



**UNIVERSIDAD  
DE GRANADA**

**TRABAJO FIN DE GRADO**  
**INGENIERÍA DE TECNOLOGÍAS DE TELECOMUNICACIÓN**

# **Estudio de Resonancias de Schumann a través de la Inteligencia Artificial supervisada**

**Realizado por:**

Ángel Bermúdez Caballero

**Dirigido por:**

D. Jesús Francisco Fornieles Callejón

D. Franciso Jesús Lara Abelenda

**Departamento de:**

Electromagnetismo y Física de la Materia



Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación

—  
Granada, noviembre de 2024

# **Estudio de Resonancias de Schumann a través de la Inteligencia Artificial supervisada**

Ángel Bermúdez Caballero

**Palabras clave:** Resonancias de Schumann, inteligencia artificial, frecuencias ELF, tormentas geomagnéticas, aprendizaje supervisado, cavidad Tierra-Ionosfera, magnetómetros, cambio climático, ondas electromagnéticas, índices K, SVR, XGBoost Regressor, Random Forest Regressor.

**Resumen:** En este trabajo se lleva a cabo un estudio exhaustivo sobre las Resonancias de Schumann y su relación con las tormentas geomagnéticas, empleando técnicas de inteligencia artificial supervisada. Se comienza estableciendo el contexto teórico sobre las Resonancias de Schumann, incluyendo su importancia y los aspectos históricos que justifican el estudio. A continuación, se desarrollan los conceptos teóricos necesarios, como la descripción de la estación de captación, el análisis de Fourier aplicado a las frecuencias ELF y el modelo teórico de la cavidad Tierra-Ionosfera, con un enfoque específico en la representación de los campos eléctricos y magnéticos. Luego se introduce el papel de la inteligencia artificial en el estudio de estos fenómenos, describiendo las principales técnicas y algoritmos de Machine Learning utilizados, así como su aplicación específica para analizar los datos de resonancias. En la metodología, se detallan los pasos seguidos para la recolección y el preprocesamiento de los datos, así como los enfoques utilizados para entrenar y validar los modelos. Posteriormente, se presentan y discuten los resultados obtenidos, destacando las correlaciones encontradas entre las resonancias y la actividad geomagnética. Finalmente, se concluye analizando el impacto de los resultados en el campo del monitoreo de la actividad geomagnética y proponiendo líneas futuras de investigación para la mejora de los modelos desarrollados.

# Study of Schumann Resonances through Supervised Artificial Intelligence

Ángel Bermúdez Caballero

**Keywords:** Schumann resonances, artificial intelligence, ELF frequencies, geomagnetic storms, unsupervised learning, Earth-Ionosphere cavity, magnetometers, climate change, electromagnetic waves, K indexes, SVR, XGBoost Regressor, Random Forest Regressor.

**Summary:** This work conducts a comprehensive study on Schumann resonances and their relationship with geomagnetic storms, using supervised artificial intelligence techniques. It begins by establishing the theoretical context of Schumann resonances, including their importance and the historical aspects that justify the study. Subsequently, the necessary theoretical concepts are developed, such as the description of the measurement station, the Fourier analysis applied to ELF frequencies, and the theoretical model of the Earth-Ionosphere cavity, with a specific focus on the representation of electric and magnetic fields. Then, the role of artificial intelligence in the study of these phenomena is introduced, describing the main machine learning techniques and algorithms used, as well as their specific application to analyze resonance data. In the methodology, the steps followed for data collection and preprocessing are detailed, as well as the approaches used to train and validate the models. Subsequently, the results obtained are presented and discussed, highlighting the correlations found between the resonances and geomagnetic activity. Finally, the conclusion analyzes the impact of the results on the field of geomagnetic activity monitoring and proposes future lines of research for improving the developed models.

# Índice

<b>1.Introducción</b>	<b>5</b>
1.1. Resonancias de Schumann	5
1.2. Importancia de las frecuencias ELF	12
1.3. Justificación histórica	14
1.4. Objetivos del proyecto	15
<b>2. Desarrollo Teórico</b>	<b>15</b>
2.1. Descripción de la estación	15
2.2. Análisis de Fourier en el procesamiento de señales ELF	19
2.3. Modelo Teórico de la cavidad Tierra-Ionosfera	21
2.4. Representación de los campos eléctricos y magnéticos	22
2.5. Inteligencia artificial	24
2.5.1. Procesamiento del Lenguaje natural	25
2.5.2. Machine Learning	29
2.5.3. Algoritmos y métricas de evaluación	32
<b>3. Metodología</b>	<b>43</b>
<b>4. Resultados</b>	<b>53</b>
<b>5. Conclusiones</b>	<b>65</b>
<b>6. Bibliografía</b>	<b>69</b>

# 1.Introducción

En este TFG se realiza un análisis de las Resonancias de Schumann empleando técnicas de inteligencia artificial supervisada para estudiar su relación con las tormentas geomagnéticas. Las Resonancias de Schumann son ondas electromagnéticas de baja frecuencia generadas en la cavidad entre la Tierra y la ionosfera, un fenómeno que actúa como un indicador de la actividad eléctrica global y que ha cobrado relevancia en el monitoreo de cambios climáticos. Las señales utilizadas en este estudio fueron captadas por magnetómetros ubicados en Sierra Nevada, Granada, orientados en direcciones norte-sur y este-oeste. Este trabajo abarca la recolección y preprocesamiento de datos de estas señales, analizando sus frecuencias ELF (frecuencias extremadamente bajas) en el rango de 3-30 Hz, donde se encuentran los modos principales de las resonancias.

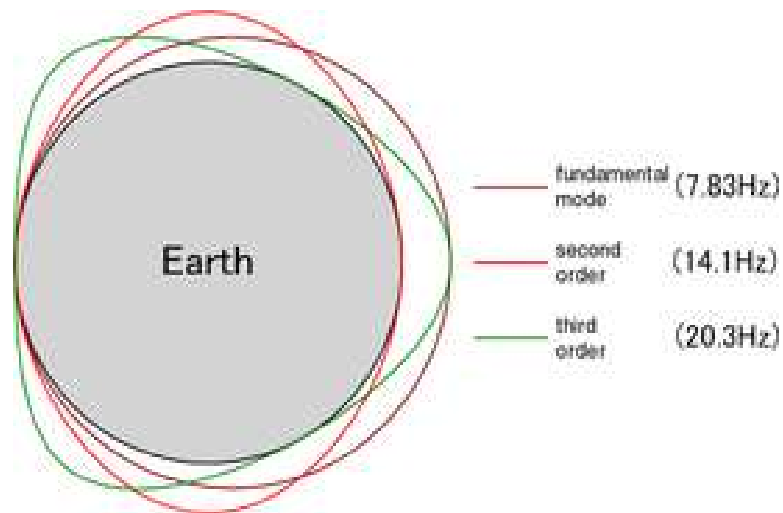
Para el análisis, se emplearon algoritmos de aprendizaje automático supervisado, como el Support Vector Regression (SVR), XGBoost y Random Forest, seleccionados por su efectividad en la identificación de patrones en datos. Estos modelos ayudan a estimar la intensidad de las tormentas geomagnéticas y a detectar correlaciones significativas entre la actividad geomagnética y las Resonancias de Schumann, lo cual resulta de gran valor en la investigación de los efectos de las tormentas solares en el campo magnético terrestre. Este estudio no solo profundiza en la aplicabilidad de las frecuencias ELF en la detección geofísica, sino que también abre nuevas posibilidades en el uso de inteligencia artificial para el análisis de fenómenos naturales de largo alcance.

## 1.1. Resonancias de Schumann

Se tiene constancia de que en 1952 el físico alemán Winfried Otto Schumann, de una manera teórica consiguió esclarecer las condiciones por las cuales en la cavidad entre la tierra y la ionosfera se podrían generar ondas estacionarias de baja frecuencia, esto se debe a que son un fenómeno electromagnético [1].

### Aspectos físicos básicos

Indagando un poco más en este campo de las resonancias, se debería saber una serie de puntos básicos sobre las mismas. Las Resonancias de Schumann como bien se ha dicho antes, son unas ondas electromagnéticas de baja frecuencia que viajan y rebotan en la cavidad entre la Tierra y la ionosfera, sin embargo, se debe tener en cuenta que la ionosfera actúa como una capa conductora, es decir, en términos simples la Tierra podría ser como una esfera conductora rodeada por un dieléctrico que en nuestro caso es el aire y que está aislada o recubierta por esta capa conductora. Otra de sus características es que actúa a frecuencias extremadamente bajas entre 7 y 60 Hz. En total, se conocen ocho modos principales de resonancia dentro de este intervalo donde el más predominante y estable tiene una frecuencia de 7,83 Hz aproximadamente, al que le siguen otros de frecuencias en torno a 14,1 y 20,3 Hz, son los que se van a estudiar en este estudio en mayor medida.



*Ilustración 1. Esfera terrestre y los tres primeros modos principales.*

### Cálculo de los modos de resonancia

Una manera más analítica y enfocada al estudio matemático de las Resonancias de Schumann parte con la base de un radio de la Tierra aproximadamente de 6.400 km y un espesor de la atmósfera que se entiende como la distancia entre la superficie terrestre y la ionosfera que oscila entre 60 y 90 km. Estas características geométricas no sólo definen las propiedades de la propia cavidad, sino que también son parte de la resolución de la ecuación de onda de la cavidad esférica. Otro concepto básico que no se puede pasar por alto es la resonancia fundamental que ocurre cuando la longitud de onda de la señal electromagnética es aproximadamente la circunferencia de la Tierra lo que resulta como el primer modo con una frecuencia de 7,83 Hz, que a través de la cual, se considera la conclusión de que a medida que aumentan las frecuencias, los modos superiores corresponden a fracciones más pequeñas de la circunferencia terrestre. Todos estos cálculos se pueden comprobar y calcular a través de la fórmula general que se puede observar a continuación:

$$f_n = \frac{n \cdot c}{2 \cdot \pi \cdot R}$$

Donde

$f_n$  es la frecuencia del modo  $n$

$c$  es la velocidad de la luz en el aire (aproximadamente  $3 \cdot 10^8$  m/s)

$R$  es el radio de la Tierra

$n$  es el número de modo, un entero positivo.

### Importancia del fenómeno

Después de haber indagado en la parte matemática y física de estas resonancias ahora se procede con la importancia de este fenómeno. Se debe tener en cuenta que actúan como un indicador de la actividad eléctrica global de nuestro planeta. El motivo es que los

rayos actúan como excitadores y esta excitación en forma de descargas eléctricas generadas por las tormentas, hacen que se puedan monitorear la distribución y frecuencia de las tormentas a lo ancho y largo de nuestro planeta haciendo hincapié en las regiones donde las tormentas ocurren más a menudo o con mayor intensidad.

### Propiedades de la propagación de ondas ELF

A continuación, se estudian las características de propagación únicas que tienen estas ondas electromagnéticas de frecuencia extremadamente baja. Estas ondas tienen la capacidad de penetrar en materiales densos lo que hace que sean útiles para aplicaciones tecnológicas en entornos submarinos y subterráneos. Hemos clasificado en diferentes tipos a estas ondas dependiendo de sus condiciones de propagación.

#### Propagación cuasi-libre

Se da cuando la longitud de onda es mucho menor que la circunferencia de la Tierra ( $2\pi r \gg \lambda$ ), en este caso las ondas viajan a través de la cavidad con mínimas pérdidas de energía.

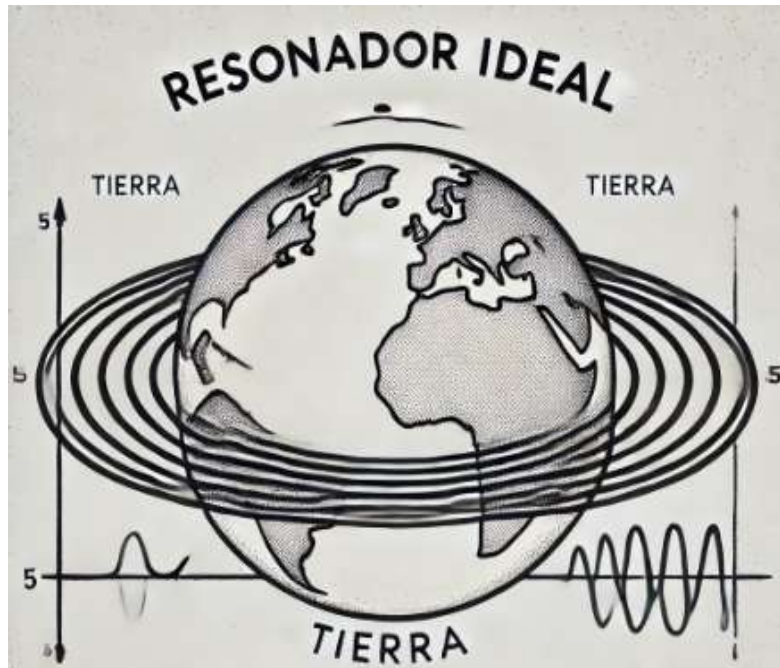


*Ilustración 2. Esquema de propagación cuasi-libre.*

#### Propagación resonador ideal

Aquí es cuando la longitud de onda es igual o muy próxima a la circunferencia de la Tierra ( $2\pi r \approx \lambda$ ), en este caso como bien dice su nombre, la cavidad actúa como un resonador ideal amplificando las señales electromagnéticas.

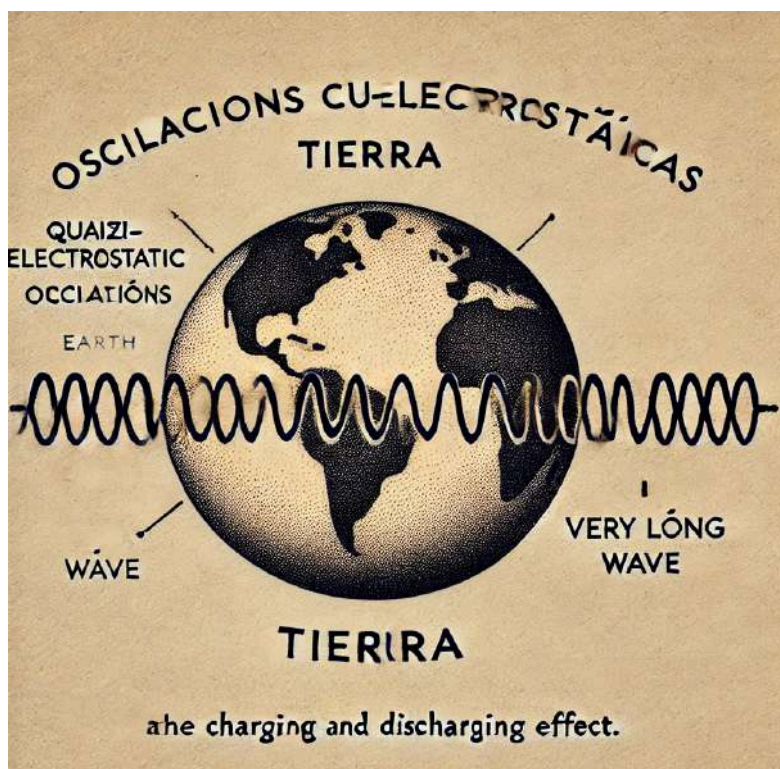




*Ilustración 3. Esquema de resonador ideal.*

Propagación Cuasi-electroestática:

Por último, se encuentran las oscilaciones cuasi-electroestáticas que se dan cuando la longitud de onda es mucho mayor a la circunferencia de la Tierra ( $2\pi r \ll \lambda$ ), en este caso la actividad actúa como un condensador donde se cargan y descargan.

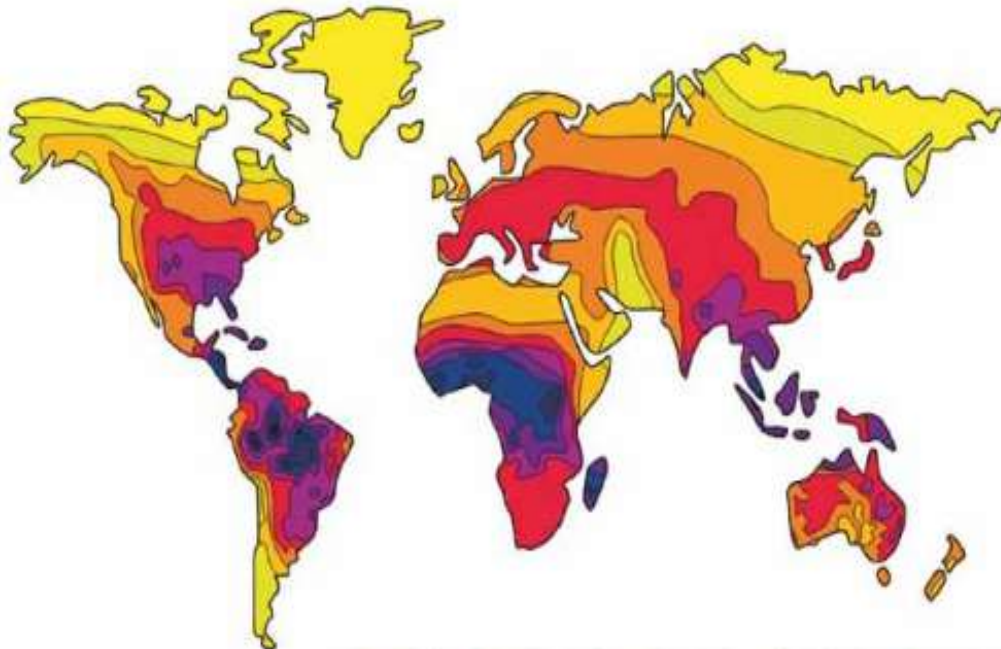


*Ilustración 4. Esquema de Propagación Cuasi-electroestática*



## Factores que afectan a las Resonancias de Schumann

En el siguiente apartado se comenta qué factores pueden influir en la frecuencia y amplitud de las Resonancias de Schumann ya que estas no son constantes. La actividad eléctrica atmosférica produce la excitación de las resonancias que están a la orden del día debido a las tormentas eléctricas por la generación de rayos que actúan como excitación. Esto se puede observar en regiones como el Sudeste Asiático, África y América del Sur o por ejemplo durante los monzones que son vientos que producen lluvias torrenciales y fuente de fuertes inundaciones. En cuanto a las tormentas solares, se deben a estas perturbaciones en el campo magnético terrestre, las propiedades de la ionosfera son alteradas y en consecuencia también se ven alteradas los modos de resonancia. Por último, las variaciones estacionales, cuyos cambios en las condiciones climáticas y la posición del sol respecto a la tierra también influyen [2].



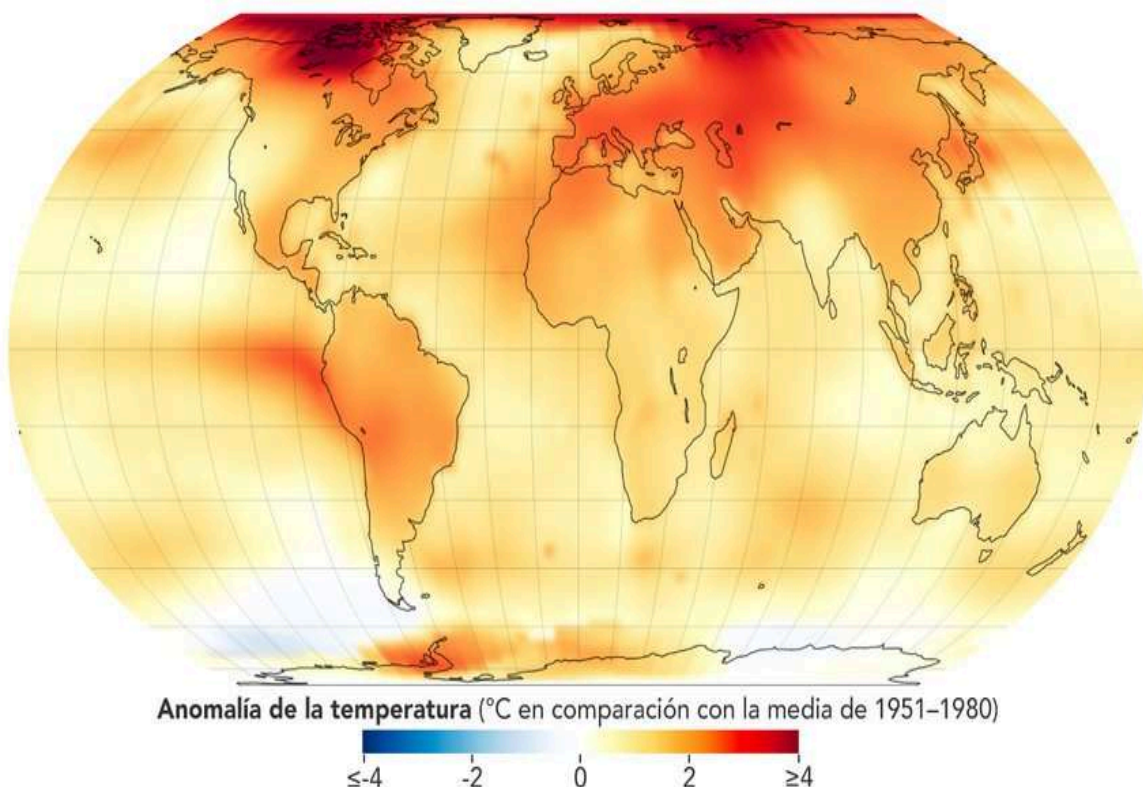
*Ilustración 5. Mapa de la actividad eléctrica global.*

## Resonancias de Schumann como indicador del cambio climático global

En el siguiente apartado miraremos un poco más al presente y futuro ya que se comenta como las Resonancias de Schumann han empezado a llamar la atención por su capacidad para monitorear el estado del planeta. Como se ha mencionado anteriormente, el hecho de haberse dado cuenta que al estar estas resonancias muy ligadas a los cambios en las condiciones atmosféricas y de la ionosfera hace que sean un claro indicador del cambio climático. El cambio climático ha modificado nuestro planeta y lo sigue haciendo, entre otras cosas debido al aumento de la temperatura global. El calentamiento tiene relación y puede modificar la frecuencia e intensidad de las tormentas en regiones tropicales, por lo que puede ser una variación en nuestras resonancias al igual que la ubicación de la actividad de

tormentas eléctricas dado que el cambio climático también está variando estos patrones de lluvia.

Otro factor clave es la conductividad de la ionosfera que cambia dependiendo de la radiación solar y otros factores climáticos. La relación de la temperatura global y la ionosfera hace que, debido al calentamiento global, haya un aumento de los gases efecto invernadero o un incremento de la evaporación del agua lo que hace que haya más cantidad de vapor de agua en la atmósfera lo cual a su vez afecta a los rayos, que como se ha mencionado, excitan las tormentas eléctricas y en consecuencia excitan a las resonancias y a la propagación de las ondas electromagnéticas. Todo lo comentado anteriormente, hace abrir una ventana interesante y con un paso más hacia adelante, los científicos podrían usar las Resonancias de Schumann para estudiar el cambio climático a lo largo del tiempo. Al estar estas resonancias tan vinculadas y relacionadas a los cambios en la atmósfera podrían aportar datos históricos sobre cómo ha variado el clima a lo largo de los años. En la siguiente sección se hablará sobre las frecuencias extremadamente bajas, o ELF (Extremely Low Frequencies), que también guardan una fuerte relación con los aspectos mencionados anteriormente.



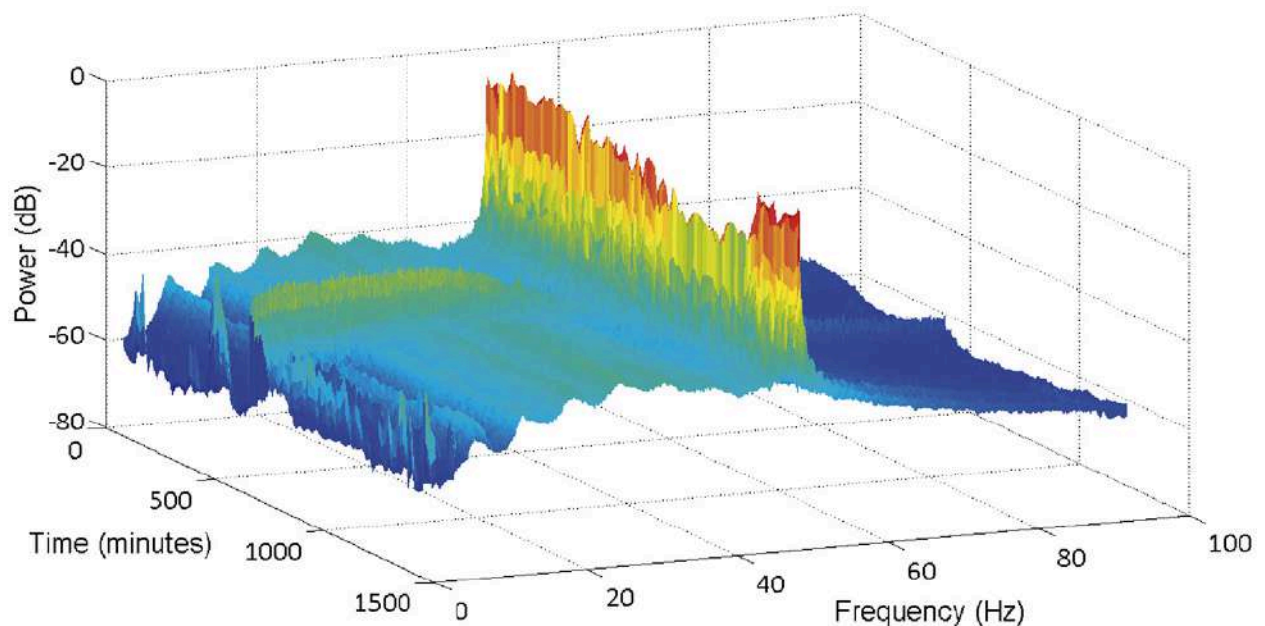
*Ilustración 6. Este es el impactante mapa que ha publicado la NASA que muestra una comparación entre la temperatura de los años 1951-1980 y la actualidad (2023) [3].*

## 1.2. Importancia de las frecuencias ELF

Las frecuencias extremadamente bajas, o ELF (por sus siglas en inglés), juegan un papel importante en la geofísica. Una de las características de estas ondas es que tienen la capacidad de viajar por grandes distancias sin perder demasiada energía en el camino. Por lo que, en el campo de la geofísica, la habilidad de viajar a través de la cavidad natural que se forma entre la Tierra y la ionosfera las convierte en una herramienta valiosa.

### Aplicaciones en la detección geofísica

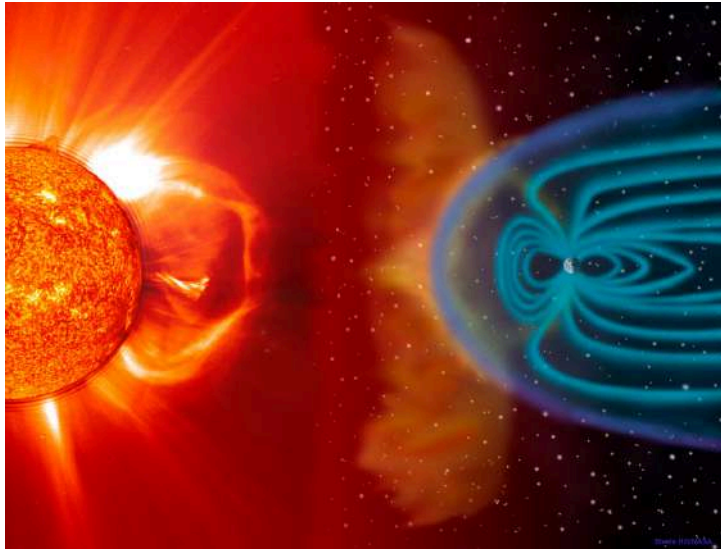
En este campo, dado que las ondas de baja frecuencia tienen longitudes de onda enormes, permiten atravesar materiales conductores como el agua y las rocas. La ciencia emplea esta capacidad para buscar recursos naturales como por ejemplo localizar yacimientos minerales, detectar acuíferos. Además, las ondas ELF también son útiles para monitorear terremotos y un campo digno de investigación es el de estudiar ciertos patrones para predecir o poder dar pistas sobre la actividad sísmica que se avecina. También se usan en estudios volcánicos estudiando el comportamiento de esta onda alrededor de volcanes activos dado que los científicos pueden detectar cambios en las propiedades eléctricas del suelo, lo cual, está relacionado con el movimiento de magma o la actividad sísmica antes de la erupción.[4]



*Ilustración 7. Gráfica de la actividad de las ondas ELF.*

## Contribución a la investigación del campo magnético terrestre

Otro punto que no se podría quedar en el tintero es que usando estas ondas ELF, se podría obtener información valiosa sobre el campo geomagnético. Particularmente para investigar fenómenos relacionados con las auroras boreales o cómo el campo magnético se ve afectado por partículas cargadas que vienen del espacio, como las tormentas solares.



*Ilustración 8. Recreación de una tormenta solar realizada por la NASA [5].*

### 1.3. Justificación histórica

En la siguiente sección veremos como las Resonancias de Schumann tienen una historia bastante interesante que se remonta al siglo XIX. A lo largo del tiempo, gran cantidad de científicos han ido aportando su granito de arena para entender y poder utilizar mejor este fenómeno. La evolución de este campo de investigación es la siguiente:

#### Orígenes y primeras observaciones

El primer hito en cuanto a las resonancias en la atmósfera terrestre se refiere comenzó en 1893, cuando el físico irlandés George Francis FitzGerald notó que las capas superiores de la atmósfera podrían comportarse como conductor eléctrico. Bien es cierto que en su época no causó mucho revuelo, sin embargo, fue de los primeros en sugerir que podría haber resonancia electromagnética en la cavidad formada entre la Tierra y la ionosfera, sin embargo, en ese entonces no había tecnología suficiente para desarrollar mucho más su teoría. Entonces, en 1899 entra en escena Nikola Tesla, que con su famoso experimento demostró que la Tierra podría resonar con onda electromagnética. Sin saberlo, estaba jugando con lo que más tarde serían las Resonancias de Schumann. Lo tacharon de loco al insinuar que la Tierra podría usarse como un conductor para enviar señales a grandes distancias sin cables.

## Desarrollo de la teoría de las Resonancias de Schumann

El siguiente gran paso en la evolución fue en 1925 cuando Edward Appleton y Barnett lograron demostrar que la ionosfera realmente existía. Aunque parezca simple fue muy importante, aunque no sabían que esta capa era conductora. Finalmente, en 1952 el físico alemán Winfried Otto Schumann publicó sus predicciones matemáticas sobre las resonancias que ocurren en esta cavidad, afirmando que podían resonar a frecuencias súper bajas. Schumann, junto con Herbert L. König, hicieron cálculos más detallados, logrando así demostrar experimentalmente que esas resonancias existían de verdad [2].

## Avances experimentales y estudios recientes

Ya en los años 90, la expectación por las Resonancias de Schumann siguió creciendo sobre todo por los científicos que descubrieron que estas resonancias podrían usarse para monitorear ciertos cambios en el planeta, hasta el punto de que investigadores japoneses encontraron anomalías en la frecuencia de resonancia antes de grandes terremotos como el de Niigata en 2004 y el de Noto en 2007 [6]. Hoy en día todavía sigue en fase de estudio la predicción de estas actuaciones sísmicas.

### 1.4. Objetivos del proyecto

El objetivo principal de este TFG es detectar patrones entre las Resonancias de Schumann y las tormentas geomagnéticas a partir de algoritmos de inteligencia artificial supervisada. La finalidad de este estudio es saber qué nivel de intensidad de tormenta geomagnética existe a partir de los parámetros de estas resonancias, además de encontrar relaciones entre los diferentes parámetros y características. Se debe tener en cuenta que se estudian dos grandes conjuntos de datos, pero ambos al mismo tiempo, es decir, se trabaja con los datos referidos a un año desde el 1 de abril de 2013 hasta el 31 de marzo de 2014 del magnetómetro este-oeste y del magnetómetro norte-sur.

## 2. Desarrollo Teórico

### 2.1. Descripción de la estación

#### Ubicación y características geográficas

La estación de medición de frecuencias extremadamente bajas se encuentra ubicada en el Refugio Poqueira, Sierra Nevada, Granada, España. Su altitud es de 2.500 metros sobre el nivel del mar y se ubica en un entorno ideal para la medición de las Resonancias de Schumann y otros fenómenos electromagnéticos de baja frecuencia debido a su lejanía de zona habitadas, camino principales y fuente de ruido electromagnético antropogénico como pueden ser actividades humanas o líneas de transmisión eléctrica. El acceso puede ser complicado sobre todo durante los meses de invierno cuando esta zona está cubierta de nieve. Por un lado, la hace perfecta para la captación de señales ya que evita gran parte de interferencia humana que normalmente afectaría a la medición, no obstante, limita el mantenimiento y la posibilidad de resolver problemas técnicos de manera rápida y eficaz.





*Ilustración 9. Mapa de España dónde la "x" muestra la ubicación del Refugio Poqueira, Sierra Nevada, Granada, España*

#### Estructura de la estación y dispositivos seleccionados

La estación está constituida por una alimentación mediante paneles solares, los cuales, para Maximizar la recepción de luz solar están orientados a 50° y son útiles sobre todo durante los meses de invierno. Estos paneles cargan dos baterías de gel que proporcionan energía a la estación y sobre todo a los amplificadores, sistema de digitalización y registro de datos y a la antena de transmisión de datos.



*Ilustración 10. Paneles solares ubicados en el Refugio Poqueira, Granada.*

Las baterías pueden recargarse completamente en unos cinco o seis días de invierno cuando hay menos disponibilidad de luz solar. El consumo medio total de energía es de aproximadamente 10 vatios lo que permite que la estación funcione de manera autónoma durante aproximadamente 22 días sin luz solar. Cuando los datos son capturados, se transmiten automáticamente a un servidor ubicado en la UGR la Universidad de Granada utilizando una red de telefonía móvil GPRS/GSM. Se ha instalado una antena Yagi para garantizar una transmisión eficiente de los datos debido a la baja intensidad de la señal en el área. El sistema de medición consta de varios componentes clave:

#### Magnetómetros

Dos magnetómetros de bobina de inducción diseñados para medir señales en el rango de 3 a 30 Hz están orientados en direcciones norte-sur y este-oeste respectivamente y así poder captar todas las posibles variaciones en los campos electromagnéticos en ambas direcciones. Cada uno está equipado con un núcleo ferromagnético de alta permeabilidad y un alto número de espiras para garantizar una alta sensibilidad a frecuencia baja.

#### Preamplificadores

Los magnetómetros están conectados a preamplificadores colocados a una distancia de aproximadamente un metro, cuya funcionalidad es amplificar la señal para mejorar la relación señal ruido (SNR). Están conectados a través de cables coaxiales de baja pérdida para minimizar la atenuación de la señal y proporcionarán una amplificación aproximadamente de 34 dB.

#### Filtro paso bajo

Utiliza el conocido filtro de Butterworth de décimo orden para eliminar el ruido de 50 Hz causado por la línea eléctrica y que podría contaminar e interferir en nuestras mediciones. Dicho filtro sigue siendo necesario, aunque la estación, como se ha comentado anteriormente, está ubicada en una excelente localización y eso hace que dichas interferencias sean mínimas.

#### Conversor A/D

La señal amplificada y filtrada se logra digitalizar gracias a un conversor A/D de 16 bits, lo que nos proporciona una alta predicción en la medición de las señales EFL.

#### Sistema de almacenamiento y transmisión de datos

Por último, estos datos digitalizados se almacenan temporalmente en una memoria flash antes de ser transmitidos a la Universidad de Granada donde finalmente se procesarán y analizarán.



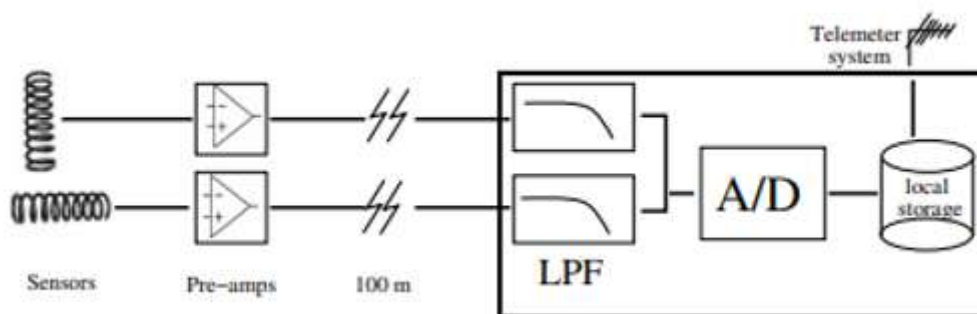


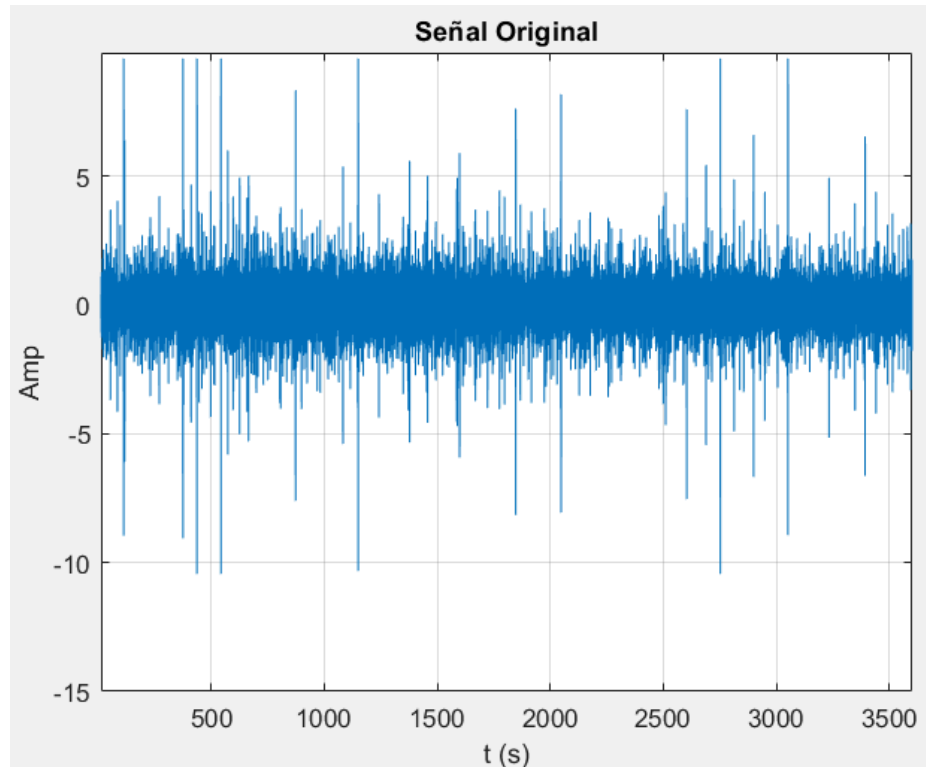
Ilustración 11. Diagrama de bloques de la estación de medición.

Problemas técnicos encontrados en la instalación y mantenimiento.

Como se ha comentado anteriormente, el mayor desafío es su ubicación, que, aunque por un lado es beneficiosa, por otro lado, el hecho de que esté en una localización remota en la montaña dificulta su instalación inicial y su mantenimiento, esto junto a los ajustados presupuestos y recursos, hace que el mantenimiento en sí mismo requiera un gran esfuerzo y constituya a veces un desafío. Un problema técnico fue la baja disponibilidad de señal de los repetidores de la red GSM en la zona. Contratiempo que se solucionó instalando una antena Yagi que mejoró tanto la recepción como la transmisión de la señal. Pese a que está situada en un entorno difícilmente alcanzable por el ser humano, ciertos factores ambientales como el viento, el movimiento de los árboles o la propia fauna hace que se pueda generar ruido en las mediciones y haya que tener cierta precaución a la hora del estudio de los datos. Se consiguió reducir considerablemente la interferencia causada por el viento enterrando a un metro los magnetómetros y preamplificadores.

## 2.2. Análisis de Fourier en el procesamiento de señales ELF

En la ingeniería de tecnologías de telecomunicación y en muchas otras, el análisis de Fourier es una herramienta esencial para el procesamiento de señales y especialmente clave para estudiar las Resonancias de Schumann. Las señales ELF varían en frecuencia y amplitud a lo largo del tiempo, no son estacionarias y, además, son extremadamente bajas en frecuencia. A partir de las series temporales obtenidas por el magnetómetro, se obtienen los espectros de estas. El análisis de Fourier permitirá descomponerlas en sus componentes de frecuencia fundamentales facilitando su estudio y monitoreo. Utilizaremos la transformada discreta de Fourier que es capaz de convertir una señal en el dominio del tiempo en una representación en el dominio de la frecuencia. Lo que se refiere a que permite aislar las frecuencias clave de las resonancias y obtener información sobre su comportamiento a lo largo del tiempo [7].



*Ilustración 12. Señal original.*

## Fundamentos de la Transformada de Fourier

La transformada de Fourier es una herramienta matemática que descompone una señal continua o discreta en una suma de funciones sinusoidales de diferente frecuencia. En su versión discreta (DFT) toma muestras de una señal y las convierte al dominio de la frecuencia. Su expresión matemática es la siguiente:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi kn}{N}}$$

Donde

$X(k)$  es el espectro de frecuencias de la señal.

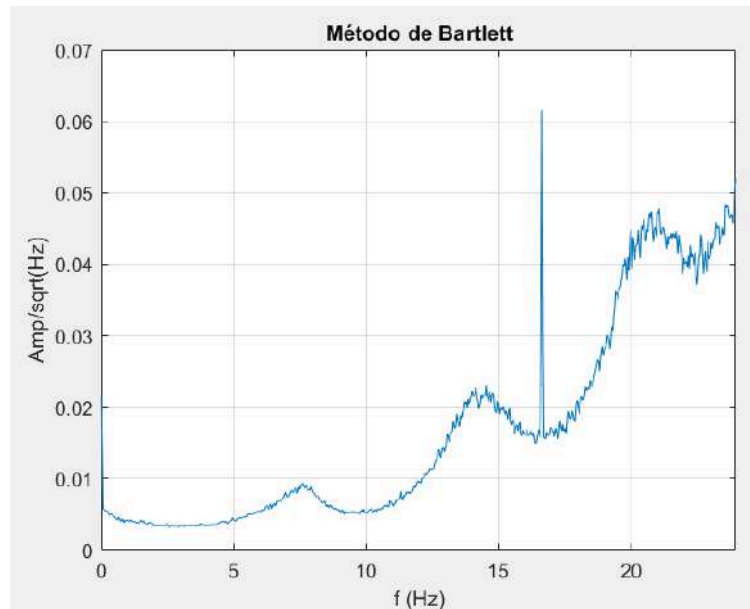
$x(n)$  es la señal en el dominio del tiempo.

$N$  es el número de muestras.

## Uso de la técnica de Bartlett en la medición de señales no estacionarias.

Las señales no estacionarias presentan variaciones a lo largo del tiempo, esto complica su análisis, sin embargo, la técnica de Bartlett permite procesar este tipo de señales haciendo que sea una herramienta útil para mejorar la estimación espectral mediante el uso de ventanas de tiempo. Divide la señal en segmentos de longitud fija y

aplica una ventana (Hamming) a cada uno de estos segmentos. Seguidamente, se aplica la transformada FFT de cada segmento y se promedian los espectros resultantes. Esta técnica permite obtener un aspecto más estable y representativo de la frecuencia de interés mejorando la precisión y reduciendo el ruido. Se ha realizado un segmentado de manera que se hace una FFT cada 30 segundos, sin solapamiento y con una ventana Hamming.



*Ilustración 13. Resultado de la señal después de haber empleado el método de Bartlett.*

### Comparación con otros métodos de procesamiento de señales

Pese a que se usarán estas técnicas, es decir, DFT y la técnica de Bartlett, hay algunas que se mencionan a continuación debido al interés que presentan. Sin embargo, dado que el objetivo de estos métodos de procesamiento de señales es adaptar las señales y obtener espectros más estables y representativos, comentaremos únicamente la principal característica diferencial de estos métodos.

- Transformada de Wallet: Utiliza ondas sinusoidales de amplitud constante para descomponer las señales, la transformada de Wallet usa “ondas” que tienen tanto una frecuencia como una localización temporal específicas.
- Transformada de Fourier de tiempo corto (STFT): es una variación de la DFT ya que aplica la transformada de Fourier a segmentos de la señal delimitados por una ventana deslizando.

## 2.3. Modelo Teórico de la cavidad Tierra-Ionosfera

El modelo considera a dos esferas concéntricas, la más pequeña siendo la tierra y la más grande la tierra más la ionosfera siendo la capa intermedia una capa de aire dieléctrico conductora que permite la propagación de ondas electromagnéticas entre ambas.

## Modelos matemáticos y físicos de la cavidad electromagnética

El comportamiento de la cavidad tierra-ionosfera se describe mediante un modelo basado en las ecuaciones de Maxwell que trata de las relaciones entre los campos eléctricos y magnéticos en el vacío. Gracias a estas ecuaciones se estudia cómo se propagan las ondas ELF en un medio dieléctrico y cómo interactúan con la ionosfera y la superficie terrestre.

$$\nabla^2 \vec{E} - \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} = 0$$

Donde

$\nabla^2 \vec{E}$ : Laplaciana del campo eléctrico  $\vec{E}$ , describe la variación espacial del campo.

$\vec{E}$ : Campo eléctrico.

$\mu_0$ : Permeabilidad del vacío, valor fijo de  $4\pi \times 10^{-7}$  H/m.

$\epsilon_0$ : Permitividad del vacío, valor fijo de  $8.854 \times 10^{-12}$  F/m.

$\partial^2 \vec{E} / \partial t^2$ : Segunda derivada temporal del campo eléctrico, describe la aceleración del campo.

La solución de la ecuación en coordenadas esféricas utiliza funciones de Bessel para describir las variaciones radiales y funciones de Legendre para describir las variaciones angulares. Los modos transversales magnéticos son los que corresponden con las resonancias y los armónicos de estos modos con las frecuencias de resonancia [8].

## 2.4. Representación de los campos eléctricos y magnéticos

Los campos eléctricos y magnéticos tienen una distribución particular que es clave para la medición de la Resonancia de Schumann. Este comportamiento permite que las ondas se propaguen a lo largo de la superficie terrestre reflejándose entre la ionosfera y la tierra.

- Campo eléctrico radial ( $E_r$ ): Se extiende desde la ionosfera hacia la Tierra y se debilita a medida que desciende hacia la superficie.
- Campo magnético tangencial ( $H_\theta$ ): Rodea la Tierra y es perpendicular al campo eléctrico. Este campo es el que generalmente se mide en las estaciones de monitoreo, ya que es más fácil de detectar que el campo eléctrico.

### Campos radiales y tangenciales en la cavidad.

Para el estudio de la Resonancia de Schumann, el campo magnético tangencial es el componente más relevante ya que es el que nuestros magnetómetros tienen la habilidad de detectar con mayor precisión. Permiten estudiar los efectos de las tormentas eléctricas y otro fenómeno atmosférico capturando las variaciones del campo magnético tangencial.

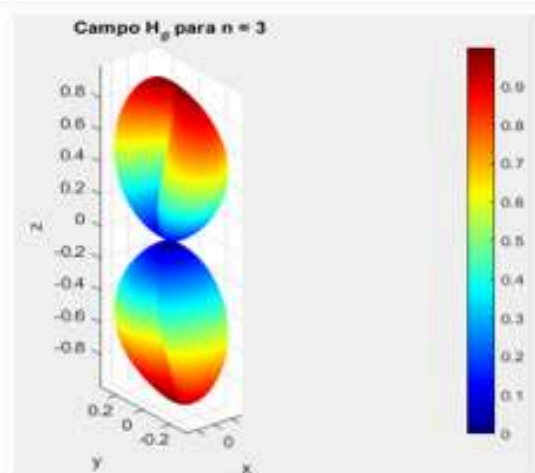


Ilustración 14. Campo magnético tangencial ( $H_\theta$ ) siendo  $n=3$ .

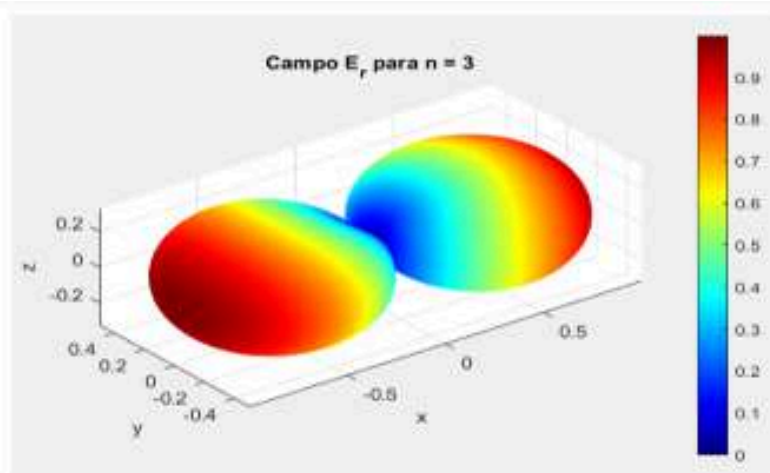


Ilustración 18. Campo eléctrico radial ( $E_r$ ) siendo  $n=3$ .

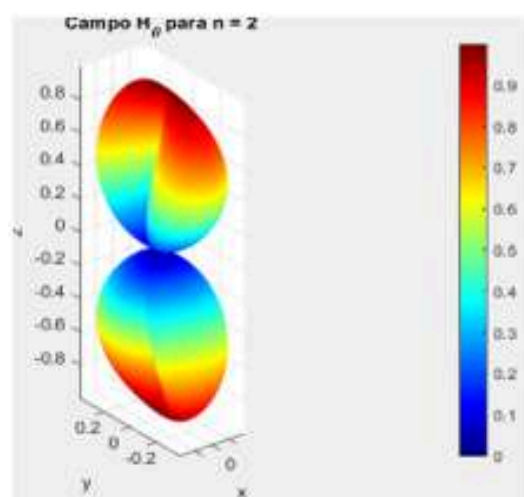


Ilustración 16. Campo magnético tangencial ( $H_\theta$ ) siendo  $n=2$ .

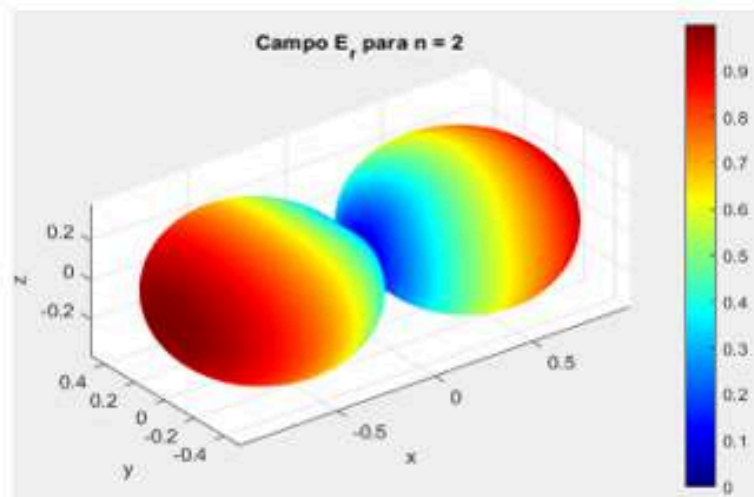


Ilustración 19. Campo eléctrico radial ( $E_r$ ) siendo  $n=2$ .

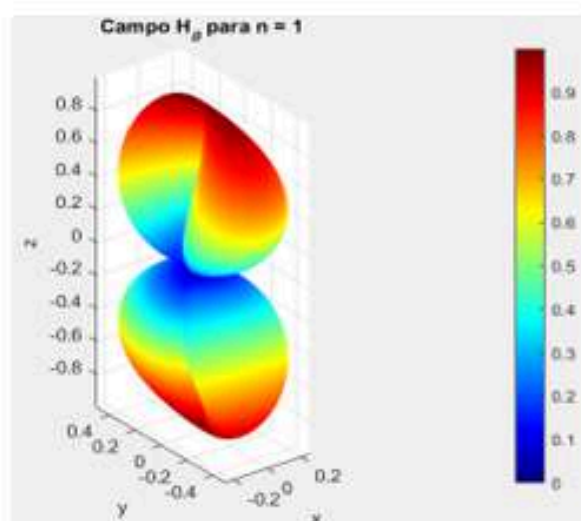


Ilustración 15. Campo magnético tangencial ( $H_\theta$ ) siendo  $n=1$ .

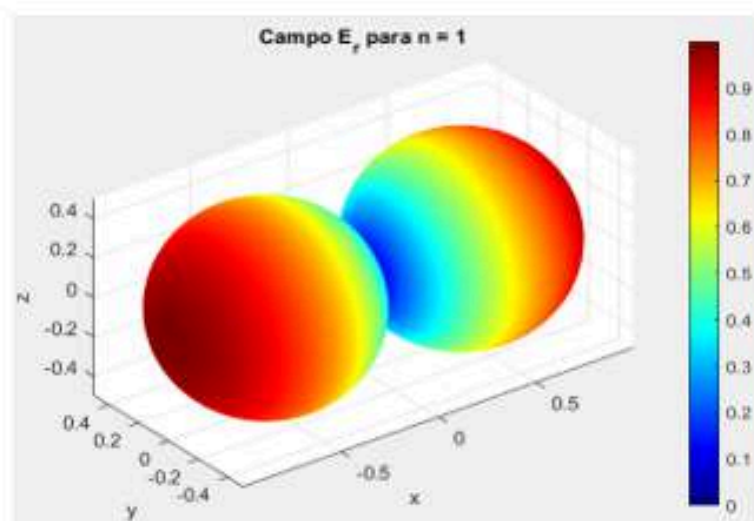
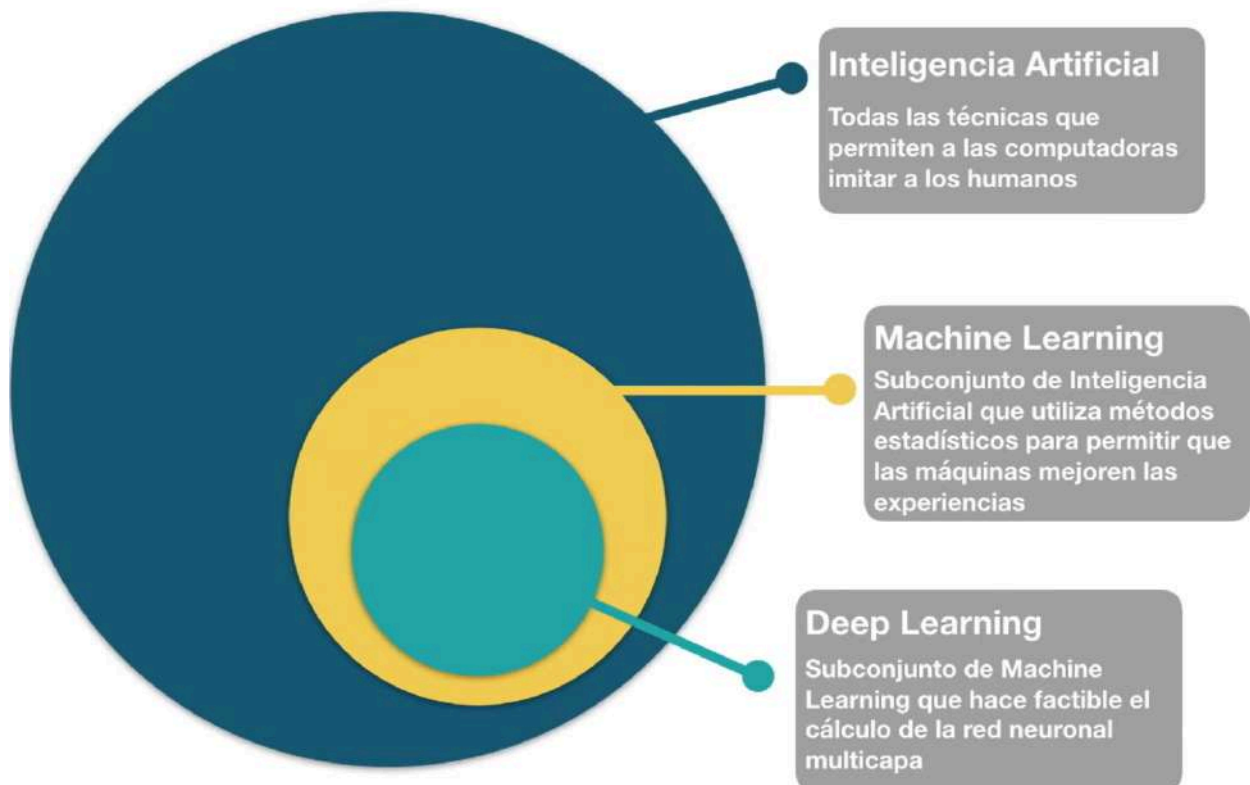


Ilustración 20. Campo eléctrico radial ( $E_r$ ) siendo  $n=1$ .

## 2.5. Inteligencia artificial

En este TFG resulta fundamental tener conocimientos previos sobre inteligencia artificial, ya que más adelante se utilizarán algoritmos de inteligencia artificial para analizar los datos obtenidos de la estación. La inteligencia artificial, es uno de los campos más innovadores de los últimos años. Puede definirse como una rama de la ciencia de los datos que busca desarrollar sistemas con la capacidad de realizar tareas que normalmente requieren intervención humana. Aunque es un campo relativamente nuevo, tiene un amplio alcance y se puede dividir en varias subramas importantes.



*Ilustración 21. Esquema sobre la historia de la Inteligencia artificial.*

### 2.5.1. Procesamiento del Lenguaje natural

Una rama de investigación derivada del campo de la inteligencia artificial es el procesamiento del lenguaje natural. Esto no es algo nuevo que haya surgido en los últimos años de investigación, sino que se viene desarrollando a lo largo de la historia. Se trata de realizar o mejorar la comunicación entre personas y máquinas mediante el uso del conocimiento del lenguaje natural. Nació en 1950 con el objetivo de solventar ciertos problemas de generación y comprensión automática del lenguaje humano, es decir, del lenguaje natural. En estos años Alan Turing dio a conocer al mundo el artículo con el que comenzamos a llamar con nombre propio a la inteligencia artificial. Realizó lo que hoy en día conocemos como la prueba de Turing, mediante la cual, demuestra la inteligencia de las máquinas en un marco experimental y de investigación. En esta prueba, la máquina aprueba o lo supera si la persona no era capaz de distinguir si se está comunicando con otra persona o con un sistema informático [9].



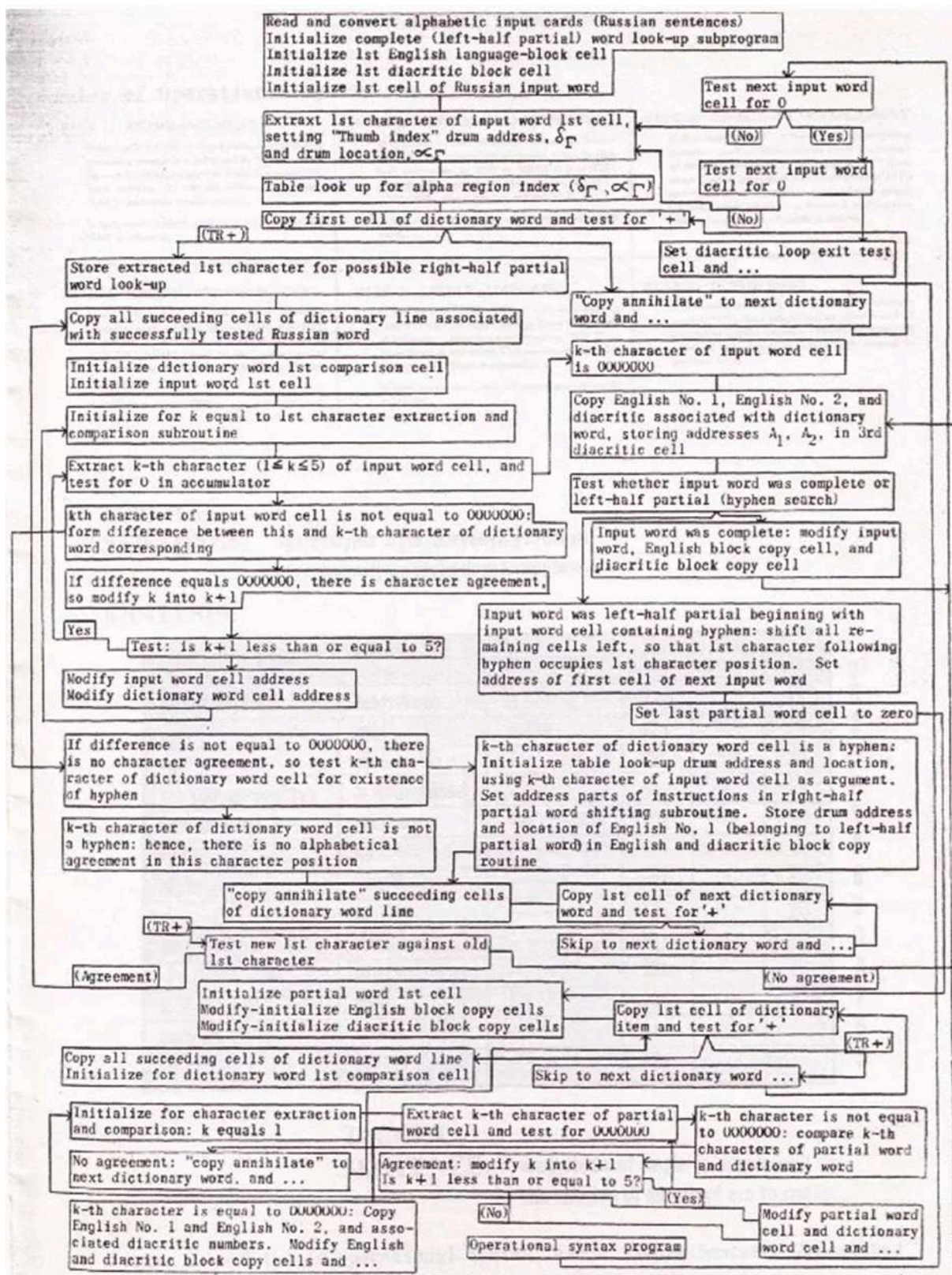


Ilustración 17. Periódico donde se publicó el Test de Turing.



## Multidisciplinar

Popularmente, el concepto de inteligencia artificial no suele ser de mucho agrado además de que se cree que únicamente toman parte los informáticos que entienden de programación. Se debe enfocar dicho concepto como una herramienta que nos ayudará a progresar y obviamente no puede ser construida únicamente por informáticos. Se necesitan diferentes disciplinas para poder completar su funcionamiento, entre estas se destacan:

## Lingüistas

Como se ha comentado anteriormente, uno de los grandes retos de la inteligencia artificial es establecer la comunicación del humano con la máquina, por lo tanto, es imprescindible el trabajo de lingüistas ya que son capaces de comprender a la perfección el lenguaje humano y pueda establecer el modelo lingüístico para que, en base a él, los programadores sean capaces de implementar el código de forma eficiente y funcional.

## Matemáticos

Las matemáticas en este campo son de vital importancia debido a la necesidad de desarrollar modelos matemáticos para poder trabajar con los ordenadores e incluso el uso de la estadística, lo que hace que sea imprescindible la disciplina matemática.

## Psicólogos

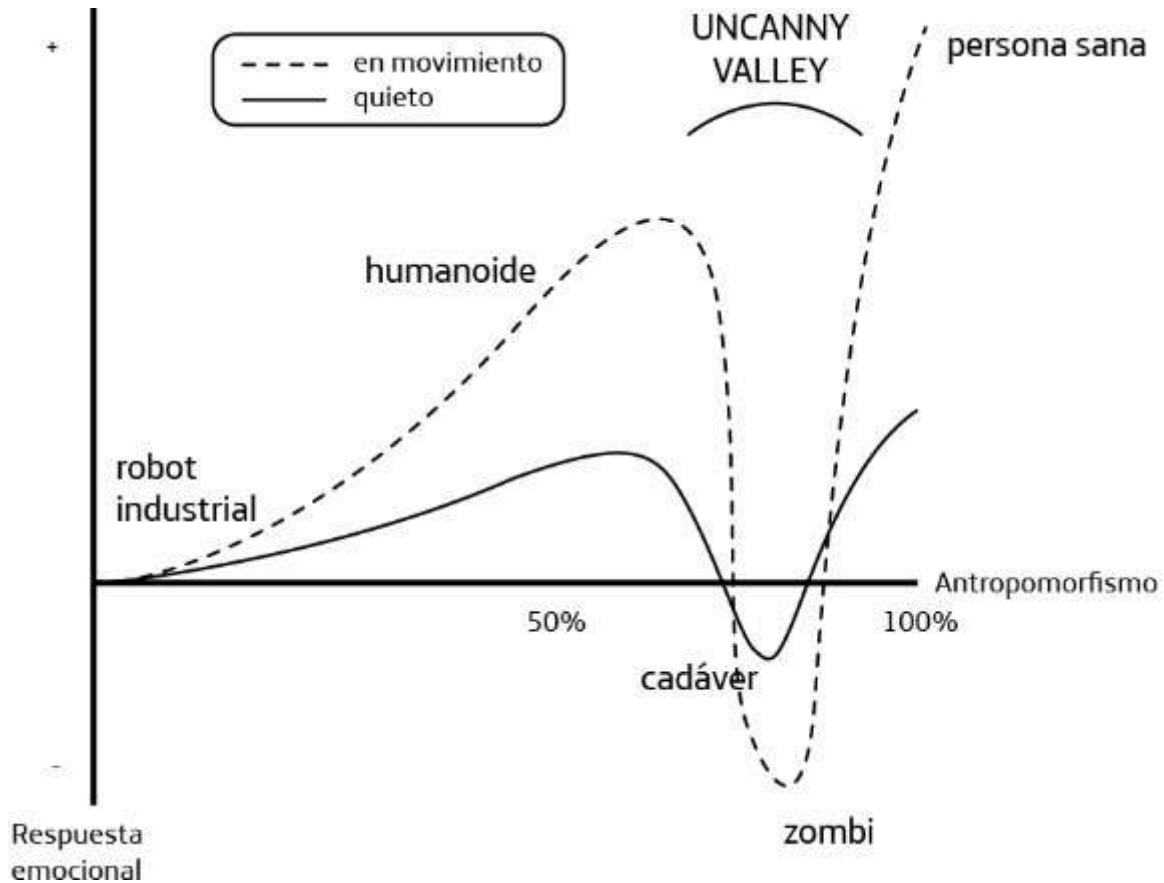
Quizás este sea el campo más difícil de tratar. Al fin y al cabo, son máquinas que pretendemos que, por ejemplo, en una conversación sean lo más parecido posible a los humanos, sin embargo, los humanos tienen algo que no es parte de una máquina y son los sentimientos. Por eso expertos psicólogos estudian la subjetividad de las palabras para analizar más allá de un enfoque lingüístico o matemático. Tomarán parte en tareas como el análisis de la toxicidad, ironía o humor.

Otro campo que es digno de comentar, aunque no interfiera directamente con la inteligencia artificial como los anteriores, es la filosofía y la ética. La inteligencia artificial como anteriormente se ha mencionado, no es del todo bien aceptada popularmente ya que ni siquiera los más expertos saben cuáles son sus límites. Es aquí donde entra el estudio filosófico y ético en el cual se estudia, por ejemplo, hasta qué punto se debe ser dependiente de la inteligencia artificial o si el avance de la inteligencia artificial puede ralentizar el avance como sociedad humana.

## ¿Para qué sirve el PLN?

El procesamiento del lenguaje natural está en pleno auge al igual que la inteligencia artificial en el mundo actual. Es cierto que muchas personas que usan esta herramienta no logran comprender cómo es capaz de interactuar una máquina con ellos mismos en su propio lenguaje. Hoy en día, son capaces de escuchar la entrada del usuario, procesarla y generar una salida correcta tanto en el ámbito de conocimiento como de expresión. ¿Cómo es capaz de entender el lenguaje humano? Aquí es donde entra el juego la rama de procesamiento de lenguaje natural. Se encarga de realizar y guiar los pasos necesarios para poder dar una respuesta totalmente coherente e inteligente. Se “enseña” a las máquinas el lenguaje humano mediante procesos de modelización matemática ya que, en una primera instancia, los ordenadores sólo entienden de bytes y dígitos. Las máquinas tienden a un desarrollo independiente y cada vez más alejado del control humano que podemos ver

reflejado en la siguiente imagen, donde a medida que el parecido a una persona aumenta, la familiaridad de las personas cae en picado. La siguiente imagen representa una teoría originalmente desarrollada por el profesor japonés M. Mori a principios de la década de 1970, Uncanny Valley (Valle inquietante) que es un valle o área de respuesta repulsiva que aparece entre un robot casi humano y una persona en buen estado de salud [10].



*Ilustración 19. Esquema de Uncanny Valley.*

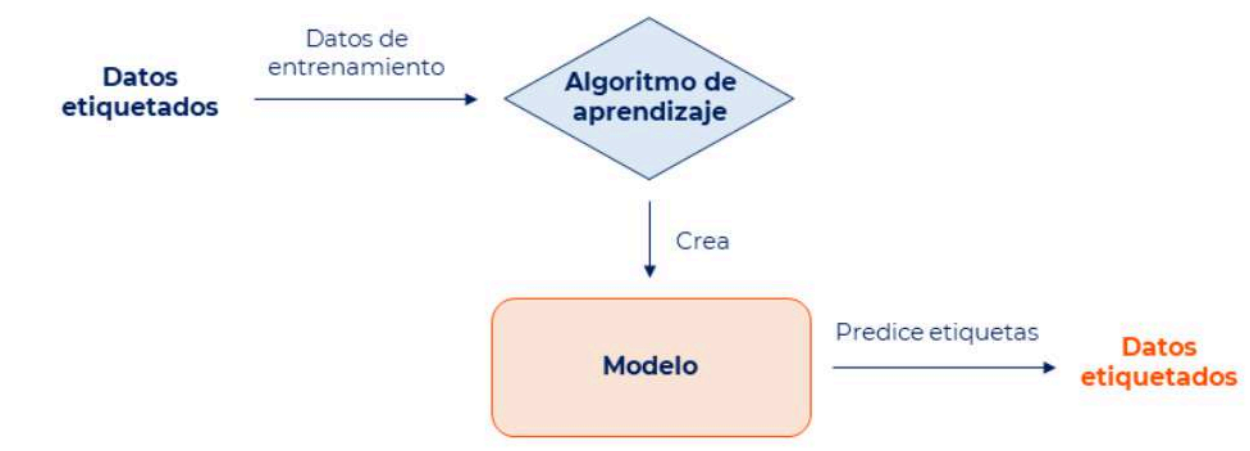
## 2.5.2. Machine Learning

Aunque pueda llegar a parecer moderno, surgió en 1950 con el test de Turing que ya hemos explicado anteriormente y en 1952 fue cuando se desarrolló el primer algoritmo capaz de aprender gracias a Arthur Samuel, que se basaba en una partida de damas donde la inteligencia cada vez era más avanzada, es decir, aprendía continuamente. El aprendizaje automático es como su propio nombre indica, cuando las máquinas adquieren un aprendizaje óptimo que las lleve a realizar distintas tareas con buenos resultados. Es una de las ramas más importantes de la inteligencia artificial y la definición técnica sería una aplicación de distintas prácticas de inteligencia artificial para convertir un sistema computacional en un sistema capaz de reconocer patrones y convertir un conjunto de datos en un programa que pueda adquirir un nuevo conocimiento a partir del mismo. La novedad es que trata de reconocer patrones y simular el comportamiento humano que, aunque es bastante complicado y no cuentan con la capacidad de razonar y aprender, tratan de simular esa experiencia en la medida de lo posible por lo que tiene la capacidad de adaptarse a los nuevos cambios que van surgiendo en los conjuntos de datos representados como entrada a

nuestro sistema. El hecho de que trabaje de esta manera hace que esté por delante de distintas técnicas aplicadas anteriormente. Entre estos algoritmos podemos diferenciar dos tipos principales:

### Aprendizaje supervisado

Estos tipos principales se diferencian y se caracterizan de manera diferente principalmente en la forma en la que se produce el entrenamiento de sus algoritmos, en este caso, al entrenamiento se le proporciona un conjunto de datos etiquetado por lo que el algoritmo será consciente de la clasificación que contiene cada una de las partes del conjunto de entrenamiento. Esto significa que este tipo de aprendizaje tiene un entrenamiento guiado gracias a las características implicadas como entrada, es decir, tiene un seguimiento. Este será el aprendizaje que utilizaremos en los siguientes apartados y en este estudio, sin embargo, explicaremos brevemente los demás tipos a continuación [11].



*Ilustración 18. Esquema explicativo del aprendizaje supervisado en Machine Learning.*

### Aprendizaje no supervisado

La principal diferencia con respecto al aprendizaje anterior es que sus algoritmos se basan en datos no etiquetados por lo que su objetivo es encontrar ciertos patrones específicos que se encuentran dentro del conjunto de entrenamiento. Como se observa en las imágenes anteriores, se ve cómo a partir del cambio en el conjunto de entrada surge la diferencia entre estos. En el aprendizaje no supervisado solo cuenta con datos en bruto y no contiene esa definición o etiqueta que le simplifica el trabajo de entrenamiento, por lo que es el propio modelo el que tiene que establecer la diferencia entre clases por sí mismo. Por otro lado, en el aprendizaje supervisado, contienen su definición en forma de etiqueta por lo que el modelo no debe pararse a entender las diferencias entre las clases [11].

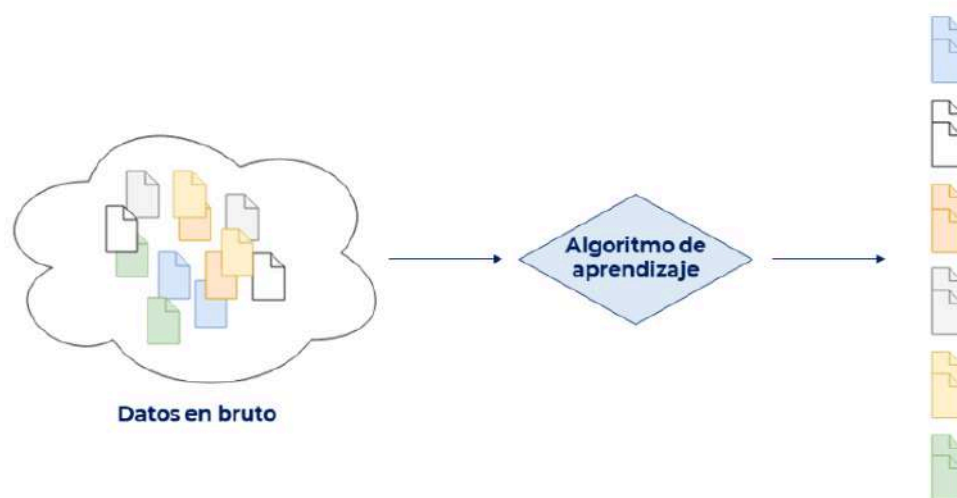


Ilustración 19. Esquema explicativo del aprendizaje no supervisado en Machine Learning.

### Aprendizaje semi supervisado

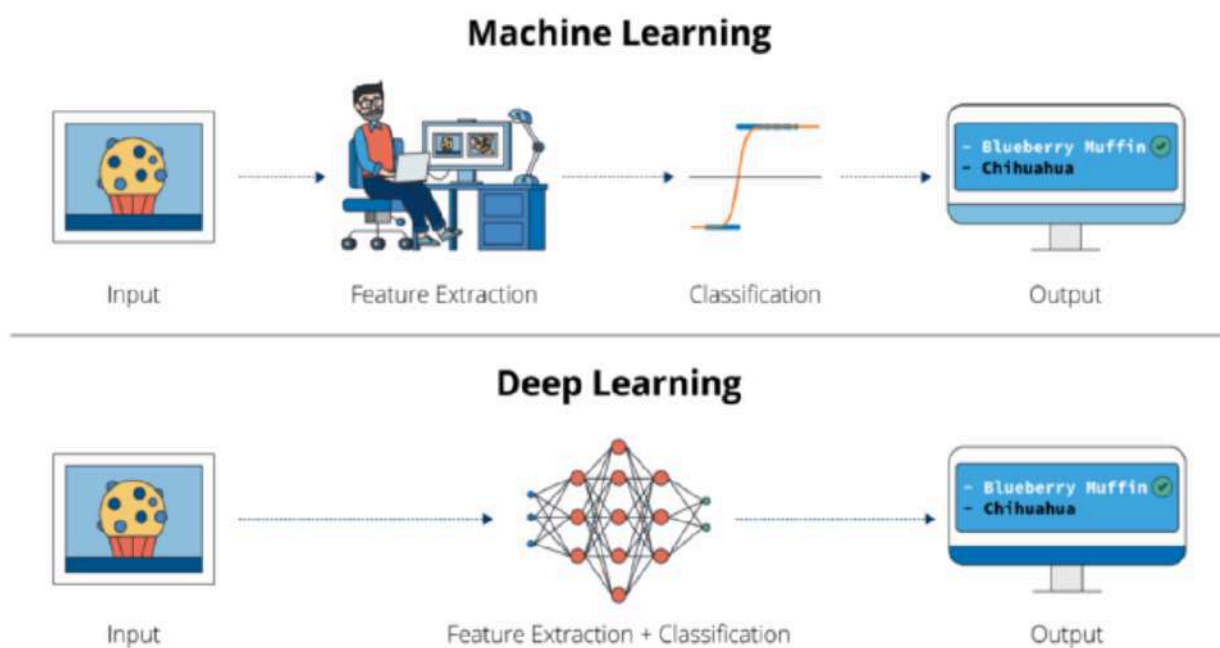
El aprendizaje semisupervisado se sitúa entre los métodos supervisados y no supervisados, aprovechando las ventajas de ambos enfoques. En este caso, el entrenamiento se realiza sobre un conjunto de datos que contiene tanto muestras etiquetadas como muestras no etiquetadas. El objetivo de este tipo de aprendizaje es utilizar una cantidad limitada de datos etiquetados junto con una gran cantidad de datos no etiquetados para mejorar la precisión del modelo, manteniendo un costo reducido en la generación de etiquetas. Así, se combina el entrenamiento guiado del aprendizaje supervisado con la capacidad de encontrar patrones del aprendizaje no supervisado. En resumen, el modelo recibe una parte de la información con etiquetas explícitas y aprende a inferir las etiquetas para el resto, aprovechando al máximo la estructura y relaciones presentes en los datos [11].



Ilustración 20. Esquema explicativo del aprendizaje semi supervisado en Machine Learning.

## Deep Learning

Aunque se van a usar principalmente las técnicas de Machine Learning, me gustaría comentar brevemente qué es Deep Learning. Es un término usualmente introducido dentro del Machine Learning cuyo objetivo se podría denominar como similar, sin embargo, la principal diferencia es que sus sistemas tratan de imitar las características arquitectónicas del sistema nervioso de un humano haciendo que esté constituido por diferentes redes de procesamiento y cada una de ellas especializada en distintas características de los datos de entrenamiento. Deep Learning, por lo tanto, es una aproximación de la propia percepción humana y son sistemas mayormente orientados al aprendizaje no supervisado [12].



*Ilustración 22. Diferentes fases entre Machine Learning y Deep Learning.*

## Algoritmos y métricas de evaluación

### Z-score

Z-score es una herramienta cuya finalidad es detectar outliers, lo que se conoce como valores atípicos, dentro de un conjunto de datos. Si nos situamos dentro del mundo de Machine Learning, es tremendamente eficaz y de lo que se encarga básicamente es de medir cuántas desviaciones estándar se aleja un valor del promedio. La fórmula básica es la siguiente [13]:

$$Z = \frac{x - \mu}{\sigma}$$

Donde

$x$  es el valor de la observación  
 $\mu$  es la media del conjunto de datos  
 $\sigma$  es la desviación estándar

## Detecting Outliers with z-Scores

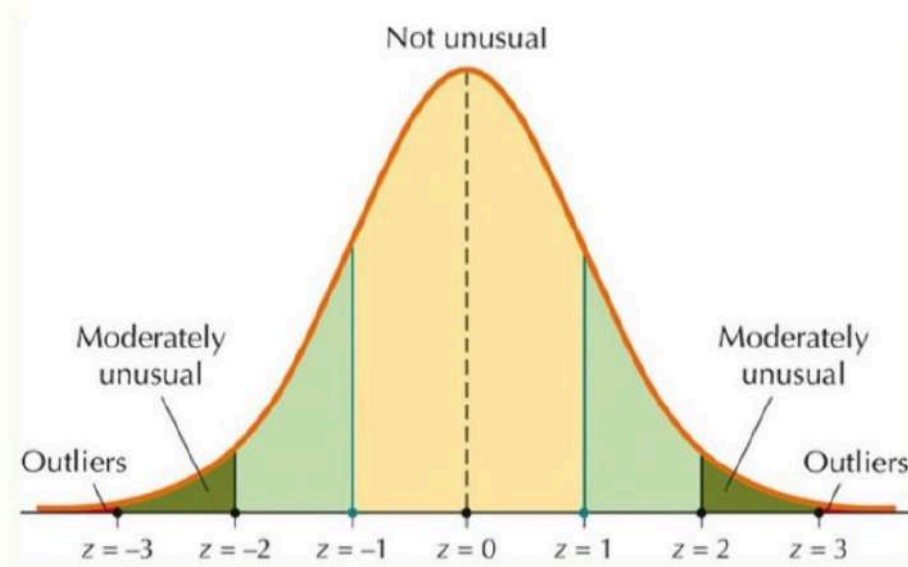


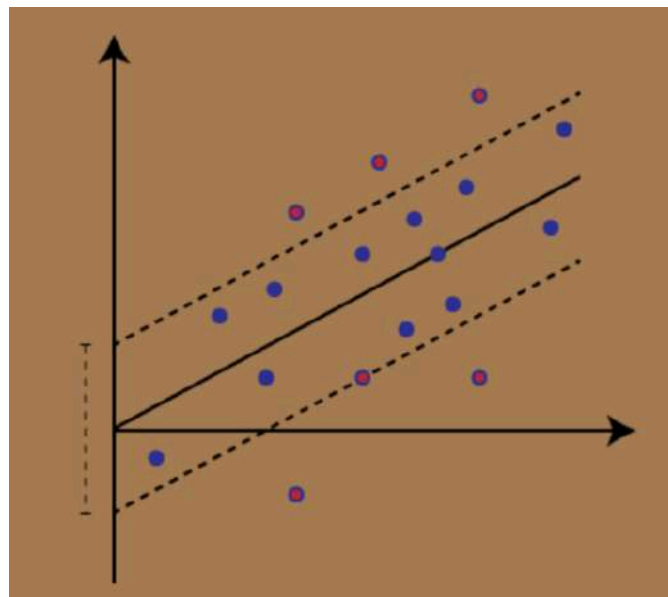
Ilustración 21. Esquema gráfico del funcionamiento de la herramienta Z-score

A continuación, se desarrollan diferentes tipos de algoritmos dentro del campo del aprendizaje automático y en este caso dentro del tipo supervisado. Son capaces de realizar un aprendizaje, lo que implica que son capaces de mejorar a medida que ganan experiencia, mejoran automáticamente. Su objetivo es proporcionar predicciones a partir de un entrenamiento donde a partir de un conjunto de datos que se le transmite como entrada, explorarán, analizarán y aprenderán patrones específicos. Los modelos que estudiaremos son los siguientes: SVR, XGBoost Regressor y Random Forest Regressor.

## SVR

Se ha elegido Support Vector Regression (SVR) que es una extensión del Support Vector Machines (SVM). La principal diferencia entre estos dos es que mientras que en Support Vector Machine se busca una línea de separación entre clases, en Support Vector Regresión se busca un hiperplano que prediga valores continuos. Por lo tanto, podemos afirmar que el objetivo del modelo SVR es minimizar el error dentro de un margen de tolerancia y al mismo tiempo mantener el modelo lo más simple posible, lo que se logra con el uso de vectores de soporte. Básicamente, un modelo especialmente útil a la hora de ajustar la relación entre la variable de entrada y salida de manera más flexible que otros modelos de regresión lineal. Entonces algunas características clave del algoritmo de regresión SVR son:

- Un margen de error controlado, es decir, un modelo que permite un margen de error en las predicciones siempre que estas estén dentro de un margen.
- Únicamente los puntos que se encuentran fuera de este margen de error se consideran vectores de soporte y afectan a la predicción. Ventaja que beneficia en cuanto a que el modelo sea eficiente en términos de memoria.
- El uso de Kernels, que se pueden definir como diferentes formas de aplicar el hiperplano, por lo tanto, dependiendo de si se busca un hiperplano lineal o no lineal podremos usar diferentes tipos de Kernel para ver cuál se ajusta mejor al modelo [16], esto se comenta en una de las secciones futuras que se enfoca más el propio trabajo



*Ilustración 22. Esquema de funcionamiento de SVR [15].*

## XGBoost Regressor

XGBoost (eXtreme Gradient Boosting) será el segundo modelo con el que se trabaje en este estudio. Es una de las herramientas más potentes dentro del mundo del Machine Learning y se basa en la implementación avanzada del algoritmo de gradiente Boosting. Significa, que en cuanto al principio de su funcionamiento se refiere, emplea una serie de árboles de decisión que se entrenan secuencialmente, por lo que, al entrenar el primer árbol, estudia los resultados y analiza los errores para que, a la hora de entrenar el siguiente árbol, corrija los errores anteriores y así de manera secuencial. En concreto, XGboost optimiza su modelo mediante la minimización del error a través de técnicas de gradiente descendiente, básicamente ajusta los pesos de cada uno de los datos para mejorar las predicciones futuras. Se entiende que después de haber explicado su funcionamiento, es un algoritmo altamente eficiente para problemas de regresión debido a su capacidad para producir resultados precisos [17]. Gracias a la siguiente figura, observaremos el funcionamiento descrito previamente de manera gráfica y dinámica:



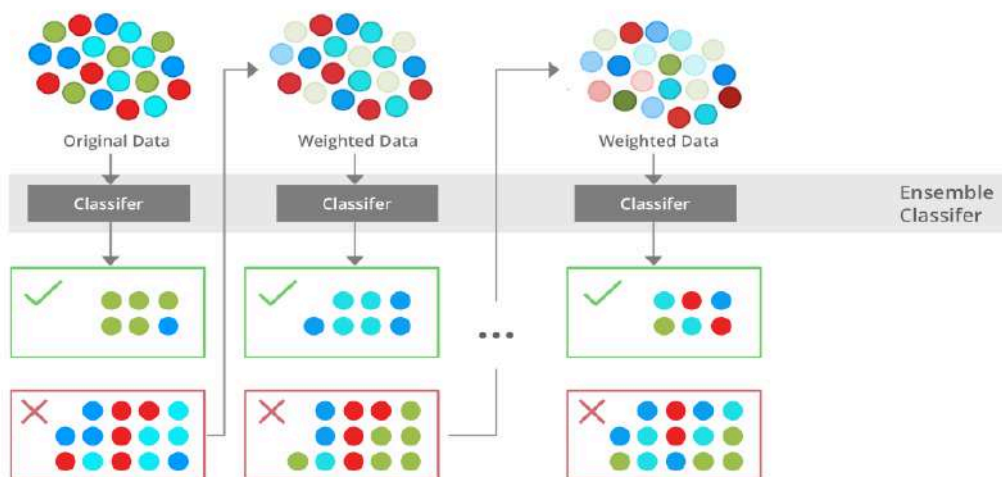


Ilustración 23. Esquema gráfico de funcionamiento de XGBoost Regressor [18].

### Random Forest Regressor

Por último, se utiliza este conocido algoritmo, basado también en árboles de decisión como el anterior modelo con la de diferencia clave de que este modelo se construye sobre la idea de Ensemble Learning, que significa que combina las predicciones de múltiples árboles de decisión para mejorar la precisión de las predicciones finales. El XGboost Regressor lo que hacía era crear secuencialmente árboles de decisión y cada uno de ellos aprendiendo de los errores del anterior, sin embargo, este algoritmo selecciona múltiples subconjuntos de los datos de entrenamiento de manera aleatoria y entrena un árbol de decisión en cada uno de ellos. Además, en cada nodo del árbol, son seleccionadas un subconjunto de características de manera completamente aleatoria para encontrar la mejor división, algo que introduce aleatoriedad y diversidad entre los árboles. Por último, el modelo promedia las predicciones de todos los árboles y esto ayuda a reducir el sobreajuste que ocurriría utilizando solo un árbol [19]. La siguiente imagen muestra el funcionamiento explicado anteriormente de manera visual e intuitiva:

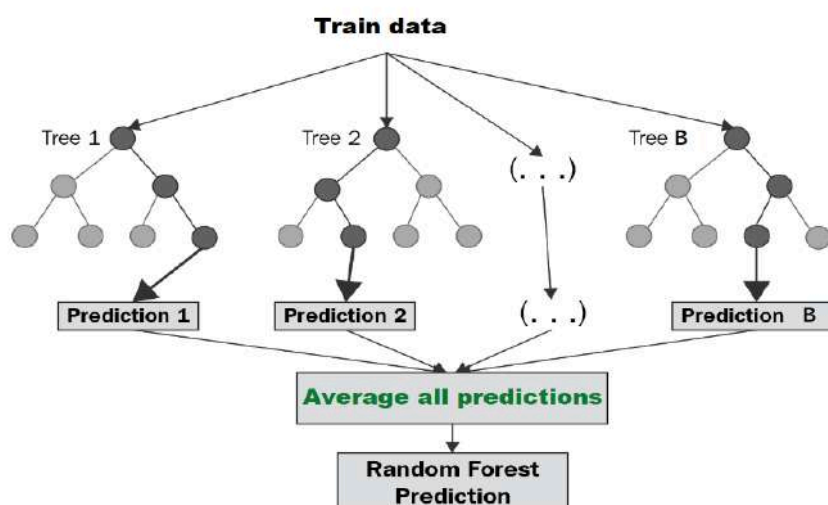


Ilustración 24. Esquema gráfico de funcionamiento de Random Forest Regressor.

## Métricas de evaluación

Para poder evaluar estos modelos y comparar resultados, se utilizan una serie de métricas de evaluación. Dichas métricas son, en general, la base de la evaluación de los modelos actuales del procesamiento del lenguaje natural, ya que son capaces de analizar y evaluar el modelo, sin la presencia de posibles medidas engañosas, que finalmente no presenten los resultados esperados.

### MSE (Mean Squared Error)

El error cuadrático medio es una métrica muy conocida dentro del mundo de Machine Learning, especialmente en tareas de regresión. Principalmente, mide el promedio del error al cuadrado entre los valores predichos por el modelo y los valores reales. Como se puede deducir, al elevar los errores al cuadrado, hace que se penalicen considerablemente y de manera más rotunda los errores grandes o los errores con desviaciones significativas [20]. Después de la ecuación y sus parámetros, se observa una gráfica representativa sobre el funcionamiento de esta métrica:

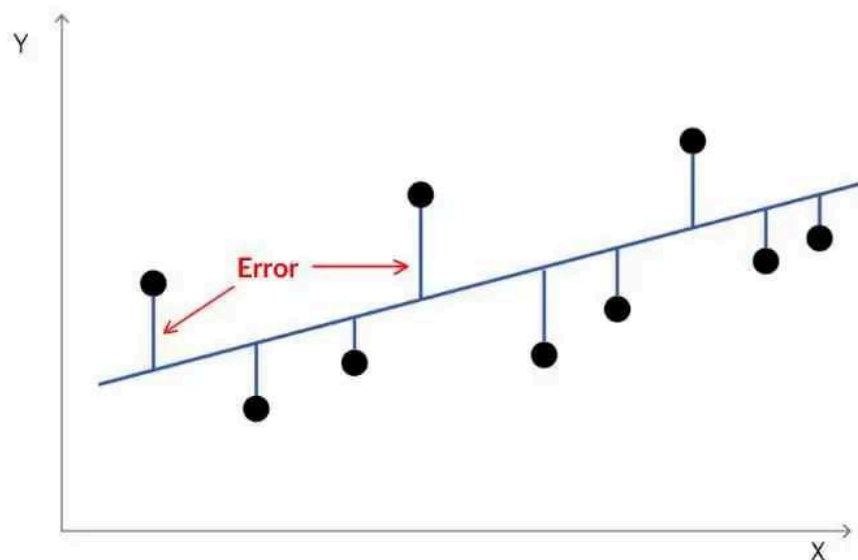
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Donde

n: Número de puntos de datos

$Y_i$ : Valores observados

$\hat{Y}_i$ : Valores predichos



*Ilustración 25. Esquema gráfico sobre el funcionamiento de las métricas de evaluación MSE y MAE.*

#### MAE (Mean Absolute Error)

Otra métrica de evaluación que sirve de gran ayuda es la conocida como MAE, muy usada también. Su funcionamiento se basa en medir la diferencia promedio entre los valores reales y los valores predichos por un modelo, dicho con otras palabras, es el promedio de las diferencias absolutas entre los valores observados y las predicciones. La diferencia más clara con respecto a la métrica anterior es que esta no eleva los errores al cuadrado, por lo que es menos sensible a los valores atípicos, sin embargo, esta diferenciación es la que buscamos entre las métricas para poder obtener enfoques y resultados diferentes que nos proporcionen información relevante [20]. Por lo tanto, se puede entender visualizando la fórmula que toma el valor absoluto de cada diferencia para cada uno de los puntos de datos, suma todas las diferencias absolutas y, por último, divide la suma por el número total de puntos de datos. En cuanto a la gráfica representativa de esta métrica, puede servir la gráfica anterior simplemente teniendo en cuenta que los errores de esta no son elevados al cuadrado.

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

Donde

n: Número de puntos de datos

Y<sub>i</sub>: Valores observados

Ŷ<sub>i</sub>: Valores predichos

#### R<sup>2</sup> (R-squared)

R<sup>2</sup> o coeficiente de determinación como también es conocida en una métrica que se basa en medir qué tan bien se ajusta el modelo de regresión al conjunto de los datos observados. Se encarga de evaluar la proporción de variabilidad total en la que la variable dependiente o etiqueta [22].

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

Donde

SS RES: suma del cuadrado de la distancia entre el punto real y el punto previsto en la línea de mejor ajuste

SS TOT: suma del cuadrado de la distancia entre el punto real y la media de todos los puntos

Y<sub>i</sub>: Valores observados

Ŷ<sub>i</sub>: Valores predichos

#### Cross validation

La validación cruzada, en español, es una técnica que permite evaluar el rendimiento de tu modelo dividiendo los datos en varios subconjuntos a los que se conocen como folds y poder entrenar así nuestro conjunto de datos en cada uno de los modelos en diferentes particiones y lograr que la evaluación sea más precisa y que no sufra de sobreajuste, ya que normalmente se ajusta demasiado con una única partición [23].

#### SHAP

Básicamente el algoritmo SHAP (Shapley Additive exPlanations) es una técnica encargada de interpretar y explicar el impacto de cada característica en las predicciones de un modelo de Machine Learning. Nos permite identificar de manera precisa y eficiente cuánto y cómo influye cada variable en una predicción específica [24].

#### Matriz de correlación

La matriz de correlación es una herramienta que mide la relación entre dos variables cuantitativas, ya sea la relación entre características como la relación entre características y nuestra salida o etiqueta.

## 3. Metodología

En este apartado se describe la metodología, con el objetivo de proporcionar una guía clara y detallada del proceso llevado a cabo. La metodología constituye una parte fundamental del trabajo, ya que define el conjunto de técnicas, herramientas y procedimientos utilizados para alcanzar los resultados obtenidos.

#### Índice K

##### Definición

Como bien se ha explicado y mencionado anteriormente, las tormentas geomagnéticas no son otra cosa que perturbaciones del propio campo magnético de la Tierra. La duración de estas puede ser desde varias horas hasta incluso varios días. Usando un lenguaje algo más técnico, podríamos decir que son producidas por un aumento brusco de las partículas emitidas en el ámbito solar, por lo que su origen es externo. Una apreciación que hay que tener en cuenta es que tienen un carácter global, lo que significa que comienza de manera simultánea en todos los puntos de la Tierra. No obstante, esto no quiere decir que todos los registros de todas las centrales u observatorios atmosféricos o geomagnéticos que captan estos sucesos produzcan los mismos datos ya que un factor que es clave en este estudio es que los valores obtenidos sobre las tormentas son mayores cuanto más altas son las latitudes.

Una vez que tenemos esto presente y en cuenta, el índice Kp, que data del año 1932, es el parámetro elegido y crucial para la investigación, por lo que lo podemos definir como un índice geomagnético de tipo cuasi-logarítmico, el cual, indica la perturbación del campo geomagnético a nivel local. El proceso de medición no es complejo, sino que simplemente

toma como referencia la curva de variación diaria de un día en calma del observatorio en el que se está midiendo durante intervalos de tres horas. (El sistema Kp de tres horas fue introducido por Julius Bartels en 1949 [25]).

## Instituciones y clases

La agencia NOAA de Estados Unidos [26] (National Oceanic and Atmospheric Administration) es la institución con los datos más completos que se han encontrado. Esta obtiene el índice Kp [27] por el cálculo de la media ponderada de los índices K observados en una red de 13 observatorios geomagnéticos repartidos por el mundo. Esta institución ha definido una escala que hoy en día es la utilizada oficialmente a nivel mundial que cuantifica la intensidad y los efectos de las tormentas geomagnéticas. A continuación, se muestra una imagen proveniente de su página web oficial que describe esta clasificación que consta de cinco posibles valores (G1 a G5) relacionados con los valores del índice Kp alcanzado, además de indicar la frecuencia promedio con que aparecen en cada ciclo solar:

Category		Effect	Physical measure	Average Frequency (1 cycle = 11 years)
Scale	Descriptor	Duration of event will influence severity of effects		
<b>Geomagnetic Storms</b>			Kp values* determined every 3 hours	Number of storm events when Kp level was met; (number of storm days)
<b>G 5</b>	Extreme	<u>Power systems:</u> widespread voltage control problems and protective system problems can occur, some grid systems may experience complete collapse or blackouts. Transformers may experience damage. <u>Spacecraft operations:</u> may experience extensive surface charging, problems with orientation, uplink/downlink and tracking satellites. <u>Other systems:</u> pipeline currents can reach hundreds of amps, HF (high frequency) radio propagation may be impossible in many areas for one to two days, satellite navigation may be degraded for days, low-frequency radio navigation can be out for hours, and aurora has been seen as low as Florida and southern Texas (typically 40° geomagnetic lat.).**	Kp=9	4 per cycle (4 days per cycle)
<b>G 4</b>	Severe	<u>Power systems:</u> possible widespread voltage control problems and some protective systems will mistakenly trip out key assets from the grid. <u>Spacecraft operations:</u> may experience surface charging and tracking problems, corrections may be needed for orientation problems. <u>Other systems:</u> induced pipeline currents affect preventive measures, HF radio propagation sporadic, satellite navigation degraded for hours, low-frequency radio navigation disrupted, and aurora has been seen as low as Alabama and northern California (typically 45° geomagnetic lat.).**	Kp=8, including a 9-	100 per cycle (60 days per cycle)
<b>G 3</b>	Strong	<u>Power systems:</u> voltage corrections may be required, false alarms triggered on some protection devices. <u>Spacecraft operations:</u> surface charging may occur on satellite components, drag may increase on low-Earth-orbit satellites, and corrections may be needed for orientation problems. <u>Other systems:</u> intermittent satellite navigation and low-frequency radio navigation problems may occur, HF radio may be intermittent, and aurora has been seen as low as Illinois and Oregon (typically 50° geomagnetic lat.).**	Kp=7	200 per cycle (130 days per cycle)
<b>G 2</b>	Moderate	<u>Power systems:</u> high-latitude power systems may experience voltage alarms, long-duration storms may cause transformer damage. <u>Spacecraft operations:</u> corrective actions to orientation may be required by ground control; possible changes in drag affect orbit predictions. <u>Other systems:</u> HF radio propagation can fade at higher latitudes, and aurora has been seen as low as New York and Idaho (typically 55° geomagnetic lat.).**	Kp=6	600 per cycle (360 days per cycle)
<b>G 1</b>	Minor	<u>Power systems:</u> weak power grid fluctuations can occur. <u>Spacecraft operations:</u> minor impact on satellite operations possible. <u>Other systems:</u> migratory animals are affected at this and higher levels; aurora is commonly visible at high latitudes (northern Michigan and Maine).**	Kp=5	1700 per cycle (900 days per cycle)

\* Based on this measure, but other physical measures are also considered.

\*\* For specific locations around the globe, use geomagnetic latitude to determine likely sightings (see [www.swpc.noaa.gov/Aurora](http://www.swpc.noaa.gov/Aurora))

Ilustración 26. Tabla de escalas del índice Kp.

- G1 - Menor (Kp = 5):

Pueden ocurrir fluctuaciones débiles en la red eléctrica, posibles impactos menores en operaciones de satélites y puede que los animales migratorios se vean afectados. Además, la aurora es visible en latitudes altas y en cuanto a la frecuencia promedio obtenemos 1720 por ciclo solar, que son unos 900 días.

- G2 - Moderada (Kp = 6):

En latitudes altas los sistemas electrónicos pueden experimentar alarmas de voltaje e incluso tormentas de larga duración pueden afectar a los transformadores. Se pueden requerir correcciones de orientación espacial por parte del control en la tierra y se puede ver afectadas las predicciones orbitales. La propagación de radio HF puede desvanecerse en

latitudes altas y la aurora puede hacerse aún más visible, de hecho, sigue cada vez haciéndose más visible en latitudes más bajas a medida que vayamos avanzando de gravedad en esta escala. Los datos de la frecuencia promedio son 600 eventos por ciclo solar (360 días).

- G3 - Fuerte ( $K_p = 7$ ):

En este nivel pueden ser necesarias correcciones de voltaje y se activarán alarmas en algunos dispositivos de protección. Los satélites pueden sufrir una carga de manera superficial y el arrastre atmosférico aumenta en satélites de órbita baja. En cuanto a la frecuencia promedio son 200 eventos por ciclo solar (130 días).

- G4 - Severa ( $K_p = 8$ ):

En el G4, los problemas de control de voltaje se ven generalizados y pueden ocasionar desconexiones activas clave en la red. En cuanto a las operaciones espaciales no varían respecto a la anterior etapa, pero, sin embargo, sí afectan a las corrientes inducidas en tuberías y la navegación por satélite se degrada por horas. Como se ha comentado anteriormente, la aurora sigue siendo cada vez más visible y hay 100 eventos por ciclo solar (60 días) de frecuencia.

- G5 - Extrema ( $K_p = 9$ ):

Por último, este es el tipo de tormenta geomagnética más fuerte que hay y se puede encontrar problemas generalizados tanto en el control del voltaje como en los sistemas de protección e incluso la red eléctrica podrían colapsar completamente causando apagones. Puede haber fallo en la comunicación con los satélites y las corrientes inducidas en tuberías pueden alcanzar cientos de amperios. Esta vez la aurora puede ser visible tanto en el sur de Florida como en el sur de Texas. Ahora la frecuencia promedio es de 4 eventos por ciclo solar (4 días).

El evento geomagnético más grande registrado en la historia es conocido como el evento Carrington que ocurrió entre el 1 y 2 de septiembre de 1859 provocando efecto geomagnético sin precedentes [28]. Hoy en día, eso supondría consecuencias devastadoras afectando a la red eléctrica, sistema de satélites, navegación aérea y comunicaciones globales. Estos eventos son extremadamente raros y los científicos aseguran que podría ocurrir cada 100-200 años.



*Ilustración 27. Eyecciones de masa coronal emergiendo de la superficie del Sol.*



A continuación, se observa otro gráfico de representación de escala del índice K, pero proporcionado por el Instituto Nacional Geográfico de España [32], que hace una definición más resumida y concreta para cada uno de los valores del índice K.

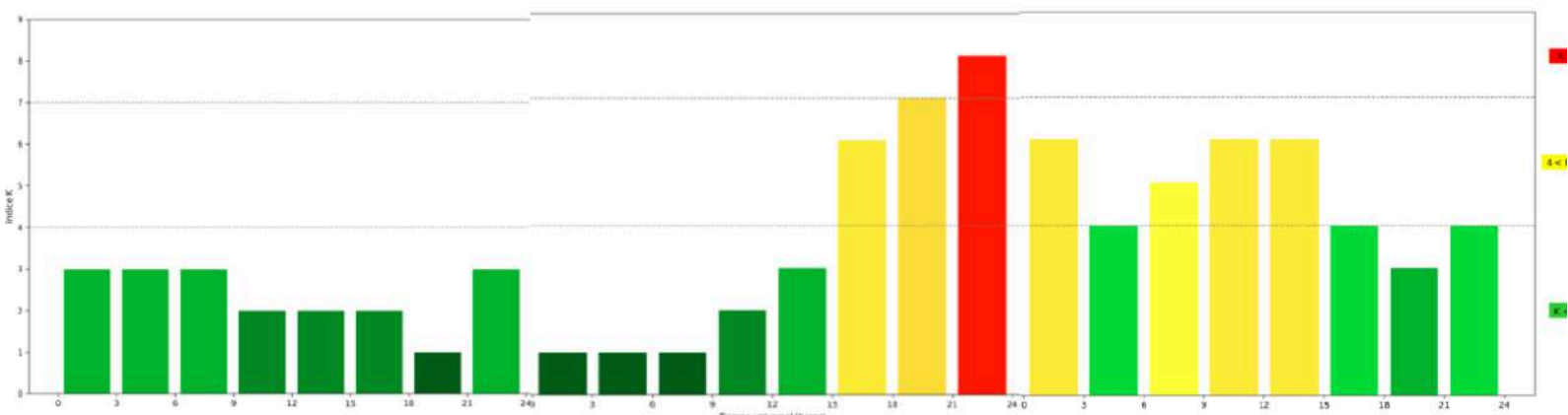
## Índice K

### Valor y significado

Valor del índice K	Significado
0	Campo geomagnético inactivo
1	Campo geomagnético muy tranquilo
2	Campo geomagnético tranquilo
3	Campo geomagnético intranquilo
4	Campo geomagnético activo
5	Tormenta solar menor
6	Tormenta solar mayor
7	Tormenta solar severa
8	Tormenta solar muy severa
9	Tormenta solar extremadamente severa

*Ilustración 28. Escala del índice K por el Instituto Nacional Geográfico de España [29].*

Una vez que ya se está familiarizado con el índice K y sus escalas y posibles repercusiones, se ha querido analizar la diferencia entre los datos obtenidos en diferentes instituciones y en diferentes bases de medición por todo el mundo. Las tres imágenes corresponden aproximadamente a los días 9, 10 y 11 de octubre de 2024. Como se puede observar y aunque estas tormentas geomagnéticas afectan de manera global, se ven pequeñas diferencias en los datos obtenidos cada tres horas. Esto se debe, como se ha comentado anteriormente, a que depende de la latitud y al haber tomado los datos desde diferentes puntos geográficos, es normal tener pequeñas variaciones en estas mediciones. Obviamente afectará directamente al estudio y al trabajo que se está realizando. A continuación, se presentan las gráficas correspondientes a en primer lugar, a la institución NOAA, que como he dicho previamente es la más fiable ya que recoge los datos de 13 observatorios repartidos por todo el mundo y hace la media. En cuanto a la segunda gráfica, corresponde al observatorio Geoforschungszentrum de la ciudad de Potsdam en Alemania. Por último, encontramos la gráfica correspondiente al Observatorio de San Pablo, ubicado en Toledo, provincia de España.





PLANETARY K-INDEX

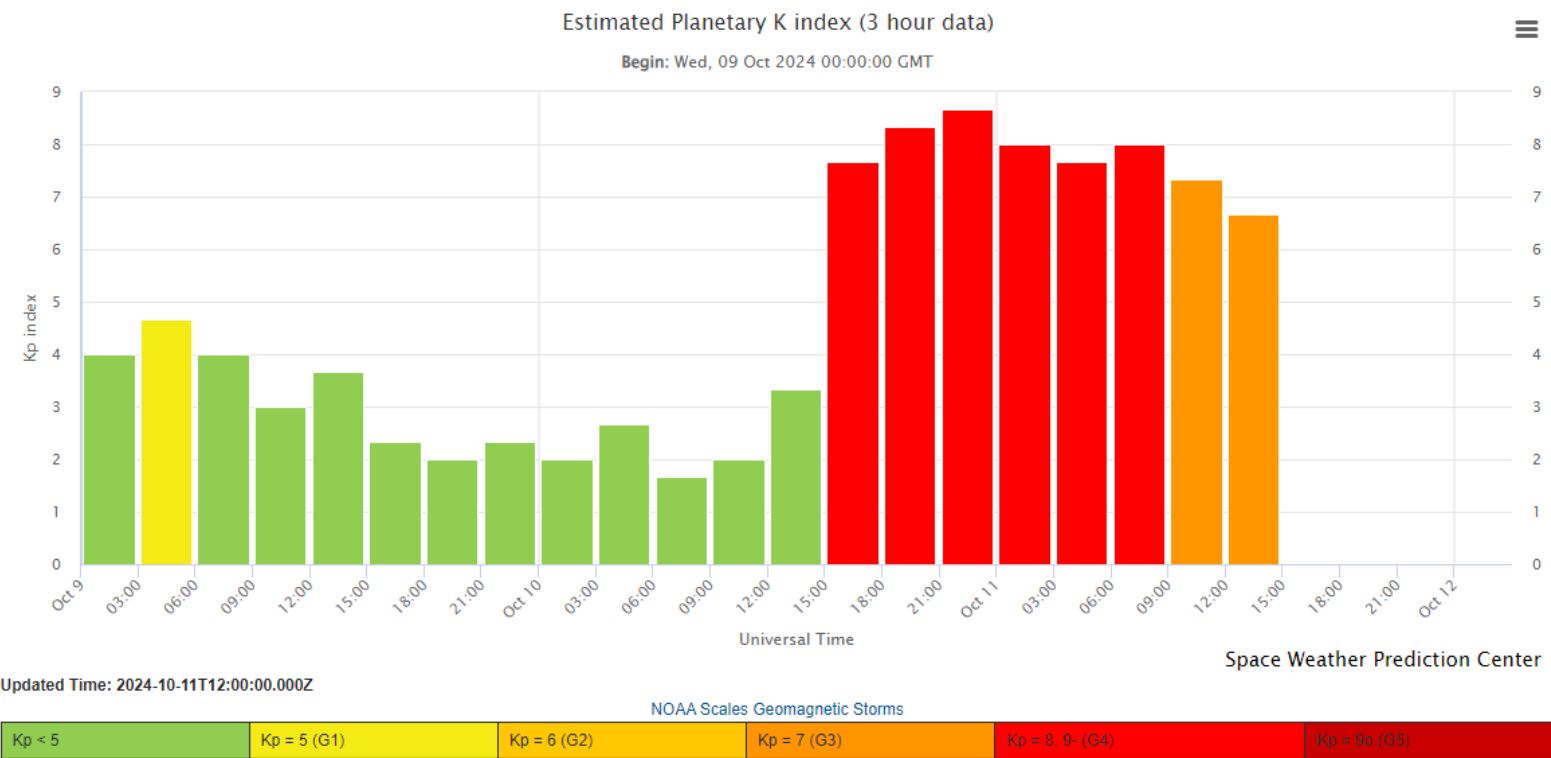


Ilustración 31. Índices K captados los días 9, 10 y 11 de octubre de 2024 por la institución NOAA [30].

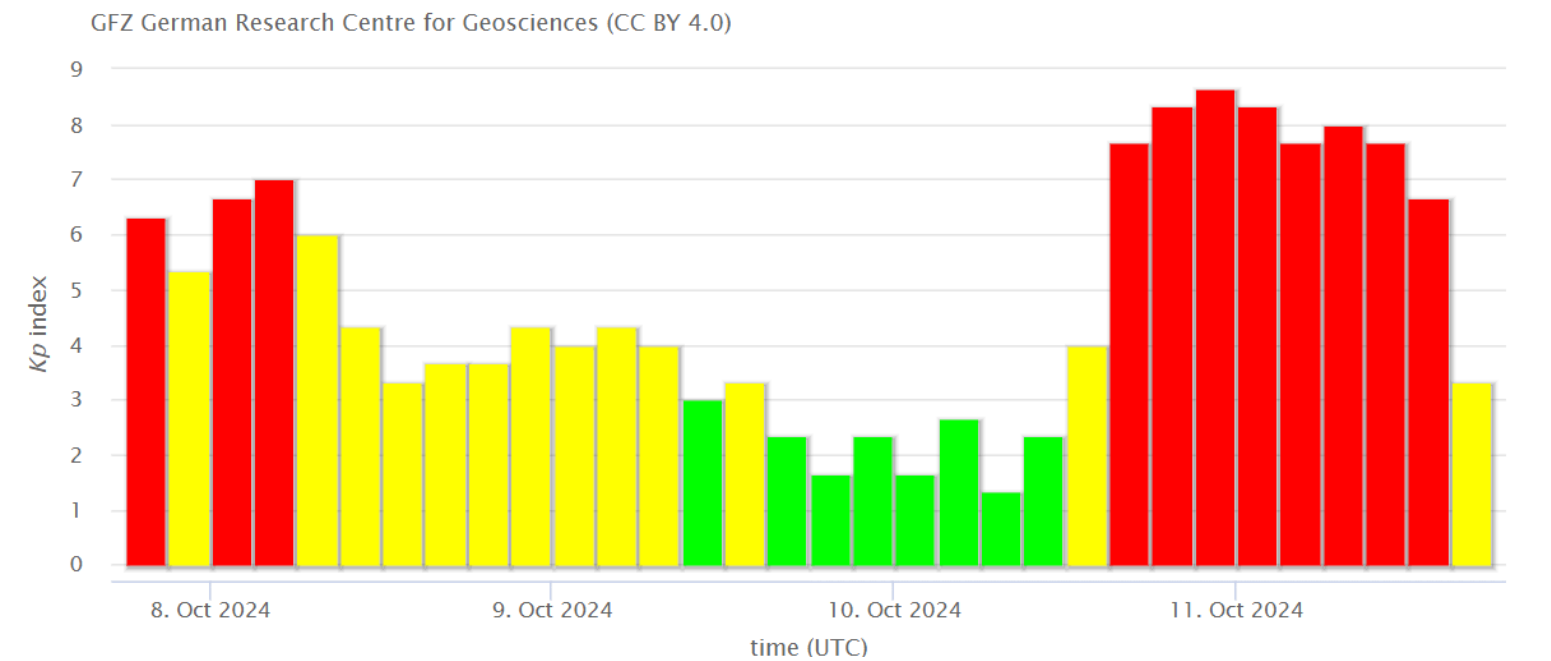


Ilustración 32. Índices K captados los días 9, 10 y 11 de octubre de 2024 por el observatorio de Potsdam, Alemania [31].

Tras una observación visual y rápida, la diferencia entre las tres gráficas es más que evidente. Observamos leves diferencias y esto es relevante en el proyecto debido a que desgraciadamente, no se ha podido obtener el conjunto de datos de índices K ni de NOAA ni del Observatorio de Toledo, España. Simplemente resulta que el conjunto de datos de los parámetros obtenidos de la central atmosférica de Granada hace referencia al intervalo de tiempo entre el 1 de abril de 2013 hasta el 31 de marzo de 2014 y en los sitios web oficiales de estas instituciones y observatorios no son públicos en el caso de NOAA y solo estaban disponibles a partir del año 2018 en el observatorio de Toledo. Es por este motivo que se han tenido que investigar otras fuentes de datos: el Observatorio de Potsdam en Alemania y el observatorio Hartland, Devon, Reino Unido. Esto interferirá como es de esperar en los resultados finales de este trabajo. En cuanto al observatorio Hartland, lo encontramos situado en el límite noroeste del pueblo de Hartland en Reino Unido. Fue construido con la intención específica de realizar trabajos magnéticos y lleva en funcionamiento de forma continua desde el año 1957 [33]. Por otro lado, nos encontramos con el centro Alemán de Investigación de Geociencias, Potsdam, Alemania, situado en el parque científico "Albert Einstein" en el Telegrafenberg de Potsdam. Lugar dónde se encuentran diferentes centros de estudios y que, desde hace unos 150 años, se han emitido y difundido información a escala mundial [34].

### Recolección de los datos

Se ha automatizado la obtención de los datos para todo un año teniendo así 21 parámetros físicos diferentes para cada uno de los magnetómetros. Estos parámetros son la energía y amplitudes y frecuencias de cada uno de los tres nodos. Para cada una de estas medidas se han obtenido el valor máximo, mínimo y la media de cada día durante un año. El siguiente paso será añadir a una tabla de Excel los datos comentados anteriormente y, además, una columna de clase para el índice K. Esto se realiza haciendo la media, ya que se debe tener cuenta el hecho de que los índices K, son obtenidos mediante 8 datos numéricos diferentes durante todo el día, cada uno referido a un periodo de 3 horas por lo que como los datos se encuentran asignados día a día, se hace la media de estos valores y así obtener la media del índice K para cada uno de los días. El resultado es una tabla con tantas filas como días tiene un año y 42 columnas de características (21 características cada magnetómetro) más la columna adicional de índices K.

Para ir comenzando y adentrándonos en un ámbito algo más técnico dentro de la inteligencia artificial, lo básico que se debe saber es que el conjunto de datos está dividido en características y etiquetas o salidas. Las características son aquellas columnas de datos referidos a los diferentes parámetros que hemos obtenido de nuestra central atmosférica de Granada, estos son: amplitudes, frecuencias y energías. Por otro lado, tenemos en nuestro caso una etiqueta, que es la columna de los K índices que hemos obtenido tanto del Observatorio de Hartland como del Observatorio de Potsdam y que utilizaremos como una salida y que, además, son los valores que tienen que predecir los algoritmos a través de las características ya dadas.

### Librerías, carga de datos y Z-score

El lenguaje de programación utilizado ha sido Python por lo que las diferentes librerías que se han utilizado han sido básicamente las necesarias para el funcionamiento del código usando mayoritariamente Numpy, Pandas y Sklearn entre otras. Por otro lado, para el cometido de cargar nuestros conjuntos de datos, se decide crear un archivo .ipynb

para guardar el código que se encarga de pasar las columnas que interesan del Excel a un archivo.csv, que es el que se puede leer dentro de Python de manera más sencilla. A continuación, se puede observar una muestra de las primeras cinco filas de cada una de las características y etiqueta además del tamaño completo del conjunto de datos, una vez que se han cargado correctamente.

Datos cargados correctamente. Tamaño: (365, 43)							
	NS_AS1_avg	NS_AS1_max	NS_AS1_min	NS_AS2_avg	NS_AS2_max	NS_AS2_min	\
0	0.382473	0.507315	0.246794	0.280533	0.398689	0.173649	
1	0.398728	0.538261	0.322621	0.286725	0.383550	0.221710	
2	0.405653	0.500604	0.308560	0.295767	0.429374	0.202853	
3	0.373040	0.494461	0.277721	0.298564	0.463861	0.192015	
4	0.366670	0.440001	0.309034	0.290928	0.387298	0.217184	
	NS_AS3_avg	NS_AS3_max	NS_AS3_min	NS_F1_avg	NS_F1_max	NS_F1_min	\
0	0.277095	0.381522	0.168621	7.831410	8.087190	7.468158	
1	0.274105	0.350953	0.195913	7.833635	8.030127	7.668224	
2	0.287091	0.396004	0.187405	7.774799	8.021052	7.301865	
3	0.286934	0.404189	0.172158	7.702141	8.165761	7.278110	
4	0.280243	0.360016	0.197132	7.731804	7.919654	7.525811	
	NS_F2_avg	NS_F2_max	NS_F2_min	NS_F3_avg	NS_F3_max	NS_F3_min	NS_E_avg
0	14.145274	14.521150	13.816782	20.648228	21.761773	20.203163	0.984126
1	14.270897	14.654839	13.995811	20.477838	20.842516	20.002941	0.968326
2	14.242236	14.651597	13.871505	20.489407	21.199519	19.956352	1.067938
3	14.144181	14.603731	13.805561	20.775133	22.129704	20.136640	1.038548
4	14.180970	14.495899	13.932230	20.618972	21.951913	20.177956	0.949963
	NS_E_max	NS_E_min	EW_AS1_avg	EW_AS1_max	EW_AS1_min	EW_AS2_avg	\
0	1.628977	0.392036	0.279278	0.493008	0.161037	0.271557	
1	1.502448	0.647136	0.270303	0.340182	0.214915	0.270582	
2	1.773943	0.568538	0.297832	0.450714	0.201613	0.270281	
3	1.900001	0.475337	0.290051	0.514776	0.174581	0.302756	
4	1.356466	0.624455	0.286243	0.411090	0.203144	0.318797	
	EW_AS2_max	EW_AS2_min	EW_AS3_avg	EW_AS3_max	EW_AS3_min	EW_F1_avg	\
0	0.487718	0.115658	0.262043	0.471839	0.121639	7.603861	
1	0.370583	0.182938	0.265970	0.335003	0.202916	7.576281	
2	0.466308	0.174192	0.254233	0.419559	0.168409	7.649983	
3	0.600665	0.151591	0.291753	0.570075	0.128003	7.551389	
4	0.437938	0.183774	0.318451	0.418857	0.213477	7.510949	
	EW_F1_max	EW_F1_min	EW_F2_avg	EW_F2_max	EW_F2_min	EW_F3_avg	\
0	7.889263	7.026908	13.962667	14.307406	13.705846	20.526839	
1	7.819992	7.075143	13.962911	14.365643	13.520482	20.488859	
2	8.061557	7.094105	14.109723	14.471599	13.722850	20.516616	
3	7.971114	7.067626	14.029450	14.394677	13.656657	20.815602	
4	7.944729	7.024607	13.976704	14.507190	13.401925	20.759043	
	EW_F3_max	EW_F3_min	EW_E_avg	EW_E_max	EW_E_min	K_avg	
0	22.643476	19.808904	0.762515	2.183683	0.170785	1.125	
1	21.233372	19.896046	0.673415	1.109763	0.409363	0.750	
2	22.618413	19.567505	0.744738	1.810996	0.323238	0.500	
3	22.766352	20.016181	0.928469	2.923666	0.221072	0.875	
4	22.576925	20.049261	0.932153	1.536104	0.450354	0.375	
Tamaño del conjunto de entrenamiento: (292, 42)							
Tamaño del conjunto de prueba: (73, 42)							

Ilustración 29. Primeras 5 filas de nuestro conjunto de datos cargado y su tamaño.

NS y EW se refieren a los magnetómetros norte-sur y este-oeste. AS, F y E corresponden a amplitudes, frecuencias y energía, mientras que los números (1,2,3) hacen referencia al modo en el que se encuentran. Por último, avg, max y min se refieren a la media, el valor máximo y el valor mínimo. Como se observa, obtenemos un conjunto total de 365 filas y 43 columnas donde la última se trata de nuestra salida, el índice K. También es necesario comentar que se ha dividido el conjunto de datos en un 80% de entrenamiento (292 filas de 42 columnas) y un 20% de validación (73 filas de 42 columnas). A continuación, se ha aplicado la herramienta anteriormente comentada Z-score que para poder detectar outliers en el conjunto de datos se calcula Z-score de cada observación y así filtrar aquellos con Z-score fuera del rango  $[-3, 3]$ .

**Filas con outliers: 36**

**Tamaño de X\_train después de aplicar clipping: (292, 42)**

*Ilustración 30. Resultado de haber implementado la herramienta de Z-score.*

Como se observa, después de la aplicación de esta herramienta en nuestro conjunto de datos, hay un total de 36 filas que son consideradas como outliers. Una vez conocido este dato, hay dos opciones, o eliminar estas filas de datos o minimizar su importancia y que no afecten de manera relevante al análisis de nuestro conjunto de datos. Finalmente, se ha decidido minimizar su importancia y no eliminar dichas filas debido a que el conjunto de datos no es muy extenso y podría suponer un mal mayor. La técnica que se ha utilizado para restarle importancia a estas filas se llama clipping. Es por eso por lo que el tamaño del 80% de entrenamiento sigue siendo de 292 filas y 42 columnas. Significa que no se ha disminuido el número de filas pese a que 36 de ellas hayan sido clasificadas como outliers. No se han eliminado sino reducido su importancia.

### Los conjuntos de datos empleados

Primeramente y antes de indagar de lleno en la parte más resolutiva y quizás menos teórica del estudio, se va a presentar cada uno de los conjuntos de datos seleccionados. Son un total de cuatro: dos de ellos contienen las 42 características con las que contaban originalmente y se diferencian en el origen del índice K, es decir, unos datos provienen del observatorio de Hartland, Reino Unido y otros del observatorio de Potsdam, Alemania. De los dos conjuntos de datos restantes utilizados e implementados, uno de ellos se basa en un conjunto donde en vez de tener los máximos, mínimos y las medias de cada uno de los nodos, para cada uno de los parámetros, esta vez únicamente seleccionaremos las medias de cada uno de los parámetros, eliminando así todas las columnas de máximos y mínimos. Este conjunto de datos tendrá 14 características más la columna de la salida del índice K. Por último, el conjunto de datos que nos queda es en el que usamos únicamente las frecuencias, es decir, por razones que se explican más adelante, se ha decidido seleccionar únicamente las columnas que hacen referencia a los datos de las frecuencias y eliminando así las columnas que hacen referencia a las características relacionadas con los parámetros de amplitud y energía, quedando 18 columnas de características más la salida.

## Elección de los algoritmos de aprendizaje automático

En este momento surge una cuestión, utilizar algoritmos de regresión o de clasificación. Básicamente la principal diferencia entre algoritmos de regresión y de clasificación es que los algoritmos de regresión predicen un resultado numérico y los algoritmos de clasificación te dicen un resultado de clases, es decir, se predice de qué clase son los índices (como hemos visto anteriormente hay desde la clase G1 hasta la G5). Esta última era la primera idea, sin embargo, tras observar los conjuntos de datos finales, se encuentran desbalanceados. Significa que alrededor del 85% de los datos hacen referencia a la clase G1 y el 15% restante hacía referencia a las clases G2, G3, G4 y G5. Como se puede intuir, para que un conjunto de datos esté balanceado, tiene que haber un porcentaje similar para cada una de las clases y nuestro conjunto de datos no lo cumple. Por lo cual, llegados a este punto teníamos que elegir entre seguir usando algoritmo de clasificación, pero tener que añadir algunos algoritmos simples como SMOTE o CTGAN, que se encargan de ampliar el conjunto de datos, en este caso de las clases que menos porcentaje tendrían, hasta llegar a un conjunto de datos global balanceado. No obstante, no se selecciona esta opción ya que se considera que gran valor de este estudio reside en que se ha realizado puramente con datos reales y básicamente lo que hacen estas herramientas es generar o ampliar una base de datos a partir de datos artificiales, aunque estén estos sacados de los reales siguen siendo artificiales. Aun así, es una técnica popularmente usada pero finalmente se usan algoritmos de regresión. Básicamente, estos algoritmos predicen a través de las características, una etiqueta de índice K y es un valor numérico. La elección de los algoritmos no ha sido sencilla ya que se ha probado alrededor de diez modelos y tras ver cómo funcionan con nuestros conjuntos de datos se han seleccionado los que explicaremos a continuación. El motivo del descarte ha sido porque en cuanto a su funcionamiento, o son bastante parecidos a los que se han elegido o porque directamente el rendimiento del modelo frente a nuestro conjunto de datos era menor que el de los modelos escogidos.

## Estandarización de los datos

Antes de implementar el modelo a nuestro conjunto de datos, hemos tenido que realizar una estandarización de los datos. Es esencial en este modelo ya que se debe hacer cuando las características tienen diferentes unidades, como las nuestras, que en el caso de que los valores sean más grandes entre una característica y otra, pueden hacer que se desbalancee el modelo. El hecho de estandarizar las características nos asegura que todas ellas contribuyen por igual a la predicción y no que algunas puedan tener mayor impacto del que deben.

```
# Estandarización de Los datos
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

*Ilustración 31. Líneas de código referidas a la estandarización de los datos.*

## SVR

Este modelo, como se ha definido en secciones anteriores, es especialmente eficaz en espacios de alta dimensión y cuando el número de características es significativo, en este caso 42 características y 365 muestras por lo que lo hace ideal para nuestro caso. El kernel RBF (Radial Basis Function), también conocido como kernel Gaussiano, es el que se ha seleccionado y es uno de los más usados tanto en SVM como en SVR. Se ha elegido dado que es el que mejor rendimiento nos daba y está especializado cuando los datos están en un espacio de mayor dimensión, lo que permite encontrar patrones más complejos de los datos. Junto al kernel RBF, se encuentra el parámetro C, que controla el grado de penalización del error en el modelo SVR. Originalmente este parámetro se encuentra con el valor 1.0 y se ha elevado a 3.0, lo que significa que el modelo penalizará los errores más de lo habitual, pero no de manera demasiado estricta.

```
# Definir el modelo SVR
svr_model = SVR(kernel='rbf', C=3.0) # kernel (rbf, linear o poly) y el valor de regulación C
```

*Ilustración 32. Líneas de código referidas a la elección de parámetros del modelo SVR.*

## XGBoost Regressor

En cuanto a los parámetros de este modelo, se ha variado el `learning_rate` decreciéndolo considerablemente hasta el valor de 0,05. El parámetro se encarga de controlar cuánto cambia el modelo en respuesta a cada nuevo árbol que se añade, por lo que, al disminuir dicho valor, el modelo avanzará más lentamente (algo que no es de relevancia ya que este modelo es muy rápido y ágil computacionalmente hablando y no se ha notado apenas diferencia). Otra ventaja de haber modificado el modelo con un `learning_rate` bajo, es que reduce el riesgo de sobreajuste. En cuanto al parámetro `max_depth = 3`, define la profundidad máxima del árbol en el modelo y el valor 3 indica que no son muy profundos, por lo que este modelo capturará iteraciones simples entre las características al contrario que el modelo anterior para así tener un rango de estudio de nuestro conjunto de datos diferente. No ha sido necesario variar ninguno de los parámetros restantes.

```
# Definir el modelo XGBoost Regressor
xgboost_model = XGBRegressor(n_estimators=100, learning_rate=0.05, max_depth=3, random_state=42)
```

*Ilustración 33. Líneas de código referidas a la elección de parámetros del modelo XGBoost Regressor.*

## Random Forest Regressor

En cuanto a los parámetros, el único que hemos observado que mejora en modelo es el número de estimaciones (`n_estimators`). Dicho parámetro se refiere a que en vez de que el modelo generase 100 árboles de decisión, como originalmente sucedía, ahora genera 1000 árboles de decisión. La modificación implica que se mejora la estabilidad y precisión del modelo ya que cuantos más árboles tiene el modelo, más robusto es ante las variaciones de los datos. Sin embargo, el modelo que estamos estudiando, no se caracteriza por ser ágil



computacionalmente hablando y mucho menos en este caso que además aumentamos de 100 a 1000 el número de árboles que tiene que generar. No obstante, con el número de estimaciones original tardaba en obtener los resultados alrededor de 25 segundos y ahora, con la actualización de los parámetros, alrededor de 2 minutos. Un cambio que siempre hay que tener en cuenta pero que en este caso no supone más allá de un minuto y medio debido a que el conjunto de datos tampoco es precisamente extenso. Finalmente, se ha decidido aceptar esta penalización o desventaja, pero a su vez aumentando el rendimiento del modelo. Se ha probado a estudiar los demás parámetros, pero la modificación de estos no ayudaba al resultado final, por lo que es lo único que se ha variado.

```
# Definir el modelo Random Forest Regressor
rf_model = RandomForestRegressor(n_estimators=1000, max_depth=10, random_state=42)
```

*Ilustración 34. Líneas de código referidas a la elección de parámetros del modelo Random Forest Regressor.*

Después de la explicación de los parámetros y aplicación de cada uno de los modelos de Machine Learning, se habla a continuación sobre la interpretación de cada una de las métricas de evaluación previamente descritas.

## MSE

La manera de interpretar este modelo es que cuanto más bajo sea el valor numérico del resultado, más fiable y eficiente será el modelo. La peculiaridad más particular de esta métrica de evaluación es que es muy sensible a los valores atípicos, también llamados outliers, por lo que gracias a que previamente se ha aplicado en nuestro conjunto de datos la herramienta Z-score, que reducía el peso de 36 filas clasificadas como outliers, esta métrica ha mejorado de manera considerable.

## MAE

Al igual que antes, cuanto mayor cercano a cero esté el resultado mayor precisión le corresponde. Un apunte importante que se debe tener en cuenta a la hora de leer estos resultados es que se expresan en las mismas unidades que los datos originales, lo que quiere decir, que si tenemos un error de 0,5 significa que, de media, el error para cada uno de los puntos de datos es de 0,5 en la unidad de la salida o etiqueta.

## R<sup>2</sup>

En cuanto a su interpretación, si el resultado es cero, indica que el modelo no es capaz de aplicar ninguna variabilidad en los datos y no ha conseguido esa relación que se busca, de otro modo, si el resultado es uno o cercano, indica que el modelo relaciona perfectamente toda la variabilidad de los datos. Dicho de otro modo, cuanto más esté el resultado cerca del valor cero, más bajo será el poder predictivo del modelo. También se debe comentar que la principal desventaja de esta métrica es que cuantas más características tenga el conjunto de datos con el que esté trabajando nuestro algoritmo, más dificultad tendrá el algoritmo para relacionar cada una de las características con el resultado. En lo que al conjunto de datos respecta, al tener gran cantidad de características, no es de extrañar tener un resultado, teóricamente hablando, en esta métrica.

## Cross Validation

En este caso, hemos vuelto a calcular tanto la métrica MSE como la métrica  $R^2$  para comprobar si pudiéramos mejorar aún más sus resultados. Como se puede observar en la captura de las dos líneas del código, se ha integrado esta técnica con un número de cinco folds (cv), lo que implica que los datos se dividirán en cinco subconjuntos con un tamaño parecido y cuatro de ellos se entrenan y uno se evalúa. El proceso a su vez se repetirá cinco veces cambiando el conjunto de evaluación. Para obtener el valor numérico del resultado final, se calculan las métricas para cada fold y finalmente se hará la media para obtener una estimación del rendimiento del modelo. Se han escogido que sean cinco folds porque es lo que mejor se ajustaba y mejores resultados proporciona a nuestros modelos y conjuntos de datos. En el caso de que se aumente aún más dicho valor, ya no se obtienen mejores resultados y lo único que se consigue es incrementar la carga computacional.

```
# Aplicar cross-validation con 5 folds
scores_mse = cross_val_score(rf_model, X_train, y_train, cv=5, scoring='neg_mean_squared_error')
scores_r2 = cross_val_score(rf_model, X_train, y_train, cv=5, scoring='r2')
```

*Ilustración 35. Líneas de código referidas a la elección de parámetros de la herramienta de Cross Validation.*

## SHAP

Un dato relevante y que se debe apreciar antes de comenzar con el análisis de los gráficos, es que solo existen los gráficos representando a los modelos XGBoost y Random Forest Regressor ya que son los más compatibles con SHAP. Se debe, principalmente, a que pueden descomponer sus predicciones como la suma de contribuciones de cada árbol, lo cual es similar a cómo funciona la técnica de SHAP. Sin embargo, SVR utiliza un enfoque de separación de margen en el espacio de características, lo que hace que esta relación entre características y predicciones sea menos directa y para la herramienta gráfica, sea difícil su interpretación, además, que generalmente es menos fiable. Dichos motivos han servido para tomar la decisión de únicamente representar estos dos algoritmos y como los resultados han sido similares en cada uno de los modelos, no aportaría más información la representación del modelo SVR de la que ya tenemos.

## Matriz de correlación

En cuanto a su funcionamiento, los valores pueden ir de -1 a 1, donde 1 expresa una correlación perfecta positiva y relación directa. Sin embargo, cuando es -1 indica una correlación perfecta negativa lo que implica que cuando una variable aumenta la otra disminuye y viceversa. En el caso de que este valor sea 0 significa que no hay ninguna correlación entre ambas variables, es decir, no existe una relación lineal entre ambas.

## 4.Resultados

Se pone en común en tablas los resultados de todos los algoritmos de aprendizaje automático del conjunto de datos de ambos magnetómetros y de diferentes conjuntos de datos. La primera tabla que se visualiza será la obtenida del Observatorio de Potsdam, Alemania. Se refiere al conjunto completo sin reducción de filas, es decir, están las 42 columnas de características. (Se sombreadá de amarillo los mejores resultados para cada una de las métricas de evaluación en cada una de las tablas).

Modelo	MSE cross-validation	R <sup>2</sup> cross-validation	MSE	MAE	R <sup>2</sup>
Support Vector Regressor	0,54309	0,44266	0,46865	0,53543	0,44962
XGBoost Regressor	0,57704	0,41633	0,53413	0,55844	0,37272
Random Forest Regressor	0,57452	0,41732	0,53761	0,57229	0,36864

*Tabla 1. Resultados con el conjunto de datos del observatorio de Potsdam, Alemania.*

La tabla siguiente cumple con los mismos principios que la anterior (sin eliminación de características), simplemente los datos de índices K obtenidos son del observatorio de Hartland, Reino Unido.

Modelo	MSE cross-validation	R <sup>2</sup> cross-validation	MSE	MAE	R <sup>2</sup>
Support Vector Regressor	0,51539	0,37943	0,34750	0,43638	0,59446
XGBoost Regressor	0,50848	0,38407	0,43942	0,49092	0,48720
Random Forest Regressor	0,53767	0,35078	0,38667	0,48990	0,54875

*Tabla 2. Resultados con el conjunto de datos del observatorio de Hartland, Reino Unido.*

La tabla a continuación, se basa en una reducción de columnas respecto al conjunto de datos completo original que hemos utilizado anteriormente. Como se indica en secciones anteriores, se trata de únicamente de las 18 columnas de características referidas a las frecuencias. Más adelante se explica el por qué se ha elegido justamente esta clasificación:

Modelo	MSE cross-validation	R <sup>2</sup> cross-validation	MSE	MAE	R <sup>2</sup>
Support Vector Regressor	0,51465	0,37892	0,53253	0,54937	0,37854
XGBoost Regressor	0,50787	0,38274	0,47017	0,53090	0,45130
Random Forest Regressor	0,53884	0,34732	0,41581	0,51031	0,51475

*Tabla 3. Resultados con el conjunto de datos de frecuencias.*

La siguiente tabla se refiere al conjunto sin las columnas que hacen referencia a los máximos y a los mínimos, dejando únicamente las medias de cada uno de los parámetros: amplitud, frecuencia y energía. Los resultados obtenidos son:

Modelo	MSE cross-validation	R <sup>2</sup> cross-validation	MSE	MAE	R <sup>2</sup>
Support Vector Regressor	0,58224	0,30040	0,43031	0,49108	0,49783
XGBoost Regressor	0,49688	0,39825	0,44810	0,51073	0,47707
Random Forest Regressor	0,53244	0,35651	0,41011	0,49728	0,52140

*Tabla 4. Resultados con el conjunto de datos de medias.*

En la tabla que se muestra a continuación, se observan los mejores resultados de cada una de las tablas anteriores haciendo referencia a cada uno de los conjuntos de datos anteriores y lo volveremos a comparar de nuevo sombreando los mejores resultados.

Conjuntos de datos	MSE cross-validation	R <sup>2</sup> cross-validation	MSE	MAE	R <sup>2</sup>
Conjunto de datos Alemania	0,54309	0,44266	0,46865	0,53543	0,44962
Conjunto de datos UK	0,50848	0,35078	0,34750	0,43638	0,59446
Conjunto de datos de únicamente frecuencias	0,50787	0,38274	0,41581	0,51031	0,51475
Conjunto de datos de únicamente las medias	0,49688	0,39825	0,41011	0,49108	0,52140

*Tabla 5. Comparativa de resultados entre los mejores de cada conjunto de datos.*

Finalmente, en la tabla anterior se puede observar visualmente los valores referidos a las métricas de Cross-Validation de MSE y R<sup>2</sup> son peores que los propios MSE y R<sup>2</sup>. Por lo tanto, en la siguiente tabla se hace de nuevo una última distinción con la que se obtienen esta vez los mejores resultados posibles. Se puede visualizar que el conjunto de datos completo del Observatorio de Hartland Reino Unido, es el que mejores resultados otorga, representados en la siguiente tabla:

	MSE	MAE	R <sup>2</sup>
Mejor valor final	0,34750	0,43638	0,59446

*Tabla 6. Mejores resultados finales para cada modelo.*

En cuanto a los modelos más distinguidos y que mejor rendimiento han proporcionado, principalmente en las métricas con Cross-Validation ha sido el modelo XGBoost Regressor, aunque con alguna excepción. Sin embargo, en cuanto a las métricas sin la herramienta de Cross-Validation y las que obtienen los mejores resultados finales en la tabla anterior, han sido enteramente proporcionadas por el modelo SVR.

## Comparaciones de gráficas de resultados y de las relaciones obtenidas

En el siguiente apartado se usan, gracias a Python, diferentes técnicas de representación gráfica que nos ayudan a poder entender la interacción entre las características y la salida de índice K además de la importancia en el modelo de cada una de ellas. Se aprecian diferentes imágenes y gráficos sobre nuestros modelos. Se va a tratar con el algoritmo SHAP y con la herramienta gráfica de matriz de correlación, ambos definidos y explicados en apartados anteriores. Ya dentro del algoritmo de SHAP, el primer gráfico se trata del llamado Bar Chart y obtenemos información sobre la importancia media de cada una de las características.

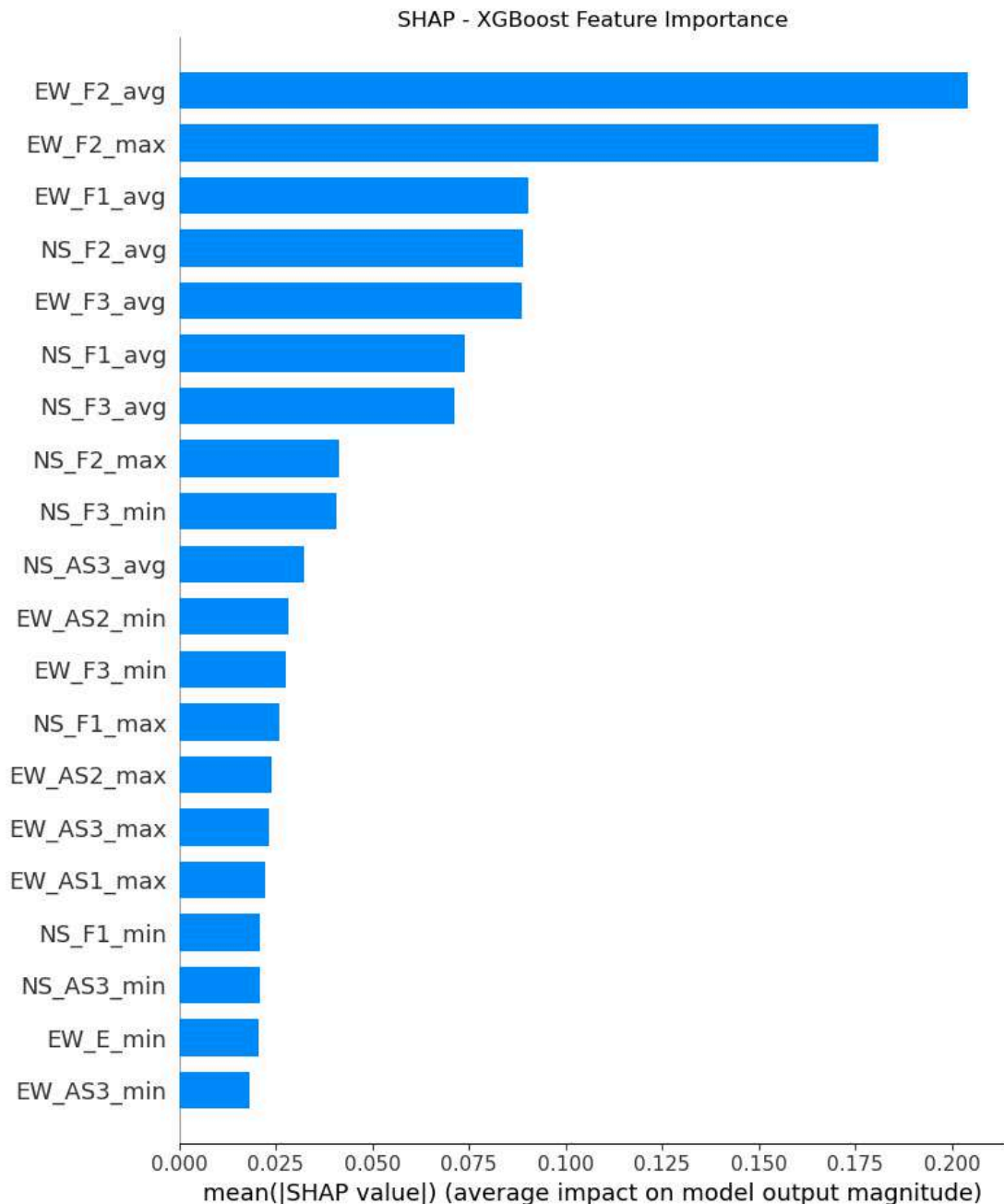
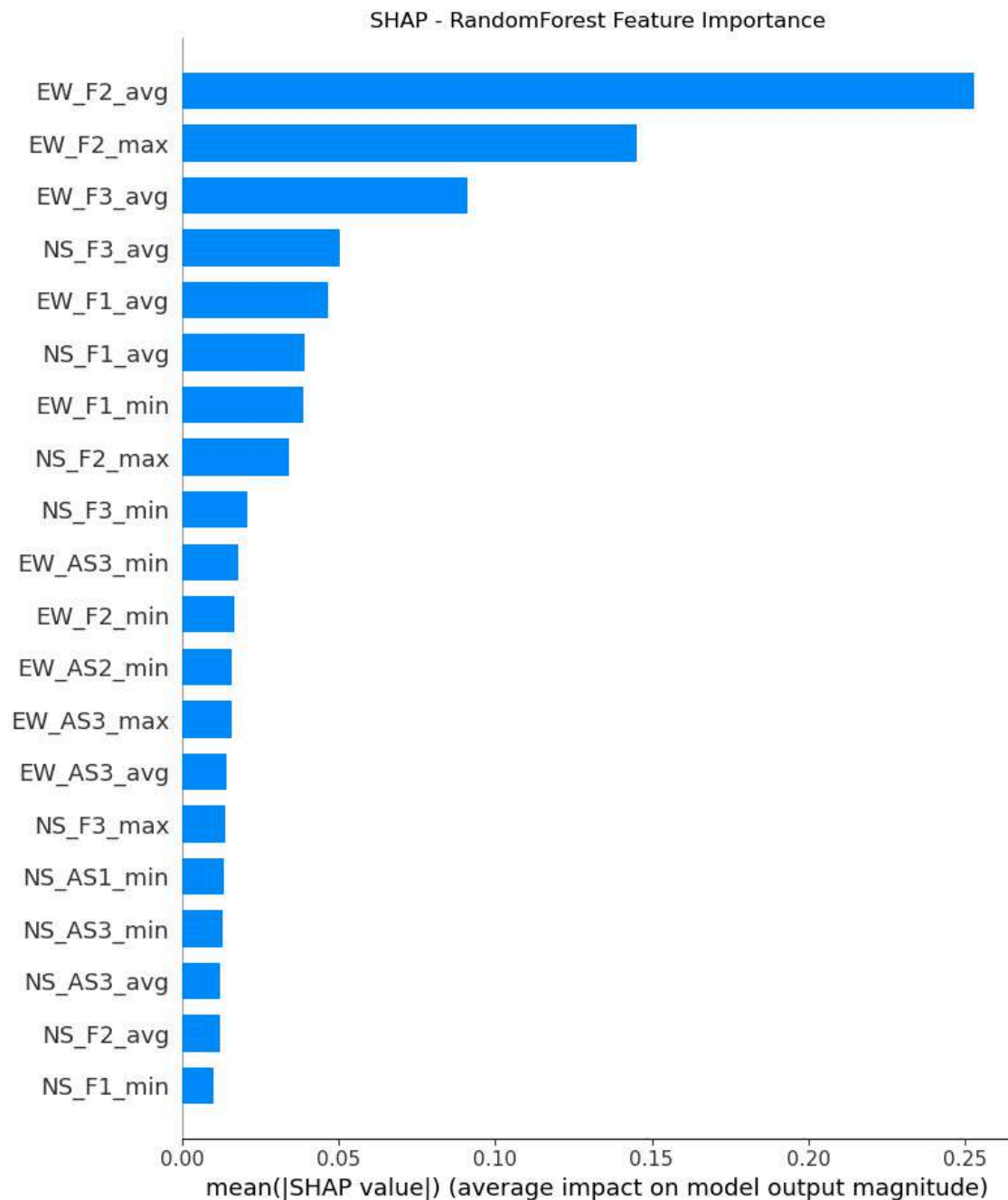


Ilustración 36. Gráfico Bar Chart referido al modelo XGBoost Regressor.





*Ilustración 37. Gráfico Bar Chart referido al modelo Random Forest Regressor.*

Como se observa, el gráfico muestra las características más importantes dentro de nuestros modelos. Prácticamente todas las características relacionadas con la frecuencia están por encima del resto, es por esto por lo que se tomó la decisión anteriormente de crear un subconjunto a estudiar reduciendo el conjunto de datos completo a solo las características de frecuencia para ver si así mejora la eficiencia de nuestro modelo o se puede sacar alguna conclusión más clave. No obstante, aunque se obtienen resultados cercanos, no se obtiene una mejoría en cuanto a los valores de las métricas de evaluación, pero sí sirve para alguna figura representativa más adelante. El siguiente par de gráficos siguen siendo de la técnica SHAP, pero esta vez son los llamados Summary Plot o Dot Plot de SHAP. Este gráfico nos proporciona una vista más detallada de cómo cada característica influye en el modelo.

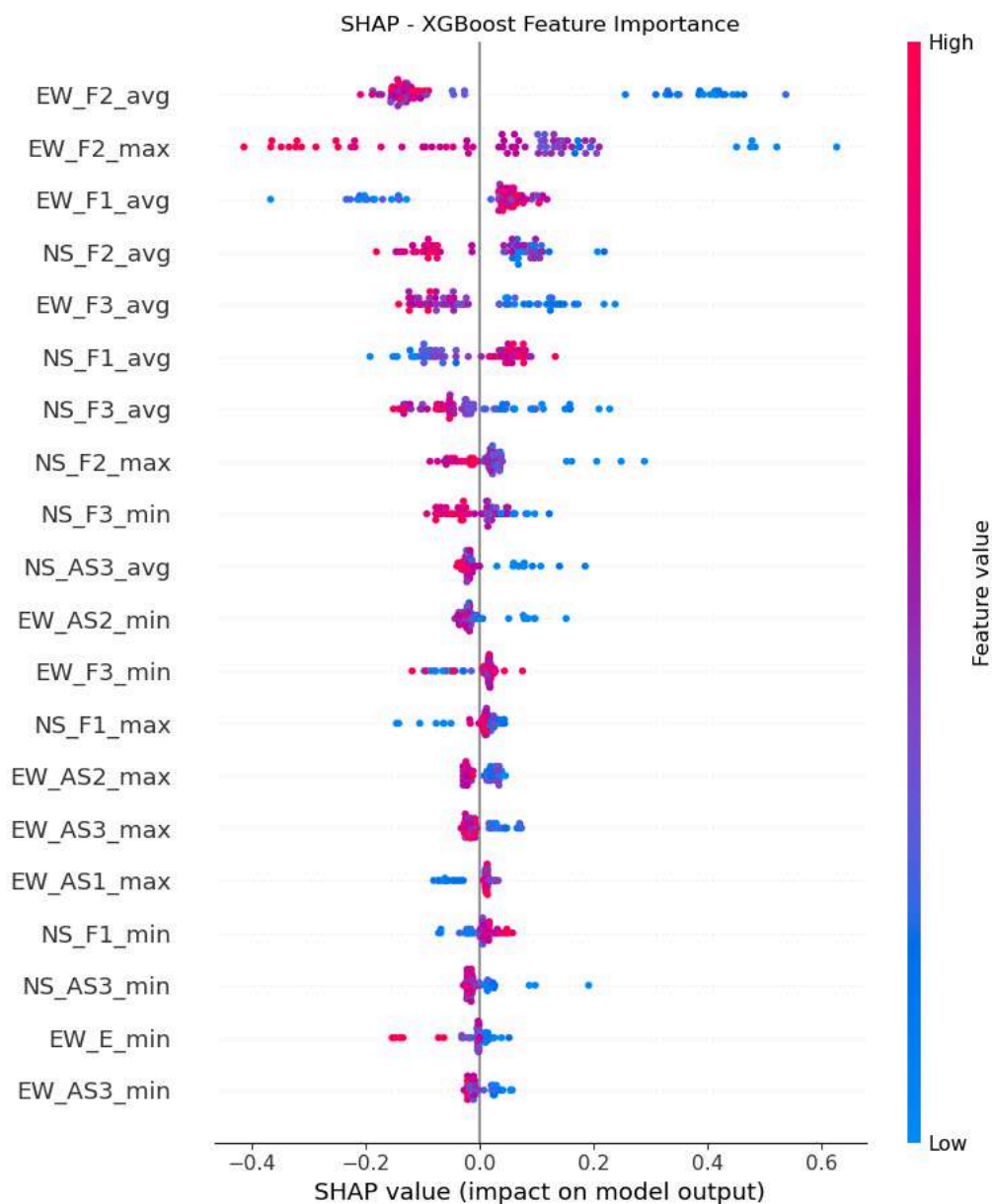


Ilustración 38. Gráfico Summary Plot o Dot Plot referido al modelo XGBoost Regressor.

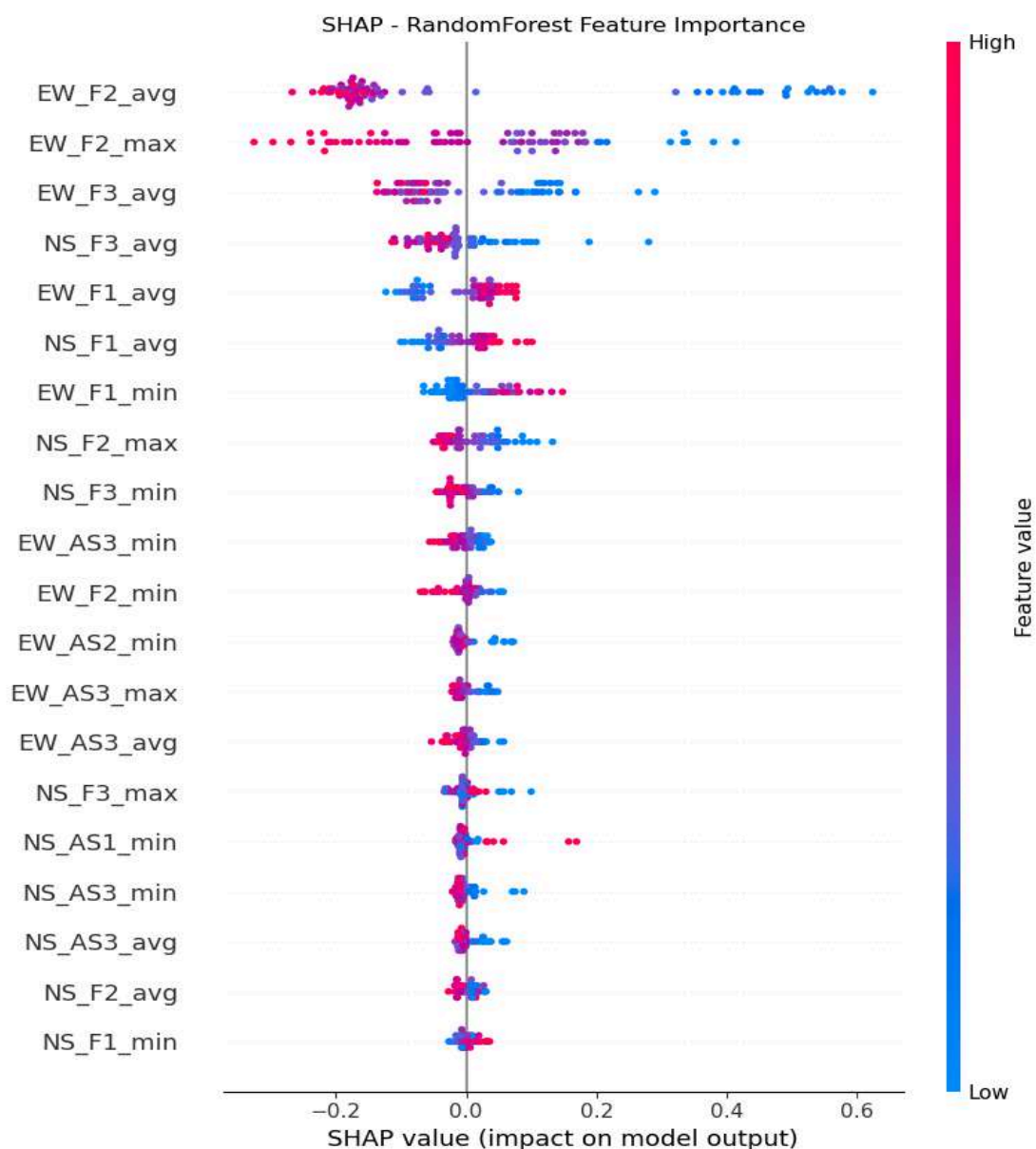


Ilustración 39. Gráfico Summary Plot o Dot Plot referido al modelo Random Forest Regressor.

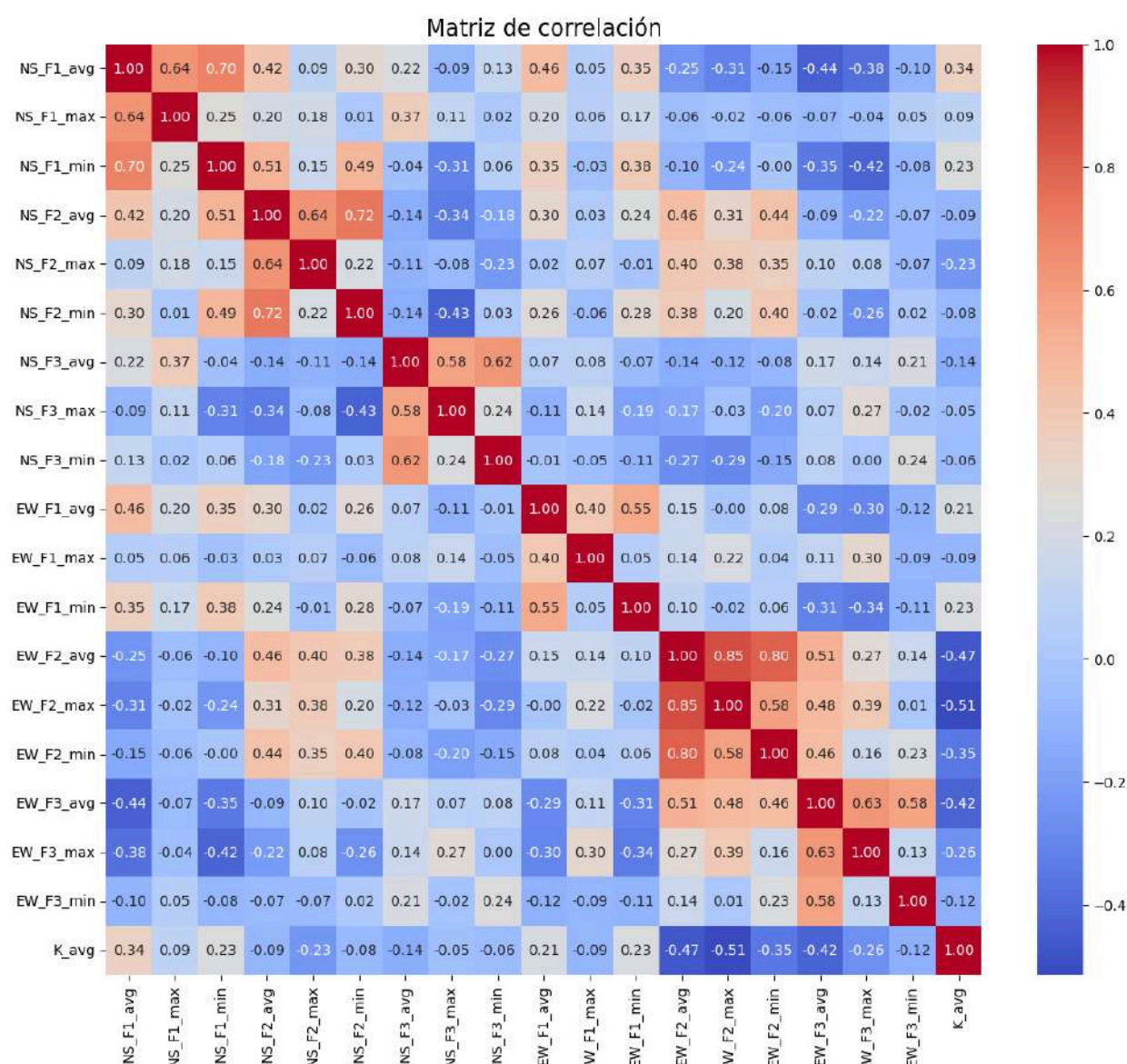
Para interpretar las dos gráficas anteriores, es preciso entender lo siguiente: cuantos más puntos en el eje x haya dibujados para cada una de las características, más importancia tendrá cada una de ellas en el modelo. Como se observa en esta parte, las características de más puntos y en consecuencia de más peso, son de nuevo mayoritariamente las frecuencias. Por otro lado, una información aún más relevante y curiosa es que en el caso de que los puntos rojos están a la derecha significa que para valores altos de esa característica aumenta la predicción, es decir, aumenta la salida. Mientras que los puntos azules a la izquierda indican que a valores bajos de esa característica tiende a disminuir la salida. Lo podemos observar mayoritariamente en los valores de las frecuencias del nodo 1. Por otro lado, tenemos la situación en el que los puntos rojos están a la izquierda y los puntos azules a la derecha, lo que indica lo contrario, es decir, una relación inversa. Cuantos más puntos rojos haya a la izquierda significa que a medida que aumente esta característica, la salida disminuye y con los puntos azules a la derecha, cuanto más disminuyen los valores de estas características más aumenta el valor de la salida. En este caso, se puede observar que el resto de las frecuencias tienen esta relación inversa entre características y etiqueta (frecuencias del nodo 2 y 3). Las siguientes imágenes corresponden a la herramienta de representaciones gráficas matriz de correlación:



Matriz de correlación

NS_AS1_avg	1.00	0.85	0.92	0.88	0.65	0.82	0.92	0.81	0.85	0.34	0.06	0.22	-0.25	-0.24	-0.24	0.12	0.10	0.21	0.90	0.59	0.84	0.76	0.51	0.73	0.57	0.37	0.63	0.65	0.61	0.58	0.29	0.17	0.13	-0.56	-0.52	-0.42	-0.44	-0.31	-0.15	0.68	0.47	0.63	0.23
NS_AS1_max	-0.85	1.00	0.71	0.77	0.78	0.63	0.79	0.87	0.67	0.27	0.15	0.08	0.25	0.22	0.33	0.11	0.12	0.04	0.84	0.83	0.66	0.63	0.62	0.57	0.48	0.45	0.48	0.55	0.63	0.44	0.21	0.21	0.05	0.42	0.35	-0.34	-0.32	-0.19	-0.10	0.62	0.63	0.48	0.14
NS_AS1_min	-0.92	0.71	1.00	0.76	0.49	0.83	0.83	0.67	0.84	0.30	0.01	0.21	-0.25	0.23	0.23	0.12	0.13	0.20	0.80	0.48	0.87	0.70	0.43	0.74	0.49	0.30	0.61	0.59	0.52	0.56	0.30	0.14	0.13	-0.56	-0.54	-0.45	-0.46	0.30	-0.15	0.60	0.39	0.62	0.24
NS_AS2_avg	-0.88	0.77	0.76	1.00	0.78	0.87	0.93	0.85	0.86	0.08	-0.05	0.00	-0.36	-0.33	-0.26	0.03	0.06	0.17	0.94	0.61	0.83	0.65	0.48	0.62	0.59	0.43	0.62	0.67	0.65	0.59	0.13	0.24	-0.03	-0.42	0.35	-0.30	-0.27	-0.15	-0.07	0.69	0.51	0.62	0.10
NS_AS2_max	-0.65	0.78	0.49	0.78	1.00	0.52	0.71	0.88	0.56	0.00	0.13	-0.15	-0.33	-0.19	-0.41	0.09	0.16	0.00	0.80	0.87	0.53	0.51	0.71	0.38	0.52	0.65	0.41	0.54	0.72	0.39	0.03	0.30	-0.15	-0.26	0.15	-0.20	-0.04	0.06	0.01	0.63	0.76	0.39	-0.00
NS_AS2_min	-0.82	0.63	0.83	0.87	0.52	1.00	0.84	0.66	0.89	0.12	-0.12	0.08	-0.30	-0.36	-0.16	0.06	0.07	0.18	0.81	0.43	0.91	0.61	0.33	0.68	0.52	0.27	0.65	0.60	0.49	0.60	0.17	0.13	0.05	-0.43	-0.39	-0.31	-0.35	-0.26	-0.11	0.60	0.34	0.65	0.09
NS_AS3_avg	-0.92	0.79	0.83	0.93	0.71	0.84	1.00	0.88	0.92	0.09	-0.08	0.00	-0.43	-0.35	-0.35	0.10	0.14	0.26	0.95	0.61	0.86	0.70	0.49	0.70	0.55	0.39	0.65	0.68	0.66	0.64	0.18	0.19	-0.00	-0.53	0.44	-0.42	-0.35	-0.19	-0.11	0.70	0.51	0.67	0.10
NS_AS3_max	-0.81	0.87	0.67	0.85	0.88	0.66	0.88	1.00	0.73	0.10	0.09	-0.06	-0.38	-0.27	-0.43	0.12	0.19	0.15	0.90	0.84	0.68	0.61	0.64	0.55	0.51	0.52	0.51	0.61	0.72	0.50	0.13	0.25	-0.09	-0.43	0.31	-0.35	-0.23	-0.08	-0.08	0.67	0.68	0.52	0.06
NS_AS3_min	-0.85	0.67	0.84	0.86	0.56	0.89	0.92	0.73	1.00	0.06	-0.12	0.01	-0.39	-0.31	-0.29	0.13	0.16	0.25	0.86	0.49	0.95	0.64	0.38	0.71	0.49	0.28	0.65	0.62	0.54	0.65	0.20	0.14	0.03	-0.48	0.39	-0.39	-0.34	-0.20	-0.12	0.62	0.39	0.68	0.06
NS_F1_avg	-0.34	0.27	0.30	0.08	0.00	0.12	0.09	0.10	0.06	1.00	0.64	0.70	0.42	0.09	0.30	0.22	-0.09	0.13	0.08	0.07	0.07	0.28	0.10	0.23	0.12	-0.10	0.11	0.09	-0.03	0.03	0.46	0.05	0.35	0.25	0.31	-0.15	-0.44	-0.38	-0.10	0.10	-0.02	0.08	0.34
NS_F1_max	-0.06	0.15	0.01	-0.05	0.13	-0.12	-0.08	0.09	-0.12	0.64	1.00	0.25	0.20	0.18	0.01	0.37	0.11	0.02	0.01	0.21	-0.07	0.09	0.17	-0.03	0.08	0.10	-0.06	0.02	0.05	-0.10	0.20	0.06	0.17	-0.06	-0.02	-0.06	-0.07	-0.04	0.05	0.06	0.13	-0.09	0.09
NS_F1_min	-0.22	0.08	0.21	0.00	-0.15	0.08	-0.00	-0.06	0.01	0.70	0.25	1.00	0.51	0.15	0.49	-0.04	-0.31	0.06	-0.05	-0.12	0.01	0.16	-0.08	0.15	0.05	-0.21	0.05	0.02	-0.16	0.00	0.35	-0.03	0.38	0.10	0.24	-0.00	-0.35	-0.42	-0.08	0.00	0.20	0.04	0.23
NS_F2_avg	-0.25	0.25	-0.25	-0.36	-0.33	-0.30	-0.43	-0.38	-0.39	0.42	0.20	0.51	1.00	0.64	0.72	-0.14	-0.34	-0.18	-0.40	-0.26	-0.36	-0.16	-0.21	-0.18	0.23	-0.32	-0.26	-0.31	-0.43	0.27	0.30	0.03	0.24	0.46	0.31	0.44	-0.09	0.22	-0.07	0.28	-0.30	-0.26	0.09
NS_F2_max	-0.24	0.22	-0.23	-0.33	-0.19	-0.36	-0.35	-0.27	-0.31	0.09	0.18	0.15	0.64	1.00	0.22	-0.11	-0.08	-0.23	-0.26	-0.13	-0.26	-0.22	-0.07	-0.27	-0.21	-0.08	-0.30	-0.28	-0.22	-0.27	0.02	0.07	-0.01	0.40	0.38	0.35	0.10	0.08	-0.07	-0.22	-0.10	-0.29	0.23
NS_F2_min	-0.24	0.33	-0.23	-0.26	-0.41	-0.16	-0.35	-0.43	-0.29	0.30	0.01	0.49	0.72	0.22	1.00	0.14	-0.43	0.03	-0.39	-0.43	-0.28	-0.13	-0.36	-0.10	-0.14	-0.39	-0.13	-0.22	-0.43	-0.17	0.26	0.06	0.28	0.38	0.20	0.40	-0.02	-0.26	0.02	0.23	-0.44	-0.16	0.08
NS_F3_avg	-0.12	0.11	0.12	0.03	0.09	0.06	0.10	0.12	0.13	0.22	0.37	-0.04	-0.14	-0.11	-0.14	1.00	0.58	0.62	0.08	0.12	0.11	0.16	0.15	0.11	0.23	0.21	0.15	0.21	0.20	0.13	0.07	0.08	-0.07	-0.14	-0.12	-0.08	0.17	0.14	0.21	0.18	0.13	0.10	-0.14
NS_F3_max	-0.10	0.12	0.13	0.06	0.16	0.07	0.14	0.19	0.16	-0.09	0.11	-0.31	-0.34	-0.08	-0.43	0.58	1.00	0.24	0.17	0.21	0.18	0.00	0.16	-0.01	0.03	0.20	-0.00	0.09	0.22	0.05	-0.11	0.14	-0.19	-0.17	-0.03	-0.20	0.07	0.27	-0.02	0.09	0.22	0.03	-0.05
NS_F3_min	-0.21	0.04	0.20	0.17	0.00	0.18	0.26	0.15	0.25	0.13	0.02	0.06	-0.18	-0.23	0.03	0.62	0.24	1.00	0.12	-0.11	0.15	0.22	0.05	0.21	0.31	0.18	0.29	0.33	0.26	0.28	-0.01	0.05	-0.11	0.27	0.29	-0.15	0.08	0.00	0.24	0.25	0.03	0.24	-0.06
NS_E_avg	-0.90	0.84	0.80	0.94	0.80	0.81	0.95	0.90	0.86	0.08	0.01	-0.05	-0.40	-0.26	-0.39	0.08	0.17	0.12	1.00	0.75	0.85	0.66	0.57	0.62	0.52	0.45	0.57	0.63	0.68	0.56	0.14	0.27	-0.01	-0.46	-0.34	-0.39	-0.31	-0.12	-0.15	0.70	0.63	0.60	0.08
NS_E_max	-0.59	0.83	0.48	0.61	0.87	0.43	0.61	0.84	0.49	0.07	0.21	-0.12	-0.26	-0.13	-0.43	0.12	0.21	-0.11	0.75	1.00	0.49	0.43	0.68	0.35	0.35	0.52	0.31	0.40	0.60	0.30	0.05	0.28	-0.09	-0.26	0.13	-0.27	-0.14	0.03	-0.08	0.54	0.78	0.33	0.01
NS_E_min	-0.84	0.66	0.87	0.83	0.53	0.91	0.86	0.68	0.95	0.07	-0.07	0.01	-0.36	-0.26	-0.28	0.11	0.18	0.15	0.85	0.49	1.00	0.61	0.37	0.69	0.46	0.26	0.61	0.57	0.49	0.58	0.19	0.15	0.07	-0.44	-0.35	-0.38	-0.34	-0.19	-0.14	0.59	0.38	0.64	0.06
EW_AS1_avg	-0.76	0.63	0.70	0.65	0.51	0.61	0.70	0.61	0.64	0.28	0.09	0.16	-0.16	-0.22	-0.13	0.16	0.00	0.22	0.66	0.43	0.61	1.00	0.64	0.88	0.81	0.45	0.79	0.85	0.62	0.76	0.31	0.07	0.19	0.43	0.40	-0.35	-0.33	-0.33	-0.02	0.90	0.49	0.81	0.20
EW_AS1_max	-0.51	0.62	0.43	0.48	0.71	0.33	0.49	0.64	0.38	0.10	0.17	-0.08	-0.21	-0.07	-0.36	0.15	0.16	0.05	0.57	0.68	0.37	0.64	1.00	0.43	0.60	0.81	0.42	0.57	0.78	0.39	0.11	0.22	-0.04	-0.25	-0.16	-0.24	-0.08	0.04	0.01	0.70	0.91	0.41	0.09
EW_AS1_min	-0.73	0.57	0.74	0.62	0.38	0.68	0.70	0.55	0.71	0.23	-0.03	0.15	-0.18	-0.27	-0.10	0.11	-0.01	0.21	0.62	0.35	0.69	0.88	0.43	1.00	0.66	0.27	0.85	0.76	0.50	0.83	0.31	0.05	0.19	0.48	0.45	-0.39	-0.41	-0.39	-0.10	0.76	0.36	0.91	0.17
EW_AS2_avg	-0.57	0.48	0.49	0.59	0.52	0.52	0.55	0.51	0.49	0.12	0.08	0.05	-0.23	-0.21	-0.14	0.23	0.03	0.31	0.52	0.35	0.46	0.81	0.60	0.66	1.00	0.69	0.80	0.94	0.74	0.77	-0.08	-0.01	-0.09	-0.37	-0.37	-0.19	-0.02	-0.13	0.21	0.90	0.51	0.72	0.02
EW_AS2_max	-0.37	0.45	0.30	0.43	0.65	0.27	0.39	0.52	0.28	-0.10	0.10	-0.21	-0.32	-0.08	-0.39	0.21	0.20	0.18	0.45	0.52	0.26	0.45	0.81	0.27	0.69	1.00	0.41	0.62	0.88	0.40	-0.19	0.13	-0.31	-0.23	-0.17	-0.14	0.17	0.17	0.19	0.65	0.83	0.33	-0.08
EW_AS2_min	-0.63	0.48	0.61	0.62	0.41	0.65	0.65	0.51	0.65	0.11	-0.06	0.05	-0.26	-0.30	-0.13	0.15	-0.00	0.29	0.57	0.31	0.61	0.79	0.42	0.85	0.80	0.41	1.00	0.84	0.58	0.92	0.05	-0.01	0.05	0.48	0.45	-0.30	-0.26	-0.28	0.01	0.79	0.38	0.93	0.08
EW_AS3_avg	-0.65	0.55	0.59	0.67	0.54	0.60	0.68	0.61	0.62	0.09	0.02	0.02	-0.31	-0.28	-0.22	0.21	0.09	0.33	0.63	0.40	0.57	0.85	0.57	0.76	0.94	0.62	0.84	1.00	0.78	0.86	-0.04	-0.02	-0.10	-0.49	-0.45	-0.33	-0.16	-0.23	0.13	0.94	0.52	0.83	0.04
EW_AS3_max	-0.61	0.63	0.52	0.65	0.72	0.49	0.66	0.72	0.54	-0.03	0.05	-0.16	-0.43	-0.22	-0.43	0.20	0.22	0.26	0.68	0.60	0.49	0.62	0.78	0.50	0.74	0.88	0.58	0.78	1.00	0.58	-0.08	0.15	-0.25	-0.45	-0.36	-0.33	-0.06	-0.01	0.10	0.80	0.84	0.55	0.03
EW_AS3_min	-0.58	0.44	0.56	0.59	0.39	0.60	0.64	0.50	0.65	0.03	-0.10	-0.00	-0.27	-0.27	-0.17	0.13	0.05	0.28	0.56	0.30	0.58	0.76	0.39	0.83	0.77	0.40	0.92	0.86	0.58	1.00	-0.01	-0.04	-0.04	-0.46	-0.41	-0.32	-0.23	-0.26	0.04	0.79	0.37	0.94	0.01
EW_F1_avg	-0.29	0.21	0.30	0.13	0.03	0.17	0.18	0.13	0.20	0.46	0.20	0.35	0.30	0.02	0.26	0.07	-0.11	-0.01	0.14	0.05	0.19	0.31	0.11	0.31	-0.08	-0.19	0.05	-0.04	-0.08	0.01	1.00	0.4											





*Ilustración 41. Matriz gráfica de correlación entre las características de las frecuencias y la salida.*

El principal cometido es buscar relaciones entre las diferentes características y nuestro índice K. El motivo por el cual hay dos matrices de correlación, es que la primera corresponde al conjunto de datos entero de 42 características, es por eso por lo que se forma una matriz tan grande y es difícil de apreciar visualmente las relaciones entre nuestra salida de índice K y las frecuencias. Es por eso por lo que se ha decidido insertar una matriz de correlación con el conjunto de datos de solo las columnas referidas a las frecuencias ya que, además, son las que más importancia y peso tienen, como hemos visto en los gráficos del apartado anterior y es lo que más interés representa. Como bien se puede observar, se observa que mientras las características referidas al nodo 1 tienen una correlación positiva, en las características referidas a los nodos dos y tres tienen una correlación negativa. Se extiende más la explicación de este resultado en el apartado de conclusión.

## 5. Conclusiones

Se han estudiado diferentes campos para poder obtener los resultados del último apartado, ahora se concretan las diferentes conclusiones que se han logrado obtener.

- **MSE (Mean Squared Error = 0,34750):**

El valor es bajo y lo cual sugiere en promedio que los errores cometidos por el modelo no son muy grandes. Sin embargo, tampoco se puede considerar como un modelo perfecto de cohesión entre el conjunto de datos y nuestros modelos. Aun así, se puede ver como un buen resultado y más teniendo en cuenta que los valores son completamente reales.

- **MAE (Mean Absolute Error = 0,43638):**

Esta métrica mide la magnitud promedio de los errores, lo que quiere decir que el modelo predice con un error medio de aproximadamente 0,43 unidades. En este caso, se debe tener en cuenta cómo afecta esta variación a los valores de nuestra salida que es el índice K. Dado que esta diferencia en nuestra etiqueta no es de gran consideración, sino que, además, es relativamente común ver estas pequeñas variaciones entre las medidas del índice del mismo día entre diferentes observatorios. Entonces, se puede concluir que es un buen resultado y aún más considerando los datos y valores puramente reales.

- **R<sup>2</sup> (Coeficiente de determinación = 0,59446):**

Por último, el valor de esta métrica se refiere a que aproximadamente el 59% de la variabilidad de los datos de salida se explica por el modelo. No es un valor particularmente alto, pero sugiere que el modelo ha encontrado relaciones complejas en los datos. Se puede ver también en las imágenes anteriores, donde se observa que prácticamente relacionan los resultados con las frecuencias en una mayor medida en comparación con el resto de las características. Lo ideal sería un coeficiente de determinación más cercano a 1.

En cuanto a la selección del mejor modelo, es fácil distinguir con los resultados anteriores, que el algoritmo SVR es el que mejor se adapta a los conjuntos de datos y, por lo tanto, mejores resultados han dado. Bien es cierto que el modelo XG Boost ha tenido los mejores valores con la herramienta Cross Validation, pero, aun así, los valores de SVR han sido mejores. Como bien se ha comentado previamente, en la definición del modelo SVR, el modelo es conocido por su capacidad para encontrar relaciones complejas con muchas dimensiones o con una estructura complicada de características y aunque los otros modelos tienden a ser muy potentes y frecuentemente superando a SVR en rendimiento, la característica singular de nuestro conjunto de datos ha hecho que SVR haya sido más eficiente. En adición, se aprecia que en todos y cada uno de los mejores valores han sido proporcionados por el conjunto de datos completo del Observatorio de Hartland, Reino Unido. No hay ninguna discusión ya que en todos y cada uno de los posibles valores ha obtenido un mayor rendimiento.

Se han estudiado diferentes campos para poder obtener los resultados del último apartado, ahora se concretan las diferentes conclusiones que hemos logrado obtener. Se ha hecho hincapié en la relación tan importante que hay entre las frecuencias de cada uno de los nodos y la salida. Es por eso por lo que se decide hacer un subconjunto de los datos para estudiar solo los modelos con las columnas que hacen referencia a las frecuencias, pero que, sin embargo, no se obtienen mejores resultados. Por lo tanto, se puede sacar la conclusión de que, pese a que las características que más peso tienen sean las frecuencias, tiene aún más peso la cantidad de datos con el que trabajamos, por lo que cuanto mayor sean los conjuntos, mejores resultados se obtienen. Por otro lado, ya que se está hablando



sobre las características, hay que hacer énfasis en la correlación positiva que hay entre las frecuencias del nodo 1 y la correlación negativa que hay entre las frecuencias del nodo 2 y 3 de nuestras características de las Resonancias de Schumann con los índices K.

Otro punto para tener en cuenta es que dentro de las características con más peso en nuestros modelos (las frecuencias), se aprecia que los parámetros del magnetómetro este-oeste tienen más peso e importancia en general que los parámetros del magnetómetro norte-sur.

Por último, y quizás la conclusión más prometedora es que los mejores resultados que se han obtenido han sido gracias a los datos usados a partir del observatorio de Hartland, en Reino Unido, aunque los resultados obtenidos a partir de los datos del observatorio de Potsdam en Alemania son parecidos. Uno se encuentra en el sur de Reino Unido y otro cerca de Berlín y hay que tener en cuenta, como ya se comentó en apartados anteriores, una característica muy importante sobre la toma de medidas de índice K, es que pueden variar con la latitud. De tal modo que supuestamente las tormentas geomagnéticas afectan al planeta Tierra de manera global, dependiendo de la latitud que tenga el punto donde se estén tomando las medidas para el índice K, estas pueden variar dando lugar a un pequeño error. A continuación, se muestra la distancia y una hoja de características en la que se puede observar perfectamente las comparaciones entre el refugio Poqueira de Sierra Nevada de nuestra central atmosférica de donde se han obtenido los datos de las características y los dos observatorios de donde se han obtenido los valores de los índices K.

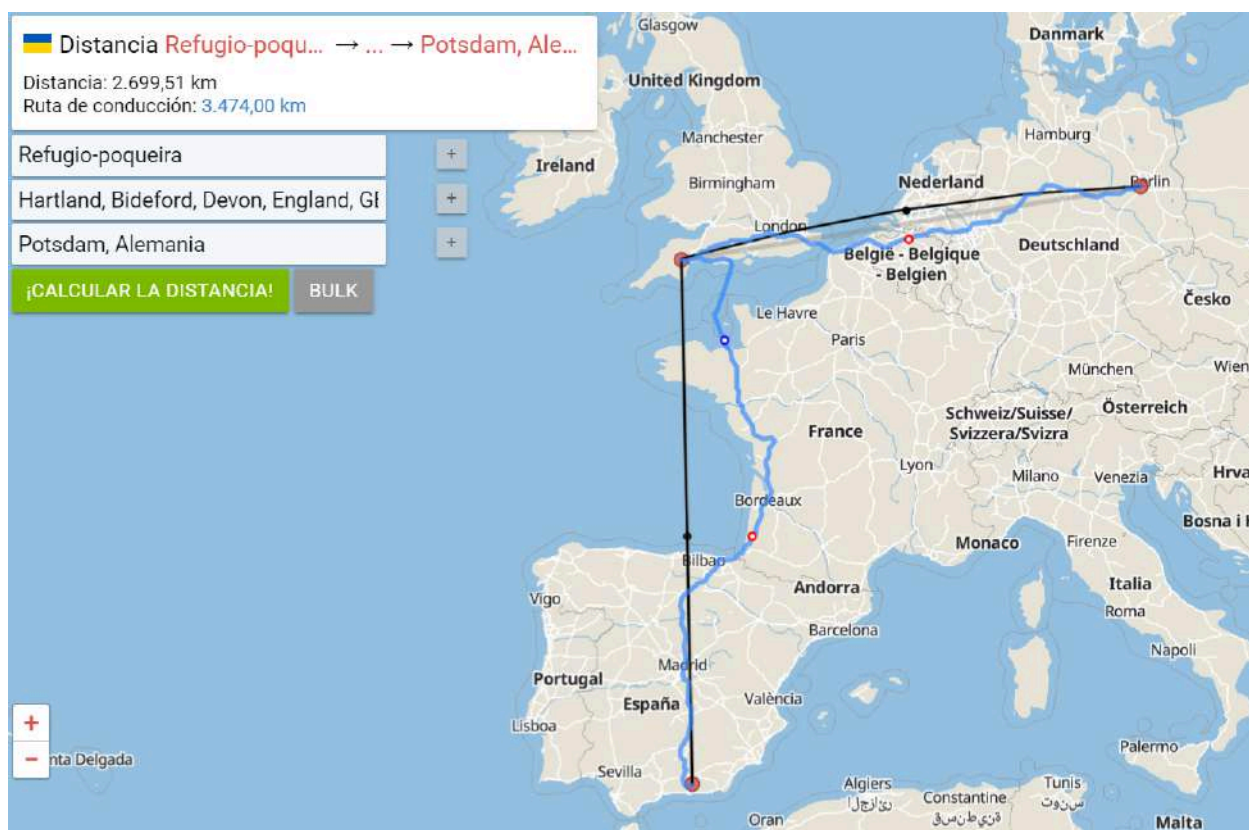


Ilustración 42. Mapa entre el Refugio Poqueira, el observatorio de Hartland y el observatorio de Potsdam.

#1 **Refugio-poqueira** 37.027569,-3.323662

Capileira, Granada, Andalucía, España  
Capileira, Granada, Andalusia, Spain

Línea recta: 1.526,24 km

Ruta de conducción: 2.115,60 km (26h 51min)

#2 **Hartland, Bideford, Devon, England, GBR**

50.750000,-3.750000

Crediton Hamlets, Mid Devon, Devon, England, United Kingdom  
Crediton Hamlets, Mid Devon, Devon, England, United Kingdom

Línea recta: 1.173,27 km

Ruta de conducción: 1.358,40 km (14h 50min)

#3 **Potsdam, Alemania** 52.400931,13.059140

Innenstadt, Potsdam, Brandenburg, Deutschland  
Innenstadt, Potsdam, Brandenburg, Germany

*Ilustración 43. Hoja de características entre el Refugio Poqueira, el observatorio de Hartland y el observatorio de Potsdam.*

Como bien se observa, la latitud en el observatorio de Potsdam, Alemania es de 52,4 grados, mientras que la latitud en el observatorio de Hartland es de 50,75 grados. La latitud de la estación atmosférica está en unos 37 grados aproximadamente. En cuanto a los datos, se observa que los resultados de Alemania con respecto a los resultados obtenidos con los índices K de Reino Unido son algo peores. Lo que quiere decir que con poco más de 1,5 grados de latitud de diferencia, ya se ha visto mejora en los resultados. Ahora si se plantea que se hubiera podido obtener los datos de algún observatorio de España o si hubiera alguno en Granada, según esta relación, los resultados deberían haber sido muchos mejores.

#### Estrategias de mejora y trabajos futuros

Se ha realizado un estudio novedoso aplicando técnicas de inteligencia artificial en el reconocimiento de patrones en un año de datos correspondientes a diferentes métricas asociadas a las Resonancias de Schumann. Es por eso por lo que las conclusiones obtenidas son más que prometedoras, sin embargo, hay muchas cosas que también se podrían mejorar. Por un lado, se pueden implementar algoritmos de aprendizaje profundo, que, aunque se ha hablado en algún apartado sobre ellos, solo se han utilizado algoritmos de aprendizaje automático. Otro punto que tampoco se puede pasar por alto es que se han obtenido los datos de observatorios de Reino Unido y Alemania y en el caso de que se hubieran obtenido los datos de un observatorio con una latitud más cercana a la de la estación atmosférica de Sierra Nevada, los resultados podrían haber sido aún mejores.

OBSERVATORIO	LATITUD
SAN PABLO	39° 32' 50" N
GUÍMAR	28° 19' 15" N
EBRO	40° 49' 14" N

*Ilustración 44. Tabla de valores de las latitudes de los diferentes observatorios en España.*

Observando la tabla anterior, destacan los tres observatorios más importantes en este campo de España y sus latitudes. En el caso de que se hubiera podido obtener los datos del Observatorio de San Pablo o de Gúímar, la diferencia de latitud entre la estación atmosférica en Sierra Nevada sería mínima de apenas 2 o 3 grados de latitud. En cuanto a los trabajos y estudios posteriores a este trabajo de fin de Grado, las conclusiones abren puertas a una posible ventana de estudio como puede ser buscar el motivo o explicación física por el cual se han obtenido estas correlaciones entre los índices K y las frecuencias de cada uno de los modos o por qué han sido más relevantes las características de los datos obtenidos por el magnetómetro este-oeste que por el magnetómetro norte-sur o cuales serían los resultados implementando datos de índices K obtenidos en observatorios con latitudes más similares a la del refugio de Poqueira, la estación atmosférica en Sierra Nevada, Granada.

## 6. Bibliografía

- [1] KULAK, A. (2019). Winfried Otto Schumann—An Unfinished Biography. *BOOK OF EXTENDED ABSTRACTS*, 35.
- [2] Price, C. (2016a). ELF Electromagnetic Waves from Lightning: The Schumann Resonances. *Atmosphere*, 7(9), 116. <https://doi.org/10.3390/atmos7090116>
- [3] Sekiguchi, M., Hayakawa, M., Nickolaenko, A. P., & Hobara, Y. (2006). Evidence on a link between the intensity of Schumann resonance and global surface temperature. *Annales Geophysicae*, 24(7), 1809-1817. <https://doi.org/10.5194/angeo-24-1809-2006>
- [4] Wang, Z., Zhou, C., Zhao, S., Xu, X., Liu, M., Liu, Y., . . . Shen, X. (2021). Numerical Study of Global ELF Electromagnetic Wave Propagation with Respect to Lithosphere–Atmosphere–Ionosphere Coupling. *Remote Sensing*, 13(20), 4107. <https://doi.org/10.3390/rs13204107>
- [5] Ciencia, N. (2018, 4 septiembre). Ondas ELF, ¿las primeras señales de un terremoto? *Nova Ciencia*. <https://novaciencia.es>
- [6] Hermeticus Network - Schumann Resonance. (s. f.). <https://hermeticus.net/schumann.html>
- [7] Carducci, L. M., Luna, F., Alonso, R., Julio, J., Zothner, E., & Fano, W. G. MEDICIÓN Y PROCESAMIENTO DE SEÑALES PARA LA DETECCIÓN DE ANOMALÍAS DE CAMPO MAGNÉTICO DE MUY BAJAS FRECUENCIAS.
- [8] Cano Domingo, C. (2022). Analysis of the earth-ionosphere resonances and its application to the study of other natural phenomena.
- [9] Vicci, A. C. (2018). ¿Qué pasó con el Test de turing? *Teorías, Enfoques y Aplicaciones en las Ciencias Sociales*, 11(23), 61-69.
- [10] Administrador, & Administrador. (2018, 15 febrero). La aceptación de los robots sociales en entornos humanos | Telos. <https://telos.fundaciontelefonica.com/archivo/numero104/la-aceptacion-de-los-robots-sociales-en-entornos-humanos/>
- [11] decide4AI. (2022, 16 noviembre). Tipos de aprendizaje que usan los algoritmos de Machine Learning. <https://decidesoluciones.es/tipos-de-aprendizaje-algoritmos-machine-learning/>
- [12] Tcastator. (2022, 29 julio). A Beginner's Guide to Deep Learning | Enthought, Inc. <https://www.enthought.com/blog/a-beginners-guide-to-deep-learning/>
- [13] Shiffler, R. E. (1988). Maximum Z scores and outliers. *The American Statistician*, 42(1), 79-80.

- [14] Preet, A. (2022, 2 septiembre). Dealing with outliers using the Z-Score method. <https://www.analyticsvidhya.com/blog/2022/08/dealing-with-outliers-using-the-z-score-method/>
- [15] Alakh. (2024, 9 octubre). Support Vector Regression Tutorial for Machine Learning. <https://www.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machine-learning/>
- [16] Sun, Y., Ding, S., Zhang, Z., & Jia, W. (2021). An improved grid search algorithm to optimize SVR for prediction. *Soft Computing*, 25, 5633-5644.
- [17] Shehadeh, A., Alshboul, O., Al Mamlook, R. E., & Hamedat, O. (2021). Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. *Automation in Construction*, 129, 103827.
- [18] GeeksforGeeks. (2023, 6 febrero). XGBoost. <https://www.geeksforgeeks.org/xgboost/>
- [19] El Mrabet, Z., Sugunaraj, N., Ranganathan, P., & Abhyankar, S. (2022). Random forest regressor-based approach for detecting fault location and duration in power systems. *Sensors*, 22(2), 458.
- [20] Hodson, T. O. (2022). Root mean square error (MSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 2022, 1-10.
- [21] Van Otten, N. (2024, 8 abril). 10 Regression Metrics for Machine Learning & Practical How to Guide. <https://spotintelligence.com/2024/03/27/regression-metrics-for-machine-learning/>
- [22] Palma, R. P. (2022). Análisis crítico del coeficiente de determinación ( $R^2$ ), como indicador de la calidad de modelos lineales y no lineales. *Matemática*, 20(2).
- [23] Berrar, D. (2019). Cross-validation.
- [24] Mangalathu, S., Hwang, S. H., & Jeon, J. S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures*, 219, 110927.
- [25] Nacional, I. G. (s. f.). Instituto Geográfico Nacional. <https://www.ign.es/web/ign/portal/gmt-que-son-tormentas-geomagneticas>
- [26] Homepage | NOAA / NWS Space Weather Prediction Center. (s. f.). <https://www.swpc.noaa.gov/>
- [27] Kp Index - Kp index. (s. f.). <https://kp.gfz-potsdam.de/en/>

- [28] Klein, C., & Klein, C. (2023, 4 agosto). A perfect solar superstorm: the 1859 Carrington Event. <https://www.history.com/news/a-perfect-solar-superstorm-the-1859-carrington-event>
- [29] Nacional, I. G. (s. f.-b). Instituto Geográfico Nacional. <https://www.ign.es/web/ign/portal/gmt-indice-k>
- [30] Planetary K-Index | NOAA / NWS Space Weather Prediction Center. (s. f.). <https://www.swpc.noaa.gov/products/planetary-k-index>
- [31] Kp Index - Kp index. (s. f.-b). <https://kp.gfz-potsdam.de/en/>
- [32] Nacional, I. G. (s. f.-c). Instituto Geográfico Nacional. <https://www.ign.es/web/ign/portal/gmt-magneto-san-pablo>
- [33] British Geological Survey 1998 - 2017 (c)NERC <http://www.bgs.ac.uk>  
mailto:www-bgs@bgs.ac.uk. (s. f.). Hartland Magnetic Observatory. <https://geomag.bgs.ac.uk/operations/hartland.html>
- [34] F. Ossing. (s. f.). History. <https://www.gfz-potsdam.de/en/about-us/organisation/history>



Yo, **Ángel Bermúdez Caballero**, alumno de la titulación INGENIERÍA DE TECNOLOGÍAS DE TELECOMUNICACIÓN de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 77375187J, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Ángel Bermúdez Caballero

A handwritten signature in black ink, consisting of a series of loops and a long horizontal stroke extending to the right.

Granada a 12 de noviembre de 2024 .



D. Jesús Francisco Fornieles Callejón,  
Profesor del Área de Departamento de  
Física de la Universidad de Granada.

D. Franciso Jesús Lara Abelenda.

**Informan:**

Que el presente trabajo, titulado **Estudio de Resonancias de Schumann a través de la Inteligencia Artificial supervisada**, ha sido realizado bajo su supervisión por **Ángel Bermúdez Caballero**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 12 de noviembre de 2024.

**Los directores:**



D. Jesús Francisco Fornieles Callejón

D. Franciso Jesús Lara Abelenda

A handwritten signature in black ink, appearing to read "F. LARA", is written over a horizontal line.

# Agradecimientos

Este agradecimiento es para todas y cada una de las personas que me han acompañado a lo largo de este camino. Pero, sobre todo, va dirigido a vosotros:

En primer lugar, estas palabras son para mi familia, sobre todo a quienes han sido mi guía constante y mis mayores ejemplos, mis padres. Gracias por el apoyo incondicional, por la educación y por el esfuerzo incansable que habéis hecho para que pudiera llegar hasta aquí.

También a Celia, persona que me ha acompañado y ayudado en los momentos más duros y ha hecho felices tantos otros, siempre ha sido capaz de alegrarme y animarme.

Finalmente, a todos mis amigos, tanto los de siempre como aquellos con quienes he compartido estos tres años tan especiales en Granada y el último en Praga, especialmente a Jorge. Sin vosotros, nada habría sido igual. Gracias por estar a mi lado en las buenas y en las malas. Los momentos que hemos vivido juntos son inolvidables.







