11}$$

donde:

- $f_m(x)$  : es la nueva versión del modelo en la iteración  $m$ .
- $h_m(x)$  : es un modelo débil que aprende a corregir los errores de la iteración anterior.

- $\nu$ : es el *learning rate*, que regula la magnitud del ajuste en cada iteración para evitar sobreajuste.

Tras  $m$  iteraciones, el modelo final es la suma acumulativa de todas las correcciones:

$$f_M(x) = f_0(x) + \nu \sum_{m=1}^M h_m(x) \quad \text{Ec. 12}$$

### Perceptrón multicapa (MLP)

Es una red neuronal artificial que modela funciones complejas combinando capas de transformación lineal y funciones de activación no lineales. Esta estructura permite capturar patrones abstractos para tareas como clasificación y regresión.

La red está compuesta por una capa de entrada, varias capas ocultas y una capa de salida. La transformación de cada capa  $l$  se define como:

$$h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}) \quad \text{Ec. 13}$$

donde:

- $h^{(l-1)}$ : salida de la capa anterior (para la primera capa,  $h^{(0)} = x$ ).
- $W^{(l)}$ : la matriz de pesos que conecta la capa  $l-1$  con la capa  $l$ .
- $b^{(l)}$ : vector de sesgos en la capa  $l$ , que permite ajustar la activación de las neuronas.
- $\sigma(\cdot)$ : es la función de activación, que introduce la no linealidad en el modelo.

El entrenamiento del MLP ajusta los parámetros de la red para minimizar una función de pérdida. Una de las funciones más utilizadas es el error cuadrático medio (MSE), definido como:

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad \text{Ec. 14}$$

donde:

- $y_i$ : valor real e  $\hat{y}_i$  la predicción del modelo.
- $N$ : número total de muestras.
- $\theta$ : representa todos los pesos y sesgos de la red.

Para minimizar  $J(\theta)$ , se emplea retropropagación con descenso de gradiente:

$$W^{(l)} \leftarrow W^{(l)} - \eta \frac{\partial J}{\partial W^{(l)}}, \quad b^{(l)} \leftarrow b^{(l)} - \eta \frac{\partial J}{\partial b^{(l)}}, \quad \text{Ec. 15}$$

donde:

- $\frac{\partial J}{\partial W^{(l)}}$  y  $\frac{\partial J}{\partial b^{(l)}}$ : representan los gradientes de la función de pérdida.
- $\eta$ : tasa de aprendizaje, que determina el tamaño del ajuste en cada iteración.

Para evitar el sobreajuste, se aplican técnicas de regularización. La regularización L2 penaliza pesos elevados:

$$J_{reg}(\theta) = J(\theta) + \lambda \sum_l ||W^{(l)}||^2 \quad \text{Ec. 16}$$

donde  $\lambda$  cantidad de penalización aplicada.

Otra técnica complementaria es el *dropout*, que desactiva aleatoriamente ciertas neuronas en cada capa oculta durante el entrenamiento, reduciendo la dependencia de características específicas. Además, *early stopping* detiene el entrenamiento cuando el error en validación empieza a aumentar, evitando que el modelo memorice los datos.

Tras el entrenamiento, la capa de salida genera la predicción:

$$\hat{y} = \sigma_{out}(W^{(L+1)}h^L + b^{(L+1)}) \quad \text{Ec. 17}$$

donde:

- $W^{(L+1)}$  y  $b^{(L+1)}$  : son los pesos y sesgos de la capa de salida.
- $\sigma_{out}$ : es la función de activación de la salida.

### XGBoost

Es una técnica avanzada de ensemble que se basa en *boosting*, en la que se construye un modelo aditivo incorporando secuencialmente árboles de decisión. Su principal ventaja radica en la optimización eficiente de la función objetivo y la introducción de mecanismos de regularización para evitar el sobreajuste [35].

El modelo minimiza una función objetivo compuesta por dos términos:

1. Función de pérdida, que mide la diferencia entre las predicciones y los valores reales.
2. Término de regularización, que penaliza la complejidad del modelo para evitar sobreajuste.

La función se expresa como:

$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(f_t) \quad \text{Ec. 18}$$

donde:

- $N$ : es el número total de observaciones en los datos.
- $l(y_i, \hat{y}_i)$ : es la función de pérdida para el dato  $i$ .
- $\Omega(f_t)$ : es la función de regularización que mide la complejidad del árbol  $f_t$ .
- $T$ : es el número total de árboles en el modelo.

El modelo se construye de forma aditiva, comenzando con una predicción inicial. Luego, en cada iteración, se ajusta el modelo mediante la optimización de segundo orden. En este proceso, se calcula el gradiente negativo (primera derivada de la función de pérdida) y el Hessiano (segunda derivada), lo que permite mejorar la eficiencia del ajuste:

$$L^{(t)} = \sum_{i=1}^N \left[ l(y_i, \hat{y}^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \quad \text{Ec. 19}$$

donde:

- $\hat{y}^{(t-1)}$ : es la predicción del modelo en la iteración anterior.
- $g_i$ : es el gradiente, que mide la tasa de cambio de la función pérdida con respecto a la predicción anterior:

$$g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} \quad \text{Ec. 20}$$

- $h_i$ : es el Hessiano, que mide la curvatura de la función de pérdida para ajustar mejor el modelo:

$$h_i = \frac{\partial^2 l(y_i, \hat{y}_2^{(t-1)})}{\partial \hat{y}_i^{(t-1)2}} \quad \text{Ec. 21}$$

A partir de estos valores, se construye un nuevo árbol  $f_t$  que minimiza la aproximación de la función objetivo, corrigiendo los errores residuales acumulados en el modelo previo. Finalmente, la predicción del modelo se actualiza sumando la contribución del nuevo árbol:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad \text{Ec. 22}$$

donde  $f_t(x_i)$  es la predicción generada por el nuevo árbol  $t$ .

Este proceso se repite hasta alcanzar el número máximo de iteraciones  $T$ , mejorando progresivamente la precisión del modelo mientras se controla su complejidad mediante la regularización.

### 2.5.2 Revisión bibliográfica del uso de la IA para la predicción de radón

En los últimos años, se ha empezado a usar cada vez más el aprendizaje automático para predecir la concentración de radón y entender qué factores influyen en su emisión. Gracias a estas técnicas, se han podido desarrollar modelos más precisos y flexibles, que ayudan a gestionar mejor el riesgo radiológico y a aplicar medidas de control más eficaces. Entre los enfoques más utilizados están los que se basan en variables ambientales y también aquellos que tienen en cuenta cómo cambian los datos a lo largo del tiempo.

Uno de los trabajos más relevantes en esta línea es el artículo científico “*Machine learning as a tool for analysing the impact of environmental parameters on the radon exhalation rate from soil*” [36], donde se emplea el algoritmo *Gradient Boosting Machine* (GBM) para analizar la influencia de variables como la temperatura del suelo, la humedad y la presión atmosférica en la tasa de exhalación de radón. Los resultados mostraron que GBM no solo alcanzaba una buena precisión predictiva, sino que también permitía interpretar con claridad la importancia relativa de cada variable. Por su adecuación a conjuntos de datos no secuenciales y su capacidad explicativa, GBM ha sido incluido como uno de los modelos considerados en este estudio.

Otro enfoque interesante es el propuesto en el artículo científico “*Prediction of radon exhalation rate based on VMD LSTM ELMAN*” [37], que combina técnicas de descomposición de señales (VMD) con redes neuronales recurrentes (LSTM y ELMAN). Este modelo híbrido fue diseñado para capturar tanto fluctuaciones rápidas como tendencias prolongadas en series temporales continuas. Dado que los datos empleados en este estudio no presentan una secuencia temporal estructurada, este tipo de modelo no resulta aplicable al caso actual.

En cuanto al potencial de las redes neuronales artificiales (ANN), el artículo “*Exploring the efficacy of machine learning models for predicting soil radon exhalation rates*” [38], presenta una comparación entre ANN y la programación genética de expresiones para estimar la tasa de exhalación a partir de concentraciones de radionúclidos. Aunque las ANN ofrecieron una mayor precisión, su menor interpretabilidad y la necesidad de una mayor cantidad de datos limitan su aplicabilidad en estudios exploratorios. Por ello, en este trabajo se ha descartado su uso, priorizando modelos más robustos frente al sobreajuste y con mayor transparencia en la interpretación de los resultados.

Finalmente, en el trabajo presentado en la 2023 18th Iberian Conference on Information Systems and Technologies (CISTI), titulado “*Forecasting Short-Term Indoor Radon: A Machine Learning Approach Using LSTM Networks*” [39], se emplearon modelos LSTM y Bi-LSTM para predecir concentraciones de radón en interiores a corto plazo a partir de datos obtenidos con sensores IoT. Este enfoque demostró ser eficaz en entornos con datos densos y continuos, adecuados para el aprendizaje secuencial. No obstante, dada la estructura discontinua de la base de datos utilizada en el presente estudio, estos modelos no resultan apropiados.

## 3 MATERIALES Y MÉTODOS

### 3.1 Ubicación y características del lugar de estudio

La fase experimental de este estudio se realizó en el edificio del CUD, que se encuentra en la Escuela Naval Militar, en Marín, Pontevedra. Este edificio está compuesto por dos plantas: en la planta superior se localizan la secretaría del CUD, varios despachos del profesorado, la sala de juntas, la sala de reuniones y la secretaría económica. En la planta baja se encuentran la sala de grados, el resto de los despachos del profesorado y aulas donde se imparten las clases.

En la Figura 3-1, se muestra una vista aérea de la Escuela Naval Militar, donde se ha señalado el Edificio del CUD con un recuadro rojo para facilitar su localización.



**Figura 3-1 Emplazamiento del CUD dentro de la Escuela Naval militar**

Dentro de la planta baja, se ha seleccionado como objeto de estudio el despacho 114, situado al final del pasillo en la cara sur del edificio, tal y como se muestra en la Figura 3-2. Esta elección se basa en investigaciones previas, recogidas en el apartado 1.2, que identifican este despacho como el lugar con las concentraciones de radón más altas registradas en todo el recinto de la Escuela Naval Militar.

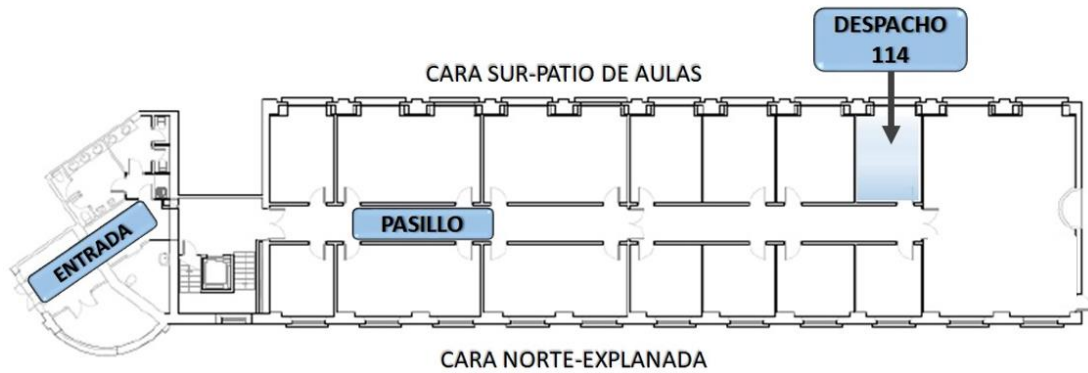


Figura 3-2 Plano de la planta baja del CUD con la ubicación del lugar de estudio

El edificio dispone de un sistema de ventilación mixto compuesto por una Unidad de Tratamiento de Aire (UTA) y un fancoil, diseñado para garantizar una óptima calidad del aire y una climatización adecuada en las estancias. Tal como se muestra en la Figura 3-3, la UTA, instalada en el falso techo, toma aire del exterior, lo filtra y lo distribuye a través de conductos que recorren ese mismo espacio, renovando así el aire interior. Posteriormente, este aire tratado es recirculado por el fancoil, también ubicado en el falso techo, el cual mezcla el aire de retorno de la estancia con el aire nuevo de la UTA, ajustando su temperatura mediante un intercambiador de calor antes de reintroducirlo en la habitación.

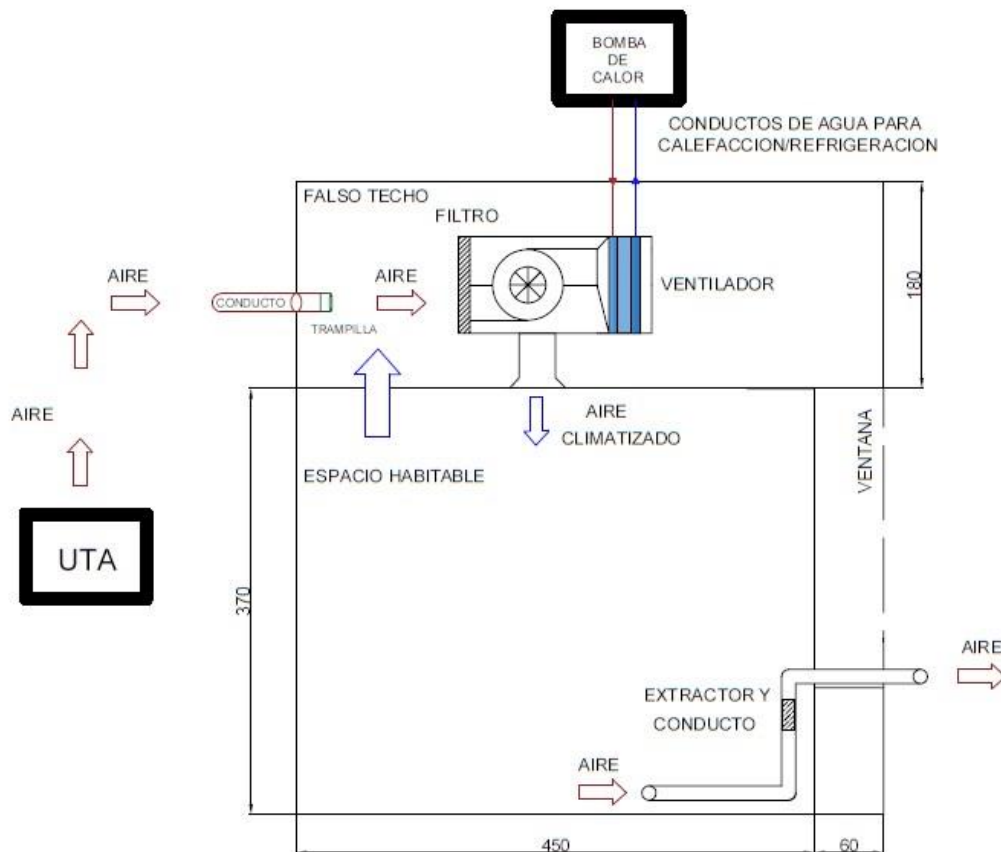


Figura 3-3 Funcionamiento general del sistema de ventilación en los despachos [19]

### 3.2 Toma de datos experimentales

El estudio se fundamenta en la creación de una base de datos que permita analizar la tasa de exhalación de radón. Para ello, se han recopilado datos previos sobre la concentración de radón, y se han complementado con variables que describen las condiciones internas y externas del entorno en el

momento de cada medición. Los registros abarcan el periodo comprendido entre septiembre de 2023 y principios de julio de 2024, lo que permite estudiar la evolución del radón en distintas condiciones y comprender con mayor precisión los factores que influyen en su concentración, dispersión y exhalación.

### 3.2.1 Tasa de exhalación de radón

Para la toma de datos se seleccionó el método de medición en continuo descrito en el apartado 2.4 debido a que permite registrar variaciones en tiempo real y obtener una mayor resolución temporal en la evolución de la concentración de radón.

#### Equipos empleados

Para la medición de radón, se ha utilizado el dispositivo RadonEye RD200 mostrado en la Figura 3-4. Este instrumento es capaz de detectar concentraciones de radón de hasta  $3696 \text{ Bq/m}^3$  y tiene una tasa de precisión del 10%. Además, permite realizar un monitoreo continuo, registrando la concentración de radón en intervalos de 10 minutos, lo que facilita un análisis detallado de cómo evoluciona la presencia del gas a lo largo del tiempo [14] [15].



Figura 3-4 RadonEye RD200 [40]

Para asegurar la fiabilidad de las mediciones realizadas, es crucial utilizar una cámara de acumulación que cumpla con las especificaciones de la norma UNE-EN ISO 11665-7. Esta norma establece requisitos como la impermeabilidad de la cámara, lo que previene la pérdida de gas y garantiza mediciones estables. Además, es importante que el material y color de la cámara no alteren la temperatura interna del aire acumulado, especialmente si está expuesta a la luz solar. Para este estudio, se ha elegido un cubo de polipropileno (ver Figura 3-5) porque cumple con todos estos requisitos: tiene un bajo coeficiente de difusión que minimiza la pérdida de gas, un espesor de 1 mm que impide la filtración del radón, y su composición asegura que la temperatura interna se mantenga estable durante la medida. Las dimensiones de la cámara se detallan en la Tabla 3-1.



Figura 3-5 Cámara de acumulación [16]

| Cámara de acumulación               | Dimensiones |
|-------------------------------------|-------------|
| Diámetro superior (cm)              | 24,90       |
| Diámetro inferior (cm)              | 23,20       |
| Diámetro medio (cm)                 | 24,05       |
| Altura (cm)                         | 16,60       |
| Volumen efectivo (cm <sup>3</sup> ) | 7540,98     |
| Volumen efectivo (m <sup>3</sup> )  | 0,00754     |
| Área (cm <sup>2</sup> )             | 454,28      |
| Área (m <sup>2</sup> )              | 0,045       |

**Tabla 3-1 Dimensiones cámara acumulación gas [16]**

Por último, para optimizar las condiciones de medición y evitar el contacto directo de los medidores con el suelo, se han utilizado soportes diseñados especialmente para elevarlos 45 mm. Estos soportes, mostrado en la Figura 3-6, están fabricados en ácido poliláctico, utilizando impresión 3D con la impresora Ultimaker 3 del edificio de investigación del CUD. Cada soporte tiene un diámetro exterior de 100 mm, un diámetro interior de 50 mm y una altura de 45 mm, lo que asegura una elevación adecuada de los medidores. Además de su funcionalidad, este diseño permite una colocación estable y repetible, reduciendo errores asociados a variaciones en la posición del detector.



**Figura 3-6 Soporte [16]**

### Comportamiento y cálculo de la tasa de exhalación

Durante la medición de radón, la concentración en la cámara de acumulación varía en función de la tasa de exhalación del suelo, el volumen de la cámara y posibles fugas que puedan producirse durante el proceso. Como se observa en la Figura 3-7, inicialmente se produce un proceso de purgado en el cual la concentración de radón es baja. Posteriormente, comienza la acumulación del gas, que sigue un crecimiento lineal como consecuencia de su progresiva exhalación en un espacio cerrado y limitado. Con el tiempo, la concentración alcanza un punto de saturación, donde se estabiliza debido al equilibrio entre la exhalación del suelo y la disipación del radón. Esta etapa final, representada en el gráfico como una meseta, coincide con el llamado valor límite. En este punto, la tasa de exhalación del gas desde el suelo se iguala con la tasa de pérdida por difusión y desintegración, asegurando que la entrada y salida del gas permanezcan constantes.

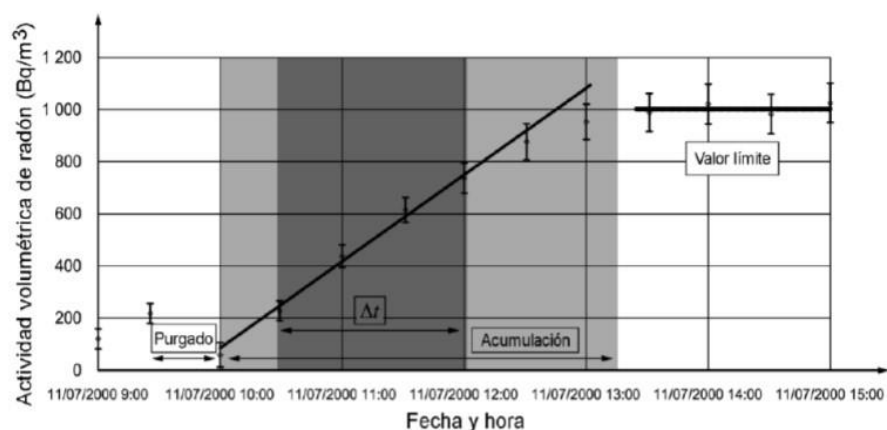


Figura 3-7 Ejemplo de la concentración de radón en la cámara de acumulación [31]

De acuerdo con la norma UNE-EN ISO 11665-7, la tasa de exhalación en superficie se determina a partir de la pendiente de la recta obtenida en la fase de acumulación, considerando la superficie de la cámara de acumulación en contacto con el suelo y el volumen efectivo del recipiente. Este cálculo se expresa con la ecuación:

$$\Phi = \frac{pte \cdot V}{S} \quad \text{Ec. 23}$$

donde:

- $\Phi$ : tasa de exhalación de gas radón en un punto de la superficie de estudio. Medido en  $\text{Bq/m}^2 \cdot \text{s}$ .
- $Pte$ : pendiente de la recta obtenida en la fase de acumulación. Medido en  $\text{Bq/m}^3 \cdot \text{s}$ .
- $V$ : volumen efectivo de la cámara de acumulación. Medido en  $\text{m}^3$ .
- $S$ : superficie efectiva de la cámara de acumulación. Medido en  $\text{m}^2$ .

### Procedimiento de medición

Tal y como se mencionó en el apartado 1.2, el TFG de Marín Alonso permitió una caracterización más precisa de la tasa de exhalación en el despacho 114, utilizando como punto de referencia el marcado en color rosa en la Figura 3-8. A partir de las mediciones realizadas en dicho punto, es posible estimar la concentración de radón en el resto del despacho.

|    | A    | B    | C    | D    | E    | F    | G    | H    | I    | J    | K    |
|----|------|------|------|------|------|------|------|------|------|------|------|
| 1  |      |      |      | 0,25 | 0,24 | 0,14 | 0,38 | 0,13 | 0,14 | 0,11 | 0,06 |
| 2  | 0,25 | 0,34 | 0,41 | 0,54 | 0,41 | 0,39 | 0,75 | 0,64 | 0,22 | 0,02 | 0,19 |
| 3  | 0,43 | 0,26 | 0,23 | 0,58 | 0,44 | 0,82 | 0,18 | 0,50 | 0,34 | 0,29 | 0,09 |
| 4  | 1,62 | 1,52 | REF  | 0,97 | 0,69 | 0,59 | 0,53 | 0,44 | 0,40 | 0,13 | 0,10 |
| 5  | 0,76 | 0,27 | 0,91 | 0,50 | 0,44 | 0,62 | 0,45 | 0,46 | 0,24 | 0,25 | 0,20 |
| 6  | 0,22 | 0,21 | 0,63 | 0,39 | 0,45 | 0,36 | 0,41 | 0,28 | 0,18 | 0,18 | 0,12 |
| 7  | 0,29 | 0,25 | 0,48 | 0,49 | 0,26 | 0,35 | 0,50 | 0,28 | 0,22 | 0,25 | 0,29 |
| 8  | 0,35 | 0,28 | 0,25 | 0,20 | 0,21 | 0,25 | 0,28 | 0,25 | 0,16 | 0,19 | 0,13 |
| 9  | 0,29 | 0,46 | 0,11 | 0,18 | 0,20 | 0,30 | 0,28 | 0,25 | 0,09 | 0,21 | 0,14 |
| 10 | 0,75 | 0,75 | 0,52 | 2,06 | 1,06 | 2,04 | 2,90 | 1,63 | 0,38 | 0,79 | 0,30 |
| 11 | 0,30 | 0,31 | 0,24 | 0,19 | 0,23 | 0,69 | 0,35 | 0,50 | 0,45 | 0,09 | 0,09 |
| 12 | 0,19 | 0,15 | 0,20 | 0,30 | 0,18 | 0,52 | 0,38 | 0,76 | 0,32 | 0,52 | 0,25 |

Figura 3-8 Relación de la concentración entre los diferentes puntos [16]

En el presente TFG se llevaron a cabo múltiples mediciones en ese mismo punto con el objetivo de recopilar datos de exhalación bajo distintas condiciones. Para automatizar el proceso y facilitar la toma de datos, se empleó un sistema de apertura y cierre remoto desarrollado en el marco del proyecto de investigación PICUD -2022-03 RnMod [41]. Este sistema, controlado mediante un canal de Telegram, permitía al operador enviar órdenes de apertura o cierre del cubo sin necesidad de presencia física.

La Figura 3-9 muestra un ejemplo del funcionamiento del sistema, donde se puede ver cómo se gestionan las órdenes de forma remota y en tiempo real.



**Figura 3-9 Chat de Telegram para controlar la apertura y cierre de cubo**

Una vez abierto el cubo, era necesario ventilar su interior para eliminar los restos de radón acumulados durante el ciclo anterior. Esto garantizaba que la siguiente medición no estuviera influenciada por residuos de radón. La ventilación se realizaba mediante un sistema instalado en la parte superior del cubo, que se activaba al abrirse el cubo. Este sistema estaba compuesto por un ventilador y un conjunto de válvulas que permitían la entrada y salida de aire, facilitando la renovación del volumen interno. En la Figura 3-10 se muestra el montaje del dispositivo.



**Figura 3-10 Montaje del dispositivo de ventilación**

Con el fin de facilitar el seguimiento y el análisis de los datos recogidos durante las distintas mediciones, se utilizó un sistema de visualización en tiempo real que mostraba la concentración de radón medida por el dispositivo RadonEye. Gracias a esta herramienta, fue posible supervisar de manera continua el proceso de acumulación del gas y detectar posibles anomalías durante la medición. Además, el sistema permitía descargar los datos registrados para su análisis e interpretación en detalle. Para ello, se empleó el software Grafana, elegido por su capacidad para representar la información de forma clara, dinámica e intuitiva. En la Figura 3-11 se muestra un ejemplo de visualización con varios ciclos consecutivos de acumulación y vaciado en la cámara

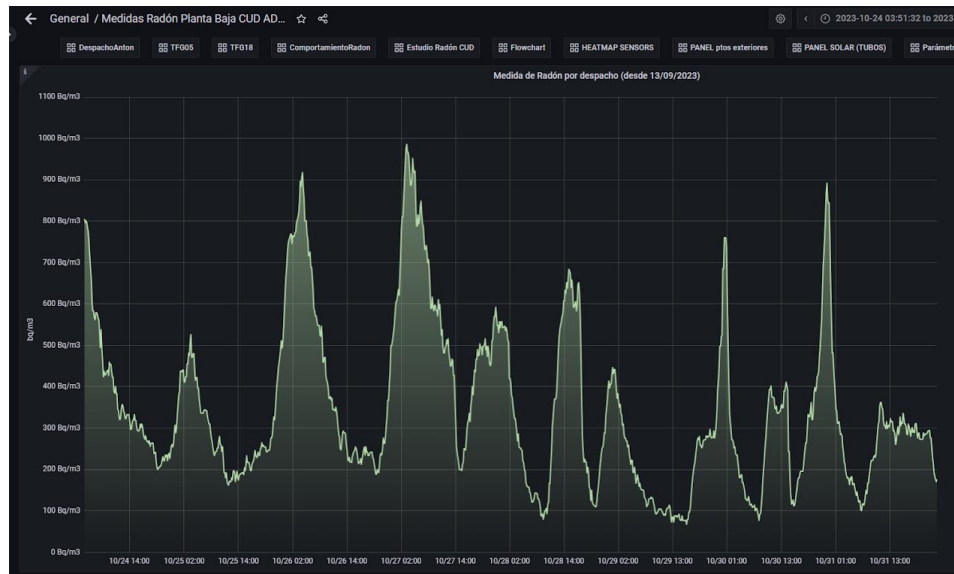


Figura 3-11 Representación en Grafana de las concentraciones de radón

### 3.2.2 Variables externas e internas relacionadas con la concentración de radón

A partir de lo expuesto en el apartado 2.2, donde se describen los principales factores que influyen en la inmisión de radón, se seleccionaron una serie de variables con el objetivo de analizar su relación con la exhalación del gas:

- **Presión atmosférica:** puede modificar el gradiente de presión entre el subsuelo y el interior del edificio, favoreciendo o dificultando la entrada de radón.
- **Temperatura exterior y temperatura interior:** se consideran por su influencia en los movimientos de aire por convección, que pueden modificar la dinámica del radón en el ambiente interior.
- **Diferencia de temperatura (interior - exterior):** permite evaluar si el gradiente térmico, ya sea positivo o negativo, tiene un efecto sobre la acumulación o liberación de radón.
- **Precipitaciones:** afectan al grado de saturación del terreno y a la permeabilidad del suelo, lo cual puede limitar o redirigir la liberación de radón.
- **Velocidad del viento y dirección del viento:** influyen en la ventilación natural del entorno y pueden afectar la acumulación o dispersión del gas.
- **Radiación solar:** puede modificar la temperatura del terreno y del aire próximo al edificio, alterando así el comportamiento del radón.
- **Humedad relativa:** influye en las condiciones del aire y del suelo, afectando el transporte y la acumulación del gas.
- **Mareas:** las oscilaciones del nivel del mar pueden generar pequeñas variaciones en la presión subterránea, especialmente en zonas costeras, influyendo en la exhalación del radón.

- **Concentración inicial de radón:** nivel de radón presente en la cámara de acumulación antes de iniciar cada medición. Esta variable se incluye para analizar si comenzar desde concentraciones más altas influye en una mayor tasa de exhalación del gas.
- **Ventilación general:** indica si el sistema de ventilación del edificio estaba activo, lo cual influye directamente en la renovación del aire y en la concentración del gas.
- **Uso del fancoil:** permite registrar si el sistema de climatización estaba funcionando, generando circulación de aire que puede modificar los niveles de radón.
- **Condición laboral:** variable binaria (0: no laborable; 1: laborable) que identifica si la medición se realizó durante una franja horaria de actividad. Durante los períodos laborales, la presencia de personas genera ventilación natural debida a la apertura frecuente de puertas y, en ocasiones, de ventanas, lo que puede influir en la dispersión del gas acumulado.
- **Mes:** a cada mes del año se le asignó un valor numérico con el objetivo de incluir esta variable en el análisis y estudiar posibles patrones estacionales en la concentración de radón.
- **Hora del día:** transformada a formato decimal para su uso como variable continua en los algoritmos de análisis. Permite estudiar la variación horaria de la exhalación.

Una vez definidas las variables de estudio, es necesario explicar su procedencia y cómo se obtuvieron. Algunas de ellas fueron extraídas directamente del sistema de medición, mientras que otras provienen de plataformas meteorológicas, registros operativos o dispositivos de control. A continuación, se describen las fuentes utilizadas, agrupadas según el tipo de variable al que corresponden.

### Variables Meteorológicas

Para analizar su impacto, se han recopilado registros detallados de dos fuentes especializadas:

- **Meteo Galicia:** de la estación que tienen en el puerto de Marín, se obtuvieron en formato de tabla (como se muestra en la Figura 3-12) los datos de presión atmosférica, temperatura, intensidad y dirección del viento, radiación solar y humedad relativa, que permiten analizar cómo afectan las condiciones climáticas a la dispersión del radón.

| Santiago-EOAS-(CO) |          |        |                     |        |
|--------------------|----------|--------|---------------------|--------|
| Fecha              | Variable | Valor  | Código (validación) | Unidad |
| 25/02/2024 00:10   | Presión  | 981.09 | 1                   | hPa    |
| 25/02/2024 00:20   | Presión  | 981.23 | 1                   | hPa    |
| 25/02/2024 00:30   | Presión  | 981.27 | 1                   | hPa    |
| 25/02/2024 00:40   | Presión  | 981.13 | 1                   | hPa    |
| 25/02/2024 00:50   | Presión  | 980.90 | 1                   | hPa    |
| 25/02/2024 01:00   | Presión  | 980.76 | 1                   | hPa    |
| 25/02/2024 01:10   | Presión  | 980.76 | 1                   | hPa    |
| 25/02/2024 01:20   | Presión  | 980.76 | 1                   | hPa    |
| 25/02/2024 01:30   | Presión  | 980.55 | 1                   | hPa    |
| 25/02/2024 01:40   | Presión  | 980.35 | 1                   | hPa    |
| 25/02/2024 01:50   | Presión  | 980.30 | 1                   | hPa    |
| 25/02/2024 02:00   | Presión  | 980.35 | 1                   | hPa    |
| 25/02/2024 02:10   | Presión  | 980.32 | 1                   | hPa    |
| 25/02/2024 02:20   | Presión  | 980.19 | 1                   | hPa    |
| 25/02/2024 02:30   | Presión  | 979.98 | 1                   | hPa    |
| 25/02/2024 02:40   | Presión  | 979.97 | 1                   | hPa    |
| 25/02/2024 02:50   | Presión  | 980.02 | 1                   | hPa    |
| 25/02/2024 03:00   | Presión  | 979.84 | 1                   | hPa    |
| 25/02/2024 03:10   | Presión  | 979.78 | 1                   | hPa    |
| 25/02/2024 03:20   | Presión  | 979.78 | 1                   | hPa    |
| 25/02/2024 03:30   | Presión  | 979.53 | 1                   | hPa    |
| 25/02/2024 03:40   | Presión  | 979.72 | 1                   | hPa    |

Figura 3-12 Datos obtenidos en Meteo Galicia [42]

- **Meteoblue:** fuente empleada para obtener datos de precipitación en unidades compatibles con las predicciones meteorológicas, lo que facilita la comparación con la evolución de la concentración de radón [43].

## Mareas

Se obtuvieron del sistema de observación de Puertos del Estado, que dispone de un mareógrafo en Marín. A través de su servidor de datos Thredds, se accedió a registros con una alta resolución temporal (cada 0,5 segundos). La variable consultada fue SLEV (*Sea Level*), que indica la altura de la superficie del mar sobre el nivel medio, expresada en metros.

## Variables Internas

- **Laborable:** se utilizó el calendario académico mostrado en la Figura 3-13 para determinar si una jornada era laborable o no.

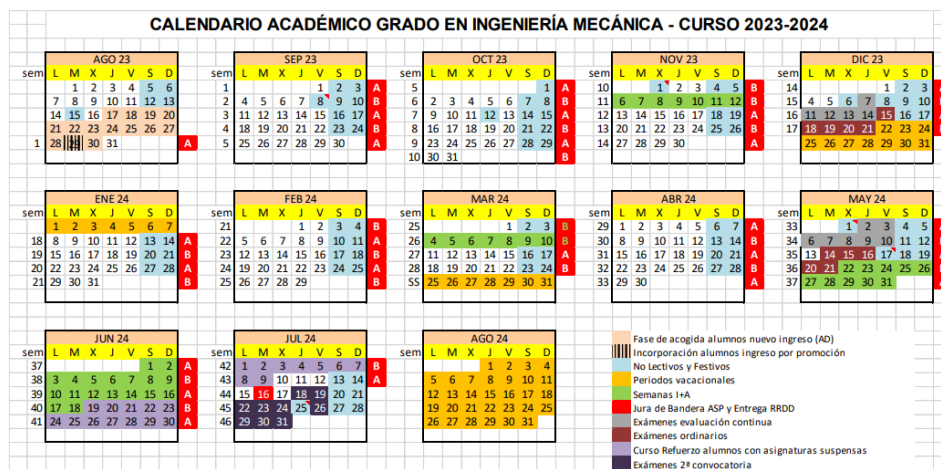


Figura 3-13 Calendario académico curso 2023-2024 [44]

- **Ventilación:** se registró el modo de funcionamiento del fancoil, el cual se puede controlar remotamente mediante el chat de Telegram mencionado en el apartado 3.2.1. Este sistema cuenta con tres posiciones activas (1 a 3), que corresponden a diferentes caudales de aire, además del estado 0, que indica que el equipo está apagado. En la Figura 3-14 se muestra un ejemplo del registro de comandos enviados al sistema a través de Telegram.

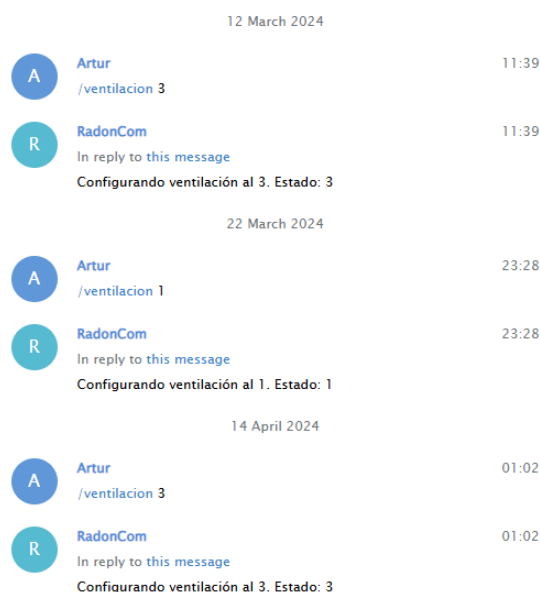


Figura 3-14 Registro del control del fancoil

### 3.3 Tratamiento de datos experimentales

#### 3.3.1 Tasa de exhalación

La ausencia de una metodología definida dificultó el análisis de la tasa de exhalación de radón. No se habían establecido tiempos mínimos para mantener el sistema cerrado, ni tampoco cuánto debía esperarse tras la apertura antes de iniciar un nuevo ciclo de medición. Cada ciclo de apertura y cierre fue distinto, reflejando variaciones considerables en la concentración de radón y dificultando la estandarización del proceso.

Para abordar esta situación, se analizó cada ciclo de manera individual, evaluando la evolución de la concentración de radón, eliminando aquellos registros que presentaban valores atípicos y estableciendo manualmente el tiempo de espera y el tiempo de medición en cada proceso. Se determinó que los tiempos de espera para comenzar a medir, una vez cerrado el cubo, debían ser lo suficientemente largos para permitir una acumulación representativa del gas, pero sin prolongarlos innecesariamente, ya que no aportaban un incremento sustancial en la concentración. A su vez, se evitó que fueran demasiado cortos para garantizar que reflejaran adecuadamente la fase inicial de acumulación del gas. Además, se buscó capturar la fase de crecimiento del radón sin llegar a un punto de saturación, donde la concentración se estabiliza debido al equilibrio entre la exhalación del suelo y la disipación del gas.

En la Figura 3-15 se muestra el tramo de la curva que se seleccionó en cada ciclo, correspondiente al intervalo en el que la concentración aumenta de forma representativa, sin llegar al punto de saturación.

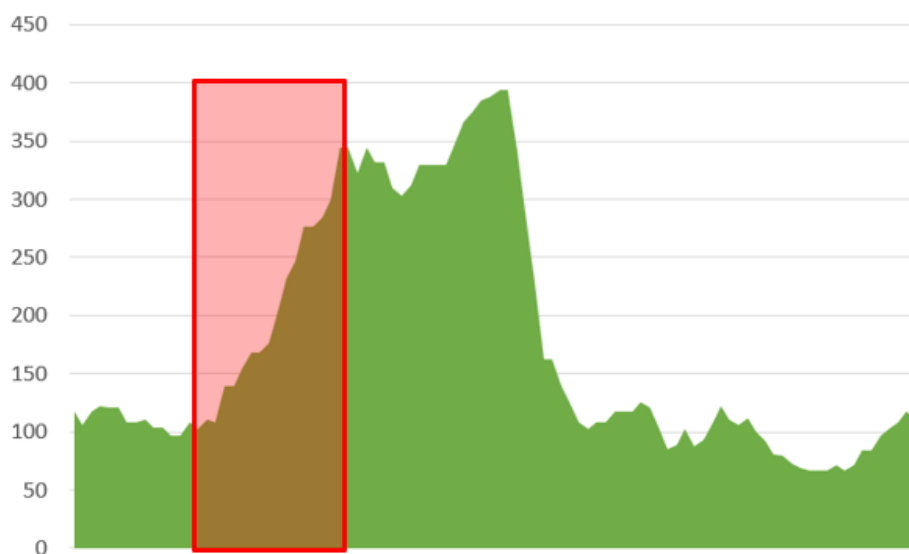


Figura 3-15 Ciclo de apertura/cierre

Con el objetivo de optimizar el procedimiento y reducir la intervención manual, se implementó un sistema automatizado utilizando Excel para el monitoreo y análisis de pendientes en los datos registrados. Excel fue elegido por ser una herramienta más visual, lo que facilita la interpretación de los datos y el análisis de tendencias.

#### 3.3.2 Variables externas e internas relacionadas con la concentración de radón

Una vez calculadas las tasas de exhalación correspondientes a los distintos periodos en los que el cubo permaneció cerrado, se determinaron los valores de las diferentes variables asociadas a cada uno de estos intervalos.

Para las variables de tipo cuantitativo, como la temperatura, presión atmosférica, velocidad y dirección del viento, precipitaciones, radiación y humedad relativa, se calculó el valor medio durante cada periodo. En la Figura 3-16 se muestra un ejemplo del cálculo de estos promedios, donde se destacan

en amarillo los valores medios de presión y temperatura obtenidos a partir de los registros individuales correspondientes a un intervalo concreto.

| Tasa exhalacion (mBq/m2/s) | Presión (hPa) | Temperatura (°C) | Presión (hPa) | Temperatura (°C) |
|----------------------------|---------------|------------------|---------------|------------------|
|                            | 1005,16       | 18,11            |               |                  |
|                            | 1005,09       | 18,05            |               |                  |
|                            | 1005,17       | 18,12            |               |                  |
|                            | 1005,16       | 18,34            |               |                  |
|                            | 1005,12       | 18,35            |               |                  |
|                            | 1005,06       | 18,42            |               |                  |
|                            | 1005,22       | 18,58            |               |                  |
|                            | 1005,48       | 18,5             |               |                  |
|                            | 1005,67       | 17,85            |               |                  |
|                            | 1005,87       | 17,54            |               |                  |
|                            | 1005,97       | 17,18            |               |                  |
|                            | 1005,91       | 17,1             |               |                  |
|                            | 1005,81       | 17,08            |               |                  |
|                            | 1005,64       | 17,02            |               |                  |
|                            | 1005,75       | 17,02            |               |                  |
|                            | 1005,88       | 17,11            |               |                  |
|                            | 1005,88       | 17,26            |               |                  |
|                            | 1005,8        | 17,29            |               |                  |
| 3,857                      | 1005,76       | 17,35            | 1006,148      | 16,896           |
|                            | 1005,75       | 17,26            |               |                  |
|                            | 1005,88       | 17,37            |               |                  |
|                            | 1006,1        | 17,03            |               |                  |
|                            | 1006,02       | 16,96            |               |                  |
|                            | 1006,11       | 16,69            |               |                  |
|                            | 1006,29       | 16,54            |               |                  |

**Figura 3-16 Valor medio de las variables durante los distintos periodos**

Por otro lado, para las variables cualitativas, como por ejemplo si era laborable o no, el modo de funcionamiento del fancoil y la configuración de la ventilación general, se utilizó la moda para calcular el valor de dichas variables durante el periodo de tiempo que se había calculado la tasa de exhalación.

La Figura 3-17 muestra un ejemplo del cálculo de estas modas, donde se destacan los valores de ventilación general y de si es laborable obtenidos a partir de los registros individuales correspondientes a un intervalo concreto.

| Tasa exhalacion (mBq/m2/s) | Ventilación general | Laborable | Ventilación general | Laborable |
|----------------------------|---------------------|-----------|---------------------|-----------|
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 0         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
| 3,857                      | 70%                 | 1         | 70%                 | 1         |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |
|                            | 70%                 | 1         |                     |           |

**Figura 3-17 Moda de las variables cualitativas durante los diferentes periodos**

Finalmente, una vez obtenidos los valores representativos de todas las variables durante los periodos en los que el cubo permaneció cerrado, se procedió a consolidar la base de datos de datos final. Tal como se muestra en la Figura 3-18, cada fila corresponde a un intervalo con su respectiva tasa de exhalación, junto con el conjunto completo de variables cuantitativas y cualitativas asociadas.

| Fecha Real       | Tasa exhalacion (mBq/m <sup>2</sup> /s) | Presión (hPa) | Temperatura (°C) | Temperatura interior (°C) | Viento (km/h) |
|------------------|-----------------------------------------|---------------|------------------|---------------------------|---------------|
| 21/09/2023 8:40  | 3,856565657                             | 1006,148      | 16,896           | 20,04                     | 4,25          |
| 21/09/2023 23:30 | 11,122                                  | 1011,472      | 17,661           | 20,93                     | 4,91          |
| 22/09/2023 18:30 | 4,815676768                             | 1016,57       | 20,501           | 21,22                     | 12,27         |
| 23/09/2023 18:10 | 3,184181818                             | 1018,473      | 21,118           | 22,23                     | 8,95          |
| 24/09/2023 17:10 | 2,379333333                             | 1018,087      | 21,07            | 21,5                      | 10,14         |
| 25/09/2023 8:40  | 1,203919192                             | 1018,34       | 16,104           | 20,5                      | 2,46          |
| 25/09/2023 22:50 | 3,085252525                             | 1018,124      | 20,683           | 20,93                     | 13,03         |
| 26/09/2023 20:20 | 0,462787879                             | 1016,818      | 20,814           | 20,8                      | 11,75         |
| 29/09/2023 11:10 | 2,486646465                             | 1021,346      | 21,283           | 20,29                     | 4,37          |
| 02/10/2023 14:40 | 4,166767677                             | 1018,628      | 24,949           | 20,68                     | 11,87         |
| 03/10/2023 20:00 | 46,41460606                             | 1022,319      | 23,315           | 20,43                     | 19,13         |
| 04/10/2023 11:40 | 7,718161616                             | 1022,509      | 20,342           | 19,49                     | 3,06          |
| 05/10/2023 8:50  | 2,736484848                             | 1021,512      | 16,308           | 19,81                     | 7,09          |
| 05/10/2023 14:10 | 2,049010101                             | 1022,715      | 21,017           | 20,91                     | 6,2           |
| 06/10/2023 13:30 | 1,883010101                             | 1020,667      | 22,707           | 20,52                     | 3,95          |
| 08/10/2023 0:00  | 3,625171717                             | 1019,287      | 21,36            | 20,98                     | 1,12          |

Figura 3-18 Valor de las variables por cada periodo de cálculo de la tasa de exhalación

### 3.4 Selección de variables

Una vez recopilada la información, se llevó a cabo un proceso de selección de variables con el objetivo de identificar aquellas más relevantes y eliminar las redundantes, optimizando así la precisión y robustez del modelo.

Para ello, se realizó un análisis de correlación entre las distintas variables, con el fin de detectar posibles problemas que pudieran afectar al rendimiento del modelo. Este análisis permitió abordar dos aspectos clave:

- **Sobreajuste:** ocurre cuando el modelo aprende tan bien los datos de entrenamiento que también incorpora errores o particularidades específicas que no se repiten en otros casos. Como resultado, el modelo pierde capacidad para generalizar con datos nuevos.
- **Colinealidad:** sucede cuando dos o más variables del modelo están fuertemente relacionadas entre sí. Esto puede confundir al modelo, dificultar la interpretación y hacerlo más inestable, ya que no queda claro cuál de las variables está influyendo realmente en el resultado.

Para identificar estas relaciones, se utilizó una matriz de correlaciones implementada en Python, que muestra la intensidad y dirección de las asociaciones entre variables. En esta matriz, los valores positivos aparecen en rojo (relación directamente proporcional), y los negativos en azul (relación inversa). La intensidad del color refleja la fuerza de la correlación, que varía entre -1 (correlación negativa perfecta) y 1 (correlación positiva perfecta), siendo 0 la ausencia de relación lineal. Esta representación facilita la detección de relaciones significativas dentro del conjunto de datos.

A partir del análisis de esta matriz, se llevó a cabo una primera criba, descartando aquellas variables con baja correlación respecto a la variable dependiente. Con el conjunto resultante, se realizaron múltiples pruebas con los distintos modelos utilizados, variando la combinación de variables en cada caso. El objetivo no era solo encontrar la configuración óptima para cada modelo, sino también analizar si ciertas correlaciones entre variables aportaban información complementaria o generaban redundancia. De este modo, se pudo evaluar cómo afecta la inclusión o exclusión de determinadas variables al rendimiento y la capacidad de generalización de cada modelo.

### 3.5 Entrenamiento de los modelos predictivos

Los modelos predictivos empleados para estimar la tasa de exhalación, descritos en el apartado 2.5.1, fueron implementados utilizando el lenguaje de programación Python. Los códigos correspondientes se encuentran disponibles en el Anexo III: códigos de

Cada uno de los modelos utilizados fue configurado con un conjunto específico de parámetros, seleccionados en función de su comportamiento habitual y de la naturaleza del problema. La Figura 3-19, la Figura 3-20, la Figura 3-21, la Figura 3-22 y la Figura 3-23 resumen los parámetros más relevantes definidos en la implementación de cada uno de ellos:

| Random Forest  |                                                                       |
|----------------|-----------------------------------------------------------------------|
| Valor asignado | Significado parámetro                                                 |
| 100            | Número de árboles que se construyen en el modelo.                     |
| 42             | Semilla para reproducibilidad en la generación de números aleatorios. |

**Figura 3-19 Parámetros del modelo Random Forest**

| SVR            |                                                |
|----------------|------------------------------------------------|
| Valor asignado | Significado parámetro                          |
| rbf            | Función del núcleo para transformar los datos. |

**Figura 3-20 Parámetros delo modelo SVR**

| XGBoost          |                                                                       |
|------------------|-----------------------------------------------------------------------|
| Valor asignado   | Significado parámetro                                                 |
| reg:squarederror | Función objetivo para regresión, minimizando el error cuadrático      |
| 100              | Número de árboles que se construyen en el modelo.                     |
| 0,1              | Tasa de aprendizaje que reduce la contribución de cada árbol.         |
| 42               | Semilla para reproducibilidad en la generación de números aleatorios. |

**Figura 3-21 Parámetros del modelo XGBoost**

| GBM            |                                                                       |
|----------------|-----------------------------------------------------------------------|
| Valor asignado | Significado parámetro                                                 |
| 100            | Número de árboles que se construyen en el modelo.                     |
| 0,1            | Tasa de aprendizaje que reduce la contribución de cada árbol          |
| 3              | Profundidad máxima de cada árbol                                      |
| 42             | Semilla para reproducibilidad en la generación de números aleatorios. |

**Figura 3-22 Parámetros del modelo GBM**

| MLP            |                                                                                                                 |
|----------------|-----------------------------------------------------------------------------------------------------------------|
| Valor asignado | Significado parámetro                                                                                           |
| 100            | Número de capas ocultas y neuronas en cada una                                                                  |
| relu           | Función de activación para las capas ocultas.                                                                   |
| adam           | Algoritmo de optimización para el entrenamiento                                                                 |
| 42             | Semilla para reproducibilidad en la generación de números aleatorios.                                           |
| 5000           | Número máximo de iteraciones durante el entrenamiento.                                                          |
| 1,00E-04       | Tolerancia para el criterio de convergencia; el entrenamiento se detiene si la mejora es inferior a este valor. |

**Figura 3-23 Parámetros del modelo MLP**

Antes del entrenamiento, fue necesario aplicar un proceso de escalado a los datos en algunos modelos, especialmente en aquellos que son sensibles a la magnitud de las variables, como la regresión lineal, SVR y redes neuronales. Este paso permitió que todas las variables numéricas compartieran una escala común, evitando que las de mayor rango dominaran el proceso de aprendizaje. En cambio, los modelos basados en árboles, como Random Forest y GBM, no precisaron este ajuste, ya que su rendimiento no se ve afectado por la escala de las variables.

Con los datos ya preparados, se dividió el conjunto total en dos subconjuntos diferenciados, con el fin de evaluar adecuadamente el rendimiento de los modelos y reducir el riesgo de sobreajuste:

- **Conjunto de entrenamiento (80%):** utilizado para ajustar los parámetros de los modelos y aprender los patrones subyacentes en los datos.
- **Conjunto de prueba (20%):** destinado a evaluar la capacidad de generalización de los modelos al aplicarse sobre datos no utilizados durante el proceso de entrenamiento.

Con esta división establecida, el rendimiento del conjunto de prueba de los modelos fue evaluado mediante distintas métricas estadísticas que permiten comparar objetivamente su precisión en la predicción de la tasa de exhalación:

### Errores Absoluto Medio (MAE)

Esta métrica permite obtener una medida clara e interpretable del error promedio cometido por el modelo, sin verse excesivamente afectado por valores atípicos. Se expresa como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \text{Ec. 24}$$

### Raíz del Error Cuadrático Medio (RMSE)

El RMSE otorga mayor peso a los errores grandes, lo cual resulta útil para identificar si el modelo incurre en desviaciones significativas en ciertos casos. Se expresa como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{Ec. 25}$$

### Coefficiente de Determinación ( $R^2$ )

Es una medida que indica qué porcentaje de la variabilidad de los datos es explicada por el modelo. Su valor varía entre 0 y 1, donde un valor cercano a 1 indica un ajuste óptimo del modelo. Se define como:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_1)^2} \quad \text{Ec. 26}$$

La combinación de estas tres métricas permite obtener una evaluación más completa del desempeño del modelo, considerando tanto su precisión como su capacidad para explicar la variabilidad de los datos. En particular, un valor bajo de  $R^2$  indica que el modelo explica solo una pequeña parte de la variabilidad observada, lo que sugiere un ajuste deficiente. Si, además, se obtienen valores elevados de MAE o RMSE, se refuerza la conclusión de que las predicciones del modelo presentan un alto nivel de error.

Por el contrario, un valor de  $R^2$  cercano a 1, junto con valores bajos de MAE y RMSE, indicaría que el modelo no solo ajusta bien los datos, sino que también realiza predicciones precisas, reflejando un buen desempeño general.

Para complementar el análisis cuantitativo, se llevó a cabo un análisis de residuos para identificar posibles errores sistemáticos o sesgos en el ajuste. Un modelo correctamente ajustado debe presentar residuos distribuidos aleatoriamente alrededor de la línea de referencia ( $y=0$ ). Si los residuos muestran patrones, significa que el modelo no captura correctamente algunas características de los datos, lo que podría requerir ajustes en las variables, la estructura del modelo o el volumen de datos.

Finalmente, para entender cómo el modelo responde al aumento de datos, se analizaron las curvas de aprendizaje. Estas curvas se construyen graficando el error en el conjunto de entrenamiento y en el conjunto de validación conforme crece el tamaño de la muestra. Su objetivo es determinar si el modelo sigue mejorando con más datos o si ha alcanzado su límite de aprendizaje. Estas curvas permiten detectar casos de sobreajuste, cuando el modelo se ajusta en exceso a los datos de entrenamiento y pierde capacidad de generalización; en estos casos, aumentar la cantidad de datos disponibles o aplicar técnicas de regularización puede contribuir a mejorar su desempeño. Por el contrario, cuando se observa infracapacidad, es decir, que el modelo resulta demasiado simple para representar los patrones subyacentes, es necesario aumentar su complejidad o mejorar la calidad de los datos empleados. El análisis de estas curvas orienta así la toma de decisiones tanto sobre la arquitectura del modelo como sobre las necesidades del conjunto de datos.

## 4 RESULTADOS

### 4.1 Filtrado y análisis de datos

A partir de lo observado durante el análisis manual, se estableció una serie de condiciones para automatizar el cálculo de la tasa de exhalación de radón, con el objetivo de replicar el criterio aplicado de forma individual y reducir la intervención manual en el proceso. Esta automatización permitirá abordar futuros análisis de manera más eficiente y uniforme, haciendo que la comparación entre resultados sea más sencilla y fiable.

Las condiciones que determinan el inicio del registro de los datos para el cálculo de la tasa de exhalación y su duración son las siguientes:

- El primer dato se registraba cuando:
  - Se observaba un incremento en la pendiente. Para ello, se estableció que tanto la pendiente calculada con regresión del valor actual como la de los dos valores anteriores debían ser positivas, asegurando una tendencia ascendente sostenida.
  - Pasaban 60 minutos desde el cierre del cubo, en caso de no haberse detectado un incremento en la pendiente.
- Fin de registro de datos:
  - Se detectaba una disminución en la pendiente. Para evitar que la medición finalizara debido a una variación aislada en un solo valor, se establecía como criterio que tanto la pendiente actual como la siguiente debían ser menores que la mitad del promedio de las tres pendientes anteriores, garantizando que la disminución sea sostenida antes de finalizar la medición.
  - Para tener en cuenta retardos en la medición, se permitía medir hasta 30 minutos después del cierre
  - Lleva 90 minutos midiendo.

Además, para garantizar una recopilación representativa de los datos, se estableció que para que la medición fuera válida debía durar al menos 30 minutos.

Con el fin de afinar el comportamiento del sistema, se realizaron pruebas variando el umbral utilizado para determinar la disminución en la pendiente, ensayando con valores del 40 %, 45 %, 50 %, 55 % y 60 %. Al comparar los resultados generados automáticamente con los ciclos definidos manualmente, se observó que el mejor rendimiento se obtenía al utilizar como umbral la mitad del promedio de las pendientes anteriores. Este enfoque reproducía con mayor fidelidad las decisiones

aplicadas en el análisis manual. Esta buena concordancia se refleja en la Figura 4-1, que muestra una gráfica de dispersión con una fuerte correlación entre ambos métodos y un coeficiente de determinación de  $R^2 = 0,9159$ , lo que valida la eficacia del sistema automatizado para el cálculo de la tasa de exhalación de radón.

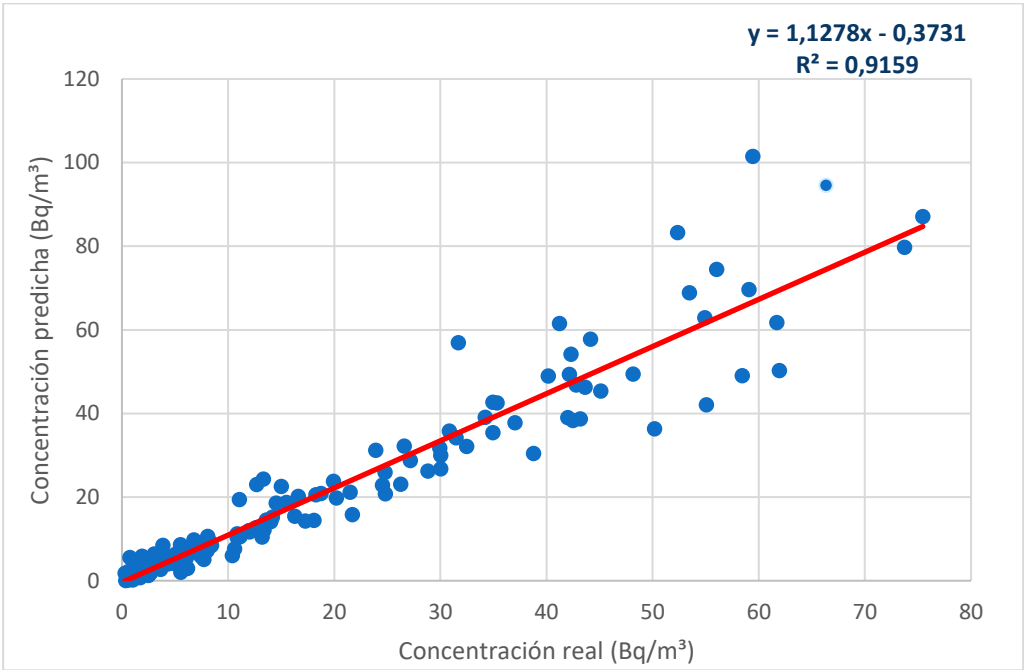


Figura 4-1 Valores obtenidos manualmente frente a los obtenidos automáticamente

4.2 Selección de variables

El análisis de la matriz de correlaciones, representada en la Figura 4-3, permitió identificar las variables con mayor impacto en la tasa de exhalación y descartar aquellas con influencia mínima. Este proceso facilitó una primera criba de variables, enfocando el análisis en los factores más determinantes.

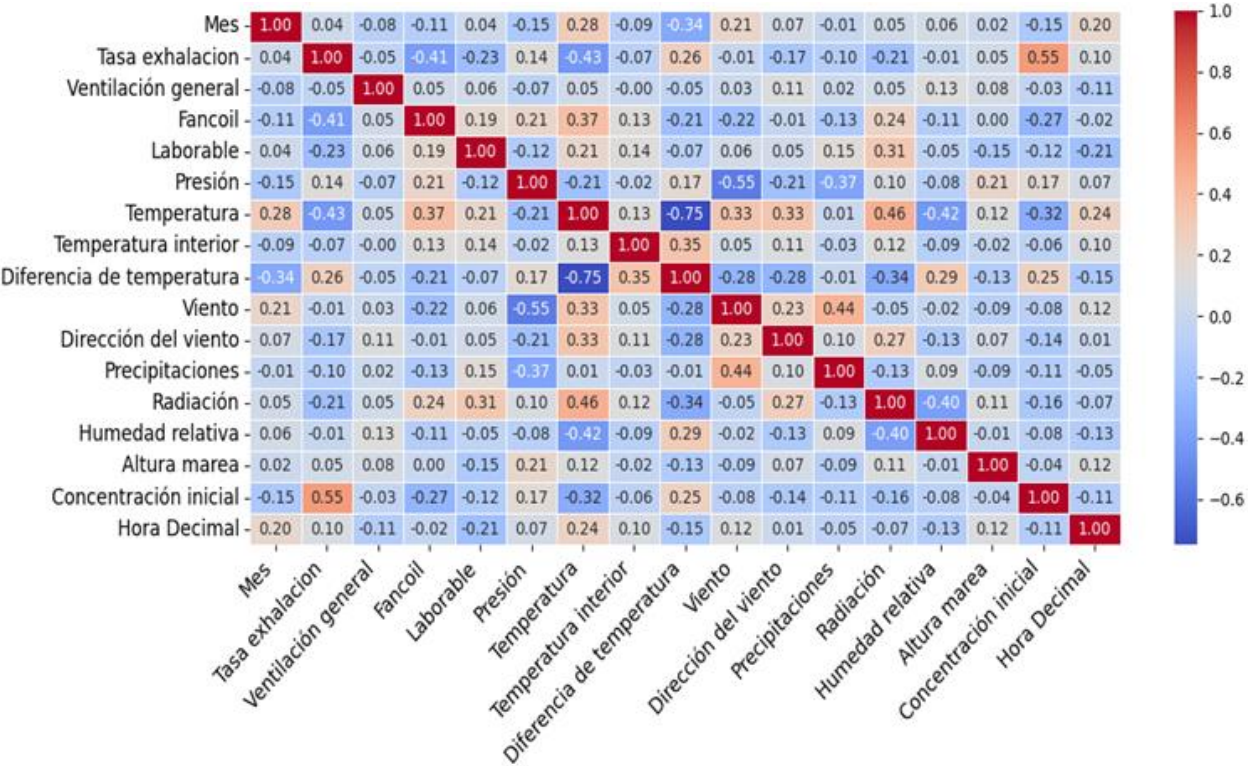


Figura 4-2 Matriz de correlaciones

La Figura 4-3 muestra una gráfica que agrupa las variables según su nivel de influencia en la tasa de exhalación. Esta representación permite identificar de manera más rápida cuáles son las más relevantes para el modelo y cuáles tienen un impacto reducido y pueden descartarse.

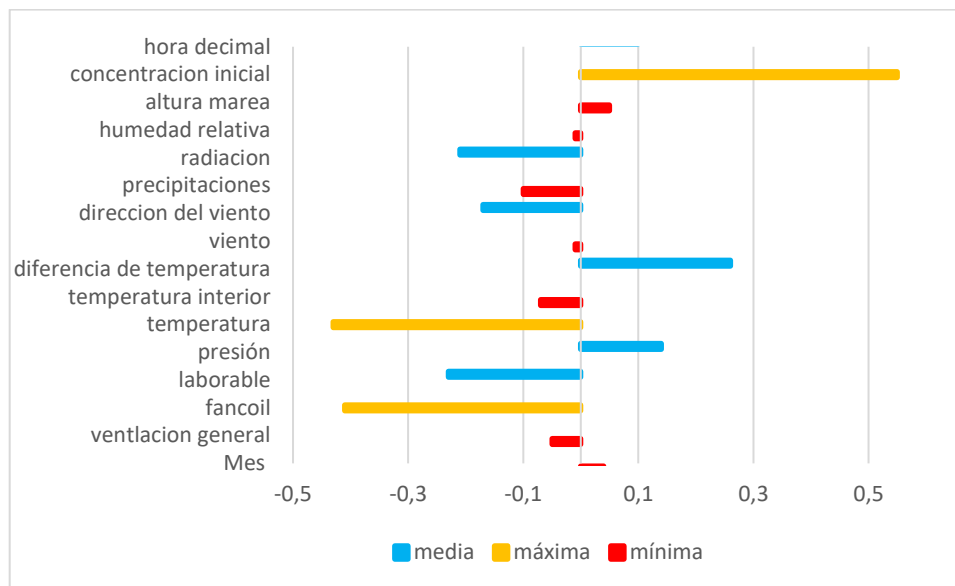


Figura 4-3 Correlación con la tasa de exhalación

Con el fin de interpretar mejor los resultados, las variables se agruparon en tres categorías según su grado de correlación con la tasa de exhalación:

- **Variables de alta influencia ( $>|0,39|$ ):** Variables con una fuerte relación con la tasa de exhalación, consideradas esenciales para el modelo final.
  - Concentración inicial (0,55): principal factor que influye en la tasa de exhalación debido a su impacto en la difusión del gas.
  - Temperatura (-0,43) y fancoil (-0,41): ambas variables muestran una correlación negativa con la exhalación del gas, indicando que el sistema de climatización juega un papel clave en la regulación del proceso.
- **Variables de influencia moderada ( $|0,1 - 0,39|$ ):** Variables con cierto impacto en la tasa de exhalación.
  - Factores atmosféricos: La radiación (-0,21) y la diferencia de temperatura (0,26) afectan la dinámica térmica del ambiente y, por ende, la liberación del gas.
  - Condiciones meteorológicas: La dirección del viento (-0,17) y la presión (0,14) pueden modificar la dispersión del gas en función de las condiciones atmosféricas.
  - Factores operacionales y temporales: Variables como día laborable (0,23) y hora decimal (0,1) sugieren que la actividad humana o patrones horarios influyen en la tasa de exhalación.
- **Variables de baja influencia ( $<|0,1|$ ):** Variables cuya relación con la tasa de exhalación es débil o insignificante, y que fueron descartadas.
  - Mes (0,04),
  - Ventilación general (-0,05)
  - Temperatura interior (-0,07)
  - Humedad relativa (-0,01)

- Viento (-0,01)
- Altura marea (0,05)

Además de analizar cada variable de forma individual, se evaluó la interacción entre ellas, ya que algunas presentan correlaciones significativas. En ciertos casos, estas relaciones son complementarias y aportan información adicional que mejora la capacidad del modelo predictivo. Sin embargo, cuando la correlación entre dos variables es demasiado alta, pueden generar redundancia, dificultando la interpretación del modelo y haciéndolo más sensible a pequeñas variaciones en los datos.

Entre las relaciones más significativas se encuentran:

- **Temperatura y fancoil:** ambos factores influyen en la climatización del espacio, lo que afecta la difusión del gas.
- **Diferencia de temperatura y radiación:** la radiación solar altera la temperatura ambiente, generando cambios en la tasa de exhalación.
- **Presión y viento:** estas variables, en conjunto, pueden modificar la dispersión del gas en el entorno

Tras este primer análisis, se redujo el conjunto inicial de 16 variables a un total de 10 variables. Con estas variables, se realizaron múltiples pruebas en las que se fue variando la configuración de estas, no solo para encontrar la combinación óptima, sino también para analizar si la correlación entre estas variables aportaba información complementaria o, por el contrario, generaba redundancia. De este modo, se examinó con detalle cómo la inclusión o exclusión de ciertas variables afectaba el desempeño del modelo y su capacidad de generalización. Curiosamente, aunque algunas variables presentaban inicialmente una baja correlación con la variable objetivo, su inclusión en determinadas combinaciones contribuyó a mejorar los resultados. Esto indica que, pese a no mostrar una relación evidente de forma aislada, algunas variables pueden aportar valor cuando interactúan con otras dentro del modelo.

En la Figura 4-5 se muestran las distintas configuraciones de las variables empleadas en cada prueba. Las marcadas con una cruz indican las variables incluidas en cada prueba, lo que permite visualizar de forma clara cómo se modificó el conjunto de variables a lo largo del proceso de experimentación.

| Prueba                 | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------------|---|---|---|---|---|---|
| Concentración inicial  | X | X | X | X | X | X |
| Fancoil                | X | X | X | X | X | X |
| Temperatura            | X | X | X | X | X | X |
| Presión                | X | X |   |   |   |   |
| Laborable              | X | X | X |   |   |   |
| Diferencia temperatura | X | X | X | X | X |   |
| Dirección viento       | X | X | X |   |   |   |
| Precipitaciones        | X |   |   |   |   |   |
| Radiación              | X | X | X | X |   |   |
| Hora decimal           | X | X | X | X | X | X |

**Figura 4-4 Variables utilizadas en cada una de las pruebas**

Para ilustrar los resultados obtenidos en las distintas pruebas, la Figura 4-5, la Figura 4-6 y la Figura 4-7 muestran el desempeño de cada uno de los modelos evaluados con los diferentes conjuntos de variables.

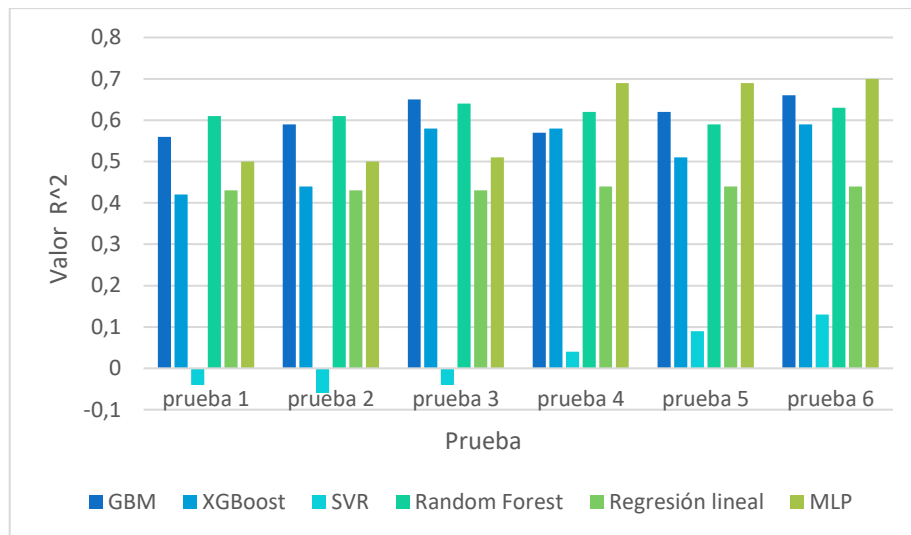


**Figura 4-5 Representación de los valores de  $R^2$  en cada una de las pruebas**

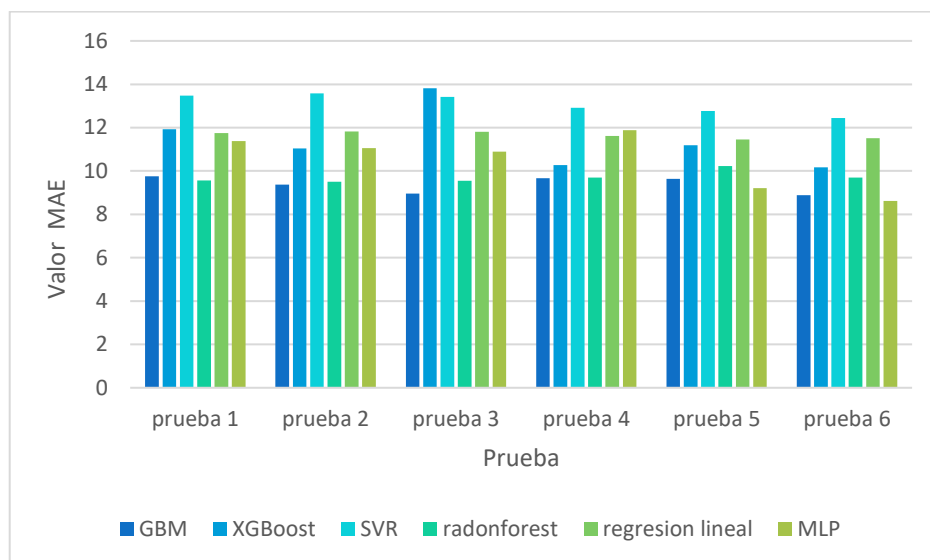


**Figura 4-6 Representación de los valores MAE en cada una de las pruebas**

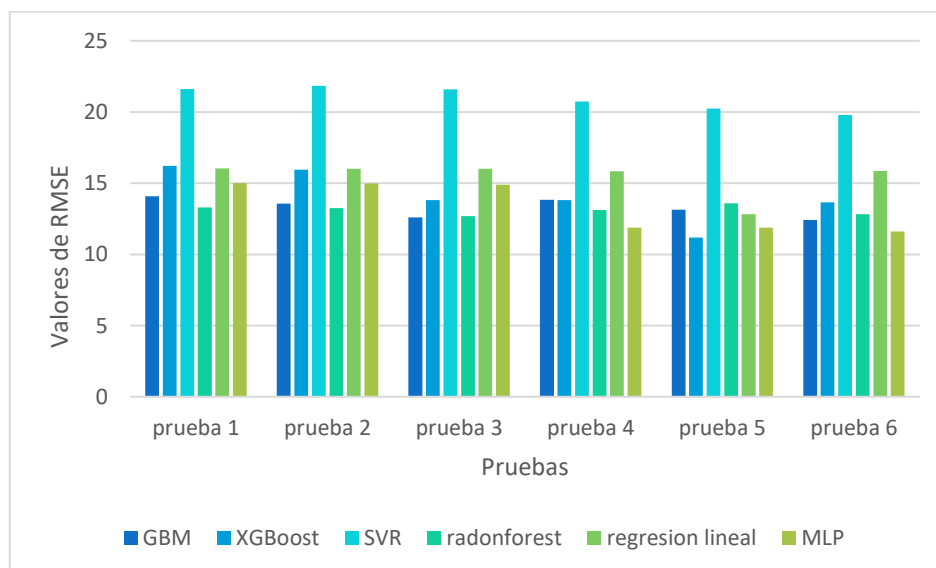


**Figura 4-7 Representación de los valores RMSE en cada una de las pruebas**

El análisis de los resultados muestra que la reducción progresiva de variables en las distintas pruebas permitió al modelo MLP mejorar significativamente su desempeño. En la última prueba, donde se utilizaron exclusivamente las variables concentración inicial, temperatura, fancoil y hora decimal, el modelo alcanzó un  $R^2$  de 0,70, el valor más alto entre los modelos evaluados. Además, logró reducir los errores (MAE y RMSE), lo que evidencia una mayor precisión en sus predicciones.

Estos resultados confirman que eliminar variables irrelevantes ayudó al modelo a generalizar mejor y reducir el riesgo de sobreajuste. Por otro lado, los modelos basados en árboles, como GBM, Random Forest y XGBoost, también experimentaron una mejora progresiva en su rendimiento, aunque en menor medida que MLP. Esto sugiere que estos modelos requieren un mayor volumen de datos para alcanzar su máximo potencial.

En contraste, la regresión lineal mantuvo un desempeño estable, con un  $R^2$  entre 0,42 y 0,44, lo que indica que la reducción de variables no impactó significativamente su precisión. Esto se debe a su capacidad limitada para capturar relaciones complejas en los datos.

Por el contrario, el modelo SVR (Máquinas de Soporte Vectorial) mostró un rendimiento inconsistente, con valores de  $R^2$  negativos en algunas pruebas y sin superar los 0,12 en ninguna configuración. Esto evidencia su baja capacidad de adaptación a los datos y su falta de efectividad en este contexto.

Con el fin de visualizar el comportamiento del modelo, la Figura 4-8 representa la relación entre los valores reales y los predichos, permitiendo apreciar su capacidad para ajustarse a los datos y capturar sus patrones de variación. Se realizó este análisis únicamente con el modelo MLP, ya que fue el que presentó mejores resultados. La gráfica ofrece una visión general del ajuste, mostrando cómo el modelo capta la tendencia general y se adapta a la mayoría de los valores; sin embargo, en casos específicos, como en la muestra 15, se observa que el modelo subestima un pico elevado, lo que indica que tiende a suavizar valores atípicos en lugar de ajustarse plenamente a ellos.

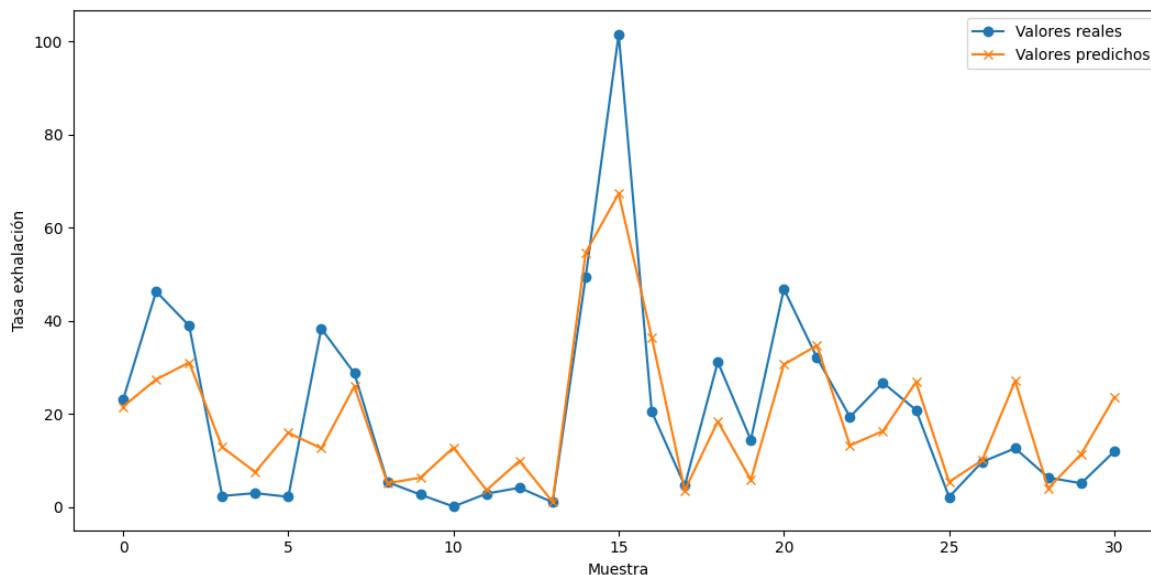


**Figura 4-8 Representación de los valores predichos frente a los valores reales**

Para identificar de forma más precisa la magnitud y localización de estos errores se complementó el análisis con el estudio de los residuos. Se analizaron los residuos únicamente de los modelos MLP, GBM, regresión lineal y SVR, ya que representan diferentes niveles de desempeño. MLP fue seleccionado como el modelo con mejor rendimiento, GBM por su buen ajuste general, pero con algunos valores atípicos, la regresión lineal como modelo base y SVR como el de menor efectividad. Modelos como Random Forest y XGBoost no se incluyeron en esta parte debido a que sus patrones de error eran similares a los de GBM y no aportaban información adicional.

La Figura 4-9 y la Figura 4-10 muestran que los modelos MLP y GBM presentan residuos distribuidos de manera mayormente aleatoria alrededor de la línea de referencia ( $y = 0$ ), lo que sugiere un buen ajuste global. Sin embargo, ambos presentan mayor dispersión en predicciones elevadas, indicando dificultades para modelar correctamente los valores extremos. Esto podría deberse a la falta de muestras en ciertos rangos de la variable objetivo, lo que limita la capacidad de los modelos para aprender esos patrones. En particular, el GBM muestra signos de sobreajuste, que podrían corregirse aumentando la cantidad y diversidad de los datos de entrenamiento.

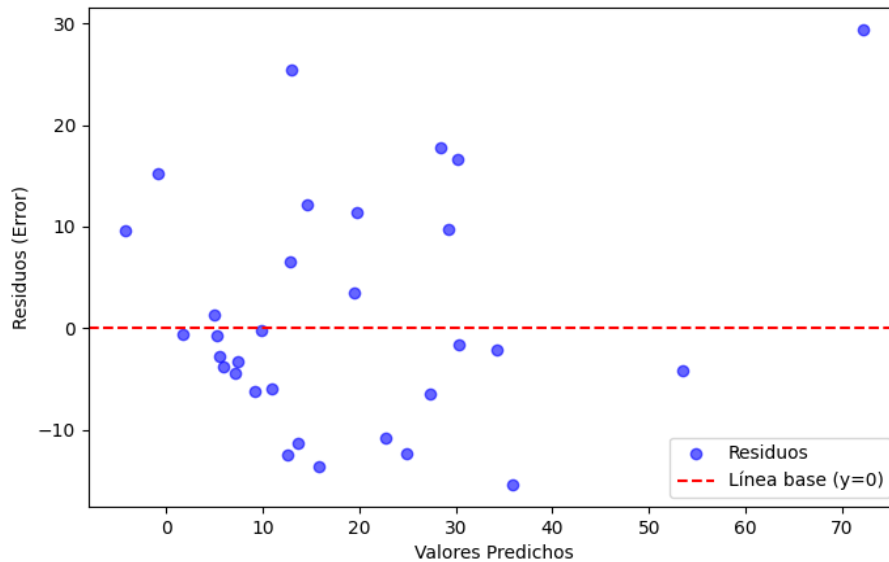


**Figura 4-9 Residuos modelo MLP**

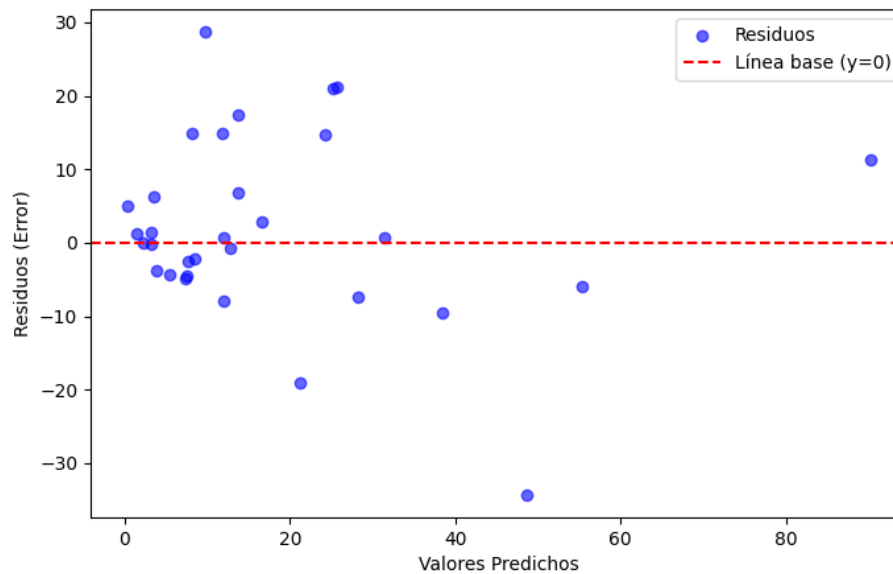


**Figura 4-10 Residuos modelo GBM**

La regresión lineal, representada en la Figura 4-11, muestra un patrón claro en los residuos, reflejando su incapacidad para capturar relaciones no lineales en los datos. Esto genera errores sistemáticos y reduce su capacidad de generalización. Aunque es posible mejorar su desempeño mediante la inclusión de términos polinómicos, seguiría estando por debajo de modelos más avanzados como MLP o GBM, que manejan mejor las interacciones complejas entre variables.

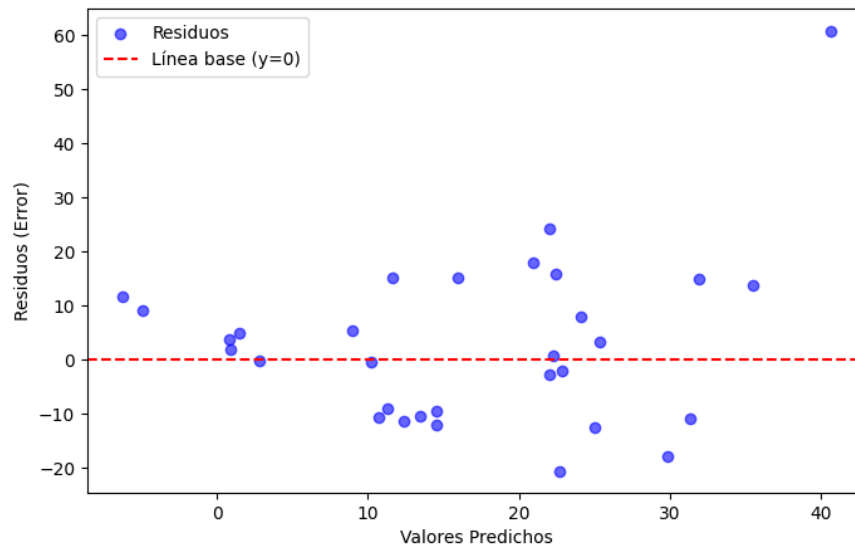


**Figura 4-11 Residuos modelo regresión lineal**

Por su parte, el modelo SVR, mostrado en la Figura 4-12, presenta los residuos más dispersos y una tendencia a subestimar los valores reales, ya que la mayoría de los errores son positivos. Además, los errores tienden a incrementarse con valores predichos más altos. Esta ineficiencia puede explicarse por la sensibilidad del modelo a la selección de hiperparámetros y a la escala de los datos, lo cual afecta negativamente su capacidad de generalización.

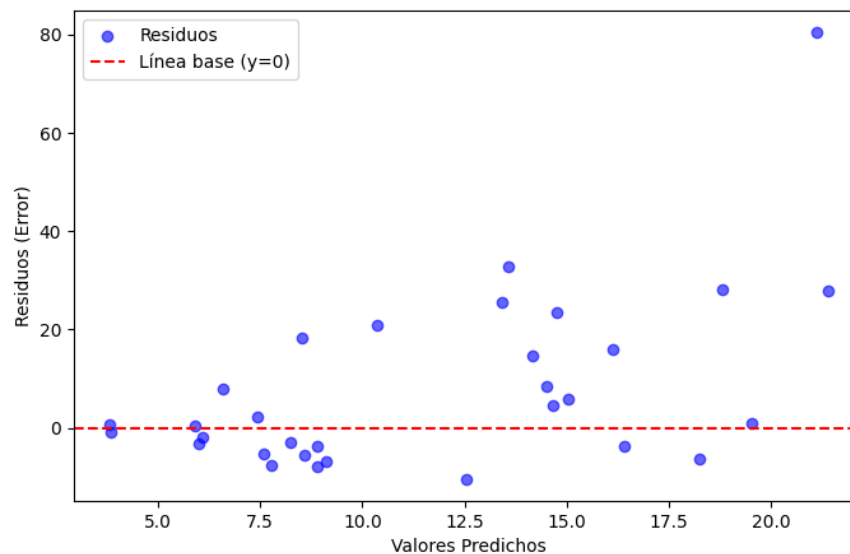


**Figura 4-12 Residuos modelo SVR**

Considerando que los residuos reflejaron problemas de sobreajuste y dificultades en la generalización en varios modelos, se procedió a analizar las curvas de aprendizaje, con el objetivo de evaluar si un mayor volumen de datos podría mejorar estos aspectos. Se construyeron curvas de aprendizaje para MLP, GBM, regresión lineal y SVR, representando cómo varía el error conforme aumenta el número de muestras en el conjunto de entrenamiento.

En primer lugar, la gráfica del modelo MLP mostrada en la Figura 4-13 refleja un error de entrenamiento muy bajo, lo que indica un fuerte ajuste a los datos iniciales. Sin embargo, el error de validación parte de valores altos y decrece paulatinamente a medida que se añaden más muestras, lo cual sugiere que el sobreajuste inicial se va mitigando y que el modelo mejora su capacidad de generalización. Este hallazgo coincide con el análisis de residuos, donde el MLP obtiene un buen desempeño en la mayoría de los casos, pero experimenta dificultades puntuales en valores extremos.

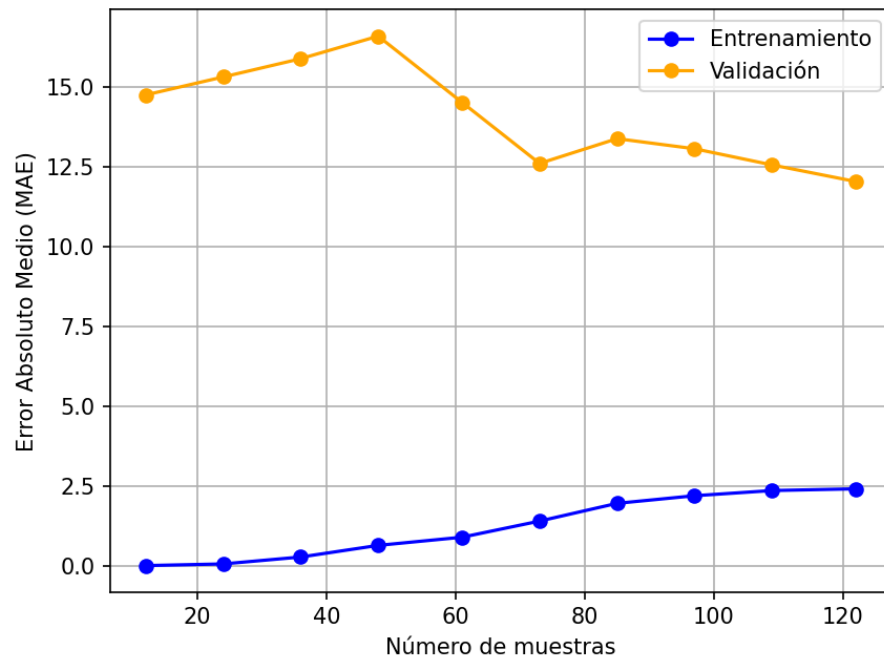


**Figura 4-13 Curva de aprendizaje MLP**

De forma similar, la Figura 4-14 muestra como el modelo GBM tiene un error de entrenamiento casi constante y muy reducido, al tiempo que el error de validación comienza en niveles elevados y disminuye de manera sostenida cuando se incrementa el volumen de datos. Esto pone de manifiesto que el modelo se encuentra sobreajustado al principio, pero responde positivamente a la inclusión de más muestras. En los residuos, el GBM también presenta un rendimiento sólido en la mayor parte del rango de valores, aunque las desviaciones más notorias se concentran en puntos extremos.

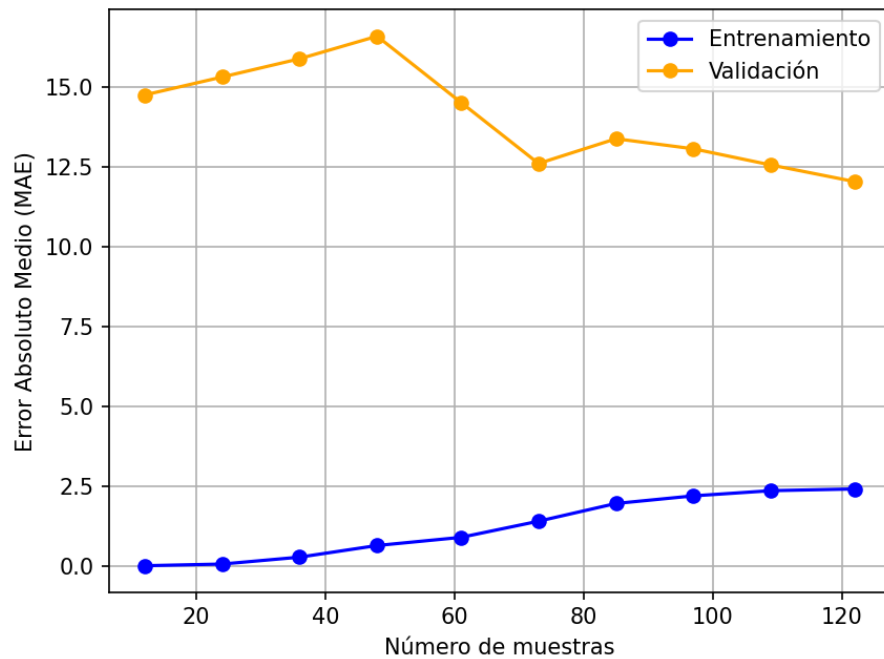


**Figura 4-14 Curva de aprendizaje GBM**

Respecto al modelo de Regresión Lineal, la Figura 4-15 muestra que el error de entrenamiento se mantiene en niveles moderadamente bajos y estables. Por otro lado, el error de validación, que se inicia en valores elevados según la figura, desciende progresivamente al aumentar el número de muestras. Esta

tendencia indica que, aunque la Regresión Lineal presenta cierto sobreajuste en condiciones de pocos datos, su capacidad para capturar la tendencia general mejora significativamente con más muestras. Los residuos analizados refuerzan esta observación, mostrando un rendimiento aceptable en la mayoría de los casos, con mayor variabilidad en los extremos.

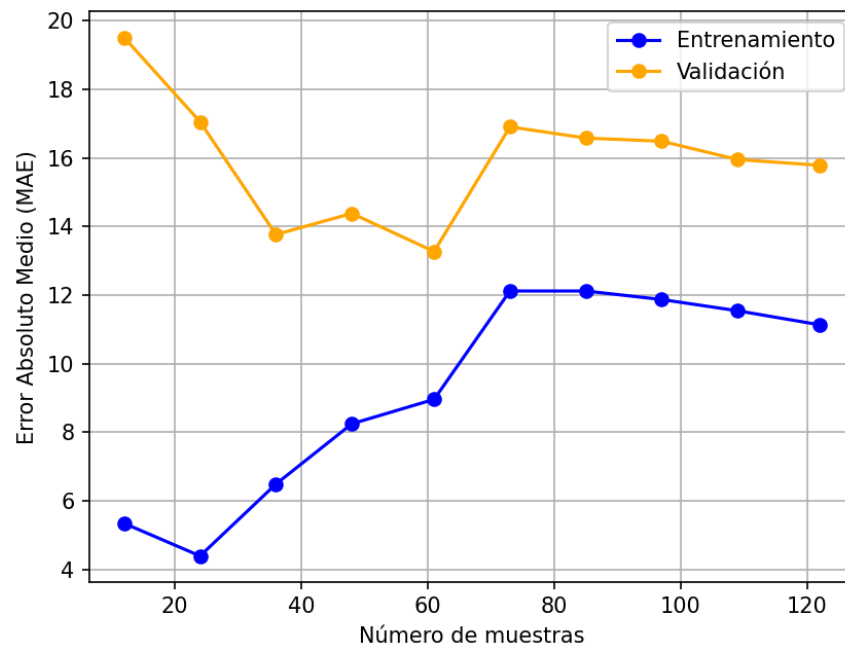


**Figura 4-15 Curva aprendizaje regresión lineal**

Finalmente, en la Figura 4-16 evidencia un comportamiento similar en el modelo SVR: el error de entrenamiento es bajo y se mantiene estable, mientras que el error de validación, inicialmente alto según la figura, disminuye de forma gradual a medida que se incrementa el tamaño del conjunto de entrenamiento. Este patrón indica que el SVR también sufre de sobreajuste cuando se trabaja con conjuntos reducidos, pero mejora su capacidad de generalización al disponer de más datos. La evolución gráfica se refleja en los residuos, donde el modelo presenta un buen rendimiento general, con algunas dificultades concentradas en los valores extremos.

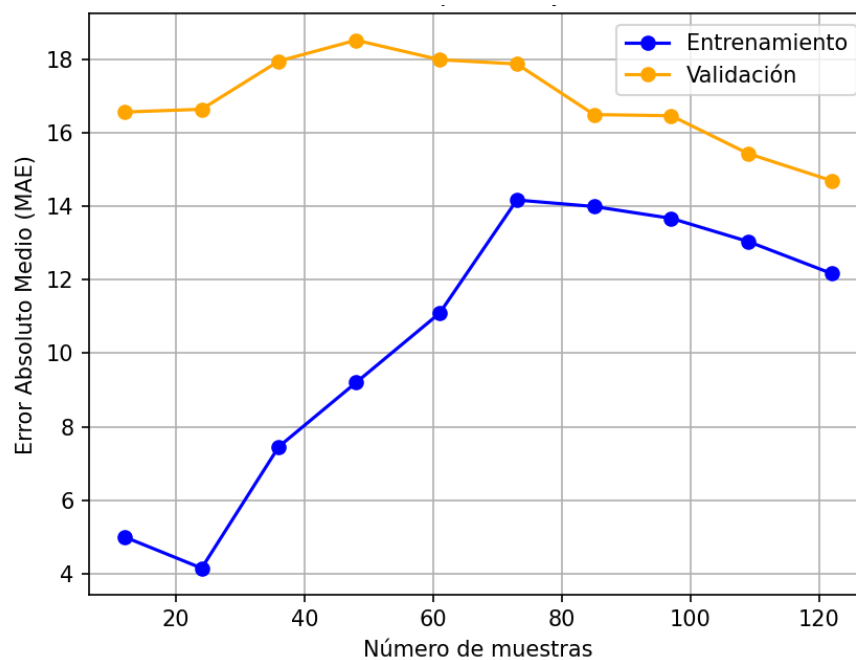


**Figura 4-16 Curva de aprendizaje SVR**

En conclusión, el análisis global de los residuos y de las curvas de aprendizaje confirma que, al incrementar el volumen de datos, todos los modelos mejoran su capacidad de generalización y mitigan el sobreajuste. No obstante, los modelos MLP y GBM se destacan por ofrecer un equilibrio óptimo entre ajuste y robustez, evidenciado en la reducción progresiva del error de validación y en la distribución favorable de los residuos. Por lo tanto, considerando exclusivamente las variables concentración inicial, temperatura, fancoil y hora decimal, el MLP y el GBM se posicionan como las opciones más efectivas para capturar la complejidad de los datos y ofrecer predicciones precisas.

## 5 CONCLUSIONES

El presente trabajo de fin de grado ha abordado la modelización de la inmisión del gas radón en espacios cerrados mediante técnicas de medición continua y modelos predictivos basados en inteligencia artificial. A través del análisis de variables meteorológicas, operativas y temporales, se ha logrado identificar patrones que permiten anticipar la concentración de radón en interiores y optimizar la gestión de medidas de mitigación.

### 5.1.1 Aportaciones del estudio

Una de las principales contribuciones de este trabajo es el desarrollo de un sistema automatizado para el cálculo de la tasa de exhalación de radón, lo cual reduce la intervención manual y garantiza la coherencia en los resultados. Asimismo, se ha demostrado la utilidad de los modelos de predicción basados en inteligencia artificial, destacando especialmente el Perceptrón Multicapa (MLP), que ha presentado un rendimiento superior gracias a su capacidad para modelar relaciones no lineales complejas presentes en los datos.

Otra aportación relevante ha sido la identificación de las cuatro variables más influyentes en la exhalación del gas radón: la concentración inicial de radón, la temperatura, el uso del fancoil y la hora decimal. Este hallazgo refuerza la necesidad de realizar una selección cuidadosa de variables al diseñar modelos predictivos, evitando redundancias y mejorando la precisión de las estimaciones.

### 5.1.2 Resumen de los hallazgos

Los resultados obtenidos han evidenciado el impacto significativo de las variables meteorológicas y temporales en la tasa de exhalación de radón, subrayando su papel en la predicción de la inmisión. Se ha observado que una selección adecuada de variables permite obtener un rendimiento óptimo, lo que demuestra la eficacia de un enfoque de modelado simplificado, pero bien fundamentado.

Además, se ha identificado una limitación clave en la escasez de datos disponibles, lo que afecta negativamente al rendimiento de modelos más simples y favorece el sobreajuste en modelos complejos, reduciendo su capacidad de generalización. Este aspecto ha sido especialmente crítico en modelos como SVR, que han mostrado un comportamiento sensible a la cantidad de datos, frente al MLP, que ha ofrecido una mejor capacidad predictiva.

También se ha constatado la necesidad de establecer criterios estandarizados y tiempos de medición definidos para garantizar la calidad de los datos recopilados. La falta de una metodología clara puede comprometer la validez de los resultados, por lo que se recomienda la implementación de protocolos sistemáticos en estudios futuros.

### 5.1.3 Implicaciones del estudio

Los hallazgos de este estudio tienen importantes implicaciones metodológicas y operativas en el análisis de la inmisión del gas radón. La metodología desarrollada proporciona una herramienta analítica para la optimización de la medición y modelización de la exhalación del gas, facilitando la toma de decisiones basada en datos históricos. Aunque su implementación requiere estudios adicionales, este modelo constituye un primer paso para predecir de manera más precisa las concentraciones elevadas de radón.

Es crucial optimizar la recopilación y almacenamiento de datos sobre la exhalación del gas radón para mejorar la precisión y fiabilidad de los modelos predictivos. La digitalización y automatización de los registros de exhalación podrían contribuir significativamente a la eficiencia del proceso de análisis, reduciendo errores y facilitando la interpretación de datos. Además, la aplicación de técnicas avanzadas de inteligencia artificial y aprendizaje automático para mejorar la predicción de la tasa de exhalación representa una línea de investigación con gran potencial. La integración de algoritmos de aprendizaje automático permitiría identificar patrones en la exhalación del gas y anticipar con mayor precisión sus variaciones en función de las condiciones ambientales.

### 5.1.4 Implicaciones y líneas futuras

La metodología desarrollada en este trabajo proporciona una base sólida para la predicción de concentraciones elevadas de radón en espacios interiores, facilitando la toma de decisiones basada en datos históricos. Su aplicación práctica puede contribuir a una gestión más eficiente del riesgo, especialmente en entornos donde la exposición al radón supone un problema para la salud pública.

De cara al futuro, resulta fundamental ampliar la base de datos mediante mediciones adicionales en diversos escenarios, lo que permitirá mejorar la capacidad de generalización de los modelos. En paralelo, será necesario explorar otras técnicas de selección de variables para afinar la precisión de las predicciones y evitar redundancias en la información utilizada.

Otra línea prometedora consiste en incorporar nuevas variables potencialmente influyentes en la dinámica del radón, como la temperatura y humedad del suelo o la concentración de CO<sub>2</sub> en interiores, con el objetivo de enriquecer los modelos y aumentar su capacidad explicativa.

La implementación de sistemas de monitoreo en tiempo real permitiría automatizar la recopilación de datos y adaptarse dinámicamente a los cambios en las condiciones ambientales. Esto podría complementarse con el desarrollo de un sistema automatizado de ventilación, que, basado en las predicciones del modelo, ajuste su funcionamiento para evitar que los niveles de radón superen los umbrales establecidos por las normativas.

## 6 BIBLIOGRAFÍA

- [1] A. Ruano-Ravina y J. M. Barros-Dios, «Radón interior. Un carcinógeno laboral olvidado».
- [2] A. Ruano-Raviña, L. Quindós-Poncela, C. Sainz Fernández, y J. M. Barros-Dios, «Radón interior y salud pública en España. Tiempo para la acción», *Gaceta Sanitaria*, vol. 28, n.º 6, pp. 439-441, nov. 2014, doi: 10.1016/j.gaceta.2014.09.003.
- [3] S. Darby *et al.*, «Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies», *BMJ*, vol. 330, n.º 7485, p. 223, ene. 2005, doi: 10.1136/bmj.38308.477650.63.
- [4] «El radón y sus efectos en la salud». Accedido: 5 de febrero de 2025. [En línea]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/radon-and-health>
- [5] A. Ruano-Ravina *et al.*, «Lung cancer mortality attributable to residential radon exposure in Spain and its regions», *Environmental Research*, vol. 199, p. 111372, ago. 2021, doi: 10.1016/j.envres.2021.111372.
- [6] H. Zeeb, F. Shannoun, y O. M. de la Salud, *Manual de la OMS sobre el radón en interiores: una perspectiva de salud pública*. Organización Mundial de la Salud, 2015. Accedido: 2 de febrero de 2025. [En línea]. Disponible en: <https://iris.who.int/handle/10665/161913>
- [7] B. Frutos Vázquez, «Estudio experimental sobre la efectividad y la viabilidad de distintas soluciones constructivas para reducir la concentración de gas radón en edificaciones», phd, E.T.S. Arquitectura (UPM), 2009. Accedido: 2 de febrero de 2025. [En línea]. Disponible en: <https://oa.upm.es/22535/>
- [8] «Mapa del potencial de radón en España - CSN». Accedido: 2 de febrero de 2025. [En línea]. Disponible en: <https://www.csn.es/mapa-del-potencial-de-radon-en-espana>
- [9] «Inicio - Laboratorio de Radón en Galicia - USC». Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <https://radon.gal/>
- [10] «Directiva 2013/59/Euratom del Consejo, de 5 de diciembre de 2013, por la que se establecen normas de seguridad básicas para la protección contra los peligros derivados de la exposición a radiaciones ionizantes, y se derogan las Directivas 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom y 2003/122/Euratom».
- [11] «DBHS.pdf». Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <https://www.codigotecnico.org/pdf/Documentos/HS/DBHS.pdf>

- [12] «Salubridad». Accedido: 2 de febrero de 2025. [En línea]. Disponible en: <https://www.codigotecnico.org/DocumentosCTE/Salubridad.html>
- [13] C. Camacho Martín, R. Devesa Rey (advisor), y L. González Gil (advisor), «Gestión de la presencia de radón en lugares de trabajo: determinación de concentraciones y evaluación de riesgos en la ENM», abr. 2021, Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <http://calderon.cud.uvigo.es/handle/123456789/569>
- [14] C. Lázaro Benavides, A. González Gil (advisor), y L. González Gil (advisor), «Evaluación de los factores determinantes en la concentración de radón en espacios interiores», abr. 2021, Accedido: 2 de febrero de 2025. [En línea]. Disponible en: <http://calderon.cud.uvigo.es/handle/123456789/589>
- [15] Á. García Denia, L. Febrero Garrido (advisor), y L. González Gil (advisor), «Medición de la concentración de radón en el Centro Universitario de la Defensa y propuesta de medidas de mitigación», abr. 2020, Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <http://calderon.cud.uvigo.es/handle/123456789/530>
- [16] A. Marín Alonso, A. González Gil (advisor), y L. Febrero Garrido (advisor), «Caracterización de los principales focos de inmisión de gas radón en despachos de la planta baja del CUD», abr. 2022, Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <http://calderon.cud.uvigo.es/handle/123456789/651>
- [17] «¿Qué es el radón y en qué lugares estamos expuestos a este gas?» Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <https://www.iaea.org/es/newscenter/news/que-es-el-radon>
- [18] «INT-04.pdf». Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <https://www.csn.es/documents/10182/27786/INT-04.20+Protecci%C3%B3n+frente+a+la+inmisi%C3%B3n+de+gas+rad%C3%B3n+en+edificios>
- [19] P. García Sobrido, A. (advisor) González Gil, y A. (advisor) Cacabelos Reyes, «Caracterización experimental del comportamiento del gas radón en el interior de edificios», abr. 2023, Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <http://calderon.cud.uvigo.es/handle/123456789/682>
- [20] A. Fernández Villar, «El mapa español de radón: análisis de factores indicativos del riesgo de exposición al radón», jul. 2017, Accedido: 5 de febrero de 2025. [En línea]. Disponible en: <https://repositorio.unican.es/xmlui/handle/10902/11553>
- [21] M. G.-T. S. Miguel y F. J. L. Acevedo, «Cartografía del potencial de radón de España».
- [22] «Guía Rehabilitación frente al radón». Accedido: 3 de febrero de 2025. [En línea]. Disponible en: <https://www.codigotecnico.org/Guias/GuiaRadon.html>
- [23] «Influence of Meteorological Conditions in the Annual Variation of Environmental Radon Activity», Accedido: 9 de febrero de 2025. [En línea]. Disponible en: <https://inis.iaea.org/records/p1s5e-3m493>
- [24] R. G. M. Crockett, G. K. Gillmore, P. S. Phillips, A. R. Denman, y C. J. Groves-Kirkby, «Tidal synchronicity of built-environment radon levels in the UK», *Geophysical Research Letters*, vol. 33, n.º 5, 2006, doi: 10.1029/2005GL024950.
- [25] «cte-db-hs6-proteccion-frente-a-la-exposicion-al-radon.pdf». Accedido: 6 de febrero de 2025. [En línea]. Disponible en: <https://assets.aldes.es/assets/documentos/cte-db-hs6-proteccion-frente-a-la-exposicion-al-radon.pdf>
- [26] F. arquitectura I, «fob arquitectura». Accedido: 25 de marzo de 2025. [En línea]. Disponible en: <https://www.fob-arquitectura.com/amplia/370/gas-radon-en-galicia.html>

- [27] «Guía Técnica RADEX», Colegio Oficial de Aparejadores y Arquitectos Técnicos de Badajoz. Accedido: 9 de febrero de 2025. [En línea]. Disponible en: <https://coaatba.com/novedades/guia-tecnica-radex/>
- [28] «RADON\_GUIA\_V1\_es\_20190308.pdf». Accedido: 9 de febrero de 2025. [En línea]. Disponible en: [https://igvs.xunta.gal/documents/11218/67016/RADON\\_GUIA\\_V1\\_es\\_20190308.pdf?version=1.0](https://igvs.xunta.gal/documents/11218/67016/RADON_GUIA_V1_es_20190308.pdf?version=1.0)
- [29] «El gas radón, un silencioso y peligroso invitado en los hogares», TÉCNICA INDUSTRIAL. Accedido: 9 de febrero de 2025. [En línea]. Disponible en: <https://www.tecnicaindustrial.es/el-gas-radon-un-silencioso-y-peligroso-invita/>
- [30] L. S. Quindós Poncela, *Radón: un gas radiactivo de origen natural en su casa*. 1995. Accedido: 9 de febrero de 2025. [En línea]. Disponible en: <https://dialnet.unirioja.es/servlet/libro?codigo=200882>
- [31] «UNE-EN ISO 11665-7:2016 Medición de la radiactividad en el amb...» Accedido: 9 de febrero de 2025. [En línea]. Disponible en: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0056714>
- [32] «Regresión lineal», *Wikipedia, la enciclopedia libre*. 11 de diciembre de 2024. Accedido: 16 de marzo de 2025. [En línea]. Disponible en: [https://es.wikipedia.org/w/index.php?title=Regresi%C3%B3n\\_lineal&oldid=164062382](https://es.wikipedia.org/w/index.php?title=Regresi%C3%B3n_lineal&oldid=164062382)
- [33] «1.4. Support Vector Machines», scikit-learn. Accedido: 17 de marzo de 2025. [En línea]. Disponible en: <https://scikit-learn/stable/modules/svm.html>
- [34] «Gradient boosting», *Wikipedia*. 27 de febrero de 2025. Accedido: 17 de marzo de 2025. [En línea]. Disponible en: [https://en.wikipedia.org/w/index.php?title=Gradient\\_boosting&oldid=1277920198](https://en.wikipedia.org/w/index.php?title=Gradient_boosting&oldid=1277920198)
- [35] «XGBoost», *Wikipedia*. 6 de marzo de 2025. Accedido: 17 de marzo de 2025. [En línea]. Disponible en: <https://en.wikipedia.org/w/index.php?title=XGBoost&oldid=1279070624>
- [36] M. Hosoda, S. Tokonami, T. Suzuki, y M. Janik, «Machine learning as a tool for analysing the impact of environmental parameters on the radon exhalation rate from soil», *Radiation Measurements*, vol. 138, p. 106402, nov. 2020, doi: 10.1016/j.radmeas.2020.106402.
- [37] S. Fang, Y. Chen, X. Wu, N. Zhao, y Y. Liu, «Radon exhalation rate prediction and early warning model based on VMD-GRU and similar day analysis», *Journal of Environmental Radioactivity*, vol. 281, p. 107593, ene. 2025, doi: 10.1016/j.jenvrad.2024.107593.
- [38] «Exploring the efficacy of machine learning models for predicting soil radon exhalation rates | Request PDF», *ResearchGate*, oct. 2024, doi: 10.1007/s00477-023-02509-x.
- [39] P.-Y. Wu, T. Johansson, C. Sandels, M. Mangold, y K. Mjörnell, «Indoor radon interval prediction in the Swedish building stock using machine learning», *Building and Environment*, vol. 245, p. 110879, nov. 2023, doi: 10.1016/j.buildenv.2023.110879.
- [40] «RadonEye (BLE) - Radon FTLab». Accedido: 17 de febrero de 2025. [En línea]. Disponible en: <https://radonftlab.com/radon-sensor-product/radon-detector/rd200/>
- [41] «Proyectos de financiación interna». Accedido: 26 de marzo de 2025. [En línea]. Disponible en: <https://cud.uvigo.es/proyectos-internos/>
- [42] MeteoGalicia, «Histórico da rede meteorolóxica», MeteoGalicia. Accedido: 26 de marzo de 2025. [En línea]. Disponible en: <https://www.meteogalicia.gal/web/home>

- [43] «history+ | Historia y archivo del tiempo», meteoblue. Accedido: 26 de marzo de 2025. [En línea]. Disponible en: <https://www.meteoblue.com/es/user/order/historyplus>
- [44] «Calendario\_academico\_CUD\_2023-24.pdf». Accedido: 26 de marzo de 2025. [En línea]. Disponible en: [https://www.uvigo.gal/sites/uvigo.gal/files/docs/universidade/institucional/calendario/Calendario\\_academico\\_CUD\\_2023-24.pdf](https://www.uvigo.gal/sites/uvigo.gal/files/docs/universidade/institucional/calendario/Calendario_academico_CUD_2023-24.pdf)

## **ANEXO I: IMPLICACIONES SOCIALES, Y/O AMBIENTALES Y/O ECONÓMICAS**

El radón es un gas radiactivo de origen natural que, a pesar de su peligrosidad, continúa siendo ampliamente desconocido por la población. Como se explica en la sección 1.1, al ser incoloro, inodoro e insípido, no puede detectarse sin instrumentos específicos, lo que dificulta su identificación. Esto, a su vez, reduce la percepción del riesgo. La Organización Mundial de la Salud lo señala como la segunda causa de cáncer de pulmón y estima que entre un 3 % y un 14 % de los casos de cáncer de pulmón a nivel mundial están asociados a su exposición. Aun así, muchas personas ignoran que pueden estar expuestas a niveles elevados de radón en sus viviendas o lugares de trabajo. Esta falta de información representa una implicación social importante, ya que obstaculiza la prevención y disminuye la presión ciudadana para que se impulsen medidas institucionales.

En respuesta a esta situación, en 2013, la Unión Europea estableció la Directiva 2013/59/EURATOM, que fijaba la obligación de que todos los Estados miembros establecieran un valor máximo de concentración anual de radón en interiores, no superior a 300 Bq/m<sup>3</sup>. España no incorporó esta recomendación hasta el año 2019, cuando modificó el Código Técnico de la Edificación (CTE) incluyendo por primera vez este límite en la normativa nacional. Esta demora pone de manifiesto la escasa prioridad que históricamente ha tenido el radón en las políticas públicas de salud ambiental, a pesar de su incidencia directa en la población.

No obstante, más allá del cumplimiento normativo, desde el punto de vista ambiental, anticipar concentraciones elevadas de radón mediante modelos predictivos permite aplicar medidas de forma más eficiente y solo cuando es necesario. Como se detalla en el capítulo 4, esta capacidad de previsión ayuda a evitar intervenciones constantes, como el uso intensivo de ventilación mecánica, que puede aumentar significativamente el consumo energético y las emisiones de gases de efecto invernadero. De este modo, anticiparse al problema también contribuye a reducir su impacto ambiental.

En línea con esta necesidad, el modelo desarrollado en este estudio, descrito en la sección 4.2, ofrece una herramienta útil para gestionar el radón de manera más eficiente y sostenible. Mejora la protección de la salud al facilitar una respuesta anticipada y también contribuye a minimizar el impacto ambiental derivado de medidas correctivas aplicadas de forma continua. En conjunto, los resultados muestran que la predicción es clave para adoptar estrategias de control más eficientes, equilibradas y respetuosas con el entorno.

Además de sus implicaciones sociales y ambientales, la gestión del radón también tiene importantes repercusiones económicas. Por un lado, la instalación de sistemas de mitigación como la ventilación mecánica o el aislamiento de suelos supone un coste considerable tanto para los particulares como para las administraciones públicas, especialmente en áreas con alta concentración de radón. Sin embargo, estos costes pueden reducirse mediante el uso de modelos predictivos como el propuesto en este estudio, que permiten actuar solo cuando es necesario, optimizando los recursos y evitando gastos innecesarios.

## ANEXO II: REFLEXIONES ÉTICAS Y SOCIALES

Uno de los aspectos que más llaman la atención al trabajar con el radón es la falta de responsabilidad que todavía existe en muchos entornos donde ya se conoce el problema. Aunque se dispone de estudios, normativas y herramientas para medirlo o predecir su comportamiento, la reacción por parte de instituciones, empresas o particulares sigue siendo limitada. No es solo una cuestión de desconocimiento, sino también de cómo se maneja la información que ya está disponible.

En algunos casos, se sabe que puede haber niveles altos de radón, pero no se toman medidas mientras no se superen los límites legales. Esto plantea una cuestión ética importante. Si ya se sabe que la exposición al radón tiene consecuencias negativas para la salud, ¿por qué esperar a que el riesgo sea oficialmente inaceptable para actuar?

También preocupa la falta de transparencia. Por ejemplo, en la compraventa o alquiler de viviendas no es habitual informar sobre los niveles de radón, a pesar de que su presencia puede depender de factores bien conocidos, como la zona geológica o las características del edificio. Esta omisión de información, aunque no siempre esté regulada por ley, sí afecta al derecho de las personas a tomar decisiones con todos los datos relevantes.

Otro aspecto importante es la desigualdad en el acceso a soluciones. No todo el mundo puede permitirse instalar sistemas de ventilación mecánica o contratar una medición profesional. Esto genera diferencias entre quienes tienen recursos para protegerse del radón y quienes, aun queriendo hacerlo, no pueden. En este sentido, sería necesario que desde las administraciones se ofreciera más apoyo a las personas que viven en zonas con mayor riesgo.

Después de realizar este trabajo, queda claro que el radón no es solo un problema técnico. También es un tema que tiene que ver con la forma en que se informa, se actúa y se prioriza la salud de las personas.

## ANEXO III: CÓDIGOS DE PYTHON

### Matriz de correlaciones

```
import pandas as pd
import seaborn as sns
import matplotlib
matplotlib.use("TkAgg")
import matplotlib.pyplot as plt

# Cargar los datos desde el archivo Excel
archivo = "datos.xlsx"
df = pd.read_excel(archivo, engine="openpyxl")

# Verificar qué tipo de datos tiene cada columna y ver si alguna columna no tiene
# el formato correcto
print("Tipos de datos en el DataFrame:")
print(df.dtypes)

# Convertir la columna "Hora" a formato numérico("Hora Decimal") para poder operar
# correctamente
if "Hora" in df.columns:
    df["Hora"] = pd.to_datetime(df["Hora"], format="%H:%M", errors='coerce')
    df["Hora Decimal"] = df["Hora"].dt.hour + df["Hora"].dt.minute / 60

# Convertir la columna "Mes" de texto a número para que el programa puede trabajar
# con ella
if "Mes" in df.columns:
    meses_dict = {
        "Enero": 1, "Febrero": 2, "Marzo": 3, "Abril": 4,
        "Mayo": 5, "Junio": 6, "Julio": 7, "Agosto": 8,
        "Septiembre": 9, "Octubre": 10, "Noviembre": 11, "Diciembre": 12
    }
    df["Mes"] = df["Mes"].map(meses_dict)

# Filtrar solo columnas numéricas para calcular la matriz
df = df.select_dtypes(include=['number'])

# Calcular la matriz de correlación y mostrarla en pantalla
correlacion = df.corr()

print("\nMatriz de Correlación:")
print(correlacion)

#Graficar la matriz de correlación
plt.figure(figsize=(16, 12))
sns.heatmap(correlacion, annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)
plt.title("Matriz de Correlación", fontsize=18)
plt.xticks(rotation=45, ha="right", fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```

## Modelo XGBoost

```
import pandas as pd
import numpy as np
from xgboost import XGBRegressor
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

# Cargar los datos desde el archivo excel
filePath = "datos.xlsx"
df = pd.read_excel(filePath, engine="openpyxl")

# Convertir la columna "Hora" a formato numérico ("Hora Decimal") para poder
trabajar correctamente con ella
if "Hora" in df.columns:
    df["Hora"] = pd.to_datetime(df["Hora"], format="%H:%M",
                                errors='coerce')
    df["Hora Decimal"] = df["Hora"].dt.hour + df[
        "Hora"].dt.minute / 60

# Variables seleccionadas para predecir tasa de exhalación
variables_finales = ["Concentración inicial", "Temperatura", "Fancoil", "Hora
Decimal"]

# Definir X (variables predictoras) e y (variable objetivo)
X = df[variables_finales]
y = df["Tasa exhalacion"]

# Dividir datos en entrenamiento y prueba; 80% entrenamiento y 20% prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Crear el modelo XGBoost para regresión, construyéndose 100 árboles
model = XGBRegressor(objective='reg:squarederror', n_estimators=100,
learning_rate=0.1, random_state=42)

# Entrenar el modelo usando los datos de entrenamiento
model.fit(X_train, y_train)

# Realizar predicciones sobre el conjunto de prueba
y_pred = model.predict(X_test)

# Evaluar el rendimiento del modelo
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

# Muestra con dos decimales las métricas de evaluación
print("\nEvaluación del Modelo:")
print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R²: {r2:.2f}")
```

## Modelo Random Forest

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from tkinter import messagebox
```

```
#Cargar los datos desde el archivo excel
filePath = "datos.xlsx"
df = pd.read_excel(filePath,engine="openpyxl")

# Convertir la columna "Hora" a formato numérico ("Hora Decimal")para poder
trabajar con ellos correctamente
if "Hora" in df.columns:
    df["Hora"] = pd.to_datetime(df["Hora"], format="%H:%M",
                                errors='coerce')
    df["Hora Decimal"] = df["Hora"].dt.hour + df[
        "Hora"].dt.minute / 60

# Variables seleccionadas para predecir tasa de exhalación
variables_finales = ["Concentración inicial", "Temperatura", "Fancoil", "Hora
Decimal"]

# Definir X (variables predictoras) e y (variable objetivo)
X = df[variables_finales]
y = df["Tasa exhalacion"]

# Crear el modelo Random Forest; se construyen 100 árboles
modelo_rf = RandomForestRegressor(n_estimators=100, random_state=42)

# Dividir datos en entrenamiento y prueba; 80% entrenamiento y 20% prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Entrenar el modelo usando los datos de entrenamiento
modelo_rf.fit(X_train, y_train)

# Realizar predicciones sobre el conjunto de prueba
y_pred = modelo_rf.predict(X_test)

# Evaluar el modelo
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

#Muestra con dos decimales las métricas de evaluación
print("\nEvaluación del Modelo:")
print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R²: {r2:.2f}")
```

## Modelo SVR

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn.preprocessing import StandardScaler
import matplotlib
matplotlib.use("TkAgg")
import matplotlib.pyplot as plt

# Cargar los datos del archivo excel
filePath = "datos.xlsx"
df = pd.read_excel(filePath,engine="openpyxl")

# Convertir la columna "Hora" a formato numérico ("Hora Decimal")para poder
trabajar con ellos correctamente
```

```

if "Hora" in df.columns:
    df["Hora"] = pd.to_datetime(df["Hora"], format="%H:%M",
                                errors='coerce')
    df["Hora Decimal"] = df["Hora"].dt.hour + df[
        "Hora"].dt.minute / 60

# Variables seleccionadas para predecir tasa de exhalación
variables_finales = ["Concentración inicial", "Temperatura", "Fancoil", "Hora
Decimal"]

# Definir X (variables predictoras) e y (variable objetivo)
X = df[variables_finales]
y = df["Tasa exhalacion"]

# Escalar los datos para evitar que las variables con rangos mayores dominen a
aquellas con rangos más pequeños
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Crear el modelo SVR con kernel RBF para capturar relaciones no lineales
modelo_svr = SVR(kernel='rbf')

# Dividir datos en entrenamiento y prueba; 80% entrenamiento y 20% prueba
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2,
random_state=42)

# Entrenar el modelo con el conjunto de entrenamiento
modelo_svr.fit(X_train, y_train)

# Realizar predicciones sobre el conjunto de prueba
y_pred = modelo_svr.predict(X_test)

# Calcular las métricas de evaluación
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

#Mostrar con dos decimales las métricas de evaluación
print("\nEvaluación del Modelo:")
print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R²: {r2:.2f}")

# Calcular los residuos para ver cómo se distribuyen los errores, y ver si el
modelo está ajustando correctamente la información
residuos = y_test - y_pred

#Realizar un gráfico con los residuos
plt.figure(figsize=(8,5))
plt.scatter(y_pred, residuos, alpha=0.6, color="blue", label="Residuos")
plt.axhline(y=0, color="red", linestyle="--", label="Línea base (y=0)")
plt.xlabel("Valores Predichos")
plt.ylabel("Residuos (Error)")
plt.legend()

# Mostrar el gráfico
plt.show()

# Calculo de la curva de aprendizaje para mostrar como varía el rendimiento a
medida que se incrementa el número de muestras.
train_sizes, train_scores, test_scores = learning_curve(
    modelo_svr,

```

```

X_scaled,
y=y,
cv=5,
scoring='neg_mean_absolute_error',
train_sizes=np.linspace(0.1, 1.0, 10)
n_jobs=-1

)

# Se convierte el error negativo a positivo
train_mae = -train_scores
test_mae = -test_scores

# Calcular la media de los MAE
train_mean = np.mean(train_mae, axis=1)
test_mean = np.mean(test_mae, axis=1)

# Graficar la curva de aprendizaje
plt.figure()
plt.plot(train_sizes, train_mean, 'o-', color="blue", label="Entrenamiento")
plt.plot(train_sizes, test_mean, 'o-', color="orange", label="Validación")
plt.xlabel("Número de muestras")
plt.ylabel("Error Absoluto Medio (MAE)")
plt.legend(loc= "best")

plt.grid()
plt.show()

```

## Modelo regresión lineal

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, learning_curve
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
import matplotlib
matplotlib.use("TkAgg")
import matplotlib.pyplot as plt

# Cargar los datos desde el archivo Excel
filePath = "datos.xlsx"
df = pd.read_excel(filePath, engine="openpyxl")

# Convertir la columna "Hora" a formato numérico ("Hora Decimal") para poder
trabajar con ellos correctamente
if "Hora" in df.columns:
    df["Hora"] = pd.to_datetime(df["Hora"], format="%H:%M", errors='coerce')
    df["Hora Decimal"] = df["Hora"].dt.hour + df["Hora"].dt.minute / 60

# Variables seleccionadas para predecir tasa de exhalación
variables_finales = ["Concentración inicial", "Temperatura", "Fancoil", "Hora
Decimal"]

# Definir X (variables predictoras) e y (variable objetivo)
X = df[variables_finales]
y = df["Tasa exhalacion"]

# Manejar valores faltantes en las variables predictoras usando imputación con la
media
imputer = SimpleImputer(strategy='mean')
X = pd.DataFrame(imputer.fit_transform(X), columns=X.columns)

```

```

# Escalar los datos para evitar que las variables con rangos mayores dominen a
aquellas con rangos más pequeños
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Definir el modelo de regresión lineal
modelo = LinearRegression()

# Dividir datos en entrenamiento y prueba; 80% entrenamiento y 20% prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Entrenar el modelo con el conjunto de entrenamiento
modelo.fit(X_train, y_train)

# Realizar predicciones sobre el conjunto de prueba
y_pred = modelo.predict(X_test)

# Calcular las métricas de evaluación
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print("\nEvaluación del Modelo:")
print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R²: {r2:.2f}")

# Calcular los residuos para ver cómo se distribuyen los errores, y ver si el
modelo está ajustando correctamente la información
residuos = y_test - y_pred

#Realizar un gráfico con los residuos
plt.figure(figsize=(8,5))
plt.scatter(y_pred, residuos, alpha=0.6, color="blue", label="Residuos")
plt.axhline(y=0, color="red", linestyle="--", label="Línea base (y=0)")
plt.xlabel("Valores Predichos")
plt.ylabel("Residuos (Error)")
plt.legend()
plt.show()

# Calculo de la curva de aprendizaje para mostrar como varía el rendimiento a
medida que se incrementa el número de muestras.
train_sizes, train_scores, test_scores = learning_curve(
    modelo,
    X_scaled,
    y=y,
    cv=5,
    scoring='neg_mean_absolute_error',
    train_sizes=np.linspace(0.1, 1.0, 10)
    n_jobs=-1
)

# Se convierte el error negativo a positivo
train_mae = -train_scores
test_mae = -test_scores

# Calcular la media de los MAE

```

```
train_mean = np.mean(train_mae, axis=1)
test_mean = np.mean(test_mae, axis=1)

# Graficar la curva de aprendizaje
plt.figure()
plt.plot(train_sizes, train_mean, 'o-', color="blue", label="Entrenamiento")
plt.plot(train_sizes, test_mean, 'o-', color="orange", label="Validación")
plt.xlabel("Número de muestras")
plt.ylabel("Error Absoluto Medio (MAE)")
plt.legend(loc= "best")

plt.grid()
plt.show()
```

## Modelo GBM

```
import pandas as pd
import numpy as np
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.model_selection import train_test_split, learning_curve
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import matplotlib
matplotlib.use("TkAgg")
import matplotlib.pyplot as plt

# Cargar los datos desde el archivo Excel
filePath = "datos.xlsx"
df = pd.read_excel(filePath, engine="openpyxl")

# Convertir la columna "Hora" a formato numérico ("Hora Decimal") para trabajar con
ella correctamente
if "Hora" in df.columns:
    df["Hora"] = pd.to_datetime(df["Hora"], format="%H:%M", errors='coerce')
    df["Hora Decimal"] = df["Hora"].dt.hour + df["Hora"].dt.minute / 60

# Variables seleccionadas para predecir tasa de exhalación
variables_finales = ["Concentración inicial", "Temperatura", "Fancoil", "Hora
Decimal"]

# Definir X (variables predictoras) e y (variable objetivo)
X = df[variables_finales]
y = df["Tasa exhalacion"]

# Definir el modelo Gradient Boosting; se van a construir 100 árboles, con una
velocidad de aprendizaje del 0.1, la profundidad máxima de cada árbol es 3
modelo = GradientBoostingRegressor(n_estimators=100, learning_rate=0.1,
max_depth=3, random_state=42)

# Dividir datos en entrenamiento y prueba; 80% entrenamiento y 20% prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Entrenar el modelo
modelo.fit(X_train, y_train)

# Hacer predicciones
y_pred = modelo.predict(X_test)

# Evaluar el modelo
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)
```

```

print("\nEvaluación del Modelo:")
print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R²: {r2:.2f}")

# Calcular los residuos para ver cómo se distribuyen los errores, y ver si el
# modelo está ajustando correctamente la información
residuos = y_test - y_pred

# Realizar un gráfico con los residuos
plt.figure(figsize=(8, 5))
plt.scatter(y_pred, residuos, alpha=0.6, color="blue", label="Residuos")
plt.axhline(y=0, color="red", linestyle="--", label="Línea base (y=0)")
plt.xlabel("Valores Predichos")
plt.ylabel("Residuos (Error)")
plt.legend()
plt.show()

# Calculo de la curva de aprendizaje para mostrar como varía el rendimiento a
# medida que se incrementa el número de muestras.
train_sizes, train_scores, test_scores = learning_curve(
    modelo,
    X,
    y=y,
    cv=5,
    scoring='neg_mean_absolute_error',
    train_sizes=np.linspace(0.1, 1.0, 10)
    n_jobs=-1
)

# Se convierte el error negativo a positivo
train_mae = -train_scores
test_mae = -test_scores

# Calcular la media de los MAE
train_mean = np.mean(train_mae, axis=1)
test_mean = np.mean(test_mae, axis=1)

# Graficar la curva de aprendizaje
plt.figure()
plt.plot(train_sizes, train_mean, 'o-', color="blue", label="Entrenamiento")
plt.plot(train_sizes, test_mean, 'o-', color="orange", label="Validación")
plt.xlabel("Número de muestras")
plt.ylabel("Error Absoluto Medio (MAE)")
plt.legend(loc="best")

plt.grid()
plt.show()

```

## Modelo MLP

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, learning_curve
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn.preprocessing import StandardScaler
import matplotlib
matplotlib.use("TkAgg")
import matplotlib.pyplot as plt
import warnings
from sklearn.exceptions import ConvergenceWarning

```

```
# Cargar los datos desde el archivo Excel
filePath = "datos.xlsx"
df = pd.read_excel(filePath, engine="openpyxl")

# Convertir la columna "Hora" a formato numérico ("Hora Decimal") para trabajar
# con ella correctamente
if "Hora" in df.columns:
    df["Hora"] = pd.to_datetime(df["Hora"], format="%H:%M", errors='coerce')
    df["Hora Decimal"] = df["Hora"].dt.hour + df["Hora"].dt.minute / 60

# Variables seleccionadas para predecir tasa de exhalación
variables_finales = ["Concentración inicial", "Temperatura", "Hora Decimal",
"Fancoil"]

# Definir X (variables predictoras) e y (variable objetivo)
X = df[variables_finales]
y = df["Tasa exhalacion"]

# Escalar los datos para evitar que las variables con rangos mayores dominen a
# aquellas con rangos más pequeños
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Dividir datos en entrenamiento y prueba; 80% entrenamiento y 20% prueba
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2,
random_state=42)

# Definir el modelo MLPRegressor; debido a que no tenemos muchos datos escogemos
# una única capa en lugar de dos ya que si no se tendríamos sobreajuste del modelo;
# tolerancia 1e-4 escogida para encontrar un buen equilibrio entre rapidez y
# precisión.
modelo = MLPRegressor(hidden_layer_sizes=(100,), activation='relu', solver='adam',
random_state=42, max_iter=5000, tol=1e-4)

# Entrenar el modelo
modelo.fit(X_train, y_train)

# Realizar predicciones sobre el conjunto de prueba
y_pred = modelo.predict(X_test)

# Evaluar el modelo
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print("\nEvaluación del Modelo MLP:")
print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R²: {r2:.2f}")

# Calcular los residuos para ver cómo se distribuyen los errores, y ver si el
# modelo está ajustando correctamente la información
residuos = y_test - y_pred

#Realizar un gráfico con los residuos
plt.figure(figsize=(8,5))
plt.scatter(y_pred, residuos, alpha=0.6, color="blue", label="Residuos")
plt.axhline(y=0, color="red", linestyle="--", label="Línea base (y=0)")
plt.xlabel("Valores Predichos")
plt.ylabel("Residuos (Error)")
plt.legend()
plt.show()
```

```

# Calculo de la curva de aprendizaje para mostrar como varía el rendimiento a
# medida que se incrementa el número de muestras.
train_sizes, train_scores, test_scores = learning_curve(
    modelo,
    X,
    y=y,
    cv=5,
    scoring='neg_mean_absolute_error',
    train_sizes=np.linspace(0.1, 1.0, 10)
    n_jobs=-1
)

# Se convierte el error negativo a positivo
train_mae = -train_scores
test_mae = -test_scores

# Calcular la media de los MAE
train_mean = np.mean(train_mae, axis=1)
test_mean = np.mean(test_mae, axis=1)

# Graficar la curva de aprendizaje
plt.figure()
plt.plot(train_sizes, train_mean, 'o-', color="blue", label="Entrenamiento")
plt.plot(train_sizes, test_mean, 'o-', color="orange", label="Validación")
plt.xlabel("Número de muestras")
plt.ylabel("Error Absoluto Medio (MAE)")
plt.legend(loc= "best")

plt.grid()
plt.show()

```