



**Escuela Superior
de Ingeniería y Tecnología**
Universidad de La Laguna

Trabajo de Fin de Grado

Predicción de series temporales mediante técnicas de aprendizaje profundo

Time series prediction using deep learning techniques

José Ramón Morera Campos

La Laguna, 13 de mayo de 2025

D. **Leopoldo Acosta Sánchez**, profesor Catedrático de Universidad adscrito al Departamento de Nombre del Departamento de la Universidad de La Laguna, como tutor

D. **Daniel Acosta Hernández**, profesor Titular de Universidad adscrito al Departamento de Nombre del Departamento de la Universidad de La Laguna, como cotutor

C E R T I F I C A N

Que la presente memoria titulada:

"Predicción de series temporales mediante técnicas de aprendizaje profundo"

ha sido realizada bajo su dirección por D. **José Ramón Morera Campos**.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 13 de mayo de 2025

Agradecimientos

XXXXX

Licencia



© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.

Resumen

XXXXX

Palabras clave: XXXXXX, XXXX, XXXX

Abstract

XXXXX

Keywords: XXXX, XXXX, XXXX

Índice general

1. Introducción	1
1.1. Motivación del proyecto	1
1.2. Planteamiento	1
1.3. Antecedentes y estado del arte	2
1.4. Objetivos	2
2. Adquisición y preprocesado de datos	4
2.1. Fuentes de los datos	4
2.2. Proceso de adquisición y almacenamiento	5
2.2.1. Flujos de adquisición de Grafcan	5
2.2.2. Flujos de adquisición de Open-Meteo	6
2.2.3. Almacenamiento	6
2.3. Preprocesado	7
2.3.1. Visualización	7
2.3.2. Manejo de datos faltantes	7
3. Modelos de predicción	10
3.1. Primera sección de este capítulo	10
3.2. Segundo apartado de este capítulo	10
3.3. Tercer apartado de este capítulo	10
4. Despliegue	11
5. Conclusiones y líneas futuras	12
6. Summary and Conclusions	13
7. Presupuesto	14
7.1. Sección Uno	14
A. Título del Apéndice 1	15
A.1. Algoritmo XXX	15
A.2. Algoritmo YYY	15
A.3. Algoritmo ZZZ	16
B. Título del Apéndice 2	17
B.1. Otro apéndice: Sección 1	17
B.2. Otro apéndice: Sección 2	17

Índice de Figuras

2.1. Mapa de las estaciones climatológicas Grafcan empleadas.	5
2.2. Flujo de adquisición de datos de Grafcan en node-red.	6
2.3. Flujo de adquisición de datos de Open-Meteo en node-red.	7
2.4. Visualización de la temperatura del aire en San Cristóbal de La Laguna. . .	8
2.5. Visualización de la temperatura del aire en San Cristóbal de La Laguna. . .	8
2.6. Visualización de la temperatura del aire en San Cristóbal de La Laguna. . .	9

Índice de Tablas

2.1. Datos faltantes por estación y fuente de datos (en horas)	9
7.1. Resumen de tipos	14

Capítulo 1

Introducción

1.1. Motivación del proyecto

La predicción de series temporales es un campo de estudio que ha cobrado gran relevancia en los últimos años, especialmente en el ámbito del aprendizaje automático y el aprendizaje profundo. La capacidad de anticipar eventos futuros a partir de datos históricos es fundamental en diversas áreas, como la economía, la meteorología, la salud y la ingeniería.

El campo de la meteorología, en particular, ha aumentado su relevancia en los últimos años debido al aumento de fenómenos climáticos extremos y su impacto en la sociedad. Prever con antelación la evolución de variables meteorológicas no solo permite planificar recursos, sino también contribuir a la prevención de desastres naturales.

En este contexto, el uso de redes neuronales ha demostrado ser una herramienta poderosa para abordar problemas complejos de predicción. Gracias a los avances en capacidad de cómputo y la disponibilidad de grandes volúmenes de datos, arquitecturas como LSTM, GRU y Transformers se han consolidado como soluciones de alto rendimiento. Estas redes no solo capturan patrones temporales de manera eficiente, sino que también pueden adaptarse dinámicamente a cambios en el comportamiento de la serie, mejorando la precisión y la robustez de las predicciones.

1.2. Planteamiento

En este trabajo se pretende emplear mediciones de múltiples estaciones meteorológicas de la isla de Tenerife con el fin de desarrollar un modelo de predicción climatológica a corto plazo. Se busca que dicho modelo sea capaz de generalizar más allá de las estaciones de entrenamiento. Esto es, que a partir de mediciones meteorológicas de cualquier origen, el modelo sea capaz de generar predicciones a corto plazo (3, 6 o 12 horas) de gran calidad. Se deciden considerar 3 variables meteorológicas: temperatura, humedad y presión atmosférica. La elección de estas variables se basa en su relevancia para la predicción del clima y la existencia de registros en las estaciones meteorológicas de Tenerife.

Se establece un especial énfasis en el tratamiento de los datos, estudiándose exhaustivamente diversas técnicas.

La idea principal pasa por la creación de ventanas de información, que son secuencias de datos consecutivos agrupados

Como parte del trabajo se contempla el despliegue de una infraestructura tecnológica

que permita la captura y el procesamiento de datos públicos, así como la creación de un modelo predictivo que extraiga información valiosa a partir de estos datos.

Así mismo, se plantea desplegar los modelos en un entorno de producción, permitiendo su uso en aplicaciones prácticas. De esta forma, cualquier usuario podrá suministrar mediciones de la variable que desee durante las últimas horas y obtener un pronóstico.

1.3. Antecedentes y estado del arte

El uso de redes neuronales y técnicas de deep learning (aprendizaje profundo) para la predicción de series temporales tiene sus orígenes en la década de 1940. Aunque el término deep learning ha ganado relevancia en los últimos años, los conceptos fundamentales y algunas de las técnicas más empleadas han existido desde hace mucho tiempo, evolucionando y perfeccionándose con el avance de la tecnología. En 1943, Warren McCulloch y Walter Pitts fueron pioneros en desarrollar un modelo computacional para redes neuronales [1], estableciendo las bases de lo que hoy conocemos como inteligencia artificial basada en redes neuronales. Más adelante, en 1949, Donald Hebb introdujo la famosa Teoría Hebbiana, una hipótesis que proponía un mecanismo de aprendizaje basado en la plasticidad neuronal [2]. Este principio se aplicó a modelos computacionales, impulsando la investigación en la simulación del aprendizaje humano.

En 1958, Frank Rosenblatt avanzó en este campo con la creación del perceptrón [3], un algoritmo pionero de reconocimiento de patrones basado en una red de aprendizaje de computadora de dos capas. Este modelo simple operaba mediante operaciones de adición y sustracción, sentando las bases para el desarrollo de redes neuronales más complejas. Durante las siguientes décadas, la investigación en redes neuronales experimentó un estancamiento debido a limitaciones técnicas y a la falta de métodos eficaces para el entrenamiento de redes con múltiples capas. Sin embargo, en 1982, un gran avance revitalizó este campo: la introducción del algoritmo de propagación hacia atrás (backpropagation) [4], que resolvía el problema del entrenamiento eficiente de redes neuronales profundas, permitiendo la formación de redes multicapa de manera más rápida y eficaz. Desde entonces, las redes neuronales han continuado evolucionando, impulsadas en gran parte por los avances en la potencia de cálculo de las GPU y el desarrollo del deep learning. Hoy en día, esta tecnología es fundamental y revolucionaria, capaz de realizar predicciones de series temporales con notable rapidez y precisión. La investigación en este ámbito sigue siendo prolífica. Una búsqueda rápida en internet revela la continua producción de proyectos y estudios enfocados en perfeccionar las técnicas de predicción temporal mediante aprendizaje profundo. Estos trabajos se caracterizan por un análisis exhaustivo y la exploración de nuevos métodos, todos con el objetivo de optimizar la precisión y la eficacia de las previsiones en aplicaciones prácticas.

1.4. Objetivos

1. Evaluar y seleccionar una fuente de datos.
2. Evaluar y seleccionar el framework de aprendizaje profundo a usar en el proyecto.
3. Diseño e implementación de la metodología de procesamiento de los datos.
4. Evaluación y selección de arquitecturas pertinentes al problema.

5. Implementación de una arquitectura adecuada.
6. Evaluación del rendimiento de la solución propuesta.
7. Comparación frente a metodologías tradicionales.
8. Despliegue de la solución en un entorno de producción.

Capítulo 2

Adquisición y preprocesado de datos

Se desea trabajar con series temporales sobre mediciones climatológicas. En concreto, se eligen las variables de temperatura del aire, humedad relativa y presión atmosférica en la superficie. Dichas variables son estudiadas con frecuencia horaria, en el intervalo comprendido entre el 1 de marzo de 2023 y el 28 de febrero de 2025.

Es relevante el uso de datos en un período múltiplo del año, para asegurar que se capturan las variaciones estacionales y que el conjunto de datos está suficientemente equilibrado.

2.1. Fuentes de los datos

Se han empleado 2 fuentes para recopilar las mediciones:

- **Grafcan:** Cartográfica de Canarias, S.A. es una empresa pública de la Comunidad Autónoma de Canarias. Dispone de una red de estaciones meteorológicas cuyas mediciones son accesibles mediante una API REST de acceso gratuito previa solicitud de una clave[5].
- **Open-Meteo:** API pública de código abierto que proporciona datos de múltiples proveedores de meteorología. Este servicio no dispone de estaciones de medición propias, sino que recopila pronósticos de diferentes modelos de predicción climatológica.

Se emplea la API de predicciones pasadas[6]. Se seleccionan los modelos ICON Global del servicio meteorológico alemán (DWD) y el modelo ARPEGE Europe de Météo-France. Ambos modelos se actualizan cada 3 horas. Se explora la posibilidad de emplear las predicciones del modelo HAROME de la AEMET, pero no están disponibles de forma pública.

Se eligen 4 ubicaciones de la isla de Tenerife con distintas características climáticas para el conjunto de entrenamiento y evaluación:

- **San Cristóbal de La Laguna 1:** La Cuesta, 35 metros de altitud.
- **San Cristóbal de La Laguna 2:** La Punta del Hidalgo, 54m.
- **La Orotava:** Camino de Chasna, 812m.
- **Arona:** Punta de Rasca, 25m.

Así mismo, se escogen 2 ubicaciones para el conjunto de test, nunca vistas en el ajuste del modelo:

- **Garachico:** La Cuesta, 35 metros de altitud.
- **Santa Cruz de Tenerife:** Camino de Chasna, 812m.

Las ubicaciones han sido elegidas al contar con estaciones de medición de Grafcan. Sus posiciones se muestran en la Figura 2.1, señaladas en rojo. De izquierda a derecha: Arona, La Orotava y San Cristóbal de La Laguna.

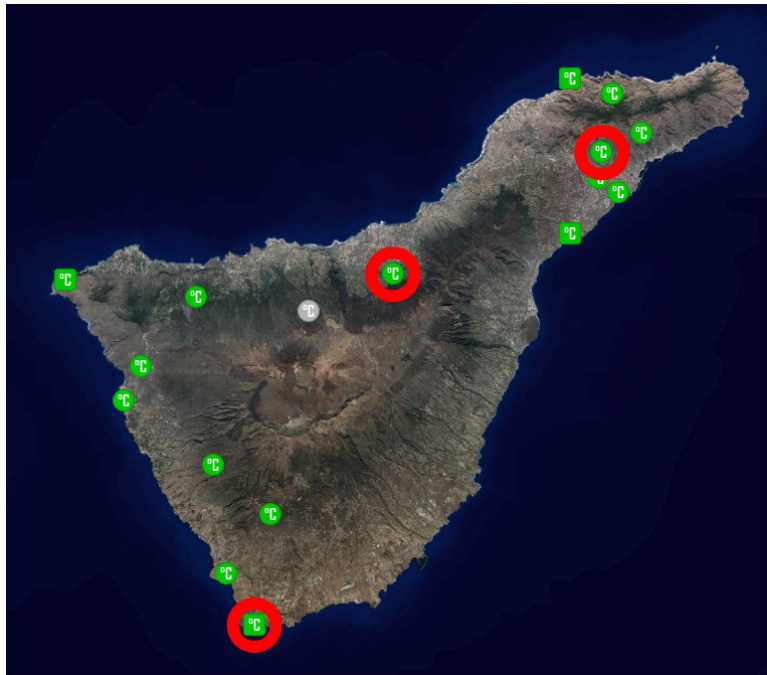


Figura 2.1: Mapa de las estaciones climatológicas Grafcan empleadas.

Inicialmente se valoró emplear las estaciones correspondientes a Los Cristianos, Santiago del Teide o la Punta de Teno, pero fueron descartadas por dos motivos: se detectó que existían períodos prolongados con datos faltantes en las mediciones de Grafcan. Algunas de ellas también exhibían poca correlación entre las mediciones del servicio Grafcan y las de Open-Meteo, lo que podría afectar la calidad de los datos.

2.2. Proceso de adquisición y almacenamiento

Para automatizar la adquisición de datos, se emplea la herramienta de orquestación node-red, que permite crear flujos de información mediante nodos que realizan tareas específicas o ejecutan código de JavaScript. En dicha herramienta se desarrollan dos paneles, uno para cada fuente de datos. Así mismo, dentro de cada panel se desarrollan dos flujos, uno para la adquisición de datos en un intervalo dado, y otro para la adquisición de datos en tiempo real, en particular, se establece la recogida de datos cada 6 horas.

2.2.1. Flujos de adquisición de Grafcan

Debido al funcionamiento de la API de Grafcan, se debe realizar una llamada para obtener la serie temporal de cada variable meteorológica de cada estación. Posteriormente,

se unen las series de cada estación en una única serie, que se almacena en una base de datos PostgreSQL. Este flujo está reflejado en la figura 2.2.

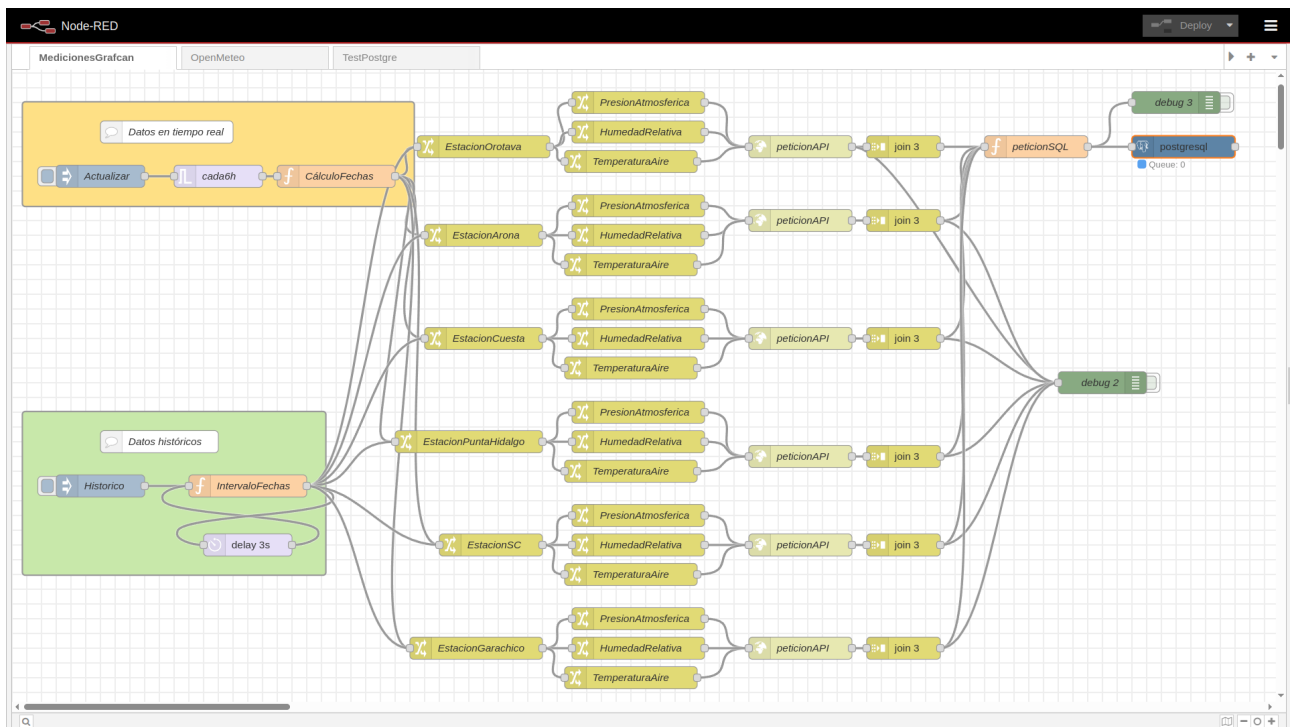


Figura 2.2: Flujo de adquisición de datos de Grafcan en node-red.

Las mediciones de Grafcan se recogen aproximadamente cada 10 minutos, si bien la frecuencia no es consistente y en ocasiones es mayor. Para manejar esta variabilidad, en este nivel los datos se agregan cada 10 minutos, usando la media de las mediciones del intervalo.

2.2.2. Flujos de adquisición de Open-Meteo

Existe una rama para obtener los datos del modelo ICON y otra para el modelo ARPEGE. Se establecen las coordenadas de cada ubicación como las de la estación de Grafcan seleccionada y se realiza una llamada a la API por cada localización y cada modelo, como se observa en la figura 2.3. Los resultados se almacenan en una base de datos PostgreSQL.

2.2.3. Almacenamiento

Se estudian distintas alternativas para el almacenamiento de los datos. Se opta por emplear TimescaleDB, una extensión del popular sistema PostgreSQL de bases de datos relacionales, especialmente adaptada para el manejo de series temporales.

Se establece un servidor TimescaleDB en un contenedor Docker. Se configura una tabla para cada estación y cada fuente: Grafcan, Open-Meteo ICON y Open-Meteo ARPEGE. Cada tabla emplea como índice y clave primaria la fecha y hora de la medición, así como su zona horaria. Las otras columnas se corresponden a la temperatura media del aire en grados Celsius, la humedad relativa en porcentaje y la presión atmosférica en superficie medida en hPa.

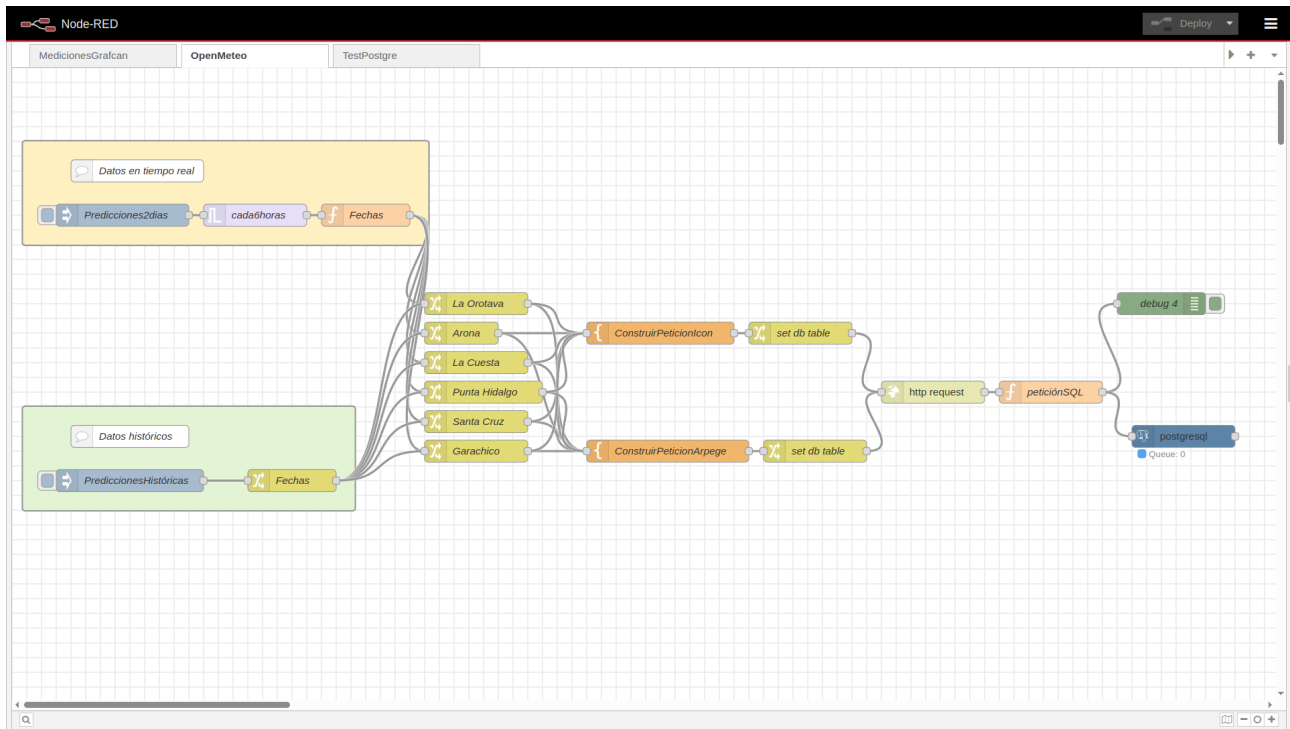


Figura 2.3: Flujo de adquisición de datos de Open-Meteo en node-red.

Es importante señalar que las mediciones de Grafcan y Open-Meteo codifican las horas en UTC, en vez de la hora local, puesto que UTC es independiente a los cambios de horario y de esta forma se mantiene la consistencia de los datos.

2.3. Preprocesado

Se desarrolla un cuaderno de Jupyter para realizar el preprocesado de los datos. Este proceso se aplica para cada estación.

En primer lugar, se obtienen las series temporales de las 3 fuentes, Grafcan y los dos modelos de Open-Meteo, para el período entre el 1 de marzo de 2023 y el 28 de febrero de 2025. Se agrgan los datos con frecuencia horaria mediante la media.

2.3.1. Visualización

Se visualizan los datos de cada variable para cada año. Podemos ver ejemplos en las figuras 2.4, 2.5 y 2.6.

Cabe destacar que en la variable de humedad relativa se observa una gran variabilidad entre los datos de las distintas fuentes, lo que puede ser ??????

2.3.2. Manejo de datos faltantes

Se detectan los datos faltantes para cada fuente. Las estadísticas se muestran en la tabla 2.1.

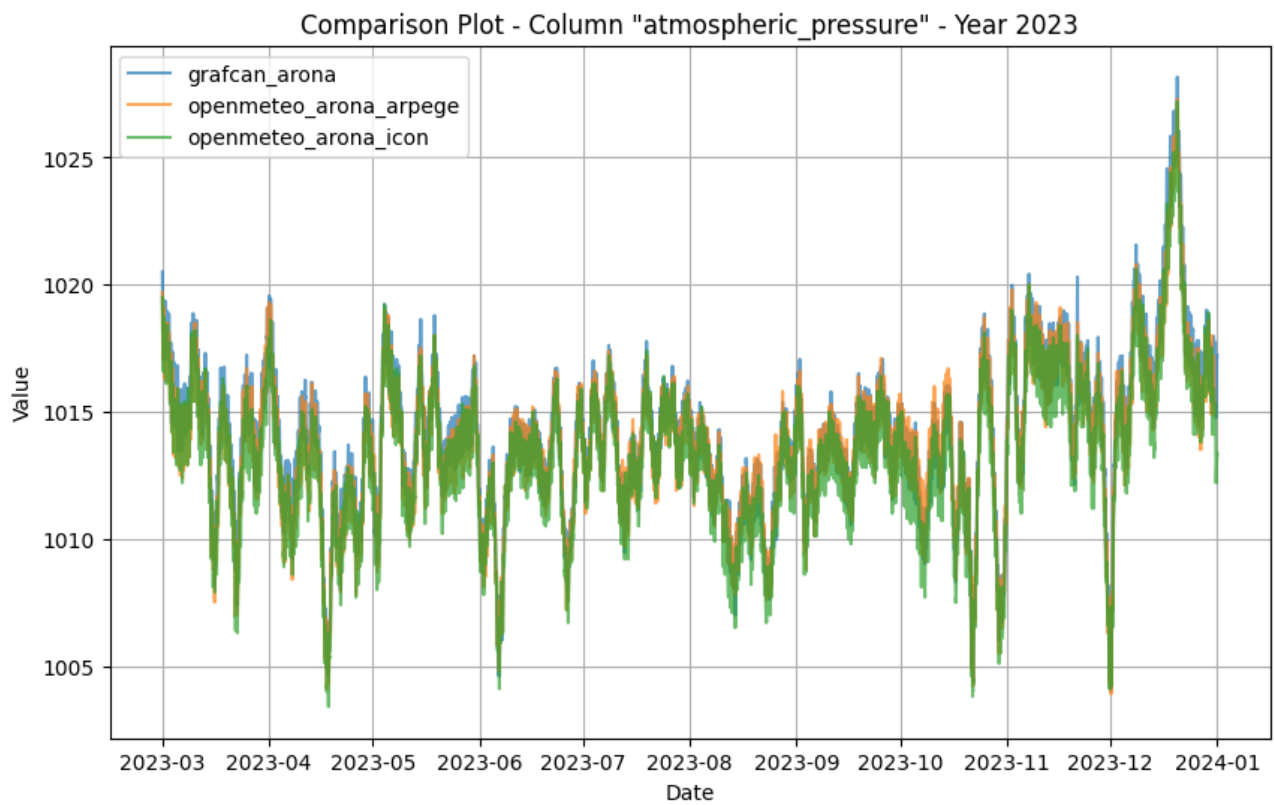


Figura 2.4: Visualización de la temperatura del aire en San Cristóbal de La Laguna.

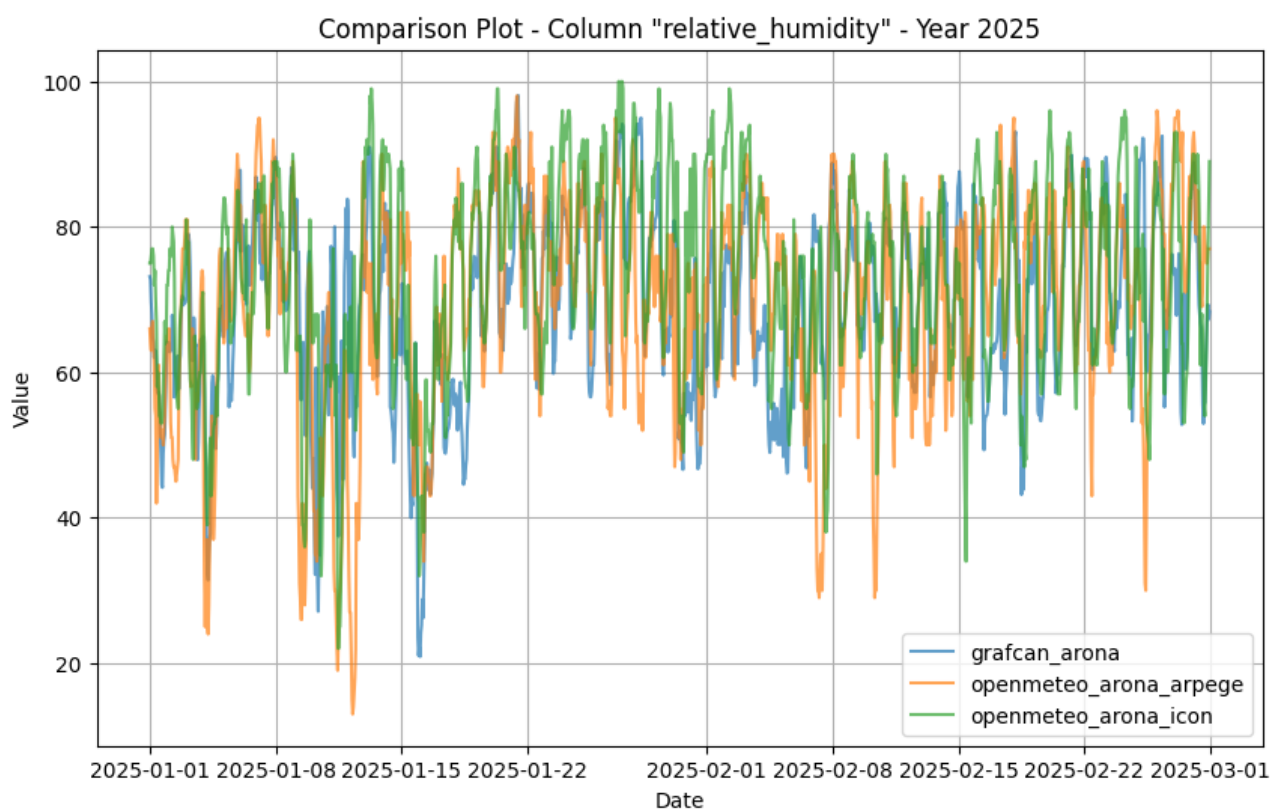


Figura 2.5: Visualización de la temperatura del aire en San Cristóbal de La Laguna.

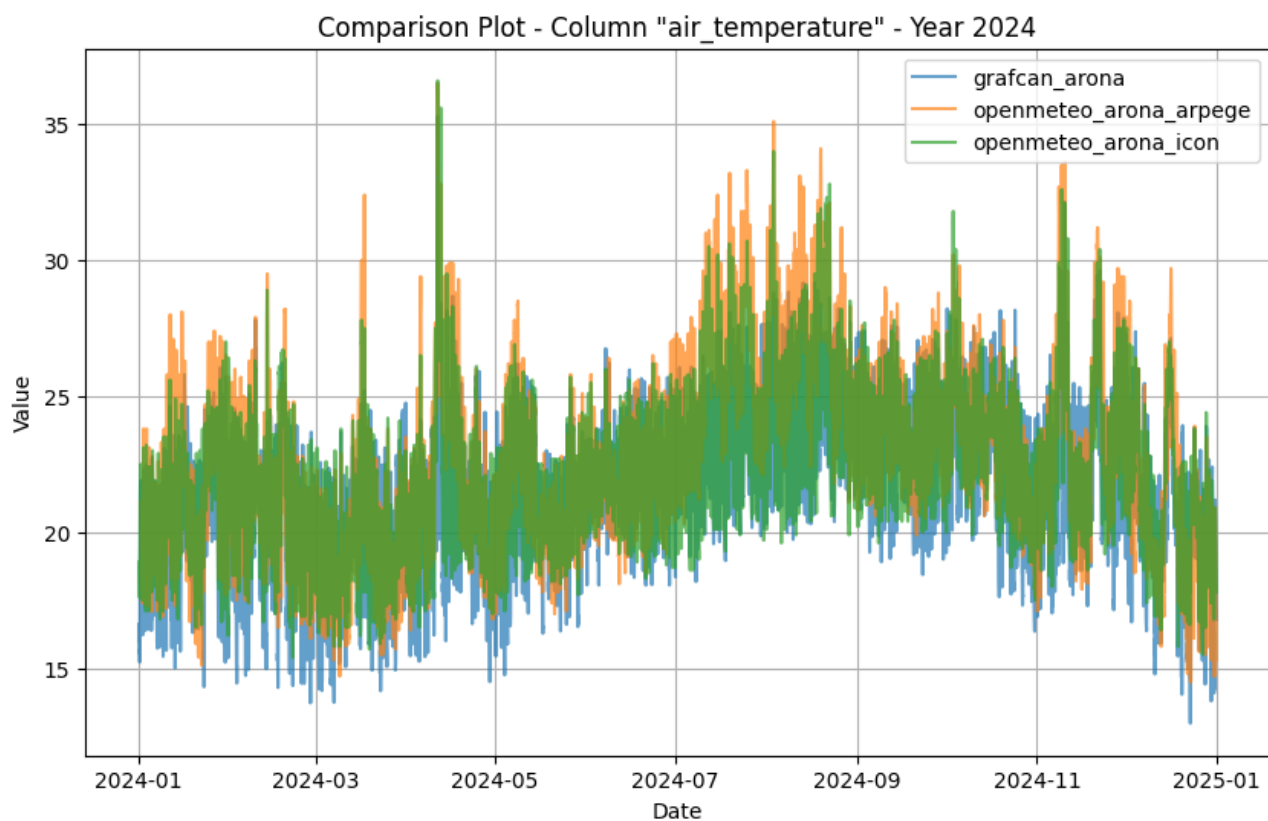


Figura 2.6: Visualización de la temperatura del aire en San Cristóbal de La Laguna.

Estación	Grafcan	Open-Meteo ICON	Open-Meteo ARPEGE
La Cuesta	0.2 %	0.1 %	0.1 %
La Punta del Hidalgo	0.3 %	0.2 %	0.2 %
Camino de Chasna	0.4 %	0.3 %	0.3 %
Punta de Rasca	0.5 %	0.4 %	0.4 %

Tabla 2.1: Datos faltantes por estación y fuente de datos (en horas)

Capítulo 3

Modelos de predicción

Los capítulos intermedios servirán para cubrir los siguientes aspectos: antecedentes, problemática o estado del arte, objetivos, fases y desarrollo del proyecto.

Bla, Bla, Bla,

3.1. Primera sección de este capítulo

3.2. Segundo apartado de este capítulo

3.3. Tercer apartado de este capítulo

Capítulo 4

Despliegue

Los capítulos intermedios servirán para cubrir los siguientes aspectos: antecedentes, problemática o estado del arte, objetivos, fases y desarrollo del proyecto.

El capítulo 1 se describió bla, bla, bla . . .

Capítulo 5

Conclusiones y líneas futuras

Este capítulo es obligatorio. Toda memoria de Trabajo de Fin de Grado debe incluir unas conclusiones y unas líneas de trabajo futuro

Capítulo 6

Summary and Conclusions

This chapter is compulsory. The memory should include an extended summary and conclusions in english.

Capítulo 7

Presupuesto

Este capítulo es obligatorio. Toda memoria de Trabajo de Fin de Grado debe incluir un presupuesto.

7.1. Sección Uno

Tipos	Descripción
AAAA	BBBB
CCCC	DDDD
EEEE	FFFF
GGGG	HHHH

Tabla 7.1: Resumen de tipos

Apéndice A

Título del Apéndice 1

A.1. Algoritmo XXX

```
/******  
*  
* Fichero .h  
*  
*****  
*  
* AUTORES  
*  
*  
* FECHA  
*  
*  
* DESCRIPCION  
*  
*  
*****/
```

A.2. Algoritmo YYY

```
/******  
*  
* Fichero .h  
*  
*****  
*  
* AUTORES  
*  
* FECHA  
*  
*  
* DESCRIPCION  
*  
*  
*****/
```


A.3. Algoritmo ZZZ

```

/*****
 *
 * Fichero .h
 *
 *****/
 *
 * AUTORES
 *
 * FECHA
 *
 * DESCRIPCION
 *
 *****/

```

Apéndice B

Título del Apéndice 2

B.1. Otro apéndice: Sección 1

texto

B.2. Otro apéndice: Sección 2

texto

Bibliografía

- [1] McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, 5(4), 115–133. <https://doi.org/10.1007/BF02478259>
- [2] Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.
- [3] Rosenblatt, F. (1958). *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- [4] Werbos, P. J. (1982). *Applications of advances in nonlinear sensitivity analysis*. In R. F. Drenick & F. Kozin (Eds.), *System Modeling and Optimization* (Vol. 38, pp. 762–770). Springer-Verlag. <https://doi.org/10.1007/BFb0006203>
- [5] Cartográfica de Canarias, S.A. *Sistema de Observación Meteorológica de Canarias* [Sitio web]. Recuperado el 12 de mayo de 2025, de <https://sensores.grafcan.es/>
- [6] Open-Meteo *Sistema de Observación Meteorológica de Canarias* [Sitio web]. Recuperado el 12 de mayo de 2025, de <https://open-meteo.com/en/docs/historical-forecast-api>