



UNIVERSITAT OBERTA DE CATALUNYA (UOC)  
MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (*Data Science*)

## TRABAJO FINAL DE MÁSTER

ÁREA: 3

# **Análisis Avanzado de Datos para la Predicción de Movimientos en el Mercado de Valores y la Identificación de Factores Influyentes en Empresas del Sector Eléctrico del IBEX-35**

---

Autor: Eduardo Mora González

Tutor: Nombre del colaborador/a docente

Profesor: Nombre del profesor responsable del área de TF

---

Barcelona, 18 de diciembre de 2023



# Créditos/Copyright

Este documento se distribuye con licencia Creative Commons Atribución Compartir Igual 4.0. El texto completo de la licencia puede obtenerse en <https://creativecommons.org/licenses/by-sa/4.0/>.

La copia y distribución de esta obra está permitida en todo el mundo, sin regalías y por cualquier medio, siempre que esta nota sea preservada. Se concede permiso para copiar y distribuir traducciones de este libro desde el español original a otro idioma, siempre que la traducción sea aprobada por el autor del libro y tanto el aviso de copyright como esta nota de permiso, sean preservados en todas las copias.





# FICHA DEL TRABAJO FINAL

Título del trabajo:	Descriptivo del trabajo
Nombre del autor:	Nombre y dos apellidos
Nombre del colaborador/a docente:	Nombre y dos apellidos
Nombre del PRA:	Nombre y dos apellidos
Fecha de entrega (mm/aaaa):	MM/AAAA
Titulación o programa:	Plan de estudios
Área del Trabajo Final:	El nombre de la asignatura de TF
Idioma del trabajo:	Catalán, español o inglés
Palabras clave	Máximo 3 palabras clave



*Momentismo absoluto, vivo minuto a minuto*  
*Al futuro lo he dejado atrás*  
*El ayer me atormentaba, el mañana me asustaba*  
*Solo creo en el momento actual*

FANGORIA





# Agradecimientos

Quisiera agradecer a todas las personas y entidades la ayuda que me han prestado en la realización de este Trabajo Fin de Máster.

En primer lugar, a mis padres, hermano y abuelos, que siempre han estado a mi lado apoyándome para seguir hacia adelante.

Quisiera hacer extensiva mi gratitud a *Diego Calvo* tutor de este T.F.M., por todo lo que me ha aportado y ayudado a juntar dos de mis grandes pasiones: la música y la economía.

Finalmente, y no por ello menos importante, quiero agradecer a *Lucia Rosa Alonso*, *Sergio Funes Olaria* y *Àlex Aguilera Martínez* todos los consejos y ayuda que me han prestado desde que nos conocemos, sin ellos todo esto no hubiera sido lo mismo.

A todos ellos, muchas gracias.

Eduardo Mora González



# Resumen

Este trabajo de investigación se enfrenta al desafío de anticipar y comprender los movimientos del mercado de valores, que son conocidos por su volatilidad y la influencia de una amplia gama de factores. Esta complejidad dificulta la toma de decisiones en inversiones y estrategias empresariales.

Los métodos tradicionales de análisis financiero a menudo no son suficientes para abordar esta complejidad, y los inversores requieren herramientas más avanzadas para analizar datos financieros y predecir tendencias del mercado con mayor precisión.

Por lo tanto, esta investigación se enfoca en la aplicación de técnicas avanzadas de análisis de datos y machine learning para mejorar la capacidad de predecir y comprender los cambios en el mercado financiero. El objetivo es utilizar tecnología y datos masivos para beneficiar a inversores y empresas en la toma de decisiones estratégicas y financieras.

**Palabras clave:** Mercado de Valores, Análisis Financiero, Análisis Avanzado de Datos, Aprendizaje Automático, Modelado Predictivo, Tendencias del Mercado, Análisis de Datos.



# Abstract

This research project addresses the challenge of anticipating and understanding movements in the stock market, which are known for their volatility and the influence of a wide range of factors. This complexity makes decision-making in investments and business strategies a significant challenge.

Traditional methods of financial analysis often fall short in addressing this complexity, and investors require more advanced tools to analyze financial data and predict market trends with greater precision.

Therefore, this research focuses on the application of advanced data analysis and machine learning techniques to enhance the ability to predict and understand changes in the financial market. The goal is to leverage technology and massive data to benefit investors and businesses in making strategic and financial decisions.

**Keywords:** Stock Market, Financial Analysis, Advanced Data Analysis, Machine Learning, Predictive Modeling, Market Trends, Data Analysis.



# Índice de figuras

1.1. Fases de la metodología <b>CRISP-DM</b> [27]. . . . .	8
2.1. Media móvil de 50-200 días (MA) de las acciones de Endesa . . . . .	15
2.2. Media Móvil de 50-200 días (MA) de las acciones de Endesa con marcadores Golden Cross y Death Cross (junio de 2018 - junio de 2021) . . . . .	16
4.1. Diagrama de las fases del estudio . . . . .	35
4.2. Crecimiento de las acciones de Endesa a lo largo de los años . . . . .	40
4.3. Crecimiento de las acciones de Iberdrola a lo largo de los años . . . . .	40
4.4. Crecimiento de las acciones de Acciona Energía a lo largo de los años . . . . .	40
4.5. Crecimiento de las acciones de Naturgy a lo largo de los años . . . . .	41
4.6. Crecimiento de las acciones de Solaria a lo largo de los años . . . . .	41
4.7. Comparación del crecimiento de las acciones de diferentes empresas a lo largo de los años . . . . .	41
4.8. Información detallada sobre la cotización del día . . . . .	42
4.9. Media móvil de 50-200 días (DMA) de las acciones de Endesa . . . . .	42
4.10. Media Móvil de 50-200 días (DMA) de las acciones de Endesa con marcadores Golden Cross y Death Cross (junio de 2018 - junio de 2021) . . . . .	43
4.11. Media móvil de 50-200 días (DMA) con marcadores Golden Cross y Death Cross del resto de empresas . . . . .	45
4.12. Descuadre acciones de Endesa por año, semana y día . . . . .	47
4.13. Noticias Endesa [16] y [17] . . . . .	48
4.14. Descuadre acciones de Iberdrola por año, semana y día . . . . .	50

4.15. Descuadre acciones de Acciona Energía por año y día . . . . .	51
4.16. Descuadre acciones de Solaria por año, semana y día . . . . .	53
4.17. Descuadre acciones de Naturgy por año, semana y día . . . . .	54
4.18. Evolución de los dividendos Endesa . . . . .	55
4.19. Comparación de la evolución mensual del precio de cierre de Endesa con y sin dividendos . . . . .	56
4.20. Evolución de los dividendos Iberdrola . . . . .	57
4.21. Comparación de la evolución mensual del precio de cierre de Iberdrola con y sin dividendos . . . . .	57
4.22. Evolución de los dividendos de Solaria . . . . .	58
4.23. Comparación de la evolución mensual del precio de cierre de Solaria con y sin dividendos . . . . .	59
4.24. Evolución de los dividendos de Naturgy . . . . .	60
4.25. Comparación de la evolución mensual del precio de cierre de Iberdrola con y sin dividendos . . . . .	60
4.26. Correlaciones entre las empresas del sector eléctrico . . . . .	61
4.27. Ejemplo salida ventana de Holt-Winters . . . . .	65
4.28. Comparación de los hiperparámetros con MSE más bajo en el periodo de TEST de 2009-09-30 a 2010-08-31 . . . . .	68
4.29. Mejores hiperparámetros I . . . . .	69
4.30. Mejores hiperparámetros II . . . . .	70
4.31. Feature Importances con todas las variables . . . . .	78
4.32. Feature Importances con todas las variables . . . . .	79
4.33. Feature Importances con el sentido incorporado en el Tipo de Noticia . . . . .	81
4.34. Feature Importances con las empresas agrupadas . . . . .	82
4.35. Feature Importances de Iberdrola . . . . .	84
4.36. Feature Importances de Acciona Energías . . . . .	85
4.37. Feature Importances de Solaria . . . . .	86



---

4.38. Feature Importances de Naturgy . . . . .	87
4.39. Impacto de las noticias . . . . .	89



# Índice de cuadros

1.1. Planificación de los módulos . . . . .	9
1.2. Cronograma de tareas . . . . .	10
4.1. Top 5 días con mayor diferencia porcentual positiva en las acciones de Endesa . . .	48
4.2. Top 5 días con mayor diferencia porcentual negativa en las acciones de Endesa . . .	49
4.3. Top 5 días con mayor diferencia porcentual positiva en las acciones de Iberdrola . .	49
4.4. Top 5 días con mayor diferencia porcentual negativa en las acciones de Iberdrola . .	49
4.5. Top 5 días con mayor diferencia porcentual positiva en las acciones de Acciona Energía	51
4.6. Top 5 días con mayor diferencia porcentual negativa en las acciones de Acciona Energía . . . . .	51
4.7. Top 5 días con mayor diferencia porcentual positiva en las acciones de Solaria . . .	52
4.8. Top 5 días con mayor diferencia porcentual negativa en las acciones de Solaria . . .	52
4.9. Top 5 días con mayor diferencia porcentual positiva en las acciones de Naturgy . . .	54
4.10. Top 5 días con mayor diferencia porcentual negativa en las acciones de Solaria . . .	54
4.11. Dividendos Endesa agrupados por año . . . . .	55
4.12. Dividendos Iberdrola agrupados por año . . . . .	56
4.13. Dividendos Solaria . . . . .	58
4.14. Dividendos Naturgy agrupados por año . . . . .	59
4.15. Matriz de correlación entre empresas del sector eléctrico. . . . .	62
4.16. MSE para cada ventana de Holt-Winters . . . . .	65
4.17. Resultados del modelo . . . . .	68
4.18. Comparación MSE entre XGBoost y Holt-Winters . . . . .	70

4.19. Representación tabular del análisis de noticias con codificador OneHot . . . . .	76
4.20. Comparación de errores MSE sin y con Noticias . . . . .	77
4.21. Datos de noticias de Endesa . . . . .	88
4.22. Datos de cotización de Endesa entre el 2014-11-19 y el 2014-11-21 . . . . .	88

# Índice general

Resumen	IX
Abstract	XI
Índice de figuras	XIII
Índice de cuadros	XVII
Índice general	XIX
<b>1 Introducción</b>	<b>3</b>
1.1. Contexto y justificación del Trabajo . . . . .	3
1.2. Descripción general del problema . . . . .	5
1.3. Motivación personal . . . . .	6
1.4. Definición de los objetivos . . . . .	7
1.4.1. Objetivo General . . . . .	7
1.4.2. Objetivos Parciales . . . . .	7
1.5. Descripción de la metodología . . . . .	8
1.6. Planificación . . . . .	9
<b>2 Conceptos Teóricos</b>	<b>11</b>
2.1. Fundamentos financieros . . . . .	11
2.1.1. Fundamentales de una empresa . . . . .	11
2.1.2. Mercado de valores . . . . .	12

2.1.3.	Trading . . . . .	13
2.1.4.	Medias Móvil (Moving Average) . . . . .	14
2.1.5.	Golden Cross y Death Cross . . . . .	15
2.1.6.	Dividendos . . . . .	16
2.2.	Fundamentos de minería de datos . . . . .	17
2.2.1.	Deep learning y las Redes Neuronales . . . . .	18
2.2.2.	Procesamiento del lenguaje natural con la API de GPT . . . . .	19
2.2.3.	XGBoost . . . . .	19
2.2.4.	Holt-Winters . . . . .	21
<b>3</b>	<b>Estado del arte</b>	<b>23</b>
3.1.	Análisis de Sentimientos en la Predicción del Mercado de Valores . . . . .	23
3.1.1.	Desarrollo del Análisis de Sentimientos . . . . .	23
3.1.2.	Aplicaciones del Análisis de Sentimientos . . . . .	24
3.1.3.	Métodos de Análisis de Sentimientos . . . . .	25
3.1.4.	Limitaciones del Análisis de Sentimientos . . . . .	26
3.1.5.	Principales artículos . . . . .	27
3.2.	Deep Learning en la Predicción del Precio de las Acciones . . . . .	29
3.2.1.	Técnicas de Deep Learning y tipos de Redes Neuronales . . . . .	29
3.2.2.	Ventajas del uso de Deep Learning . . . . .	30
3.2.3.	Limitaciones y Desafíos . . . . .	31
3.2.4.	Aplicaciones de Deep Learning en la Predicción de Acciones . . . . .	31
3.2.5.	Principales artículos . . . . .	32
<b>4</b>	<b>Diseño e implementación</b>	<b>35</b>
4.1.	Obtención de los datos . . . . .	35
4.1.1.	Justificación del uso de la API de yfinance . . . . .	36
4.1.2.	Datos de Cotización y Dividendos . . . . .	36
4.2.	Análisis de los datos . . . . .	38

4.2.1. Análisis de los Precios de Apertura y Cierre de las Acciones . . . . .	39
4.2.2. Promedio Móvil (Moving Average), Golden Cross y Death Cross . . . . .	42
4.2.3. Discrepancias en los Valores de Cierre y Apertura de Acciones . . . . .	46
4.2.4. Análisis de Dividendos . . . . .	55
4.2.5. Relación Entre Empresas . . . . .	61
4.3. Holt-Winters Vs XGBoost . . . . .	62
4.3.1. Holt-Winters . . . . .	62
4.3.2. XGBoost con Cross Validation . . . . .	66
4.3.3. Comparación de los modelos . . . . .	70
4.4. Análisis de noticias . . . . .	71
4.4.1. Justificación del uso de Web Scraping para obtener noticias . . . . .	72
4.4.2. Web Scraping . . . . .	72
4.4.3. GPT para el análisis de noticias . . . . .	73
4.5. Retroalimentación de XGBoost . . . . .	75
4.6. Pruebas del nuevo modelo XGBoost . . . . .	77
4.6.1. Feature Importances con todas las variables . . . . .	77
4.6.2. Feature Importances con el sentido incorporado a los tipos de noticias . .	79
4.6.3. Feature Importances generalizando todas las empresas distintas a Endesa	81
4.7. Extrapolación al resto de empresas . . . . .	83
4.7.1. Iberdrola . . . . .	83
4.7.2. Acciona Energías . . . . .	84
4.7.3. Solaria . . . . .	85
4.7.4. Naturgy . . . . .	86
4.8. Corroboración del estudio . . . . .	87
<b>5 Conclusiones y Líneas de trabajo futuras</b>	<b>91</b>
5.1. Conclusión del estudio . . . . .	91
5.2. Consecución de objetivos . . . . .	92

5.3. Trabajo futuro . . . . .	93
<b>Bibliografía</b>	<b>95</b>





# Capítulo 1

## Introducción

El presente capítulo tiene como objetivo proporcionar un contexto y una justificación para el trabajo de investigación que se llevará a cabo como parte del Trabajo de Fin de Máster (*TFM*).

### 1.1. Contexto y justificación del Trabajo

El presente trabajo de investigación se sitúa en el contexto de los mercados financieros globales, que desempeñan un papel crucial en la economía mundial. Los mercados de valores son el epicentro de la asignación de recursos, la toma de decisiones empresariales y las estrategias de inversión. En este entorno, la **información** es la moneda de mayor valor y la capacidad de analizarla de manera efectiva y anticipar las tendencias del mercado se ha convertido en una **ventaja competitiva** esencial.

Los mercados financieros son notoriamente volátiles y están influenciados por una multitud de factores, como eventos económicos, políticos y sociales. Las decisiones comerciales y de inversión se basan en gran medida en la interpretación de estos factores y en la capacidad de prever cómo afectarán los cambios en el futuro. La demanda de herramientas y enfoques que permitan una toma de decisiones más informada y precisa en este entorno es constante.

En la última década, hemos presenciado una explosión en la disponibilidad de datos financieros a gran escala. Las empresas cotizadas, los gobiernos y las instituciones financieras generan y comparten datos en tiempo real sobre el rendimiento de las empresas, los índices de mercado

y las condiciones económicas. Esta avalancha de datos ha llevado al surgimiento de un campo interdisciplinario conocido como *FinTech* (**tecnología financiera**), que se enfoca en el uso de tecnologías avanzadas, incluyendo análisis de datos y *machine learning*, para comprender y predecir los movimientos del mercado.

La capacidad de predecir con precisión los movimientos del mercado de valores es esencial para inversores, gestores de activos y empresas. Permite tomar decisiones informadas sobre la compra o venta de activos financieros, así como anticipar los riesgos y oportunidades emergentes. La predicción precisa puede marcar la diferencia entre el éxito y el fracaso en el mundo de las inversiones y los negocios.

Dentro de este contexto, la presente investigación se justifica por varios motivos:

- **Relevancia Económica:** Los mercados financieros influyen directamente en la asignación de recursos a nivel global, lo que afecta a la economía en su conjunto. Comprender y anticipar las tendencias del mercado tiene un impacto significativo en la toma de decisiones financieras.
- **Técnicas Avanzadas:** La investigación se basa en la aplicación de técnicas avanzadas de análisis de datos y *machine learning*, lo que representa un avance importante en la capacidad de análisis y predicción en el campo financiero.
- **Amplio Alcance Sectorial:** El análisis se realizará en una variedad de sectores empresariales, lo que garantiza la aplicabilidad de los resultados en diferentes contextos económicos y empresariales.
- **Contribución a la Investigación:** Además de su aplicación práctica, esta investigación contribuirá al cuerpo de conocimientos en el análisis financiero y la predicción de mercados, sirviendo como punto de partida para futuras investigaciones en este campo.

En resumen, este trabajo de investigación se desarrolla en un contexto financiero altamente dinámico y competitivo, donde la capacidad de anticipar los movimientos del mercado es esencial. La aplicación de técnicas avanzadas de análisis de datos y *machine learning* promete

mejorar significativamente la toma de decisiones financieras y la comprensión de los mercados de valores, lo que lo convierte en un proyecto de gran **relevancia** e **interés** tanto para el mundo académico como para el empresarial.

## 1.2. Descripción general del problema

El problema central que aborda este trabajo de investigación radica en la dificultad inherente de prever y entender los movimientos del mercado de valores. Los mercados financieros, en particular, son notoriamente **volátiles** y están influenciados por una amplia variedad de factores, desde eventos macroeconómicos y políticos hasta cambios en la percepción pública. Esta **complejidad** hace que la toma de decisiones en inversiones y estrategias empresariales sea un desafío **significativo**.

La *imprevisibilidad* de los mercados financieros a menudo conduce a resultados imprevistos y a la pérdida de oportunidades para inversores y empresas por igual. La *necesidad* de tomar decisiones fundamentadas y estratégicas se convierte en un **reto clave**, especialmente en un entorno caracterizado por la constante avalancha de información financiera y la **necesidad** de procesar y analizar grandes volúmenes de datos.

Los enfoques tradicionales de análisis financiero, basados en modelos estadísticos y técnicas convencionales, a menudo no son suficientes para capturar la dinámica compleja de los mercados modernos. Los inversores y analistas financieros necesitan herramientas más avanzadas y precisas para analizar datos financieros y predecir tendencias del mercado con mayor **confiabilidad**.

Por lo tanto, esta investigación se centra en la aplicación de técnicas avanzadas de análisis de datos y *machine learning* para abordar el problema de predecir los movimientos del mercado de valores de manera más **precisa** y **eficaz**. La pregunta fundamental que guía esta investigación es cómo utilizar los avances tecnológicos y el acceso a datos masivos para mejorar la **comprensión** y **anticipación** de los cambios en el mercado financiero, beneficiando así a inversores y empresas en sus decisiones estratégicas y financieras.

### 1.3. Motivación personal

La motivación personal para embarcarme en esta investigación proviene de una combinación de factores personales y profesionales que me han llevado a abordar este **desafiante** tema. A continuación, detallo algunas de las razones que impulsan mi interés en este proyecto:

- **Pasión por las Finanzas y la Tecnología:** Desde siempre, he tenido un profundo interés en el mundo de las finanzas y la tecnología. Observar cómo la tecnología está transformando la forma en que se toman las decisiones financieras y cómo se gestionan las inversiones me ha fascinado. Esta investigación me brinda la oportunidad de combinar estas dos pasiones y explorar cómo las técnicas avanzadas pueden revolucionar el análisis financiero.
- **Relevancia Práctica:** Entiendo la *importancia práctica* de esta investigación en un mundo cada vez más globalizado y digital. La capacidad de predecir con precisión los movimientos del mercado de valores es **esencial** tanto para inversores individuales como para empresas que buscan mantener su competitividad en un mercado en constante cambio.
- **Aprendizaje Continuo:** La investigación en análisis de datos y *machine learning* es un campo en constante evolución. Estoy motivado por la oportunidad de *aprender* y aplicar las últimas técnicas y enfoques en este campo en rápida expansión. Esta investigación me brinda la oportunidad de mantenerme *actualizado* y desarrollar habilidades valiosas.
- **Impacto Potencial:** Finalmente, lo que me motiva profundamente es la idea de que los resultados de esta investigación pueden tener un *impacto real y significativo* en la toma de decisiones financieras. Si nuestras conclusiones pueden ayudar a los inversores a tomar decisiones más informadas o a las empresas a anticipar mejor los cambios en el mercado, consideraría que este esfuerzo ha sido **altamente gratificante**.

## 1.4. Definición de los objetivos

En este apartado se mostrara el objetivo general que se pretende conseguir y los objetivos parciales en que se divide.

### 1.4.1. Objetivo General

El objetivo general de este estudio es desarrollar un enfoque avanzado de análisis financiero centrado en el sector eléctrico del IBEX-35 usando técnicas de análisis de datos y aprendizaje automático (*machine learning*) con el fin de predecir los movimientos específicos del mercado de valores en este sector, tomando en cuenta la influencia de diversos tipos de noticias.

### 1.4.2. Objetivos Parciales

- **Recopilación de Datos Financieros:** Obtener el historial de cotización y los dividendos, junto con otros indicadores relevantes, de empresas pertenecientes al sector eléctrico.
- **Análisis de Datos e Identificación de Noticias Influyentes:** Realizar un análisis detallado de los datos para identificar patrones y tendencias en el mercado de valores. Integrar el análisis de noticias utilizando modelos de lenguaje como GPT para identificar factores influyentes basados en la información disponible en distintos tipos de noticias financieras.
- **Desarrollo de Modelos de *Machine Learning*:** Crear modelos de *machine learning*, utilizando algoritmos avanzados, que sean capaces de predecir con **precisión** los movimientos futuros de las acciones en función de los datos recopilados.
- **Evaluación de la Precisión de las Predicciones:** Evaluar el rendimiento de los modelos de *machine learning* desarrollados, considerando tanto los datos financieros como la información analizada de las noticias. Utilizar métricas de evaluación adecuadas, como el error cuadrático medio (**MSE**).

- **Documentación de Resultados:** Documentar y presentar de manera clara y concisa los resultados de la investigación, incluyendo los factores identificados a través del análisis de datos financieros y noticias.

## 1.5. Descripción de la metodología

Para el desarrollo de este Trabajo de Fin de Máster (TFM), se ha optado por la metodología **CRISP-DM** [27] (*Cross-Industry Standard Process for Data Mining*), que proporciona un marco estructurado y completo para la gestión de proyectos de minería de datos.

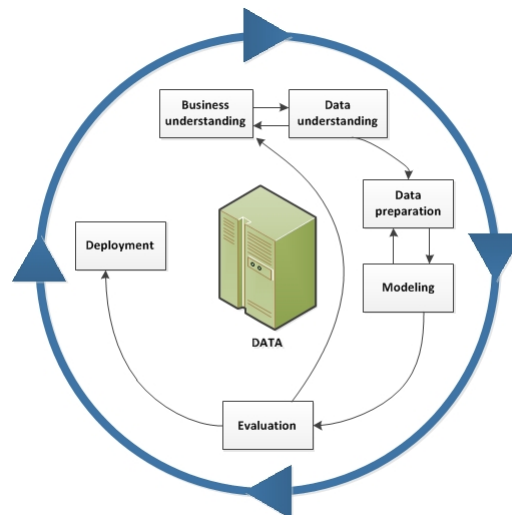


Figura 1.1: Fases de la metodología **CRISP-DM** [27].

Esta metodología se divide en las siguientes etapas clave como se muestra en la Figura 1.1:

1. **Comprensión del negocio:** En esta etapa, se definen los objetivos del proyecto y se evalúa su **relevancia** para el negocio en cuestión.
2. **Comprensión de los datos:** Aquí se recopilan y exploran los datos necesarios, evaluando su **calidad** y **pertinencia**.
3. **Preparación de los datos:** Los datos se limpian y transforman para el análisis. También se seleccionan las variables más **relevantes**.

4. **Modelado:** Se eligen las técnicas de modelado adecuadas y se construyen modelos de minería de datos.
5. **Evaluación:** Los modelos se evalúan en función de métricas específicas y se determina si cumplen con los objetivos del negocio.
6. **Despliegue:** Los resultados se implementan en un entorno de producción, junto con planes de monitoreo y mantenimiento.
7. **Documentación:** Se registra y documenta todo el proceso, incluyendo decisiones y resultados.
8. **Iteración:** Si es necesario, se puede volver a una etapa anterior para refinar el modelo o abordar nuevas preguntas.

CRISP-DM es un enfoque **flexible** y **cíclico** que permite adaptarse a cambios en el negocio o los datos a lo largo del proyecto. Esta metodología proporciona una estructura sólida para proyectos de minería de datos y garantiza un enfoque sistemático desde la definición de los objetivos hasta la implementación de soluciones prácticas [31].

## 1.6. Planificación

El proyecto que se presenta en este documento representa un compromiso significativo que abarca desde el 27 de septiembre de 2023 hasta el 16 de enero de 2024. En la tabla 1.1 se muestra la planificación de los módulos y en la tabla 1.2 el cronograma real con todas las actividades de cada módulo.

Módulo	Descripción	Fechas
1	Definición y planificación del trabajo final	27 Sep 2023 - 10 Oct 2023
2	Estado del arte o análisis de mercado del proyecto	11 Oct 2023 - 24 Oct 2023
3	Diseño e implementación del trabajo	25 Oct 2023 - 19 Dec 2023
4	Redacción de la documentación del TFM	20 Dic 2023 - 16 Ene 2024
5	Defensa del proyecto	17 Ene 2024 - 4 Feb 2024

Cuadro 1.1: Planificación de los módulos



NOMBRE DE LA TAREA	FECHA DE INICIO	FECHA DE FINALIZACIÓN
<b>MODULO 1</b>	<b>27.09.2023</b>	<b>10.10.2023</b>
Propuesta de título	27.09.2023	28.09.2023
Palabras clave	29.09.2023	29.09.2023
Resumen de la propuesta	30.09.2023	30.09.2023
Descripción y Justificación	01.10.2023	05.10.2023
Motivación Personal	05.10.2023	07.10.2023
Objetivos	08.10.2023	09.10.2023
Metodología	09.10.2023	10.10.2023
Planificación	27.09.2023	01.10.2023
<b>MODULO 2</b>	<b>11.10.2023</b>	<b>24.10.2023</b>
Conceptos Teóricos	11.10.2023	15.10.2023
Estado del arte	15.10.2023	24.10.2023
<b>MODULO 3</b>	<b>25.10.2023</b>	<b>19.12.2023</b>
Obtención de los datos	25.10.2023	31.10.2023
Análisis de los datos	01.11.2023	12.11.2023
Holt-Winters Vs XGBoos	13.11.2023	25.11.2023
Análisis de Noticias	26.11.2023	30.11.2023
Retroalimentación de XGBoost	01.12.2023	06.12.2023
Extrapolación	07.12.2023	15.12.2023
Corroboración del estudio	15.12.2023	17.12.2023
Conclusiones	18.12.2023	19.12.2023
<b>MODULO 4</b>	<b>20.12.2023</b>	<b>16.01.2024</b>
Memoria entrega preliminar	20.12.2023	02.01.2024
Memoria Final	02.01.2024	09.01.2024
Presentación Audiovisual	09.01.2024	16.01.2024
<b>MODULO 5</b>	<b>17.01.2024</b>	<b>04.02.2024</b>
Entrega Documentación	18.01.2024	18.01.2024
Defensa	?¿	?¿

Cuadro 1.2: Cronograma de tareas

# Capítulo 2

## Conceptos Teóricos

Este capítulo tiene como objetivo establecer una sólida base de conocimientos que abarque tanto los conceptos financieros fundamentales como los principios esenciales de la minería de datos.

### 2.1. Fundamentos financieros

El mundo financiero es un entorno dinámico y esencial en el contexto económico global. En este espacio, se desempeñan elementos fundamentales que incluyen los fundamentales de una empresa, la dinámica del mercado de valores y las estrategias de trading. En un escenario caracterizado por su creciente competitividad y volatilidad, la habilidad para comprender y anticipar las tendencias del mercado se ha convertido en una ventaja competitiva crucial. Esta sección proporcionará una introducción a estos pilares esenciales del mundo financiero, que servirán como base para la exploración en profundidad de estos temas a lo largo de este estudio.

#### 2.1.1. Fundamentales de una empresa

Los **fundamentos** de una empresa son aquellos aspectos que determinan su valor, su rentabilidad y su potencial de crecimiento. Para analizarlos, se pueden utilizar diferentes métodos, como el *análisis fundamental* o el *análisis por ratios* [23]. Estos métodos consisten en examinar los estados financieros de la empresa, como el *balance*, la *cuenta de resultados* y el *estado de flujos de efectivo*, así como los factores externos que la afectan, como el *entorno económico*, la *competencia* o la *regulación*.

El *análisis fundamental* busca estimar el valor intrínseco o verdadero de una empresa, comparándolo con su cotización en el mercado. Para ello, se pueden emplear distintas técnicas, como el *método top-down*, que parte de lo más general a lo más específico, o el *método bottom-up*, que hace lo contrario [4]. El objetivo es identificar las **fortalezas** y **debilidades** de la empresa, así como las **oportunidades** y **amenazas** que le ofrece el mercado.

El *análisis por ratios* utiliza indicadores numéricos que expresan la relación entre dos o más magnitudes financieras. Estos *ratios* pueden clasificarse en diferentes categorías, según lo que midan: **solvencia**, **liquidez**, **rentabilidad**, **endeudamiento**, **crecimiento**, etc. Algunos ejemplos de *ratios* son el **PER** (relación entre el precio y el beneficio por acción), el **ROE** (rentabilidad sobre el patrimonio neto) o el **D/E** (relación entre la deuda y el capital propio) [23].

Además del *análisis financiero*, los **fundamentos** de una empresa también dependen de su organización interna. Para garantizar un adecuado funcionamiento del negocio, hay que tener claros los *principios organizativos* de la empresa: **objetivo**, **especialización**, **jerarquía**, **unidad de mando**, **difusión**, **coordinación**, **equilibrio** y **flexibilidad** [21]. Estos principios permiten definir la estructura, las funciones y las responsabilidades de cada miembro de la organización.

En definitiva, los **fundamentos** de una empresa son los pilares sobre los que se sustenta su actividad y su éxito. Una empresa es una organización de personas y recursos que busca la consecución de un beneficio económico con el desarrollo de una actividad en particular [20]. Para lograrlo, debe conocer su situación actual y su proyección futura, así como adaptarse a las condiciones del mercado y a las necesidades de sus clientes.

### 2.1.2. Mercado de valores

El mercado de valores es un conjunto de mercados e instituciones donde se negocian *valores financieros*, como **acciones**, **bonos**, **derivados**, etc. El objetivo del mercado de valores es facilitar el financiamiento de las empresas y el ahorro de los inversores, ofreciendo **rentabilidad** y **liquidez** [1].

El mercado de valores se divide en dos segmentos: el **mercado primario** y el **mercado**

**secundario.** El mercado primario es aquel donde las empresas emiten nuevos valores para captar recursos. El mercado secundario es aquel donde los inversores compran y venden los valores previamente emitidos [1].

El funcionamiento del mercado de valores está regulado por las **autoridades competentes**, que velan por la **transparencia**, la **seguridad** y la **eficiencia** del mismo. En España, el organismo encargado de supervisar e inspeccionar los mercados de valores y la actividad de sus agentes es la **Comisión Nacional del Mercado de Valores (CNMV)** [1].

El mercado de valores tiene una gran **importancia** para la economía, ya que contribuye al **crecimiento**, la **innovación** y la **estabilidad financiera**. Además, permite a las empresas acceder a fuentes de financiación alternativas al crédito bancario, diversificar sus riesgos y mejorar su gestión. Asimismo permite a los inversores obtener una **renta** por su capital, diversificar su patrimonio y participar en el desarrollo de las empresas [12].

En resumen, el mercado de valores es un mecanismo que facilita el intercambio de recursos financieros entre los agentes económicos. El mercado de valores refleja las **expectativas**, las **preferencias** y la **información** de los inversores sobre el valor y el rendimiento de las empresas. El mercado de valores es un **indicador** del estado y la evolución de la economía.

### 2.1.3. Trading

El **trading** es un tipo de operativa bursátil que consiste en la compraventa de activos cotizados (activos con suficiente liquidez para ello) en un mercado financiero. Su objetivo es obtener un **beneficio positivo** en un **corto periodo de tiempo**, beneficio que se denomina *plusvalía* [15].

Existen diferentes tipos de **trading**, según el plazo temporal en el que se realizan las operaciones. El **trading de intradía** o *daytrading* es el más agresivo, ya que consiste en abrir y cerrar operaciones en el mismo día, aprovechando las fluctuaciones del precio en minutos u horas. Dentro de este tipo se encuentra el *scalping*, que es aún más rápido y se basa en operaciones que duran muy pocos minutos o incluso segundos [7].

El *swing trading* es un tipo de **trading** de medio y corto plazo, en el que las operaciones

se mantienen abiertas durante días, semanas o incluso meses. Este tipo de operación requiere menos atención al mercado y permite capturar movimientos más amplios del precio [7].

El **trading** se puede realizar con diferentes instrumentos financieros, como **acciones**, **bonos**, **divisas**, **futuros**, **opciones**, etc. Cada uno de estos instrumentos tiene sus propias características, ventajas y riesgos, por lo que es importante conocerlos bien antes de operar con ellos [15].

Para hacer **trading** se necesita una plataforma electrónica que permita acceder al mercado y ejecutar las órdenes. Esta plataforma se llama *broker* y cobra una comisión por cada operación realizada. Existen muchos *brokers* en el mercado, por lo que es conveniente comparar sus condiciones, servicios y reputación antes de elegir uno [7].

El **trading** es una actividad que implica un alto riesgo, ya que el precio de los activos puede variar de forma impredecible y generar pérdidas importantes. Por eso, es fundamental tener una buena *gestión del riesgo*, que consiste en controlar el tamaño de las posiciones, establecer límites de pérdida y beneficio, diversificar el capital y seguir una *estrategia definida* [14].

El **trading** también requiere una buena formación, *disciplina* y *psicología*. El trader debe estar al tanto de los factores que influyen en el mercado, como la situación económica, la competencia, la regulación o los eventos geopolíticos. Además, debe ser capaz de controlar sus emociones y evitar caer en errores comunes como la codicia, el miedo o la euforia [14].

En definitiva, el **trading** es una forma de invertir en los mercados financieros que busca obtener beneficios rápidos aprovechando las variaciones del precio. El **trading** tiene muchas modalidades, instrumentos y estrategias, pero también implica un alto riesgo y una gran exigencia. Por eso, es importante formarse bien y seguir unas pautas de *gestión del riesgo* y *psicología* adecuadas.

#### 2.1.4. Medias Móvil (Moving Average)

**Las medias móviles** son un componente esencial del análisis técnico, crucial para los traders. Estas líneas informativas en los gráficos se basan en cambios de precio para identificar tendencias en activos. Su impacto es significativo, influyendo en decisiones comerciales y

sirviendo como base para herramientas más avanzadas.

La clave radica en comprender que **las medias móviles de largo plazo reaccionan más lentamente que las de corto plazo ante las tendencias subyacentes** [44].

A continuación se presenta un ejemplo visual de una media móvil en acción, ilustrada en la Figura 2.1, que analiza el comportamiento de las acciones de Endesa. Este gráfico demuestra cómo la media móvil captura y suaviza las fluctuaciones de precios, proporcionando una representación más clara de la tendencia subyacente en el rendimiento de las acciones de la compañía.

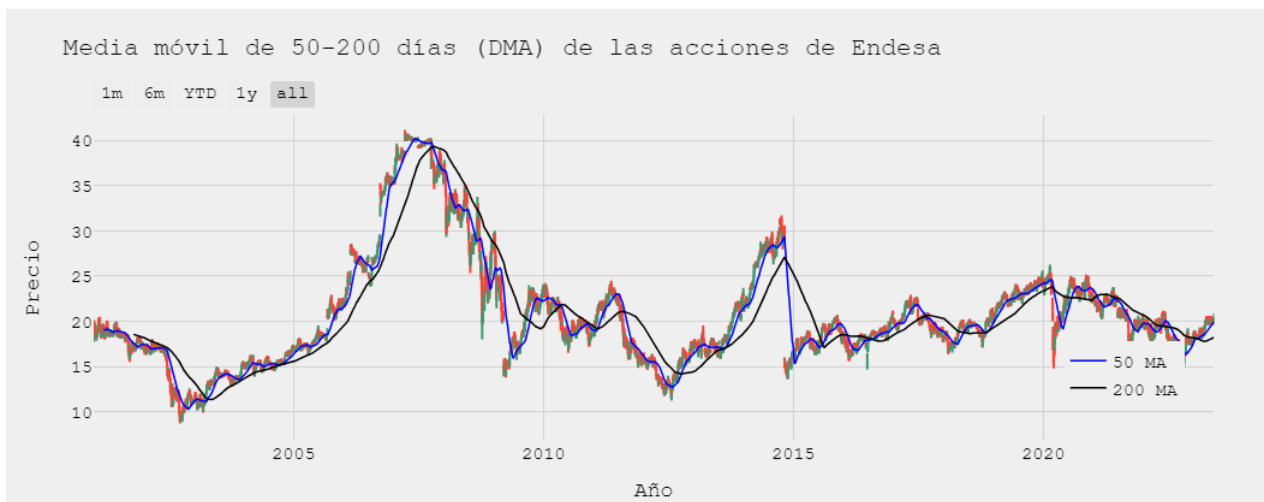


Figura 2.1: Media móvil de 50-200 días (MA) de las acciones de Endesa

### 2.1.5. Golden Cross y Death Cross

El **Golden Cross (Cruz Dorada)** se produce cuando la media móvil de corto plazo cruza por encima de la media móvil de largo plazo (comúnmente se refiere al cruce de la media móvil de 50 días sobre la media móvil de 200 días). Esto es considerado como una señal alcista, indica un posible cambio positivo en la tendencia del mercado y sugiere un potencial aumento en los precios [24].

El **Death Cross (Cruz de la Muerte)** ocurre cuando la media móvil de corto plazo cruza por debajo de la media móvil de largo plazo, especialmente el cruce de la media móvil de 50 días por debajo de la media móvil de 200 días. Esto es interpretado como una señal bajista, sugiere

un posible cambio negativo en la tendencia del mercado y puede indicar una disminución en los precios [24].

A continuación, se presenta un ejemplo visual que destaca el Media Móvil de 50-200 días (MA) aplicado al comportamiento de las acciones de Endesa. La Figura 2.2 ilustra el período comprendido entre junio de 2018 y junio de 2021, evidenciando marcadores cruciales como el **Golden Cross** y el **Death Cross**.

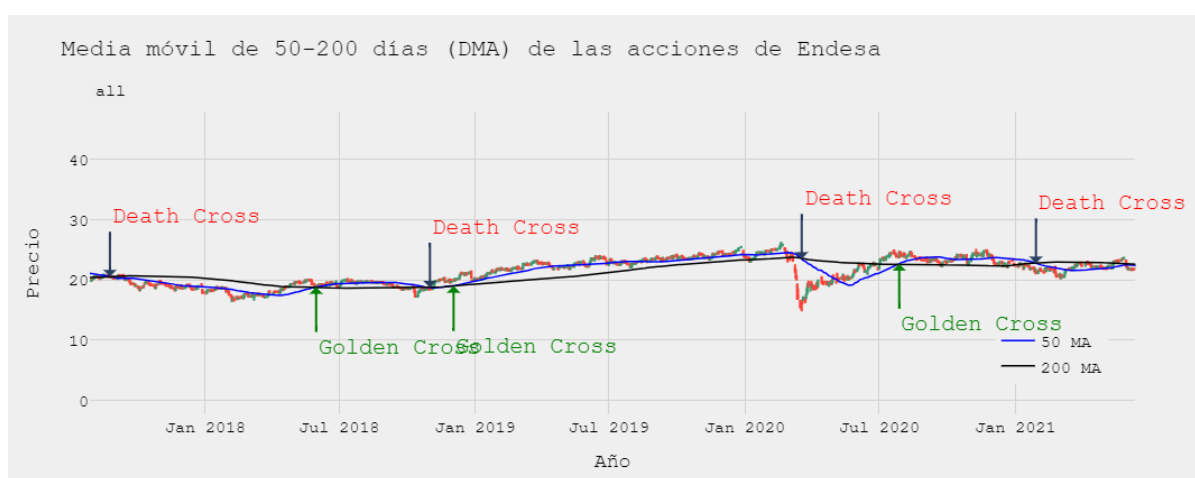


Figura 2.2: Media Móvil de 50-200 días (MA) de las acciones de Endesa con marcadores Golden Cross y Death Cross (junio de 2018 - junio de 2021)

### 2.1.6. Dividendos

Los **dividendos** [40], que representan la porción de las ganancias que las empresas distribuyen a los accionistas, desempeñan un papel esencial para atraer inversores. Estos beneficios, denominados así por su origen en «cantidad a dividir», son una recompensa para quienes invierten en la empresa.

La distribución de dividendos no es automática y está vinculada a las ganancias anuales de la empresa. La aprobación proviene de la Junta General de Accionistas tras la propuesta del Consejo de Administración.

En cuanto al cálculo, se divide la suma a repartir entre el número total de acciones, resultando en el dividendo por acción. Los accionistas reciben este dividendo según la cantidad de acciones que poseen.

Existen varios tipos de dividendos, cada uno con características particulares:

- **Dividendo a cuenta:** Basado en estimaciones, se distribuye antes del cierre del ejercicio financiero.
- **Dividendo complementario:** Se reparte después de conocer con exactitud los beneficios anuales, complementando al dividendo a cuenta.
- **Dividendo extraordinario:** Surge por ingresos excepcionales, como la venta de activos de la empresa.
- **Scrip dividend:** Permite al accionista elegir cómo recibir el dividendo: en efectivo, mediante la venta de derechos en el mercado o quedándose con nuevas acciones.
- **Dividendo fijo:** Establecido por la empresa, no depende directamente de los beneficios anuales, siendo una cantidad fija independiente del rendimiento económico.
- **Recompra de acciones:** La empresa utiliza su liquidez para comprar sus propias acciones, beneficiando a los accionistas al aumentar su participación y potencialmente el precio de cotización.

## 2.2. Fundamentos de minería de datos

La **minería de datos** es el proceso de explorar y analizar grandes conjuntos de datos para extraer conocimientos útiles para la toma de decisiones o la generación de nuevos conocimientos. La **minería de datos** se basa en técnicas estadísticas, matemáticas, informáticas e inteligencia artificial para descubrir *patrones*, *relaciones* y *tendencias ocultas* en los datos [8].

La **minería de datos** se aplica a diversos campos y sectores, como el **marketing**, las **finanzas**, la **salud**, la **educación** o la **seguridad**. Algunos ejemplos de aplicaciones de la **minería de datos** son: *segmentar a los clientes* según sus preferencias y hábitos de consumo, *detectar fraudes* o anomalías en las transacciones financieras, *diagnosticar enfermedades* o predecir su evolución, *mejorar los procesos educativos* o identificar *amenazas cibernéticas* [39].



Para realizar una **minería de datos** efectiva, se requiere seguir una serie de pasos o fases que conforman un ciclo de vida. Una metodología ampliamente utilizada es el *CRISP-DM* (*Cross Industry Standard Process for Data Mining*), que consta de seis fases: *comprensión del negocio*, *comprensión de los datos*, *preparación de los datos*, *modelado*, *evaluación* y *despliegue* [32].

La **minería de datos** implica el uso de diversas técnicas y algoritmos para procesar y analizar los datos. Algunas de las técnicas más comunes son: *clasificación*, que consiste en asignar una categoría o etiqueta a cada dato según sus características; *regresión*, que consiste en estimar una variable continua a partir de otras variables; *agrupamiento*, que consiste en agrupar los datos según su similitud; *asociación*, que consiste en encontrar reglas que relacionen los datos; y *extracción de características*, que consiste en reducir la dimensión de los datos conservando la información relevante [11].

En conclusión, la **minería de datos** es un proceso que permite obtener conocimientos valiosos a partir de grandes volúmenes de datos. La **minería de datos** se basa en técnicas avanzadas y requiere seguir una metodología estructurada. La **minería de datos** tiene múltiples aplicaciones y beneficios para diversos ámbitos y sectores.

### 2.2.1. Deep learning y las Redes Neuronales

El **deep learning** o aprendizaje profundo es una rama del **machine learning** o aprendizaje automático que se basa en el uso de **redes neuronales artificiales** para imitar el funcionamiento del cerebro humano y resolver problemas complejos de inteligencia artificial [9]. Las redes neuronales artificiales son sistemas computacionales inspirados en las neuronas biológicas, que procesan la información de forma jerárquica y se adaptan mediante un proceso de aprendizaje [37].

Existen diferentes tipos de redes neuronales según su arquitectura, su función y su aplicación. Algunos de los más conocidos son las **redes neuronales convolucionales** (CNN), que se utilizan para el procesamiento de imágenes, vídeo y audio; las **redes neuronales recurrentes** (RNN), que se emplean para el análisis de secuencias temporales como el lenguaje natural o el sonido; y las **redes neuronales profundas** (DNN), que se caracterizan por tener múltiples

capas ocultas que permiten extraer características de alto nivel de los datos [26].

El deep learning ha revolucionado el campo de la inteligencia artificial, ya que ha permitido resolver problemas que antes eran inaccesibles para las máquinas, como el reconocimiento facial, la traducción automática, la generación de texto o la conducción autónoma. Sin embargo, el deep learning también presenta algunos desafíos, como la necesidad de grandes cantidades de datos para entrenar los modelos, la dificultad para interpretar los resultados y la dependencia de los recursos computacionales [5].

### 2.2.2. Procesamiento del lenguaje natural con la API de GPT

El Procesamiento del Lenguaje Natural (PLN) es una rama de la inteligencia artificial que se enfoca en la interacción entre las computadoras y el lenguaje humano. Se centra en desarrollar algoritmos y modelos que permiten a las máquinas entender, interpretar y generar texto de manera efectiva.

El PLN abarca diversas tareas, como análisis de sentimientos, traducción automática, resumen de texto, reconocimiento de voz, entre otras. La utilización de herramientas avanzadas, como la API de GPT (Generative Pre-trained Transformer), ha revolucionado el campo al permitir el análisis contextual y la generación de texto coherente [29].

La API de GPT se emplea para analizar textos mediante avanzadas técnicas de procesamiento del lenguaje natural (NLP). Al interactuar con la API, es posible enviar un fragmento de texto como entrada y recibir respuestas generadas por el modelo GPT. Estas respuestas son coherentes y relevantes en contexto. Este proceso abarca diversas tareas, como resumen de texto, generación de contenido, traducción automática, respuesta a preguntas, entre otras aplicaciones de procesamiento del lenguaje natural. Este enfoque se utiliza para obtener una comprensión más profunda y detallada de la información contenida en las noticias recopiladas [36].

### 2.2.3. XGBoost

**XGBoost**, que significa *eXtreme Gradient Boosting*, es un algoritmo de aprendizaje automático supervisado que se utiliza para tareas de clasificación y regresión. Es una implemen-

tación eficiente y escalable de la técnica de *boosting*. El *boosting* es un enfoque de ensamblaje de modelos que combina múltiples modelos más débiles para formar uno más fuerte [47].

Las características clave de **XGBoost** son [38]:

- **Gradient Boosting:** **XGBoost** se basa en la técnica de *gradient boosting*, que optimiza iterativamente modelos débiles para minimizar la función de pérdida. Cada modelo subsiguiente se enfoca en corregir los errores del modelo anterior.
- **Árboles de Decisión:** **XGBoost** utiliza árboles de decisión como modelos base. Cada árbol se construye de manera secuencial, y se centra en las instancias que han sido mal clasificadas o cuyos residuos no han sido bien explicados por los árboles anteriores.
- **Regularización:** **XGBoost** incorpora técnicas de regularización para evitar el sobreajuste (*overfitting*) y mejorar la generalización del modelo. Incluye términos de penalización en la función objetivo para controlar la complejidad del modelo.
- **Función de Pérdida Personalizable:** Permite la personalización de la función de pérdida según los requisitos del problema, lo que hace que el algoritmo sea versátil y aplicable a diversas tareas.
- **Paralelización y Eficiencia:** **XGBoost** está diseñado para ser eficiente y escalable. Implementa técnicas de paralelización para aprovechar al máximo los recursos computacionales disponibles.
- **Selección Automática de Variables:** Puede manejar automáticamente la selección de variables importantes, lo que facilita el proceso de construcción del modelo.
- **Manejo de Datos Ausentes:** **XGBoost** puede manejar datos ausentes de manera efectiva sin requerir imputación previa.
- **Interpretabilidad:** Aunque los modelos de árboles de decisión tienden a ser menos interpretables que algunos modelos lineales, **XGBoost** proporciona funciones para evaluar la importancia de las variables, lo que ayuda en el análisis de los resultados.

### 2.2.4. Holt-Winters

El método de suavizado exponencial de **Holt-Winters** es una técnica de pronóstico utilizada en análisis de series temporales para prever patrones y tendencias futuras [42].

El método de **Holt-Winters** utiliza tres componentes principales para modelar series temporales [25]:

- **Nivel (*Level*)**: Representa el valor medio de la serie temporal.
- **Tendencia (*Trend*)**: Representa la dirección general de la serie temporal.
- **Componente estacional (*Seasonal*)**: Representa los patrones que se repiten en un período fijo de tiempo, como estaciones del año, meses o días de la semana.

Existen tres variantes del método de **Holt-Winters**, dependiendo de la combinación de los componentes mencionados [6]:

- **Suavizado exponencial simple (*Single Exponential Smoothing*)**: Utiliza solo el componente de nivel y es adecuado para datos sin tendencia ni estacionalidad.
- **Suavizado exponencial doble (*Double Exponential Smoothing*)**: Incluye el componente de tendencia y es útil cuando los datos muestran una tendencia pero no tienen estacionalidad.
- **Suavizado exponencial triple (*Triple Exponential Smoothing*)**: Incluye tanto la tendencia como el componente estacional. Es apropiado para datos que muestran tendencia y estacionalidad.



# Capítulo 3

## Estado del arte

La aplicación de técnicas de **deep learning** y minería de datos en el mercado de valores es un campo de investigación en constante crecimiento. En este capítulo, se presentan el estado del arte actual junto con una serie de artículos clave que abordan las distintas temáticas.

### 3.1. Análisis de Sentimientos en la Predicción del Mercado de Valores

El análisis de sentimientos en el mercado de valores es una disciplina que une la financiación con la tecnología de procesamiento del lenguaje natural (**NLP**). Su propósito es evaluar las emociones y puntos de vista de los inversores, analistas y el público en general sobre acciones y otros instrumentos financieros. Esta técnica se ha vuelto indispensable para inversores y operadores, ya que puede ofrecer datos valiosos sobre la tendencia futura de los precios de las acciones y otros activos financieros. Con el avance de la tecnología y el aumento de los datos disponibles, el análisis de sentimientos se ha vuelto más avanzado y exacto [46].

#### 3.1.1. Desarrollo del Análisis de Sentimientos

El análisis de sentimientos en el mercado de valores ha experimentado una notable evolución desde sus comienzos. Aquí se presentan las fases clave de su desarrollo:

1. **Análisis Convencional:** En sus primeras fases, el análisis de sentimientos en el mercado de valores se basaba principalmente en la revisión de informes financieros, noticias

y análisis de experto. Los analistas se enfocaban en evaluar el sentimiento general del mercado a través de la interpretación de los eventos económicos y financieros [18].

2. **Progresos en la Extracción de Texto:** Con el aumento de la disponibilidad de datos en línea, se produjo un cambio significativo hacia la extracción de texto y el procesamiento de lenguaje natural. Los algoritmos de **NLP** permitieron a los inversores y analistas extraer información de fuentes digitales, como noticias, redes sociales, blogs y foros financieros [18].
3. **Redes Sociales y Medios Sociales:** La popularidad de las redes sociales, como Twitter y Facebook, generó un gran interés en el análisis de sentimientos en tiempo real. Los inversores comenzaron a monitorear las redes sociales para evaluar el sentimiento del mercado y obtener información sobre eventos y noticias financieras [22].
4. **Tecnologías Avanzadas:** La adopción de tecnologías avanzadas, como el aprendizaje automático y el procesamiento de grandes volúmenes de datos (**big data**), ha impulsado la capacidad de análisis de sentimientos [19]. Los algoritmos de aprendizaje automático pueden analizar grandes conjuntos de datos de texto y identificar patrones de sentimiento de manera más precisa y eficiente.
5. **Integración con Plataformas de Trading:** Actualmente, muchas plataformas de trading e inversión incorporan análisis de sentimientos en tiempo real para ayudar a los operadores a tomar decisiones informadas. Esto incluye la identificación de eventos clave, noticias relevantes y sentimiento del mercado [13].

### 3.1.2. Aplicaciones del Análisis de Sentimientos

El análisis de sentimientos en el mercado de valores tiene una amplia gama de aplicaciones, que van desde la predicción de precios de acciones hasta la gestión de riesgos. A continuación, se detallan algunas de las aplicaciones más importantes [28]:

1. **Predicción de Precios de Acciones:** Uno de los usos más comunes del análisis de sentimientos es predecir los movimientos de precios de las acciones. Los modelos de aprendizaje automático pueden analizar el sentimiento del mercado y las noticias financieras para pronosticar si los precios de las acciones aumentarán o disminuirán.
2. **Identificación de Eventos Relevantes:** El análisis de sentimientos ayuda a identificar eventos y noticias relevantes que pueden afectar el mercado. Esto permite a los inversores estar al tanto de los acontecimientos importantes y tomar decisiones basadas en información actualizada.
3. **Gestión de Riesgos:** Las opiniones y emociones del mercado pueden influir en la volatilidad y el riesgo de los activos financieros. El análisis de sentimientos ayuda a los inversores a evaluar y gestionar mejor los riesgos al considerar el sentimiento del mercado en sus estrategias de inversión.
4. **Evaluación de la Salud de una Empresa:** El análisis de sentimientos se utiliza para evaluar la salud financiera de una empresa. La opinión pública sobre una empresa puede influir en su valor en el mercado, y el análisis de sentimientos puede ayudar a detectar señales de advertencia temprana.
5. **Seguimiento de la Competencia:** Los inversores y analistas pueden utilizar el análisis de sentimientos para realizar un seguimiento de la competencia en el mercado. Esto incluye evaluar el sentimiento del mercado hacia empresas competidoras y comparar su rendimiento relativo.

### 3.1.3. Métodos de Análisis de Sentimientos

El análisis de sentimientos se basa en una variedad de métodos y técnicas para evaluar las emociones y opiniones expresadas en el texto [46]. Algunos de los métodos más comunes incluyen [45]:



1. **Análisis de Polaridad:** Este método clasifica el texto en categorías de polaridad, como positiva, negativa o neutral. Se utilizan algoritmos de aprendizaje automático para determinar la polaridad de las palabras y frases en el texto.
2. **Detección de Emociones:** Este enfoque se centra en identificar emociones específicas, como felicidad, tristeza, miedo o enojo, en el texto. Los algoritmos de **NLP** analizan las palabras clave y las estructuras de las oraciones para detectar emociones.
3. **Análisis de Tono:** El análisis de tono evalúa el tono general del texto, como optimista, pesimista o neutral. Los algoritmos pueden determinar el tono de un artículo o comentario en función del contexto y las palabras utilizadas.
4. **Modelos de Aprendizaje Automático:** Los modelos de aprendizaje automático, como las redes neuronales, se utilizan para analizar grandes conjuntos de datos de texto y predecir el sentimiento del mercado. Estos modelos aprenden de datos históricos y mejoran su precisión con el tiempo.
5. **Diccionarios de Sentimientos:** Algunos enfoques utilizan diccionarios de palabras y frases con etiquetas de sentimiento predefinidas. Las palabras en el texto se comparan con el diccionario para determinar su sentimiento.

### 3.1.4. Limitaciones del Análisis de Sentimientos

A pesar de los avances en el análisis de sentimientos, existen desafíos y limitaciones que los analistas y los inversores deben tener en cuenta [10]:

1. **Ruido en los Datos:** Los datos de sentimiento pueden estar sujetos a ruido, ya que las emociones y opiniones pueden ser ambiguas o mal interpretadas. La ironía, el sarcasmo y la falta de contexto pueden dificultar la precisión del análisis.
2. **Cambios Rápidos en el Sentimiento:** El sentimiento del mercado puede cambiar rápidamente en respuesta a eventos inesperados. Los modelos de análisis de sentimientos deben ser capaces de adaptarse a estos cambios.

3. **Sesgo de Datos:** Las fuentes de datos pueden contener sesgos, ya sea debido a la selección de fuentes específicas o a la polarización en la generación de contenido en línea. Esto puede influir en los resultados del análisis de sentimientos.
4. **Limitaciones de Idioma y Cultura:** El análisis de sentimientos puede ser más efectivo en algunos idiomas y culturas que en otros. Los modelos de **NLP** pueden no ser igualmente precisos en todos los contextos lingüísticos.
5. **Falta de Contexto:** El análisis de sentimientos a menudo se basa en el texto sin un conocimiento completo del contexto detrás de las opiniones y emociones expresadas. Esto puede llevar a interpretaciones erróneas.

### 3.1.5. Principales artículos

Los siguientes artículos muestran algunos estudios sobre el análisis de sentimientos en el mercado de valores.

#### 3.1.5.1. Twitter sentiment analysis for stock prediction

El artículo [43] se enfoca en la aplicación de técnicas de análisis de sentimientos y aprendizaje automático con el propósito de anticipar las fluctuaciones en el mercado de valores, haciendo uso de datos recopilados de la plataforma Twitter. Los autores llevaron a cabo la obtención de datos de Twitter a través de la API de la plataforma, empleando técnicas de análisis de sentimientos para evaluar el sentimiento general presente en los tweets relacionados con el mercado de valores.

En el proceso de preprocesamiento de los datos, se llevaron a cabo tres etapas de filtrado que abarcaron la tokenización, la eliminación de palabras vacías y la supresión de caracteres especiales mediante coincidencias de expresiones regulares. Además, se implementó una técnica sencilla para abordar la ausencia de datos en la información de acciones.

Posteriormente, los autores hicieron uso de diversos algoritmos de aprendizaje automático, incluyendo el **SVM (Support Vector Machine)** y redes neuronales, para realizar pronósticos sobre los precios de las acciones. Los resultados de los distintos algoritmos fueron comparados,

y se destacó que la **red neuronal de perceptrón multicapa** exhibió el menor Error Absoluto Medio (**MAE**) y el Error Cuadrático Medio de la Raíz (**RMSE**) para los tres conjuntos de datos que contenían términos como “mercado de valores”, “StockTwits” y “AAPL”.

Además, se llevó a cabo una comparación de los resultados obtenidos a través de diversos algoritmos de análisis de sentimientos, entre los que se incluyeron el **SVM**, la regresión logística, el árbol de decisión, el árbol de regresión mejorado y el bosque aleatorio. Se concluyó que el **SVM** presentó la mayor precisión, alcanzando un nivel de precisión del 82 %.

En resumen, este artículo introduce una metodología para la utilización de datos de Twitter, combinada con técnicas de análisis de sentimientos y aprendizaje automático, con el propósito de anticipar las tendencias del mercado de valores. Los autores presentan una discusión exhaustiva sobre distintos enfoques y algoritmos utilizados y concluyen que esta metodología puede ser una herramienta valiosa para inversores y analistas del mercado de valores. El artículo también proporciona información detallada sobre el proceso de preprocesamiento de datos y los niveles de precisión alcanzados con los diferentes algoritmos empleados.

### 3.1.5.2. Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis

El artículo [30] se enfoca en el desarrollo de un modelo de predicción para evaluar el comportamiento del mercado de valores. Este enfoque se basa en la aplicación de técnicas de minería de datos y análisis de sentimientos en noticias financieras.

El objetivo principal del estudio es reducir el riesgo asociado a las inversiones en el mercado de valores y mejorar la precisión de las proyecciones mediante la implementación de un modelo de predicción eficaz. Este modelo combina datos extraídos de noticias financieras con series temporales de precios históricos del mercado de valores.

El documento introduce la técnica **n-gram**, que se utiliza para extraer palabras y términos clave de los documentos de noticias financieras, que luego se emplean en el análisis de sentimientos. Además, se aborda la técnica **TF-Idf**, un método de ponderación de características que ayuda a determinar la relevancia de las palabras en un documento o una colección de documentos.

En resumen, el documento se centra en la presentación de un modelo predictivo para evaluar el comportamiento del mercado de valores. Para lograr este propósito, se aplican diversas técnicas de preprocesamiento de datos, como la **tokenización**, **estandarización**, eliminación de palabras vacías, reducción de palabras redundantes, manejo de abreviaturas y filtrado de términos. Además, se presentan técnicas específicas como el uso de **n-gram** y **TF-Idf**, que contribuyen a identificar y extraer características clave de los datos.

## 3.2. Deep Learning en la Predicción del Precio de las Acciones

La predicción del precio de las acciones ha sido un campo de interés constante en el mundo de las finanzas y la inversión. En los últimos años, el uso de técnicas de Deep Learning y Redes Neuronales ha revolucionado la forma en que se aborda este desafío.

### 3.2.1. Técnicas de Deep Learning y tipos de Redes Neuronales

En este apartado se presentan diversos tipos de redes neuronales utilizadas en distintos estudios para la predicción del precio de las acciones [35]:

1. **Redes Neuronales Recurrentes (RNN):** Las RNN son especialmente adecuadas para modelar series temporales. Se utilizan para predecir precios de acciones a través de la captura de relaciones temporales y dependencias a lo largo del tiempo. Los tipos populares incluyen *LSTM (Long Short-Term Memory)* y *GRU (Gated Recurrent Unit)*.
2. **Redes Neuronales Convolucionales (CNN):** Aunque diseñadas para el procesamiento de imágenes, las CNN se han aplicado a la predicción de precios de acciones. Extraen características de los datos de series temporales, identificando patrones en los precios y volúmenes de negociación.
3. **Redes Generativas Adversariales (GAN):** Las GAN se utilizan para generar datos financieros sintéticos que amplían el conjunto de datos de entrenamiento. Esto permite a los modelos aprender de un espectro más amplio de condiciones del mercado.

4. **Redes Neuronales Profundas (DNN):** Las DNN se utilizan para capturar relaciones no lineales en los datos financieros. Estas redes pueden ser efectivas cuando se combinan con otras técnicas, como reducción de dimensionalidad y normalización.

Además, se emplean otras técnicas relevantes:

- **Aprendizaje Profundo para Análisis de Sentimientos [46]:** El análisis de sentimientos es crucial en la predicción de precios de acciones. Las técnicas de *Deep Learning* se utilizan para analizar noticias, informes financieros y datos en tiempo real para evaluar el sentimiento del mercado.
- **Transferencia de Aprendizaje [41]:** Esta técnica implica utilizar modelos pre-entrenados en otros conjuntos de datos (como texto financiero) para mejorar el rendimiento en la predicción de precios de acciones.

### 3.2.2. Ventajas del uso de Deep Learning

Las principales ventajas de usar técnicas de Deep Learning en la predicción de acciones son [33]:

- **Captura de Relaciones Complejas:** Las redes neuronales pueden modelar relaciones no lineales complejas presentes en los datos financieros.
- **Incorporación de Datos No Estructurados:** El análisis de sentimientos y la información textual se pueden incorporar en los modelos, lo que permite una evaluación más completa de los factores que influyen en los precios de las acciones.
- **Adaptabilidad:** Los modelos de Deep Learning pueden adaptarse a cambios rápidos en los mercados y eventos inesperados.
- **Precisión:** Cuando se entrenan adecuadamente, los modelos de Deep Learning pueden lograr altos niveles de precisión en la predicción de precios de acciones.

### 3.2.3. Limitaciones y Desafíos

Las principales limitaciones y desafíos a los que nos enfrentamos al usar las técnicas de Deep Learning en la predicción de acciones son [33]:

- **Sobreajuste:** Los modelos de Deep Learning pueden sobreajustarse a los datos, lo que reduce la capacidad de generalización.
- **Interpretación Compleja:** La interpretación de los modelos de Deep Learning puede ser difícil debido a la complejidad de las redes, lo que plantea problemas de transparencia y explicabilidad.
- **Datos Ruidosos:** Los datos financieros pueden ser ruidosos y sujetos a eventos inesperados, lo que desafía la precisión de los modelos.
- **Limitaciones de Datos Históricos:** Los modelos de Deep Learning dependen en gran medida de los datos históricos, lo que puede ser insuficiente para abordar eventos inesperados o de gran envergadura.

### 3.2.4. Aplicaciones de Deep Learning en la Predicción de Acciones

Las diferentes aplicaciones del uso de las técnicas del Deep Learning son [28]:

1. **Predicción de Precios de Acciones:** La aplicación más evidente es la predicción de los precios futuros de las acciones, lo que es de gran interés para inversores y operadores.
2. **Gestión de Cartera:** Los modelos de Deep Learning se utilizan para optimizar la composición de carteras, minimizando el riesgo y maximizando el rendimiento.
3. **Análisis de Sentimientos en el Mercado:** Los modelos analizan noticias y datos en tiempo real para evaluar el sentimiento del mercado y su influencia en los precios.
4. **Evaluación de Riesgos:** Los inversores utilizan modelos de Deep Learning para evaluar y gestionar mejor los riesgos asociados con la inversión en acciones.

5. **Simulación de Escenarios:** Las GAN se utilizan para simular escenarios y evaluar estrategias de inversión en diferentes condiciones del mercado.

### 3.2.5. Principales artículos

Los siguientes artículos muestran algunos estudios sobre el uso de Deep Learning y Redes Neuronales en la Predicción del Precio de las Acciones

#### 3.2.5.1. Dynamic Dependence Networks: Financial Time Series Forecasting & Portfolio Decisions

El artículo [48] presenta un modelo basado en **Redes de Dependencia Dinámica (DDNM)** para la predicción de series financieras y la toma de decisiones en carteras financieras. Este enfoque incorpora una metodología de aprendizaje estructurado para preservar la adaptabilidad del modelo a lo largo del tiempo y evitar la degradación de las probabilidades del modelo posterior.

El enfoque se basa de un modelo de regresión dinámica múltiple, donde cada serie temporal univariada se representa mediante una secuencia de distribuciones condicionales, conectadas a través de relaciones de regresión dinámica. La precisión predictiva de las **DDNM** se logra mediante modelos univariados independientes interconectados por relaciones de regresión dinámica.

El artículo también presenta un estudio de caso que ilustra la aplicabilidad de las **DDNM** en la predicción a corto plazo y su desempeño en diversos activos financieros. Un apéndice técnico proporciona detalles sobre la selección del modelo y la derivación de las predicciones.

En resumen, los resultados del estudio demuestran que las **Redes de Dependencia Dinámica** son una herramienta valiosa para la predicción de series financieras y la toma de decisiones en carteras financieras. Su capacidad para capturar relaciones dinámicas y no lineales, adaptarse a los cambios en los mercados y superar a los modelos de referencia las convierte en una elección prometedora en la gestión de carteras financieras.

### 3.2.5.2. On development of novel hybrid and robust adaptive models for net asset value prediction

El artículo [34] se enfoca en la creación de un modelo de predicción del valor neto de fondos de inversión mediante el empleo de una **red neuronal artificial de retropropagación (BPANN)**. Se procedió a la evaluación de dicho modelo **BPANN** en el contexto de seis fondos distintos, con el propósito de comparar su rendimiento frente a otros modelos de predicción existentes.

Los autores recolectaron datos históricos correspondientes a los seis fondos objeto de estudio, los cuales sirvieron como conjunto de entrenamiento y prueba para la implementación del modelo **BPANN**. Se utilizan métricas de error, tales como el **error medio porcentual absoluto (MAPE)** y el **error cuadrático medio (RMSE)**, para llevar a cabo una evaluación comparativa del desempeño de los diversos modelos de predicción.

Los resultados obtenidos reflejan que el modelo **BPANN** híbrido demuestra una notable mejora en la capacidad de predecir el valor neto de los fondos mutuos. Este modelo híbrido se compone de una combinación de dos modelos fundamentales: un modelo **ARMA** adaptativo y un modelo de medias móviles adaptables (**AMA**). Cabe destacar que el modelo **BPANN** híbrido es objeto de optimización mediante el empleo del algoritmo de búsqueda de recocido simulado (**SLSA**).

Adicionalmente, se destaca que el modelo propuesto se caracteriza por su resistencia frente a valores atípicos y su capacidad para gestionar eficazmente la contaminación de datos en el proceso de predicción. Asimismo, se comprueba que el modelo supera a otros enfoques tradicionales de predicción, como los modelos de **redes neuronales de propagación hacia adelante de capa única (FLANN)** y los modelos híbridos basados en estadísticas convencionales.

En resumen, el modelo presentado se distingue por su elevada capacidad predictiva y se configura como una herramienta efectiva en la anticipación del valor neto de los fondos mutuos.





# Capítulo 4

## Diseño e implementación

Este capítulo tiene como objetivo detallar exhaustivamente el proceso de diseño e implementación del estudio realizado. Con el fin de facilitar la comprensión de este proceso, se presenta el siguiente diagrama 4.1 que esquematiza el flujo seguido en la ejecución del estudio.

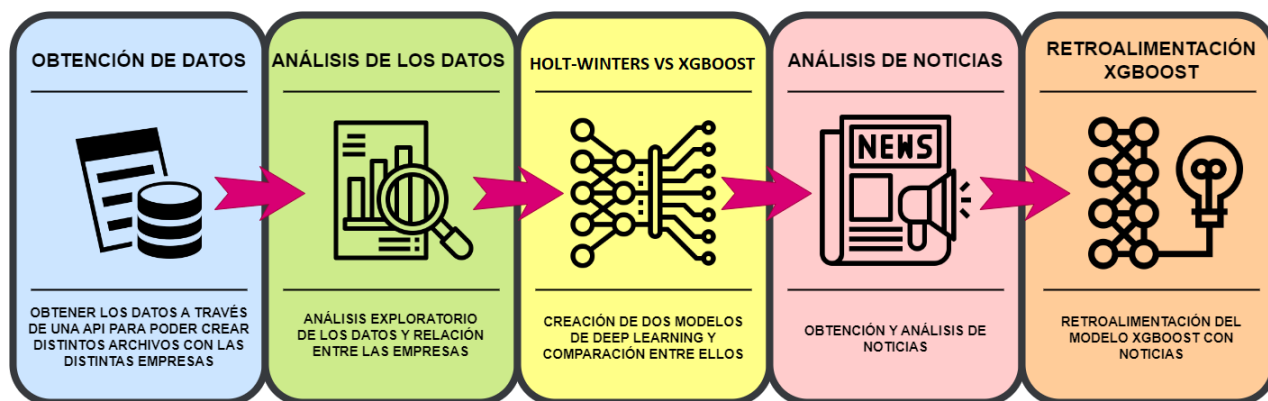


Figura 4.1: Diagrama de las fases del estudio

En las secciones siguientes, se llevará a cabo un análisis minucioso de cada una de las etapas involucradas en dicho proceso, proporcionando así una visión más detallada y clara de la metodología empleada.

### 4.1. Obtención de los datos

En esta etapa del estudio, se llevó a cabo la recopilación de datos financieros esenciales para el análisis del sector eléctrico. La plataforma *Yahoo Finanzas* y su API asociada, *yfinance*, se seleccionaron como la fuente principal de información para obtener detalles sobre el historial

de cotizaciones y dividendos de diversas empresas del sector. A continuación, se presenta un resumen del procedimiento seguido y los resultados obtenidos.

#### 4.1.1. Justificación del uso de la API de *yfinance*

La elección de utilizar *yfinance* en lugar de otras API o métodos, como el *web scraping*, se fundamenta en una serie de consideraciones que subrayan la fiabilidad y eficiencia de *yfinance* para obtener datos financieros. A continuación, se detallan algunas justificaciones para esta elección:

- **Fiabilidad de los Datos:** La API de *yfinance* ha demostrado ser una fuente confiable al proporcionar valores financieros fiables y precisos. La consistencia y exactitud de los datos son esenciales en el análisis financiero, y la capacidad de *yfinance* para cumplir con estas expectativas ha sido validada mediante pruebas comparativas con el histórico del mercado de valores de distintas fuentes.
- **Facilidad de Uso:** La API de *yfinance* ofrece una interfaz simple y accesible, lo que facilita considerablemente la implementación y extracción de información financiera relevante. La simplicidad de la API puede ser un factor determinante, especialmente al compararse con métodos más complejos como el *web scraping*.
- **Evitar la Complejidad del Web scraping:** Aunque el *web scraping* es una alternativa para obtener datos, puede ser un proceso más complejo y susceptible a cambios en la estructura de los sitios web objetivo. Además, el *web scraping* puede estar sujeto a restricciones legales o términos de servicio de las plataformas web. Optar por *yfinance* permite evitar la complejidad inherente al *web scraping*.

#### 4.1.2. Datos de Cotización y Dividendos

Dentro del ámbito de investigación centrado en el sector eléctrico, el conjunto de empresas seleccionadas pertenecientes al IBEX-35 abarca un grupo representativo que incluye a Endesa, Iberdrola, Acciona Energía, Solaria y Naturgy. Estas entidades desempeñan un papel crucial

en el panorama energético, contribuyendo significativamente al sector eléctrico y, por ende, constituyen objetos de estudio relevantes en la investigación.

El periodo de análisis abarcó desde el 2 de enero de 2001 hasta el 2 de junio de 2023. Para obtener el historial de cotizaciones de acciones de cada empresa, se utilizó la función **yf.download** de *yfinance*. Los resultados se almacenaron en archivos CSV específicos para su posterior análisis. Además, se extrajeron los datos de dividendos utilizando la misma API.

#### 4.1.2.1. Ejemplo con Endesa

Como caso representativo, se detalla el proceso seguido para la empresa Endesa. Los datos de cotización y dividendos se guardaron en dos archivos *.csv* como se muestra a continuación:

```
1 cotizacion_Endesa = yf.download('ELE.MC', start=start, end=end)
2 cotizacion_Endesa.to_csv('./data/cotizacion_Endesa.csv')
3 cotizacion_Endesa.head()
```

Listing 4.1: Descarga y almacenamiento del historial de cotizaciones de Endesa

	Open	High	Low	Close	Adj Close	Volume
Date						
2001-01-02	18.150000	18.770000	17.700001	18.68	2.334867	3900039
2001-01-03	18.379999	20.070000	18.379999	19.90	2.487358	10223480
...						

```
1 dividendos_Endesa = yf.Ticker('ELE.MC').dividends.loc[start:end]
2 dividendos_Endesa.to_csv('./data/dividendos_Endesa.csv')
3 dividendos_Endesa.head()
```

Listing 4.2: Descarga y almacenamiento de datos de dividendos de Endesa

Date	
2001-01-08	0.059
2001-07-02	0.041
...	

#### 4.1.2.2. Consideraciones para Otras Empresas

El procedimiento de obtención de datos se replicó para otras empresas seleccionadas del sector eléctrico, generalizando la metodología utilizada para Endesa. Se destaca que algunas empresas, como Acciona Energía, presentaron limitaciones, iniciando su serie temporal en fechas más recientes, como 2021-07-02. A pesar de esta variabilidad en los datos disponibles, se almacenaron todos los registros disponibles para garantizar la integridad de la información recopilada en todas las empresas del sector.

## 4.2. Análisis de los datos

En esta sección, se llevará a cabo un análisis inicial y exploratorio de los datos correspondientes a las empresas previamente mencionadas. Este análisis se estructurará en diversas secciones con el objetivo de examinar aspectos clave. A continuación, se detalla la desglose de cada sección:

1. **Análisis de los Precios de Apertura y Cierre de las Acciones:** En esta primera parte, se examinarán en detalle los datos relativos a los precios de apertura y cierre de las acciones de las empresas seleccionadas.
2. **Promedio Móvil (Moving Average), Golden Cross y Death Cross:** En esta sección, se aplicará el análisis de promedio móvil para evaluar las tendencias a largo plazo en los precios de las acciones. Además, se identificarán los Golden Cross (cruce dorado) y Death Cross (cruce de la muerte), eventos significativos en el análisis técnico que pueden indicar cambios en la dirección de la tendencia.
3. **Descuadres en el Precio de las Acciones:** Se abordarán posibles descuadres en los precios de las acciones, y se buscará explicación para entender las variaciones significativas. Este análisis incluirá factores externos e internos que podrían influir en la dinámica de los precios.

4. **Análisis de Dividendos:** Se examinarán los datos de dividendos de las empresas seleccionadas, destacando los periodos de distribución y la relación entre los rendimientos de dividendos de las distintas empresas.
5. **Relación Entre Empresas:** Finalmente, se explorará la relación entre las empresas seleccionadas, analizando posibles correlaciones en el comportamiento de los precios de las acciones. Este análisis proporcionará una perspectiva más amplia sobre la dinámica del sector eléctrico en su conjunto.

#### 4.2.1. Análisis de los Precios de Apertura y Cierre de las Acciones

El precio de apertura es el primer precio al cual se negocia una acción al comenzar el día hábil de operaciones en el mercado. Es esencialmente el valor inicial al que se compra y vende la acción cuando se inicia la sesión de trading. Tanto el precio de apertura como el de cierre son datos clave utilizados por los inversores para evaluar el rendimiento de una acción durante un período específico, proporcionando información sobre la actividad del mercado en el inicio y el cierre de cada jornada bursátil.

Las representaciones visuales que se muestran a continuación ofrecen un análisis del rendimiento histórico de las acciones de las diversas empresas a lo largo de varios años (Endesa: [4.2](#), Iberdrola: [4.3](#), Acciona Energía: [4.4](#), Naturgy: [4.5](#), Solaria: [4.6](#)). Cada representación consta de dos subgráficos: un gráfico de velas que ilustra la dinámica de precios diarios y un gráfico de barras que refleja el volumen de transacciones diarias.

Estos gráficos, como se puede comprobar en el fichero adjunto a esta memoria, son interactivos. Permiten la navegación a lo largo de los datos, facilitando la obtención de información detallada sobre los valores, tal y como se presenta en [4.8](#).

Como se puede analizar visualmente, el crecimiento y decrecimiento no sigue un patrón cíclico evidente. Aunque en ocasiones parece que sigue ciertos patrones, no es posible extraer una norma precisa de manera intuitiva para predecir el comportamiento.



Figura 4.2: Crecimiento de las acciones de Endesa a lo largo de los años



Figura 4.3: Crecimiento de las acciones de Iberdrola a lo largo de los años



Figura 4.4: Crecimiento de las acciones de Acciona Energía a lo largo de los años



Figura 4.5: Crecimiento de las acciones de Naturgy a lo largo de los años



Figura 4.6: Crecimiento de las acciones de Solaria a lo largo de los años

Figura 4.7: Comparación del crecimiento de las acciones de diferentes empresas a lo largo de los años



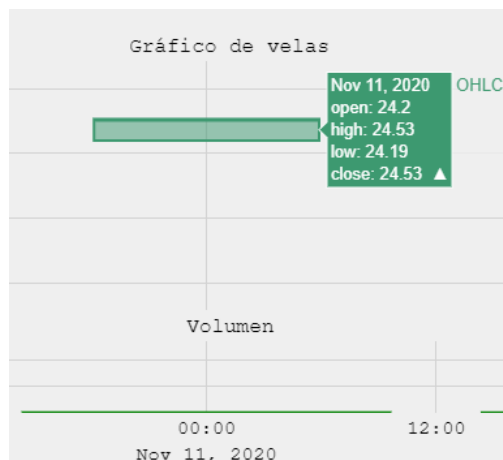


Figura 4.8: Información detallada sobre la cotización del día

#### 4.2.2. Promedio Móvil (Moving Average), Golden Cross y Death Cross

Tal y como se mencionó en los puntos 2.1.4 y 2.1.5 de este documento el **promedio móvil** o **Moving Average** (MA) es una herramienta sencilla de análisis técnico que suaviza los datos de precios al crear un promedio de precios constantemente actualizado. El promedio se calcula durante un período específico de tiempo, como 50 días, 20 minutos, 30 semanas, o cualquier periodo de tiempo que el trader elija. Como, por ejemplo, se mostrará el grafico de las acciones de Endesa 4.9.

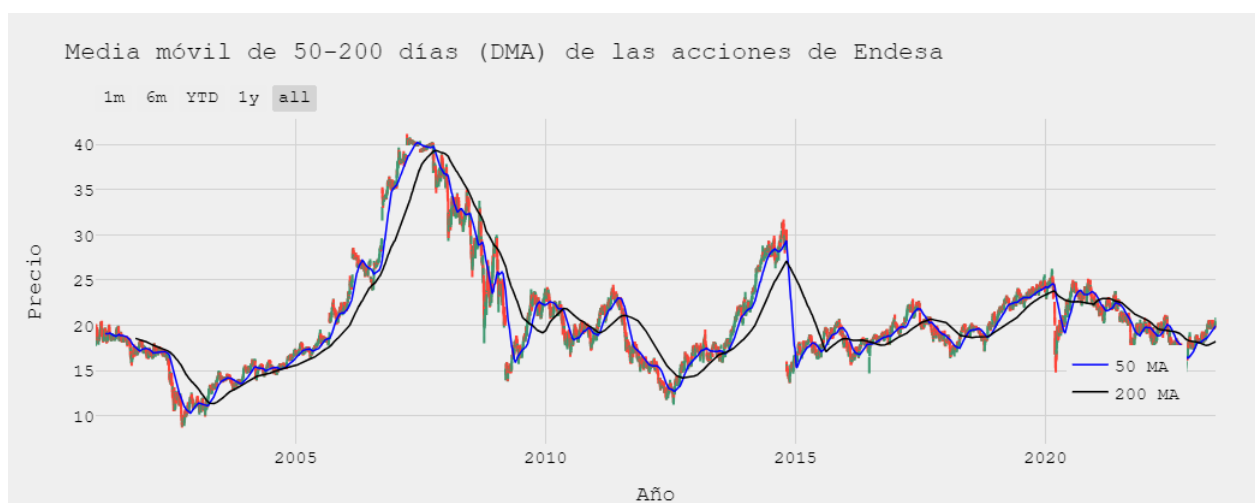


Figura 4.9: Media móvil de 50-200 días (DMA) de las acciones de Endesa

Como se puede comprobar, en algunos puntos se cortan las dos líneas; este suceso se conoce como **Golden Cross** y **Death Cross**. Estos patrones pueden tener implicaciones sobre la dirección futura de los precios y son seguidos por algunos inversores y traders como señales potenciales de cambio en la tendencia del mercado.

El **Golden Cross** (**Cruz Dorada**) se produce cuando la media móvil de corto plazo cruza por encima de la media móvil de largo plazo (comúnmente se refiere al cruce de la media móvil de 50 días sobre la media móvil de 200 días). Esto es considerado como una señal alcista, indica un posible cambio positivo en la tendencia del mercado y sugiere un potencial aumento en los precios [24].

El **Death Cross** (**Cruz de la Muerte**) ocurre cuando la media móvil de corto plazo cruza por debajo de la media móvil de largo plazo, especialmente el cruce de la media móvil de 50 días por debajo de la media móvil de 200 días. Esto es interpretado como una señal bajista, sugiere un posible cambio negativo en la tendencia del mercado y puede indicar una disminución en los precios [24].

Para determinar estos cruces, a nivel de código se han calculado las fechas en las que se han producido dichos eventos y se han añadido puntos correspondientes en el gráfico anterior 4.10.

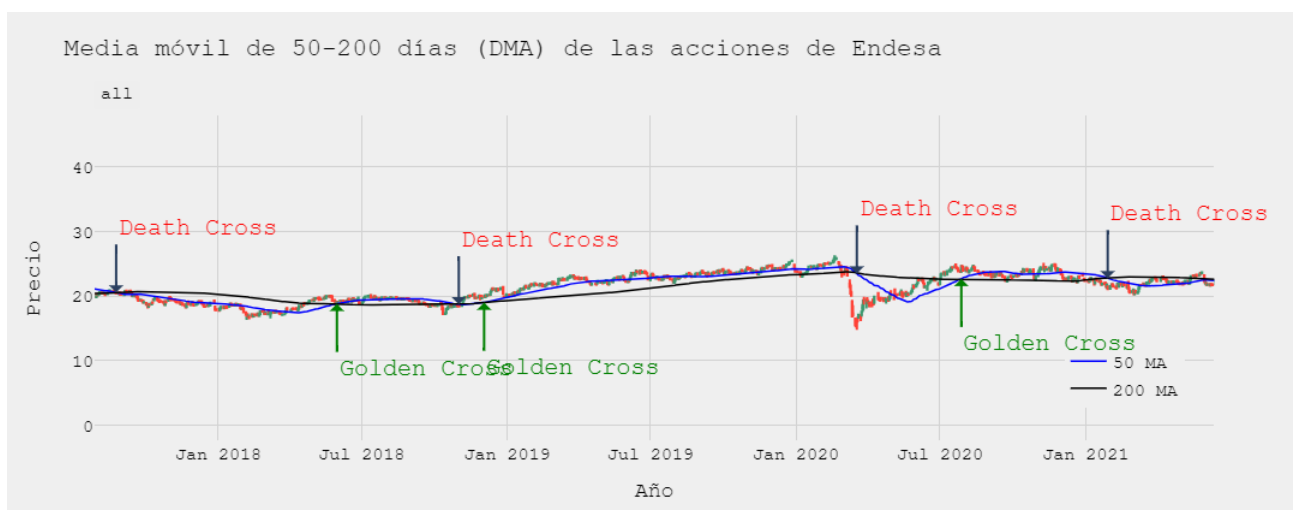


Figura 4.10: Media Móvil de 50-200 días (DMA) de las acciones de Endesa con marcadores Golden Cross y Death Cross (junio de 2018 - junio de 2021)

El algoritmo seguido de calculo ha sido el siguiente:

```

1 # Crea una columna donde 1 indica que MA50 es mayor que MA200
2 cotizacion_Endesa['Cruce'] =
3     (cotizacion_Endesa['MA50'] > cotizacion_Endesa['MA200']).astype(int)
4
5 # Encuentra las diferencias entre filas consecutivas
6 cotizacion_Endesa['Cambio'] = cotizacion_Endesa['Cruce'].diff()
7
8 # Las fechas de cruce hacia arriba son donde Cambio es 1
9 cruce_arriba = cotizacion_Endesa[cotizacion_Endesa['Cambio']==1].index
10
11 # Las fechas de cruce hacia abajo son donde Cambio es -1
12 cruce_abajo = cotizacion_Endesa[cotizacion_Endesa['Cambio']==-1].index
13
14 print("Golden_Cross:", cruce_arriba, "\nDeath_Cross:", cruce_abajo)

```

Listing 4.3: Calculo de las fecha de Golden Cross y Death Cross

```

Golden_Cross: DatetimeIndex([
    '2003-04-15', '2004-10-08', '2009-09-24', '2011-02-04', '2012-10-04',
    '2013-08-20', '2015-08-04', '2016-08-04', '2018-05-31', '2018-12-03',
    '2020-07-28', '2022-05-19', '2023-01-23'], dtype='datetime64[ns]',
    name='Date', freq=None)

```

```

Death_Cross: DatetimeIndex([
    '2004-09-17', '2007-10-19', '2010-05-06', '2011-08-18', '2013-07-29',
    '2014-11-12', '2016-01-29', '2017-08-25', '2018-11-01', '2020-03-18',
    '2021-01-29', '2022-07-20'], dtype='datetime64[ns]',
    name='Date', freq=None)

```

Todo este proceso se ha llevado a cabo para el resto de las empresas, como se puede comprobar en los siguientes gráficos [4.11](#). Además, tal y como se mencionó en el punto anterior, en el documento adjunto se puede navegar por los gráficos para obtener una visualización más clara y precisa.

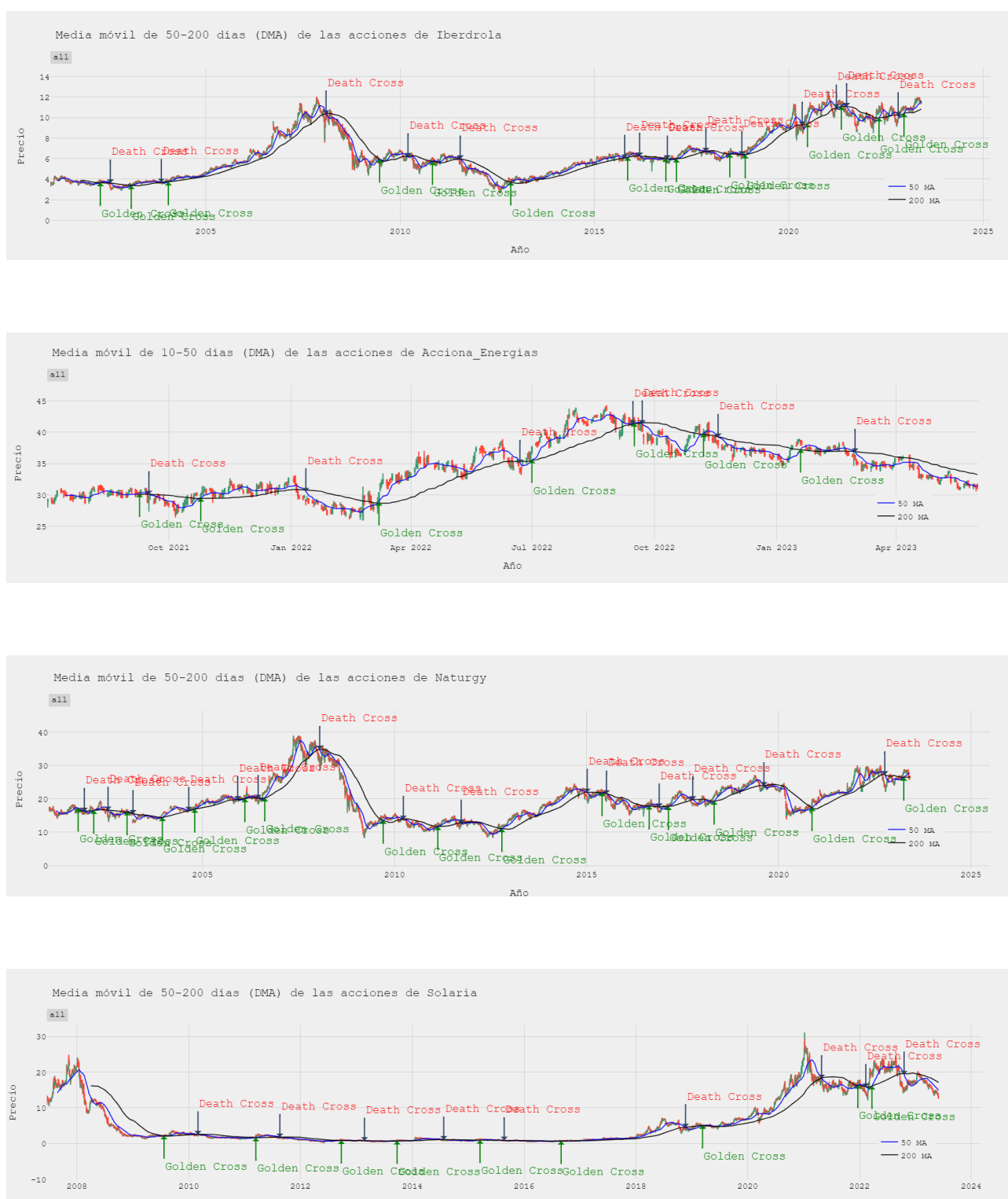


Figura 4.11: Media móvil de 50-200 días (DMA) con marcadores Golden Cross y Death Cross del resto de empresas

### 4.2.3. Discrepancias en los Valores de Cierre y Apertura de Acciones

Realizamos un análisis para detectar días en los cuales se presentan divergencias entre el precio de cierre del día anterior y la apertura del día actual en el mercado de valores. Estas discrepancias pueden ser atribuibles a información que surge después del cierre del mercado, operaciones realizadas fuera del horario bursátil (after-hours trading), acumulación de órdenes, o transacciones internacionales. En todas las circunstancias, nuestro objetivo es identificar desviaciones que superen el umbral del 2%, con el propósito de llevar a cabo un análisis detenido de posibles patrones emergentes.

El algoritmo seguido de calculo ha sido el siguiente (poniendo como ejemplo el de las acciones de Endesa y replicandolo para el resto de las empresas):

```

1 # Calcula el cambio porcentual diario en el precio de apertura con respecto
2 # al precio de cierre del dia anterior. El metodo shift(1) desplaza los
3 # datos una posicion, por lo que cotizacion_Endesa[Close].shift(1) se
4 # refiere al precio de cierre del dia anterior
5 cotizacion_Endesa['Diff_Percent'] =
6     ((cotizacion_Endesa['Open'] - cotizacion_Endesa['Close'].shift(1)) /
7      cotizacion_Endesa['Close'].shift(1)) * 100
8
9 # Selecciona las filas donde el cambio porcentual diario (en valor absoluto)
10 # es superior al 2%
11 dias_superiores_2_pc =
12     cotizacion_Endesa[cotizacion_Endesa['Diff_Percent'].abs() > 2]
13
14 # Convierte el indice del DataFrame resultante (que contiene las fechas de
15 # los dias seleccionados) a una lista de strings con formato 'YYYY-MM-DD'
16 lista_dias_superiores_2_pc =
17     dias_superiores_2_pc.index.strftime('%Y-%m-%d').tolist()
18
19 # Imprime la lista de fechas
20 print(lista_dias_superiores_2_pc)

```

Listing 4.4: Búsqueda de descuadres en las acciones

```
['2001-02-06', '2001-05-02', ... , '2023-02-21', '2023-03-08']
```

Se ha llevado a cabo un primer análisis manual de las desviaciones más destacadas en los precios de las acciones. Este estudio inicial se ha querido enfocar en identificar y examinar los eventos que tuvieron lugar durante el día en que se registraron las desviaciones, con el propósito de ofrecer una explicación.

### 4.2.3.1. Discrepancias en los Valores de Endesa

En el gráfico 4.12, se ha llevado a cabo un conteo de las discrepancias en los valores de las acciones de Endesa, categorizando la información según tres dimensiones temporales: año, semana y día de la semana.

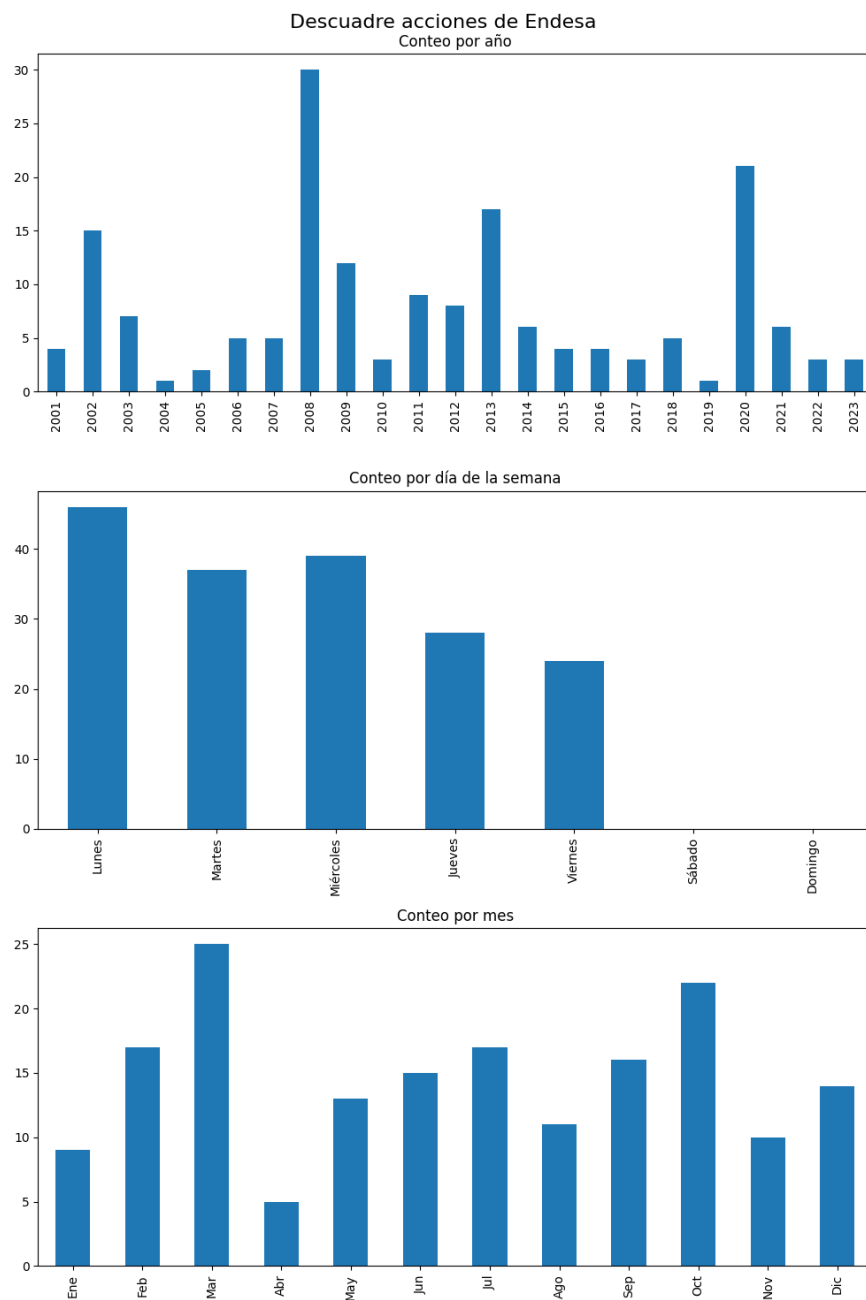


Figura 4.12: Descuadre acciones de Endesa por año, semana y día

En las tablas siguientes, se presentarán las fechas en las que se observaron las mayores diferencias porcentuales, tanto positivas 4.1 como negativas 4.2, en las acciones de Endesa. Además, se explorará, a modo de ejemplo, intentar dar una explicación para entender las razones detrás de estas variaciones positivas 4.13.

Date	Open	High	Low	Close	Adj Close	Volume	Diff Percent
2005-09-06	21.25	21.27	20.30	20.55	2.93	57627240	11.31
2006-02-21	27.50	27.94	27.40	27.55	3.99	51632444	7.93
2006-09-26	31.71	33.10	31.65	32.50	4.93	211188464	7.86
2006-09-27	35.01	35.25	34.55	35.00	5.31	35571756	7.72
2020-05-05	21.00	21.00	20.24	20.24	15.34	1102381	7.42

Cuadro 4.1: Top 5 días con mayor diferencia porcentual positiva en las acciones de Endesa

The image shows a screenshot of the El Mundo website. The top part of the page features a news article dated 2006-09-26 and 2006-09-27, titled "E.ON aumenta su oferta por Endesa 10 euros por acción tras la maniobra de Acciona". The article discusses E.ON's offer to increase its stake in Endesa. Below this, there is a section dated 2005-09-06, titled "El presidente de Endesa compró 50.000 acciones de la empresa el viernes", which reports on the purchase of shares by Endesa's president. In the bottom left corner, the date 2020-05-05 is visible, likely indicating when the screenshot was taken. The page also includes a search bar, navigation links, and a sidebar with related news.

Figura 4.13: Noticias Endesa [16] y [17]

Si observamos uno a uno los descuadres más importantes, encontramos un patrón de noticias que respaldan cada fecha, si bien pueden no ser fácilmente predecibles 4.13.

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2014-10-29	15.80	16.00	14.89	15.13	8.20	9657735	-45.70
2009-03-16	14.04	15.05	14.02	14.70	3.39	3415061	-31.38
2016-06-24	14.97	16.47	14.69	16.17	9.37	3388533	-15.12
2008-10-10	18.21	20.74	18.01	18.63	3.06	1896132	-9.27
2006-07-03	25.21	25.40	25.11	25.26	3.83	14952236	-7.28

Cuadro 4.2: Top 5 días con mayor diferencia porcentual negativa en las acciones de Endesa

#### 4.2.3.2. Discrepancias en los Valores de Iberdrola

En las tablas siguientes, se presentarán las fechas en las que se observaron las mayores diferencias porcentuales, tanto positivas [4.3](#) como negativas [4.4](#), en las acciones de Iberdrola.

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2008-10-14	6.490	6.680	6.160	6.390	3.012716	6158172	17.148011
2008-10-13	5.910	5.910	5.540	5.540	2.611963	82854064	14.534884
2006-09-27	9.4050	9.6675	9.300	9.540	4.245751	429742016	12.837429
2008-01-30	9.900000	9.980000	9.400000	9.560000	4.424663	83202880	7.258939
2001-02-08	4.275000	4.310000	4.187500	4.247500	1.518689	68618440	7.210035

Cuadro 4.3: Top 5 días con mayor diferencia porcentual positiva en las acciones de Iberdrola

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2016-06-24	4.808	5.580	4.800	5.388	3.862	71430304	-19.867
2020-03-09	9.890	10.395	9.764	9.892	8.534	41253941	-6.654
2012-04-18	3.700	3.720	3.565	3.594	2.072	339439674	-5.153
2020-03-12	9.008	9.162	8.008	8.152	7.033146	65237950	-5.039
2020-03-16	8.136	8.380	7.760	8.232	7.102166	51624713	-4.9532

Cuadro 4.4: Top 5 días con mayor diferencia porcentual negativa en las acciones de Iberdrola

En el gráfico [4.14](#), se ha llevado a cabo un conteo de las discrepancias en los valores de las acciones de Iberdrola, categorizando la información según tres dimensiones temporales: año, semana y día de la semana.



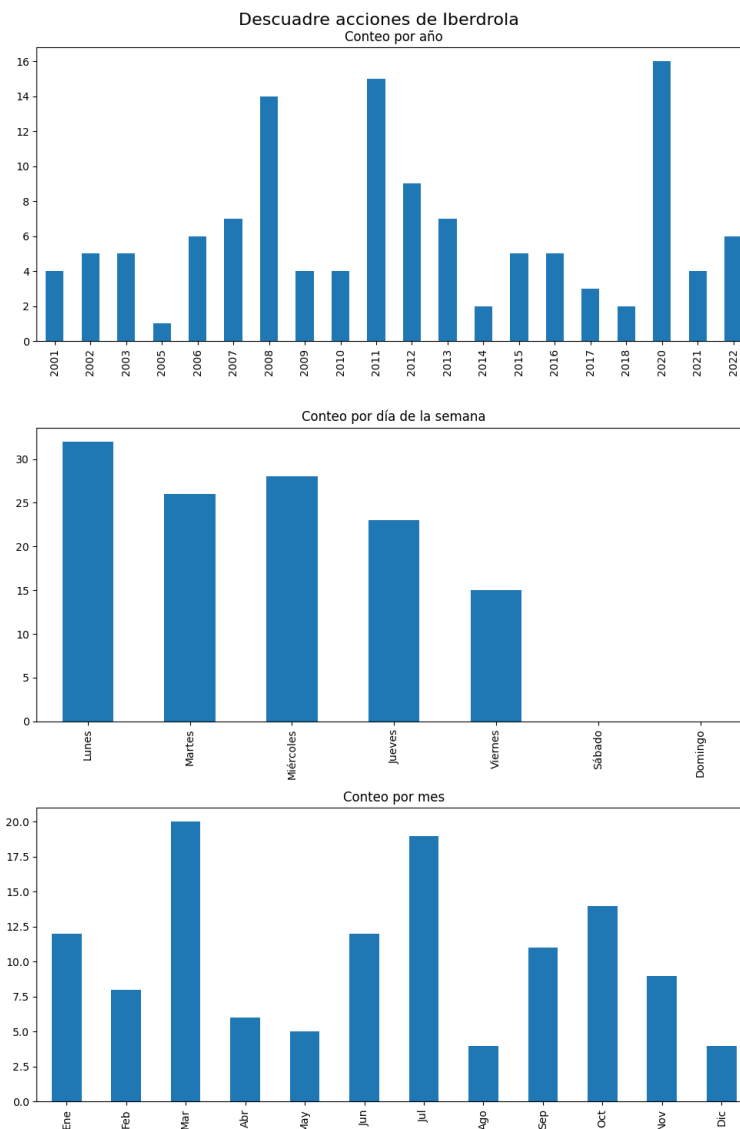


Figura 4.14: Descuadre acciones de Iberdrola por año, semana y día

#### 4.2.3.3. Discrepancias en los Valores de Acciona Energía

En las tablas siguientes, se presentarán las fechas en las que se observaron las mayores diferencias porcentuales, tanto positivas [4.5](#) como negativas [4.6](#), en las acciones de Acciona Energía.

En el gráfico [4.15](#), se ha llevado a cabo un conteo de las discrepancias en los valores de las acciones de Acciona Energía, categorizando la información según dos dimensiones temporales: año y día de la semana.

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2022-05-11	34.860	34.860	32.740	34.060	33.323	120201	5.829
2022-06-16	36.000	36.000	34.280	34.280	33.538	117517	4.167
2022-03-02	29.800	29.800	27.500	27.760	26.942	306011	3.869
2022-06-24	35.340	35.619	34.459	35.480	34.711	187320	3.636
2021-12-15	31.500	31.500	30.500	31.219	30.300	121815	3.312

Cuadro 4.5: Top 5 días con mayor diferencia porcentual positiva en las acciones de Acciona Energía

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2022-02-24	26.000	31.320	26.000	29.790	28.913	604362	-4.832
2021-10-25	29.050	30.290	29.050	30.000	29.116	107180	-3.037
2022-09-05	39.300	40.360	38.800	40.000	39.134	207838	-3.011
2022-08-04	42.400	43.060	42.000	42.160	41.247	163069	-2.930
2022-02-22	26.540	28.090	26.000	27.000	26.204	121114	-2.890

Cuadro 4.6: Top 5 días con mayor diferencia porcentual negativa en las acciones de Acciona Energía

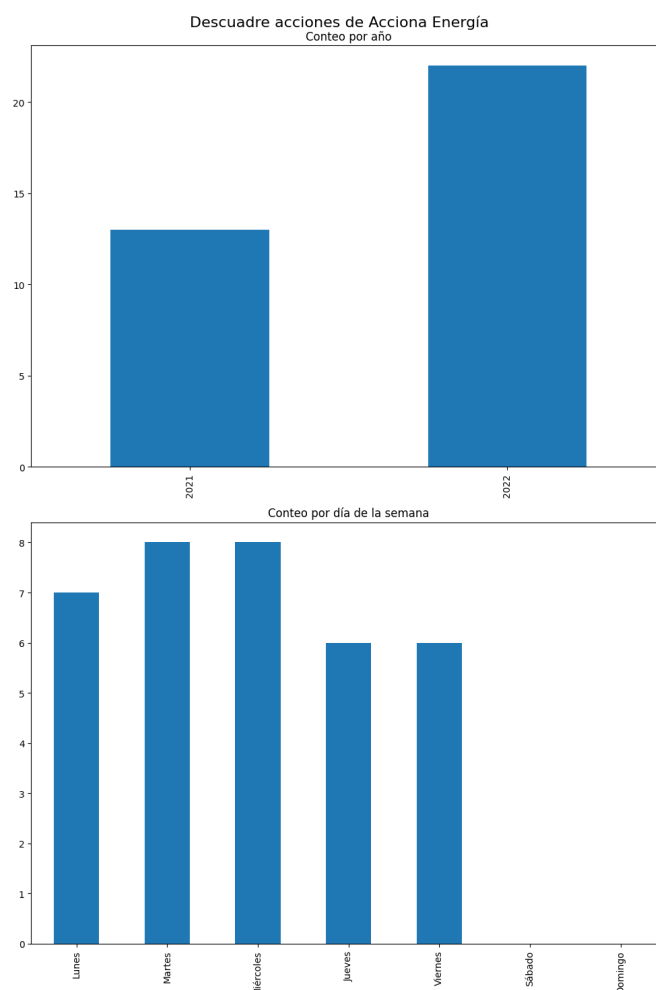


Figura 4.15: Descuadre acciones de Acciona Energía por año y día

#### 4.2.3.4. Discrepancias en los Valores de Solaria

En las tablas siguientes, se presentarán las fechas en las que se observaron las mayores diferencias porcentuales, tanto positivas 4.7 como negativas 4.8, en las acciones de Solaria.

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2012-06-07	0.950	1.060	0.815	0.950	0.950	2283333	25.000
2012-06-18	0.870	0.870	0.800	0.845	0.845	369830	13.725
2014-01-21	1.700	1.745	1.340	1.420	1.420	6435333	12.957
2008-10-29	2.370	2.370	2.000	2.120	2.0538	123811	11.792
2008-10-13	3.330	3.360	3.220	3.360	3.255	420990	11.7449

Cuadro 4.7: Top 5 días con mayor diferencia porcentual positiva en las acciones de Solaria

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2012-05-30	0.310	0.340	0.300	0.325	0.325	399493	-16.216
2018-10-23	3.165	3.410	2.850	2.850	2.850	2906767	-11.220
2020-03-09	7.980	8.575	7.900	8.150	8.150	1480580	-11.185
2020-03-12	7.100	7.365	5.800	6.160	6.160	2401987	-9.726
2016-11-09	0.650	0.720	0.650	0.715	0.715	723877	-9.722

Cuadro 4.8: Top 5 días con mayor diferencia porcentual negativa en las acciones de Solaria

En el gráfico 4.16, se ha llevado a cabo un conteo de las discrepancias en los valores de las acciones de Solaria, categorizando la información según tres dimensiones temporales: año, semana y día de la semana.

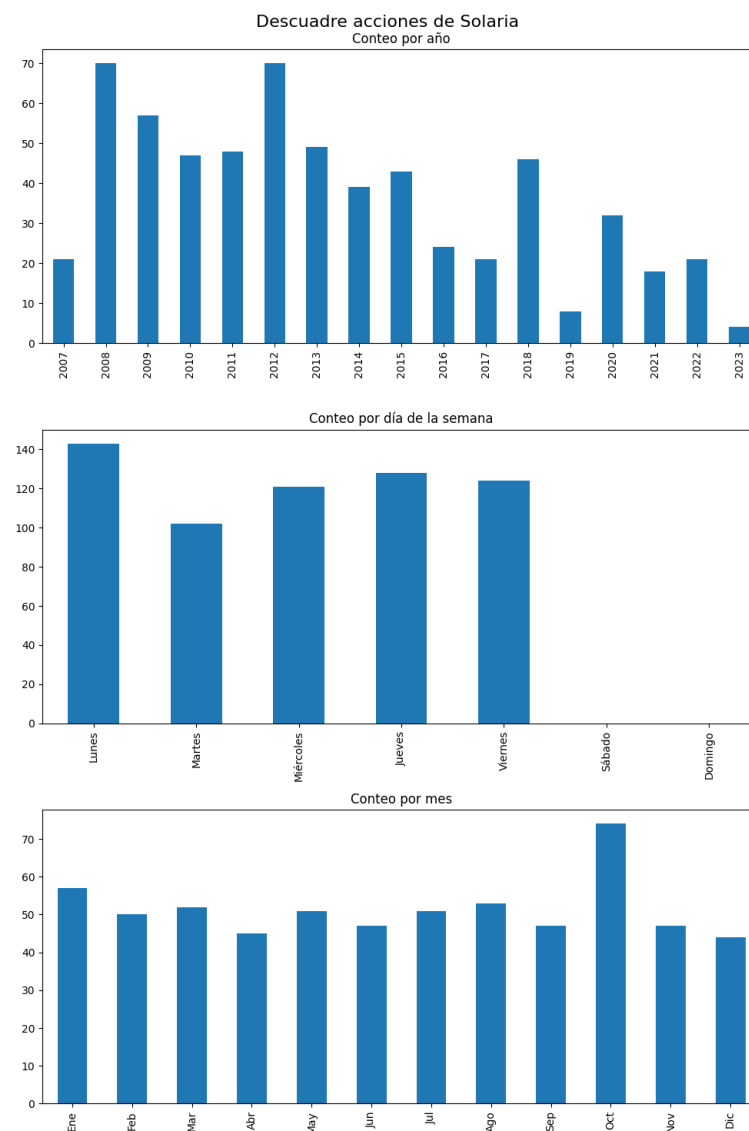


Figura 4.16: Descuadre acciones de Solaria por año, semana y día

#### 4.2.3.5. Discrepancias en los Valores de Naturgy

En las tablas siguientes, se presentarán las fechas en las que se observaron las mayores diferencias porcentuales, tanto positivas 4.9 como negativas 4.10, en las acciones de Naturgy.

En el gráfico 4.17, se ha llevado a cabo un conteo de las discrepancias en los valores de las acciones de Naturgy, categorizando la información según tres dimensiones temporales: año, semana y día de la semana.

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2021-01-26	22.630	22.710	22.020	22.190	18.857	4765095	17.773
2006-09-27	26.826	26.937	25.446	25.719	9.540	5652602	9.909
2010-05-10	12.390	12.600	12.050	12.600	5.253	5813492	7.366
2008-10-17	20.353	20.532	19.314	20.294	7.964	2086223	6.269463
2006-02-28	23.675	23.675	21.622	21.758	7.893	6450965	6.187933

Cuadro 4.9: Top 5 días con mayor diferencia porcentual positiva en las acciones de Naturgy

Date	Open	High	Low	Close	Adj Close	Volume	Diff_Percent
2008-10-10	16.615	19.451	15.763	17.875	7.016	2404926	-14.729
2003-03-10	13.030	13.192	12.706	13.030	4.329	5588376	-13.118
2016-06-24	15.550	16.490	14.685	16.010	9.541	3813881	-11.923
2020-03-23	15.760	16.830	15.100	16.020	12.96	1873586	-8.2387
2020-03-09	18.480	19.475	18.075	18.075	14.12	3127509	-7.7844

Cuadro 4.10: Top 5 días con mayor diferencia porcentual negativa en las acciones de Solaria

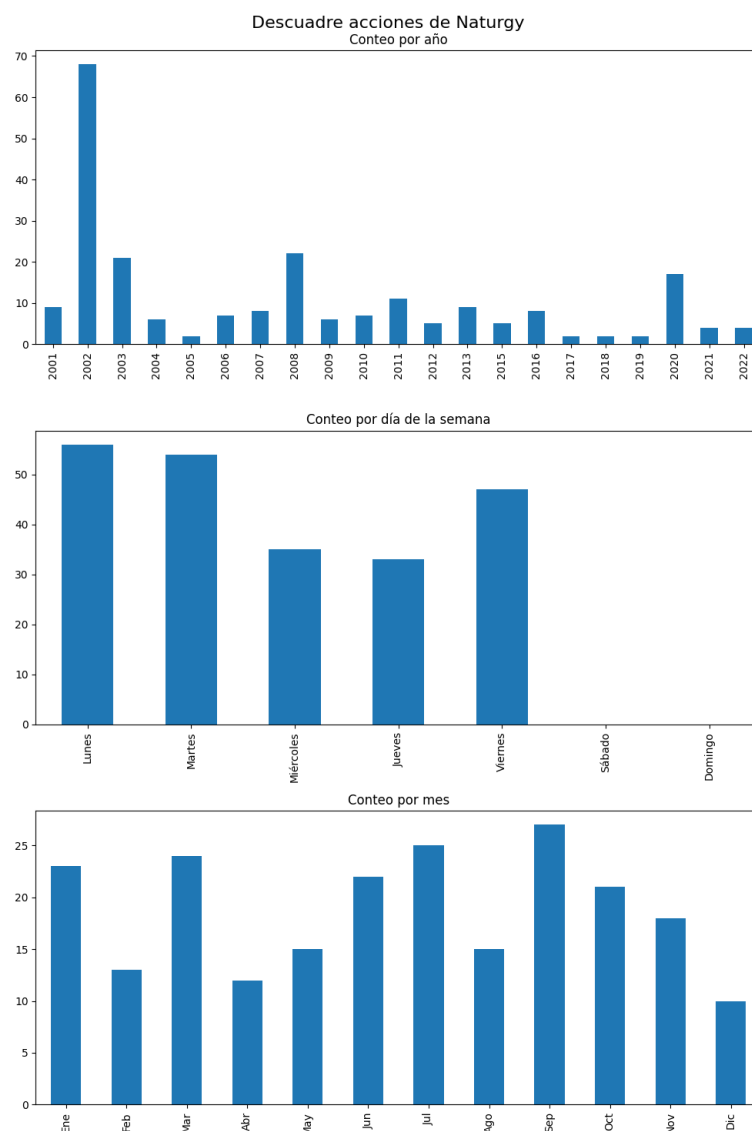


Figura 4.17: Descuadre acciones de Naturgy por año, semana y día

4.2.4. Análisis de Dividendos

Tal y como se mencionó en los puntos 2.1.6 un dividendo de una empresa es una distribución de las ganancias de esa empresa a sus accionistas. Las empresas públicas que cotizan en bolsa suelen pagar dividendos a los accionistas como una forma de compartir los beneficios generados. Estos pagos se realizan periódicamente, generalmente trimestral o anualmente, y representan una porción de las utilidades de la empresa.

4.2.4.1. Dividendos de Endesa

A lo largo de su trayectoria, Endesa ha otorgado diversos dividendos, como se evidencia en la siguiente tabla 4.2.4.1.

<b>Año</b>	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
<b>Precio</b>	0.325	0.341	0.341	0.351	0.369	0.786	0.820	0.766	5.897	0.514	0.509
<b>Año</b>	2012	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
<b>Precio</b>	0.606	7.648	0.380	0.587	0.667	0.691	0.714	0.738	1.072	0.937	1.573

Cuadro 4.11: Dividendos Endesa agrupados por año

La evolución del precio del dividendo ha sido como se ve en la figura 4.18.



Figura 4.18: Evolución de los dividendos Endesa

Si los valores de los dividendos se restan aritméticamente del precio de la acción, podríamos pensar que su efecto se puede revertir a la hora de manejar la serie temporal, teniendo en cuenta restarle el valor total de los dividendos para obtener la versión real. En la figura 4.19 se muestra el precio de la acción con y sin dividendos.

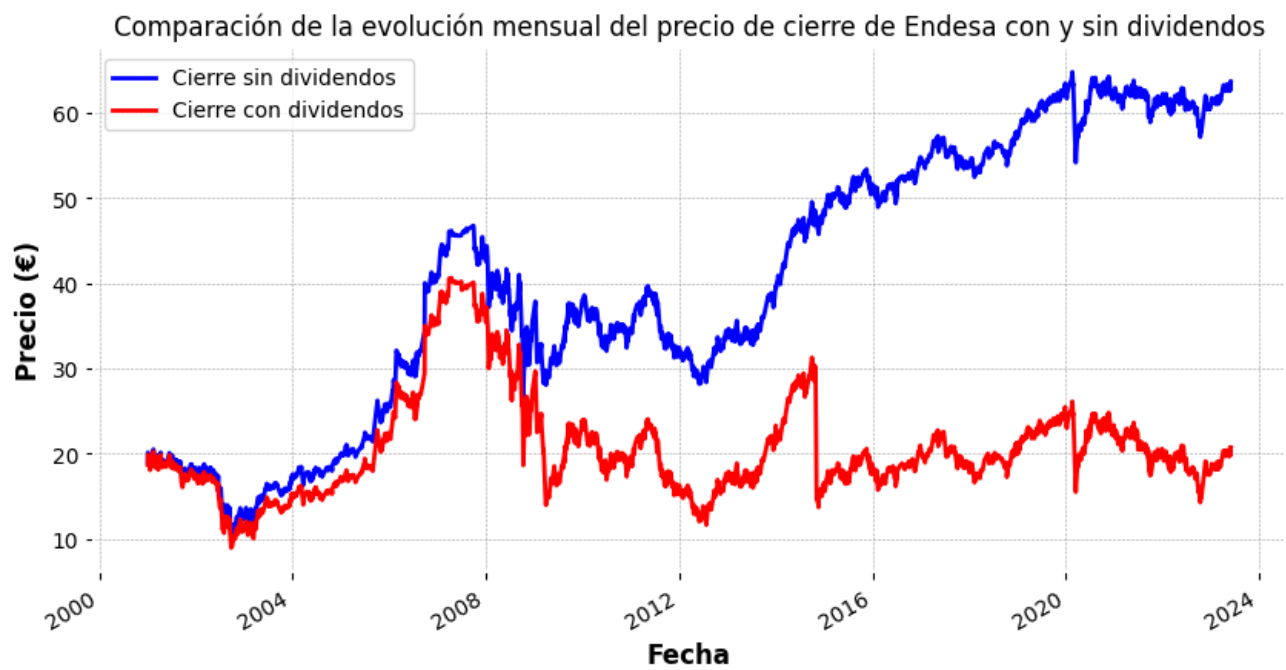


Figura 4.19: Comparación de la evolución mensual del precio de cierre de Endesa con y sin dividendos

4.2.4.2. Dividendos de Iberdrola

A lo largo de su trayectoria, Iberdrola ha otorgado diversos dividendos, como se evidencia en las siguiente tabla 4.2.4.2.

<b>Año</b>	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
<b>Precio</b>	0.069	0.074	0.076	0.084	0.096	0.111	0.130	0.136	0.164	0.172	0.181
<b>Año</b>	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
<b>Precio</b>	0.157	0.137	0.126	0.114	0.136	0.148	0.163	0.176	0.201	0.211	0.156

Cuadro 4.12: Dividendos Iberdrola agrupados por año

La evolución del precio del dividendo ha sido como se ve en la figura 4.22.

El valor total de los dividendos hasta el momento es de 6.33€

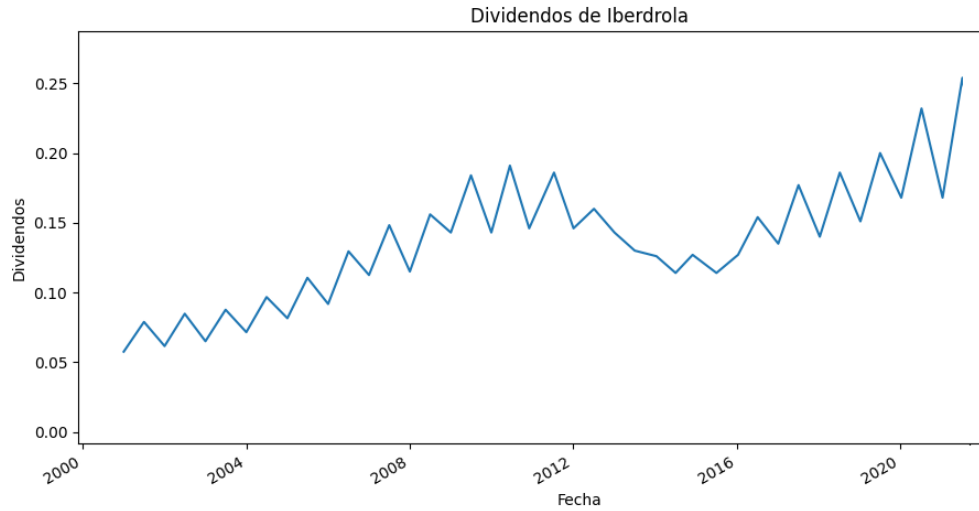


Figura 4.20: Evolución de los dividendos Iberdrola

Si los valores de los dividendos se restan aritméticamente del precio de la acción, podríamos pensar que su efecto se puede revertir a la hora de manejar la serie temporal, teniendo en cuenta restarle el valor total de los dividendos para obtener la versión real. En la figura 4.23 se muestra el precio de la acción con y sin dividendos.

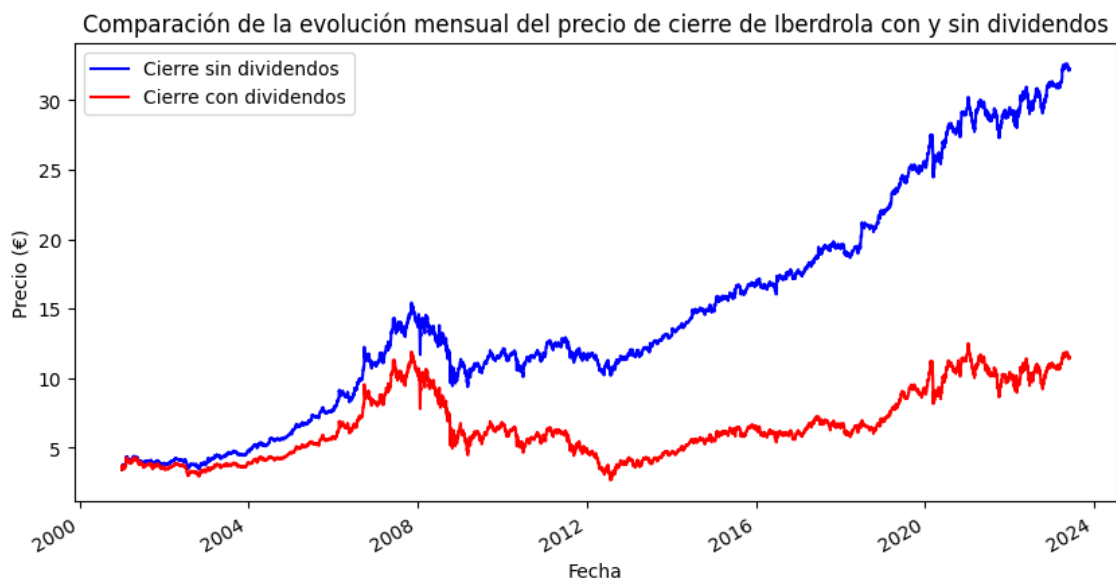


Figura 4.21: Comparación de la evolución mensual del precio de cierre de Iberdrola con y sin dividendos



#### 4.2.4.3. Dividendos de Acciona Energías

Esta empresa no ha repartido dividendo.

#### 4.2.4.4. Dividendos de Solaria

A lo largo de su trayectoria, Solaria ha otorgado diversos dividendos, como se evidencia en las siguiente tablas [4.2.4.4](#)

Fecha	Dividendos
2010-09-15	0.020
2011-05-25	0.015
2011-10-03	0.020

Cuadro 4.13: Dividendos Solaria

La evolución del precio del dividendo ha sido como se ve en la figura [4.22](#).

El valor total de los dividendos hasta el momento es de 0.055€

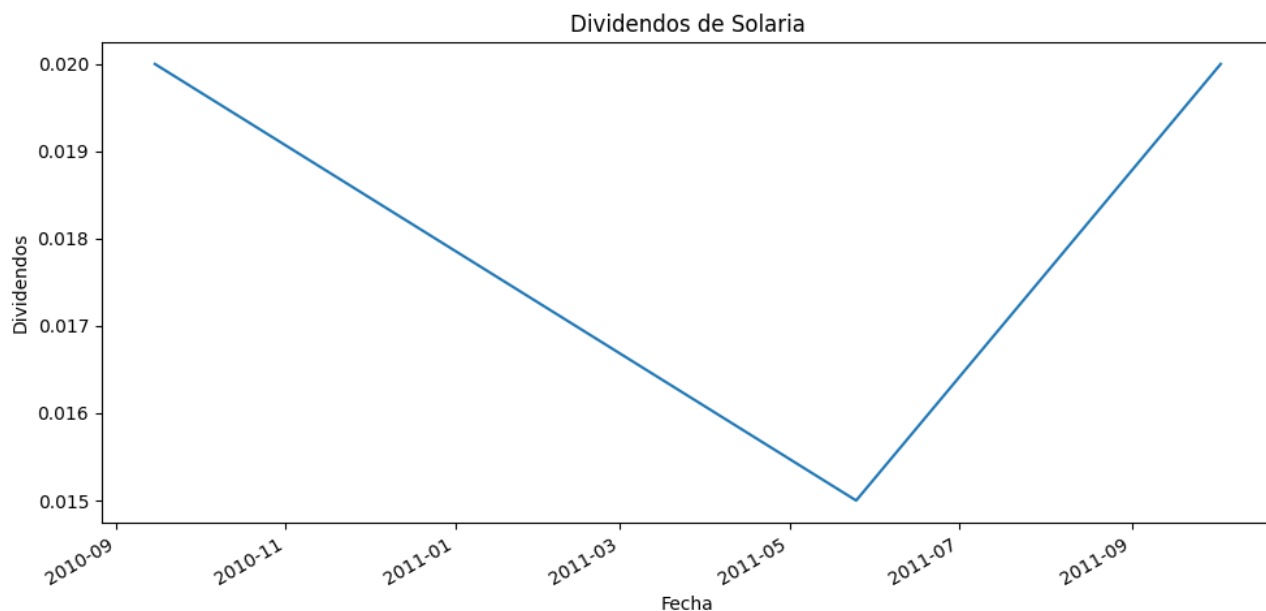


Figura 4.22: Evolución de los dividendos de Solaria

Si los valores de los dividendos se restan aritméticamente del precio de la acción, podríamos pensar que su efecto se puede revertir a la hora de manejar la serie temporal, teniendo en cuenta

restarle el valor total de los dividendos para obtener la versión real. En la figura 4.23 se muestra el precio de la acción con y sin dividendos.

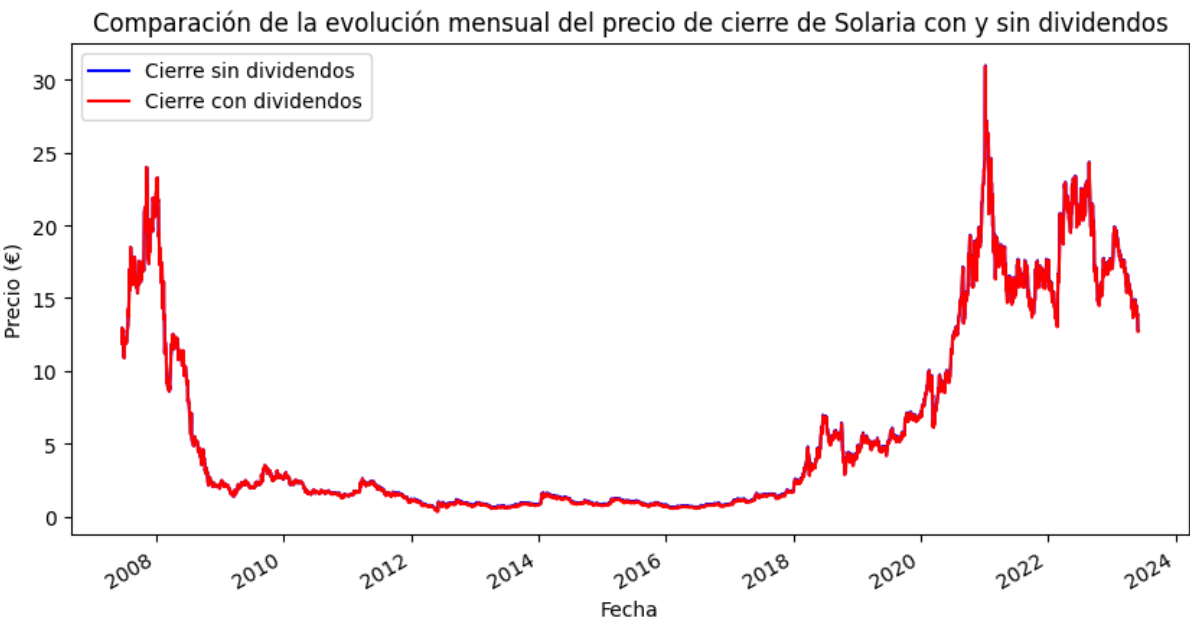


Figura 4.23: Comparación de la evolución mensual del precio de cierre de Solaria con y sin dividendos

4.2.4.5. Dividendos de Naturgy

A lo largo de su trayectoria, Naturgy ha otorgado diversos dividendos, como se evidencia en las siguiente tablas 4.14

Año	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Precio	0.238	0.281	0.340	0.511	0.604	0.715	0.834	0.970	0.000	0.790	0.805
Año	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Precio	0.830	0.894	0.897	0.908	1.330	1.330	1.950	1.92	0.913	1.730	1.00

Cuadro 4.14: Dividendos Naturgy agrupados por año

La evolución del precio del dividendo ha sido como se ve en la figura 4.24.

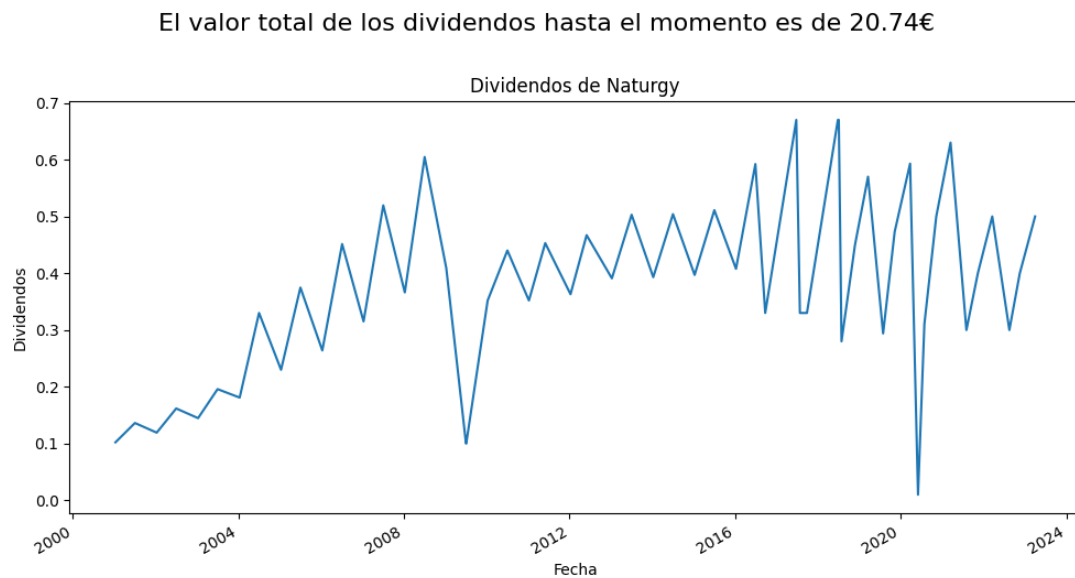


Figura 4.24: Evolución de los dividendos de Naturgy

Si los valores de los dividendos se restan aritméticamente del precio de la acción, podríamos pensar que su efecto se puede revertir a la hora de manejar la serie temporal, teniendo en cuenta restarle el valor total de los dividendos para obtener la versión real. En la figura 4.25 se muestra el precio de la acción con y sin dividendos.

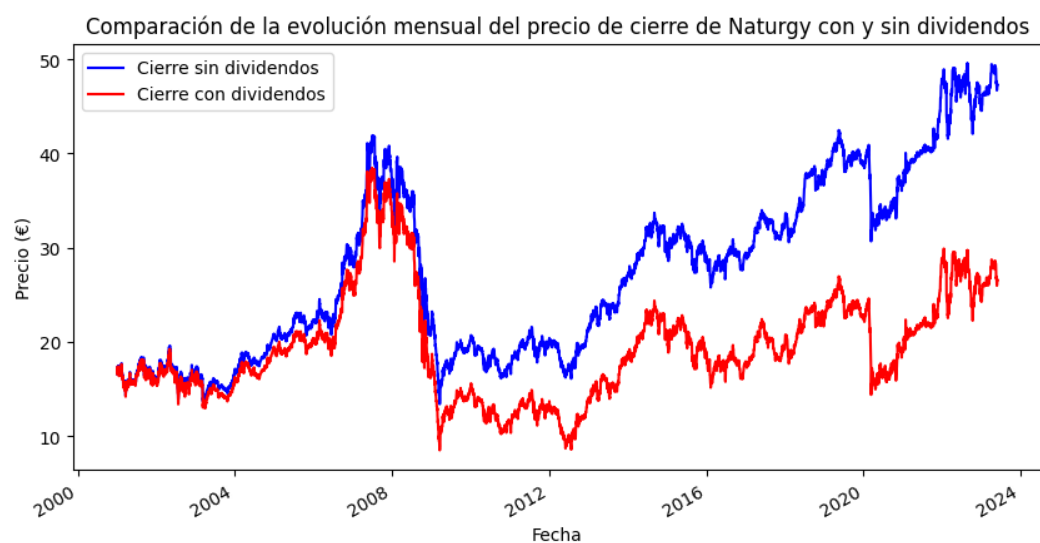


Figura 4.25: Comparación de la evolución mensual del precio de cierre de Iberdrola con y sin dividendos

### 4.2.5. Relación Entre Empresas

La correlación es una estadística que cuantifica el grado de relación entre dos variables, con un valor que oscila entre -1,0 y +1,0. Es importante destacar que la correlación mide la asociación, pero no establece causalidad ni indica si una variable causa la otra, o si la asociación es influenciada por un tercer factor.

Ahora, si deseamos analizar los rendimientos de todas las acciones en el sector eléctrico, podemos proceder creando un marco de datos que contenga las columnas *Close* para cada una de las acciones. Esto nos permitirá evaluar la relación de precios entre las acciones del sector eléctrico.

La tabla 4.15 y la Figura 4.26 presentan la matriz de correlación, proporcionando una representación visual de cómo se correlacionan las empresas entre sí en términos de rendimientos o precios.

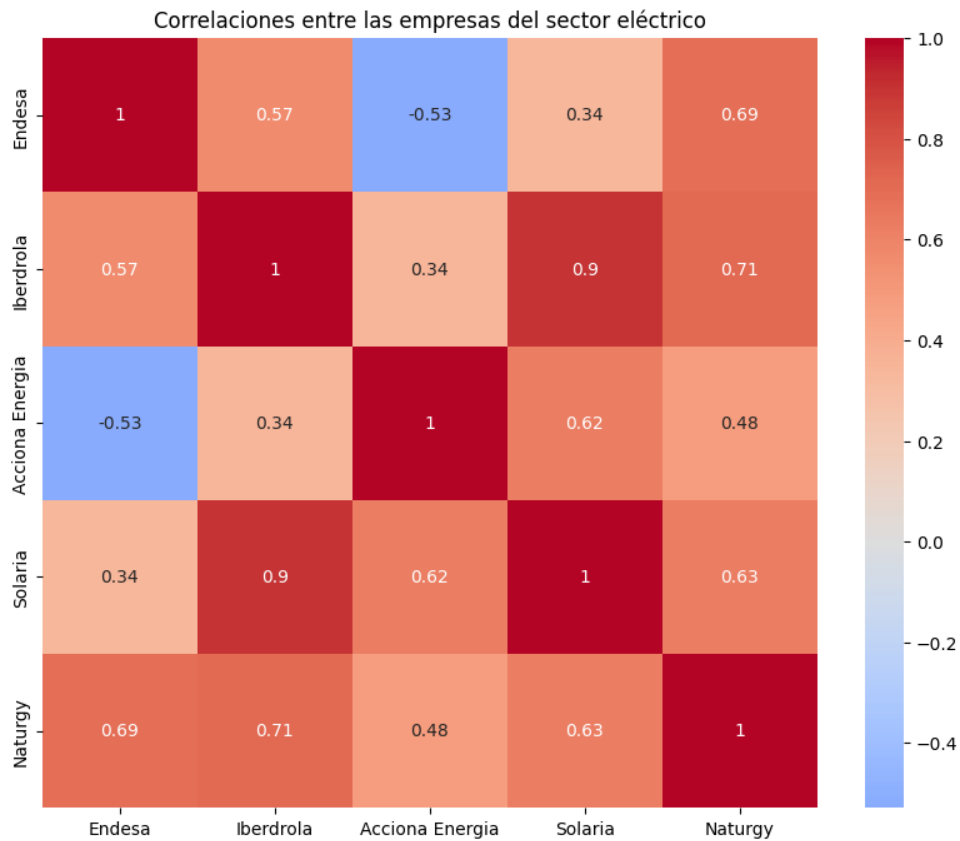


Figura 4.26: Correlaciones entre las empresas del sector eléctrico

	Endesa	Iberdrola	Acciona Energia	Solaria	Naturgy
Endesa	1.000000	0.565341	-0.529145	0.341723	0.693056
Iberdrola	0.565341	1.000000	0.344339	0.900923	0.709092
Acciona Energia	-0.529145	0.344339	1.000000	0.624007	0.476257
Solaria	0.341723	0.900923	0.624007	1.000000	0.628477
Naturgy	0.693056	0.709092	0.476257	0.628477	1.000000

Cuadro 4.15: Matriz de correlación entre empresas del sector eléctrico.

Las correlaciones varían de -0,4 a 1,0, siendo 1,0 una correlación positiva perfecta y -0,4 una correlación negativa débil. La correlación positiva más fuerte se da entre Acciona Energía y Solaria, con un valor de 0,9. La correlación negativa más fuerte se da entre Endesa e Iberdrola, con un valor de -0,53.

### 4.3. Holt-Winters Vs XGBoost

Después de llevar a cabo una exploración de los datos, se procederá a desarrollar y comparar dos modelos con el objetivo de anticipar las fluctuaciones en el precio de las acciones.

#### 4.3.1. Holt-Winters

El método de suavizado exponencial de Holt-Winters es una técnica de pronóstico utilizada en análisis de series temporales para prever patrones y tendencias futuras. Para aplicar este algoritmo se ha seguido un procedimiento estructurado que se mostrara a continuación.

Inicialmente, se ha sumado y dividido al precio de cierre de las acciones.

```

1 # Resampleamos los dividendos a una frecuencia diaria y llenamos los días
  sin dividendos con 0
2 dividendos_resampled = dividendos_Endesa.resample('D').sum().fillna(0)
3
4 # Acumulamos los dividendos a lo largo del tiempo
5 dividendos_acumulados = dividendos_resampled.cumsum()
6
7 # Eliminamos la información de zona horaria del índice
8 dividendos_acumulados.index = dividendos_acumulados.index.tz_localize(None)
9
10 # Definimos la fecha final para la reindexación
11 end_date = pd.Timestamp('2023-06-30').tz_localize(None)
12
13 # Creamos un nuevo índice que incluye todos los días hasta la fecha final
14 new_index = pd.date_range(dividendos_acumulados.index[0], end_date)
15

```

```

16 # Reindexamos los dividendos acumulados para incluir todos los dias hasta la
    fecha final
17 dividendos_acumulados =
18     dividendos_acumulados.reindex(new_index, method='ffill')
19
20 # Eliminamos la informacion de zona horaria del indice
21 dividendos_acumulados.index = dividendos_acumulados.index.tz_localize(None)
22
23 # Seleccionamos la columna 'Close' del DataFrame cotizacion_Endesa
24 cotizacion_temp = cotizacion_Endesa['Close']
25
26 # Anadimos los dividendos acumulados a la serie de precios de cierre
27 cotizacion_temp +=
28     dividendos_acumulados.reindex(cotizacion_temp.index, fill_value=0)

```

Listing 4.5: Suma de dividendo al cierre de las acciones

Posteriormente, se procedió a la creación de ventanas para la validación cruzada, asegurando una evaluación robusta del rendimiento del algoritmo.

```

1 # Define la funcion custom_time_series_split que toma un DataFrame y el
    numero de meses para entrenamiento y prueba
2 def custom_time_series_split(df, train_months, test_months):
3     # Inicializa una lista vacia para almacenar las ventanas de tiempo
4     ventanas = []
5
6     # Calcula el numero total de meses como la suma de los meses de
    entrenamiento y prueba
7     total_months = train_months + test_months
8
9     # Calcula el numero de divisiones como la longitud del DataFrame menos
    los meses de entrenamiento, dividido por los meses de prueba
10    n_splits = int((len(df) - train_months) / test_months)
11
12    # Para cada division
13    for split in range(n_splits):
14        # Calcula el inicio y el final del conjunto de entrenamiento
15        start_train = split * test_months
16        end_train = start_train + train_months
17
18        # El conjunto de prueba comienza donde termina el conjunto de
    entrenamiento
19        start_test = end_train
20        end_test = start_test + test_months
21
22        # Si el final del conjunto de prueba es mayor que la longitud del
    DataFrame, lo ajusta a la longitud del DataFrame
23        if end_test > len(df):
24            end_test = len(df)
25
26        # Anade las fechas de inicio y fin del conjunto de entrenamiento y
    prueba a la lista de ventanas
27        ventanas.append((
28            df.index[start_train],

```

```

29         df.index[end_train - 1],
30         df.index[start_test],
31         df.index[end_test - 1] ))
32
33     # Devuelve la lista de ventanas
34     return ventanas
35
36 # Define el numero de meses para el conjunto de entrenamiento y prueba
37 train_months = 36
38 test_months = 12
39
40 # Usa la funcion custom_time_series_split para obtener las ventanas de
    tiempo
41 ventanas = custom_time_series_split(cotizacion_mensual, train_months,
    test_months)
42
43 # Imprime cada ventana de tiempo
44 for i, ventana in enumerate(ventanas):
45     print(f"Ventana {i}: {[timest.strftime('%Y-%m-%d') for timest in ventana
    ]}")

```

Listing 4.6: Ventanas de Cross Validation

```

[(Timestamp('2001-01-31 00:00:00'),
  Timestamp('2003-12-31 00:00:00'),
  Timestamp('2004-01-31 00:00:00'),
  Timestamp('2004-12-31 00:00:00')),
 (Timestamp('2002-01-31 00:00:00'),
  Timestamp('2004-12-31 00:00:00'),
  Timestamp('2005-01-31 00:00:00'),
  Timestamp('2005-12-31 00:00:00')),
 ...
  Timestamp('2021-12-31 00:00:00')),
 (Timestamp('2019-01-31 00:00:00'),
  Timestamp('2021-12-31 00:00:00'),
  Timestamp('2022-01-31 00:00:00'),
  Timestamp('2022-12-31 00:00:00'))]

```

Luego de aplicar el algoritmo, se han generado predicciones para cada ventana temporal. En la Tabla 4.16 se presentan los resultados obtenidos para cada una de estas ventanas.

Ventana	MSE
0	2.141910
1	14.150930
2	92.800706
3	26.218066
4	51.785029
5	17.226519
6	6.313849
7	25.772848
8	2.807118
9	9.197091
10	50.762292
11	10.533275
12	2.091655
13	1.468497
14	4.432836
15	12.462961
16	6.744586
17	2.342448
18	2.254538

Cuadro 4.16: MSE para cada ventana de Holt-Winters

Finalmente, se ha representado de manera gráfica la predicción correspondiente a cada ventana, tal como se ilustra en la Figura 4.28 (a modo de ejemplo).

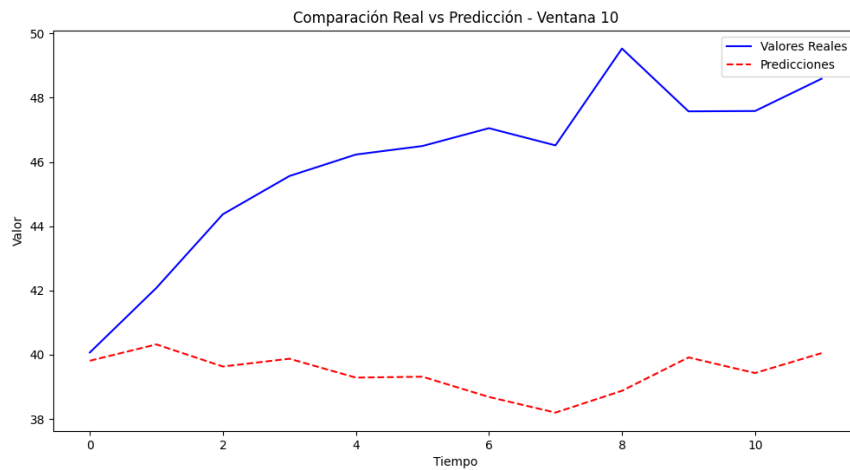


Figura 4.27: Ejemplo salida ventana de Holt-Winters



### 4.3.2. XGBoost con Cross Validation

XGBoost es un algoritmo de aprendizaje automático que se utiliza para problemas de regresión y clasificación. La validación cruzada (cross-validation) es una técnica esencial para evaluar el rendimiento de un modelo de manera más robusta. La validación cruzada divide el conjunto de datos en múltiples subconjuntos, entrena el modelo en algunos de ellos y evalúa el rendimiento en el subconjunto restante, repitiendo este proceso varias veces. Para aplicar este algoritmo se ha seguido un procedimiento estructurado que se mostrara a continuación.

Inicialmente, se llevó a cabo la preparación del dataframe, incorporando *lags* o pasos hacia atrás. La columna *Close* se ha mantenido como la variable independiente central en este proceso.

```

1 # Define el numero de lag que quieres crear
2 num_lags = 20
3
4 # Para cada numero en el rango de 1 a num_lags (inclusive)
5 for lag in range(1, num_lags + 1):
6     # Crea una nueva columna en el DataFrame llamada 'lag_{lag}'. Esta nueva
        columna es la columna Close desplazada lag periodos hacia atras .shift(
        lag) toma la columna 'Close' y "mueve" sus valores 'lag' filas hacia
        abajo
7     cotizacion_Endesa[f'lag_{lag}'] = cotizacion_Endesa['Close'].shift(lag)
8
9 # Elimina cualquier fila que tenga al menos un valor NaN (probablemente
    debido al desplazamiento anterior) y asigna el resultado a
    cotizacion_mensual
10 cotizacion_mensual = cotizacion_Endesa.dropna()
```

Listing 4.7: Creación de Lags

Posteriormente, se procedió a la creación de ventanas para la validación cruzada, asegurando una evaluación robusta del rendimiento del algoritmo.

```

1 # Define el numero de meses para el conjunto de entrenamiento y prueba
2 train_months = 36
3 test_months = 12
4
5 # Usa la funcion custom_time_series_split para obtener las ventanas de
    tiempo
6 ventanas = custom_time_series_split(cotizacion_mensual, train_months,
    test_months)
7
8 # Imprime cada ventana de tiempo
9 for i, ventana in enumerate(ventanas):
10     print(f"Ventana {i}: {[timest.strftime('%Y-%m-%d') for timest in ventana
        ]}")
```

Listing 4.8: Ventanas de Cross Validation

```
Ventana 0: ['2002-09-30', '2005-08-31', '2005-09-30', '2006-08-31']
Ventana 1: ['2003-09-30', '2006-08-31', '2006-09-30', '2007-08-31']
Ventana 2: ['2004-09-30', '2007-08-31', '2007-09-30', '2008-08-31']
...
```

Luego, se dedicó una fase a la búsqueda de hiperparámetros, afinando y seleccionando las configuraciones óptimas para garantizar un desempeño óptimo del modelo.

```
1 # Define un espacio de búsqueda para los hiperparametros del modelo
2 space = {
3     'max_depth': hp.choice('max_depth', np.arange(3, 10, dtype=int)),
4     'min_child_weight': hp.quniform('min_child_weight', 1, 6, 1),
5     'subsample': hp.uniform('subsample', 0.6, 1),
6     'n_estimators': hp.quniform('n_estimators', 100, 500, 100),
7     'learning_rate': hp.uniform('learning_rate', 0.05, 0.2)
8 }
9
10 # Genera 30 muestras del espacio de búsqueda
11 samples = [sample(space) for _ in range(30)]
12
13 # Para cada muestra, convierte 'max_depth', 'min_child_weight' y '
14     n_estimators' a enteros
15 for sample_ in samples:
16     sample_['max_depth'] = int(sample_['max_depth'])
17     sample_['min_child_weight'] = int(sample_['min_child_weight'])
18     sample_['n_estimators'] = int(sample_['n_estimators'])
19
20 # Imprime cada muestra
21 for i, params in enumerate(samples):
22     print(f"Sample {i+1}: {params}")
```

Listing 4.9: Definición espacio para los hiperparámetros

Sample 1:

```
{'learning_rate': 0.16641910935709014,
  'max_depth': 4,
  'min_child_weight': 5,
  'n_estimators': 100,
  'subsample': 0.8480975562886519}
```

...

Sample 30:

```
{'learning_rate': 0.15154446024792112,
```

```
'max_depth': 4,
'min_child_weight': 3,
'n_estimators': 100,
'subsample': 0.7586972350490817}
```

La Tabla 4.18 presenta un ejemplo de los resultados generados por el modelo en la búsqueda de hiperparámetros

Inicio Train	Fin Train	Inicio Test	Fin Test	Lags	MSE
2002-09-30	2005-08-31	2005-09-30	2006-08-31	5	56.70
2002-09-30	2005-08-31	2005-09-30	2006-08-31	10	47.13
2002-09-30	2005-08-31	2005-09-30	2006-08-31	20	57.31
2002-09-30	2005-08-31	2005-09-30	2006-08-31	5	47.28
2002-09-30	2005-08-31	2005-09-30	2006-08-31	10	45.77

Cuadro 4.17: Resultados del modelo

Al final, se visualizaron de manera gráfica los hiperparámetros óptimos para cada intervalo de fechas. Por ejemplo, para el periodo comprendido entre el 30 de septiembre de 2009 al 31 de agosto de 2010, los resultados fueron a un nivel gráfico como la figura 4.28.

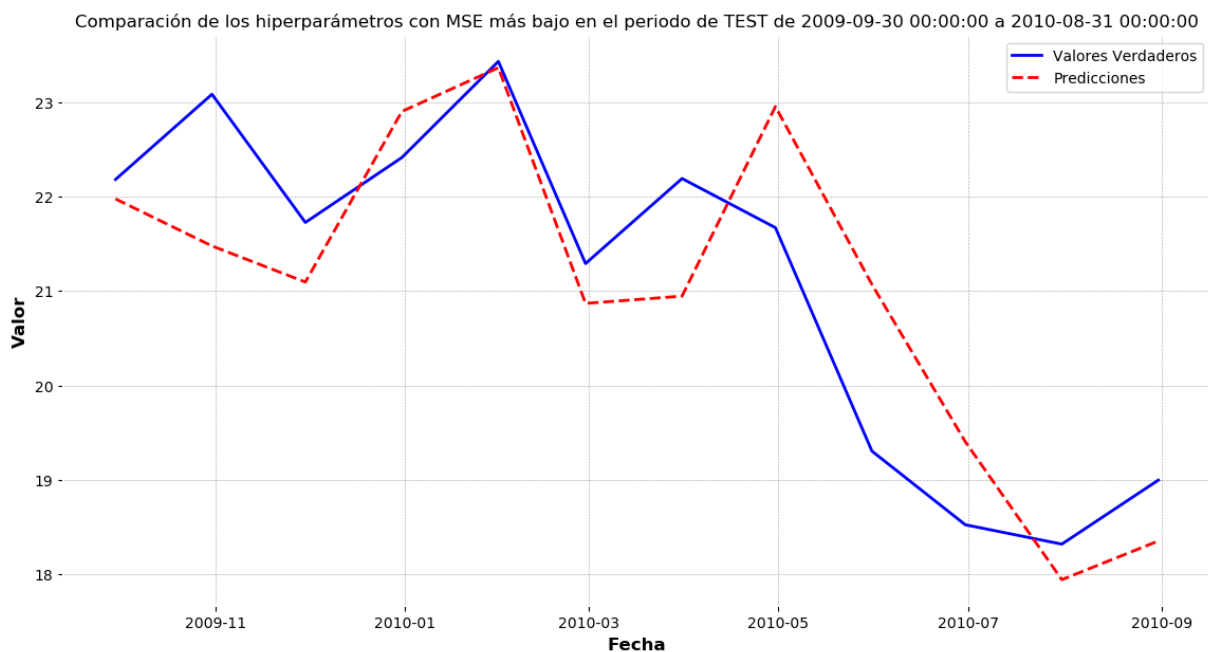


Figura 4.28: Comparación de los hiperparámetros con MSE más bajo en el periodo de TEST de 2009-09-30 a 2010-08-31

Y a nivel de hiperparámetros:

Fecha de inicio de TEST: 2009-09-30 00:00:00

Fecha de finalización de TEST: 2010-08-31 00:00:00

Lags utilizados: 10

MSE más bajo: 0.9229394206010996

Hiperparámetros:

```
{  
    'learning_rate': 0.08566207896193731,  
    'max_depth': 5,  
    'min_child_weight': 1,  
    'n_estimators': 200,  
    'subsample': 0.6210736406890162  
}
```

Examinamos los hiperparámetros que demostraron un rendimiento destacado en modelos previos, particularmente aquellos con un MSE inferior a 3. Los resultados obtenidos se presentan de manera gráfica en las Figuras 4.30 y 4.29.

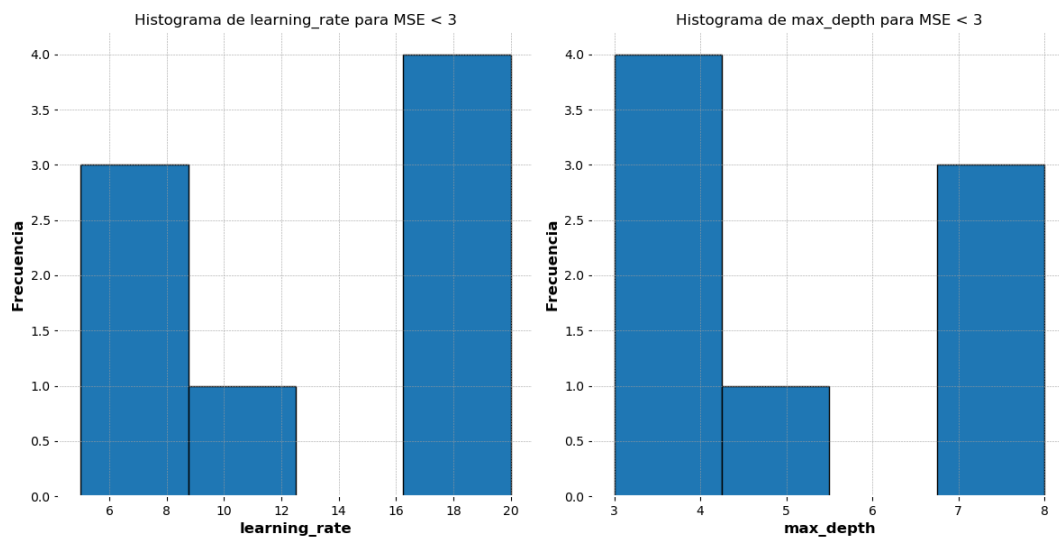


Figura 4.29: Mejores hiperparámetros I

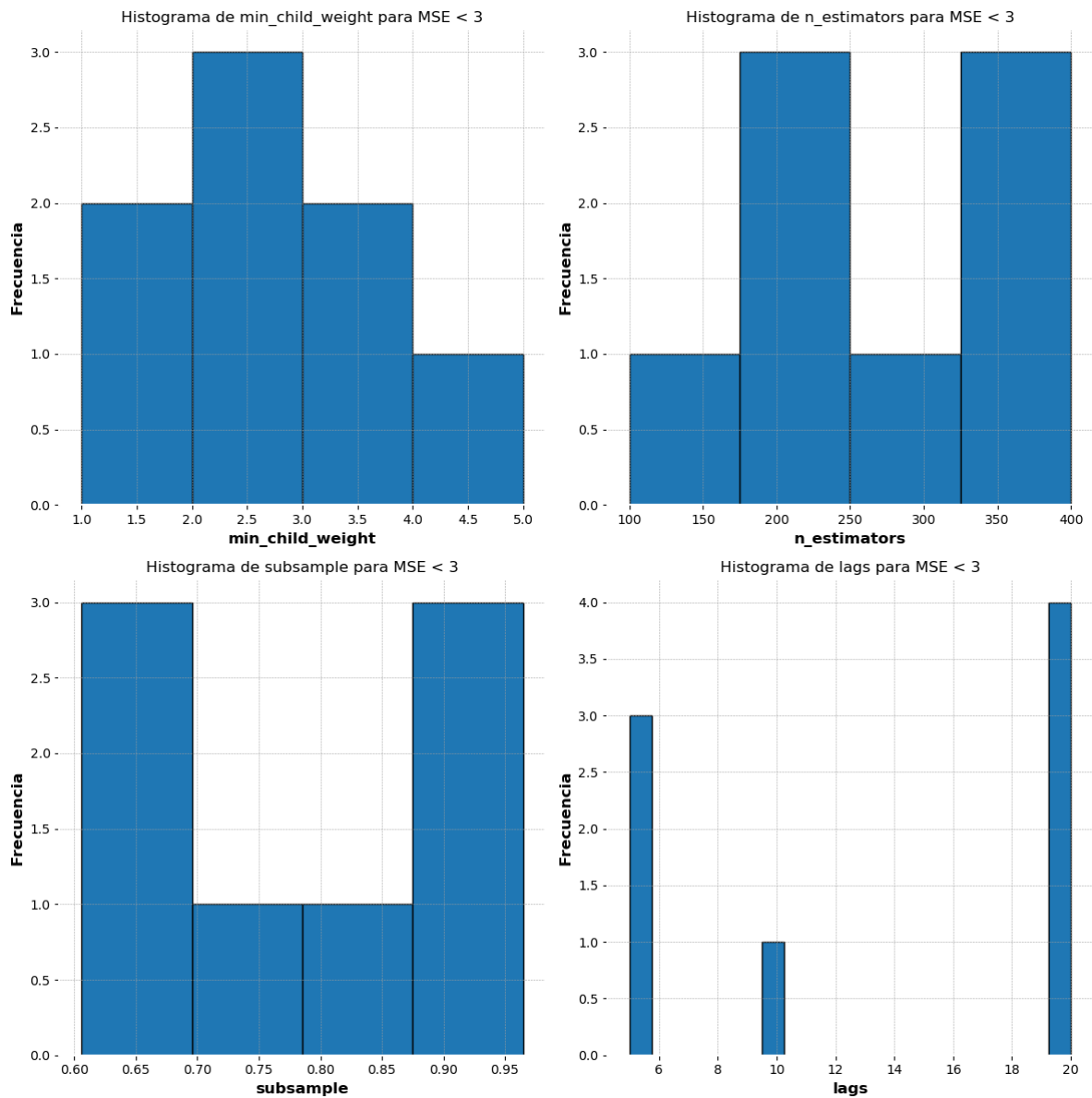


Figura 4.30: Mejores hiperparámetros II

### 4.3.3. Comparación de los modelos

Después de implementar y validar ambos modelos, se han obtenido los siguientes resultados:

Modelo	Mediana MSE	Media MSE
XGBoost	3.4718	17.1820
Holt-Winters	9.1971	17.9741

Cuadro 4.18: Comparación MSE entre XGBoost y Holt-Winters

Si analizamos los resultados, nos damos cuenta de que:

■ **XGBoost:**

- La mediana MSE es relativamente baja (3.4718), indicando consistencia en el rendimiento del modelo.
- La media MSE (17.1820) es más alta que la mediana, sugiriendo la posible influencia de outliers.

■ **Holt-Winters:**

- La mediana MSE es más alta que la de XGBoost (9.1971), indicando menos consistencia en el rendimiento.
- La media MSE (17.9741) es similar a la de XGBoost, posiblemente debido a la sensibilidad a ciertos patrones de datos.

**Comparación:** XGBoost muestra una mediana y media MSE inferiores en comparación con Holt-Winters, sugiriendo un rendimiento global más consistente y potencialmente mejor en términos de error cuadrático medio.

## 4.4. Análisis de noticias

A pesar de haber empleado diversas estrategias de aprendizaje profundo en la sección anterior con el objetivo de anticipar el valor de las acciones, hemos llegado a la conclusión de que resulta imperativo contar con información adicional más allá del mero análisis histórico de precios. Esta información adicional puede ser extraída de diversas fuentes, siendo las noticias una de las más relevantes.

Durante el minucioso análisis de los datos, hemos constatado que determinadas variaciones en el precio de las acciones se ven directamente influenciadas por eventos noticiosos relevantes. En esta sección, nos concentraremos en la recopilación y evaluación de noticias para determinar con precisión cuáles de ellas tienen un impacto significativo en las fluctuaciones del precio de las acciones.

#### 4.4.1. Justificación del uso de Web Scraping para obtener noticias

La justificación para emplear web scraping en la obtención de noticias en el ámbito financiero se fundamenta en la necesidad de acceder a información relevante y oportuna que incide directamente en las dinámicas del mercado. Tras una exhaustiva exploración y evaluación de diversas APIs, entre las que se incluyen servicios noticiosos como GNews, se ha constatado que estas no ofrecen de manera eficiente la amplitud y la velocidad requeridas para recopilar noticias financieras de manera integral. Por ejemplo, aunque las APIs proporcionan datos estructurados, la limitación en términos de número de solicitudes por día o la tardanza en la actualización de noticias cruciales en tiempo real puede comprometer la efectividad del análisis.

Además, el cambio en las políticas de acceso gratuito y la implementación de tarifas en el uso de APIs de plataformas como Twitter ha generado un problema adicional. La transición hacia modelos de pago ha impactado negativamente en la accesibilidad de los datos en tiempo real, restringiendo así el flujo constante de información. En este escenario, el web scraping se presenta como una solución práctica y eficaz, permitiendo superar estas limitaciones al extraer datos de fuentes diversas de manera más flexible y sin las restricciones impuestas por el pago de suscripciones o limitaciones de solicitudes diarias.

#### 4.4.2. Web Scraping

Para obtener las noticias, se ha optado por la sección de noticias del sector eléctrico del periódico El País [3], y a continuación se presenta el código donde se ha llevado a cabo todo el proceso:

```
1 # Lista para almacenar las noticias
2 news_data = []
3
4
5 # Funcion para extraer noticias de una pagina
6 def extract_news(url):
7     # Haz una solicitud a la pagina
8     response = requests.get(url)
9
10    # Parsea el contenido de la pagina con BeautifulSoup
11    soup = BeautifulSoup(response.content, 'html.parser')
12
13    # Encuentra todos los articulos de noticias
```

```

14 articles = soup.find_all('article', class_='c c-o c-d c--m',
15 'c c-d c--m', 'c c-o c-d c--c c--m-n', 'c c-d c--m-n'])
16
17 # Extrae la informacion relevante de cada articulo
18 for article in articles:
19     title_element = article.find('h2', class_='c_t c_t-i')
20     if title_element is None:
21         title_element = article.find('h2', class_='c_t')
22
23     title = title_element.text
24     link = title_element.find('a')['href']
25     date = article.find('time')['datetime']
26     summary = article.find('p', class_='c_d').text
27
28     # Anade la informacion a la lista de noticias
29     news_data.append([title, link, date, summary])
30
31
32 # Itera sobre las paginas
33 for i in range(280, 1, -1):
34     # Define la URL de la pagina
35     url = f'https://cincodias.elpais.com/noticias/sector-electrico/{i}/'
36
37     # Extrae las noticias de la pagina
38     extract_news(url)
39
40 # Procesa la ultima pagina
41 extract_news('https://cincodias.elpais.com/noticias/sector-electrico/')
42
43 # Crea un DataFrame con las noticias
44 df = pd.DataFrame(news_data, columns=['Title', 'Link', 'Date', 'Summary'])
45
46 # Guarda el DataFrame en un archivo CSV
47 df.to_csv('./data/noticias.csv', index=False)

```

Listing 4.10: Búsqueda de Noticias

El resultado, tal como se muestra en la línea 46, es un archivo donde se guardan todas las noticias recopiladas para su análisis posterior. El número total de noticias obtenidas es de 8305.

#### 4.4.3. GPT para el análisis de noticias

Después de recopilar las noticias, se va a usar la API de GPT para clasificar las noticias previas. La salida que obtendremos tiene el siguiente formato:

```

1 {
2     "empresa": "",
3     "tipo": "",
4     "sentido": ""
5 }

```

Listing 4.11: Salida de GPT



Donde:

- **Empresa:** Puede ser una de las siguientes opciones: 'Sector Energético', 'Iberdrola', 'Endesa', 'Naturgy', 'Abengoa'.
- **Tipo:** Indica la categoría de la noticia y puede ser una de las siguientes: Negocios y Finanzas Corporativas, Regulación y Asuntos Legales, Desarrollo de Infraestructura y Proyectos, Estrategia y Planificación Empresarial, y Dividendos.
- **Sentido:** Es un valor comprendido entre -1 y 1, reflejando la dirección positiva o negativa de la noticia.

La fecha está fusionada con el archivo original, donde se generó la información.

En nuestro contexto, la API de GPT ha sido empleada conforme al siguiente código:

```

1 # Iterar sobre cada fila del DataFrame 'noticias_df'
2 for index, fila in noticias_df.iterrows():
3     try:
4         # Concatenar el titulo y el resumen de la noticia
5         texto_concatenado = fila['Title'] + ' ' + fila['Summary']
6
7         # Definir los mensajes para la API de GPT-3.5-turbo
8         exper = f"Eres un experto financiero analizando noticias."
9         expli = f"Con el siguiente texto '{texto_concatenado}'. Determina si
            afecta a alguna de estas empresas: 'Solaria', 'Endesa', 'Naturgy', '
            Iberdrola', 'Acciona', clasif calo seg n el tipo de noticia entre las
            opciones 'Negocios y Finanzas Corporativas', 'Regulaci n y Asuntos
            Legales', 'Estrategia y Planificaci n Empresarial', 'Dividendos', '
            Desarrollo de Infraestructura y Proyectos' y dame un ratio entre -1 y 1
            que indique c mo de positiva es la noticia. Devuelve lo que te he pedido
            en un JSON con la estructura 'empresa', 'tipo', 'sentido'. No incluyas
            informaci n fuera de este JSON. Cuando haya varias empresas de las que
            te he dicho, indica 'Sector Energ tico'. Revisa que el JSON que me
            proporcionas cumpla con las condiciones que te he puesto."
10
11         # Crear una solicitud a la API de GPT-3.5-turbo
12         completion_es = openai.ChatCompletion.create(
13             model="gpt-3.5-turbo",
14             messages=[
15                 {"role": "system", "content": exper},
16                 {"role": "user", "content": expli},
17             ],
18             temperature=0.2
19         )
20
21         # Obtener la respuesta de la API
22         response_es = completion_es.choices[0].message["content"]

```

```

23
24     # Escribir la respuesta en un archivo de salida
25     with open(f"./out/{str(index)}-{datetime.now()
26               .strftime('%Y-%m-%d_%H-%M-%S')}.txt", 'w') as f:
27         f.write(str(response_es))
28
29     # Esperar 2 segundos antes de la proxima iteracion
30     time.sleep(2)
31
32 except:
33     # Manejar errores e imprimir informacion para depuracion
34     print('*'*20 + 'ERROR' + index + '*'*20)

```

Listing 4.12: Uso de la API de GPT

Al concluir, se obtuvieron un total de 8305 ficheros distintos, cada uno estructurado según el formato previamente descrito. A modo ilustrativo, un ejemplo de salida sería el siguiente:

```

{
  "empresa": "Endesa",
  "tipo": "Negocios y Finanzas Corporativas",
  "sentido": "0.8"
}

```

Y la noticia que ha tratado es la siguiente [2]:

**Endesa gana un 5,3 % más por la venta de Viesgo pero reduce su plan de inversiones**

*La compañía ha decidido remodelar su plan estratégico ante el nuevo entorno macroeconómico, para lo que reducirá en 3.300 millones sus inversiones, que se sitúan en 9.700 millones de euros, dando prioridad a la consolidación y rentabilidad de su negocio principal.*

Como se evidencia, GPT ha logrado una clasificación acertada, y dado que la noticia en su conjunto es positiva, la orientación del sentido se alinea adecuadamente con las expectativas.

## 4.5. Retroalimentación de XGBoost

En esta sección, se ha reproducido el procedimiento previamente explicado, pero ahora se ha incorporado el análisis de noticias.

Antes de integrar las noticias al modelo, se observa la presencia de diversas instancias de noticias registradas en distintas fechas. Para abordar esta variabilidad, se ha iniciado aplicando un codificador OneHot a las columnas 'empresa' y 'tipo', tal y como se muestra en la tabla 4.19.

Date	Sentido	Endesa	Iberdrola	Naturgy	S_Energético
2002-10-26	-0.500	0.0	1.0	0.0	0.0
2002-10-29	1.500	2.0	0.0	0.0	0.0
2002-10-30	-0.878	1.0	0.0	0.0	1.0
2002-11-07	2.400	1.0	0.0	0.0	2.0
2002-11-08	1.800	0.0	0.0	1.0	1.0
Date	Tipo_1	Tipo_2	Tipo_3	Tipo_4	Tipo_5
2002-10-26	0.0	0.0	0.0	0.0	1.0
2002-10-29	0.0	0.0	1.0	1.0	0.0
2002-10-30	0.0	0.0	1.0	1.0	0.0
2002-11-07	0.0	0.0	0.0	3.0	0.0
2002-11-08	0.0	0.0	0.0	1.0	1.0

Cuadro 4.19: Representación tabular del análisis de noticias con codificador OneHot

Siendo los tipos de noticias:

- **Tipo\_1:** Desarrollo de Infraestructura y Proyectos
- **Tipo\_2:** Dividendos
- **Tipo\_3:** Estrategia y Planificación Empresarial
- **Tipo\_4:** Negocios y Finanzas Corporativas
- **Tipo\_5:** Regulación y Asuntos Legales

Después de completar la preparación de las noticias, se ha seguido el procedimiento mencionado anteriormente. Los resultados posteriores a la incorporación de las noticias se muestran en la tabla 4.20.

Los resultados indican una mejora significativa al incorporar noticias en el modelo. La mediana del error MSE disminuye de 3.4718 a 1.5926, lo que sugiere una reducción en los errores medianos. Asimismo, la media del error MSE disminuye de 17.1820 a 3.85137, indicando

	Mediana MSE	Media MSE
<b>Sin Noticias</b>	3.4718	17.1820
<b>Con Noticias</b>	1.5926	3.85137

Cuadro 4.20: Comparación de errores MSE sin y con Noticias

una mejora promedio en la precisión del modelo. En resumen, la introducción de noticias ha contribuido positivamente a la capacidad del modelo para generar respuestas más precisas.

## 4.6. Pruebas del nuevo modelo XGBoost

En esta sección, se procederá a comentar y analizar las conclusiones obtenidas después de emplear el modelo XGBoost, el cual fue alimentado con las noticias. Con este fin, se llevará a cabo el reentrenamiento del modelo en varias ocasiones para extraer las *Feature Importances*.

### 4.6.1. Feature Importances con todas las variables

Vamos a reentrenar el modelo más reciente utilizando la totalidad de los datos de entrenamiento, a excepción del último período (de mayo de 2022 a abril de 2023), el cual reservaremos para la validación. Posteriormente, extraeremos las importancias de las características (*Feature Importances*) con el objetivo de comprender cuáles tipos de noticias, lags u otras variables tienen una mayor relevancia en el proceso de predicción. Utilizaremos los hiperparámetros óptimos previamente obtenidos. Los resultados numéricos de este entrenamiento son los siguientes:

Train desde 2002-09-30 hasta 2021-08-31

Test desde 2021-09-30 hasta 2022-08-31

Nº de lags: 5.

Hiperparámetros: {

    'learning\_rate': 0.15,

    'max\_depth': 5,

    'min\_child\_weight': 2,

    'n\_estimators': 350,

```
'subsample': 0.72
}
```

MSE de: 0.8672621440442416

El *Feature Importances* es el que se muestra en la imagen [4.31](#).

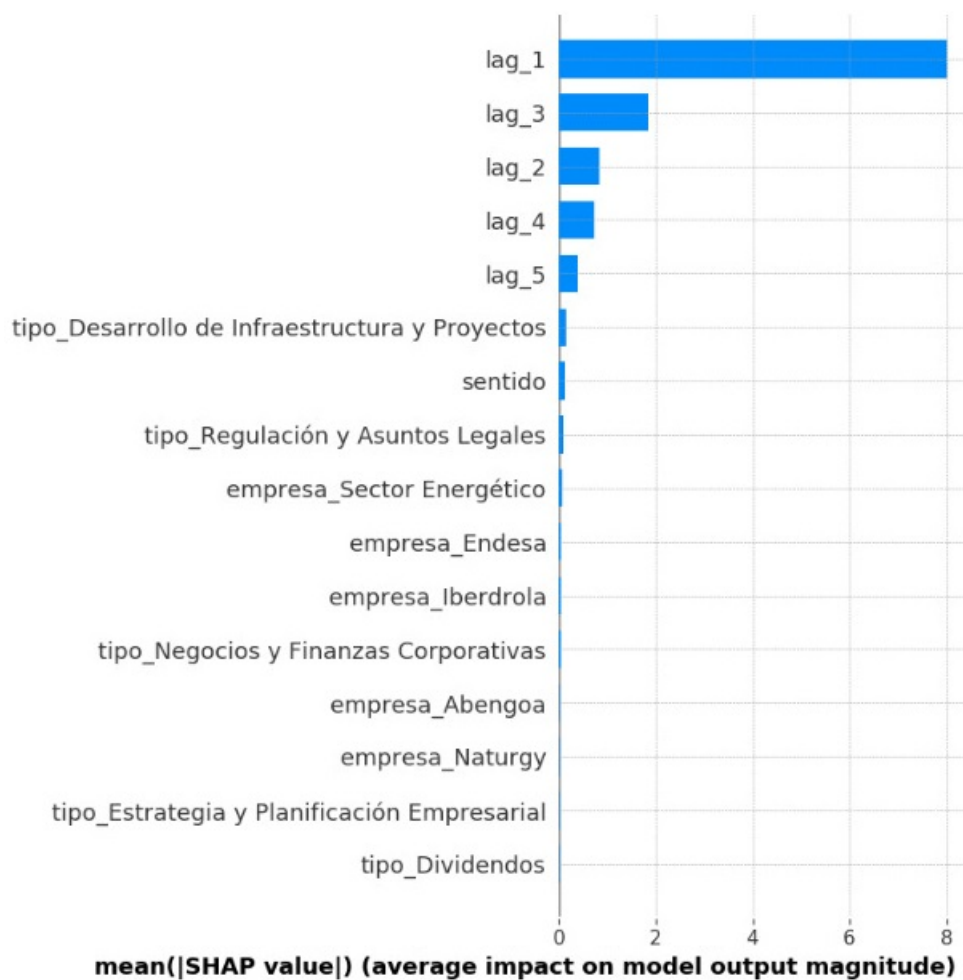


Figura 4.31: Feature Importances con todas las variables

A simple vista, es evidente que los Lags juegan un papel fundamental, ya que el precio de una acción no experimenta movimientos aleatorios, sino que se basa en gran medida en el precio de cierre previo. Al eliminar los lags y repetir el proceso, se observa el resultado reflejado en la Figura [4.32](#).

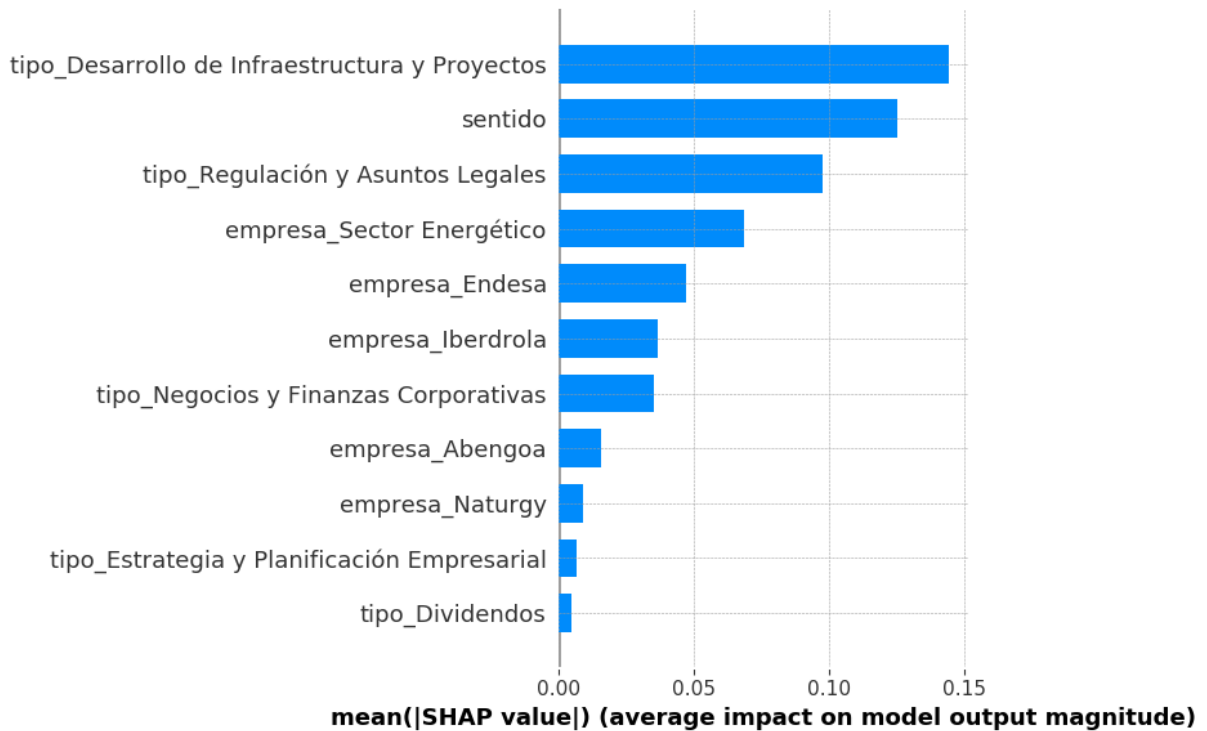


Figura 4.32: Feature Importances con todas las variables

Como se puede apreciar, las noticias relacionadas con el desarrollo de infraestructura y proyectos se consideran las más relevantes, seguidas por la evaluación del sentido, aunque esto no parece del todo preciso, ya que el sentido es el factor determinante para clasificar una noticia como positiva o negativa. No obstante, en futuras pruebas, se integrará el sentido a las noticias.

En contraste, las noticias menos influyentes son aquellas vinculadas a los dividendos, así como a la estrategia y planificación empresarial.

#### 4.6.2. Feature Importances con el sentido incorporado a los tipos de noticias

Tras observar en la sección anterior que el sentido debería estar integrado directamente en las noticias en lugar de estar por separado, se procederá a reentrenar el modelo. Este enfoque busca mejorar la coherencia y relevancia del análisis, permitiendo una evaluación más precisa de cómo el sentido de las noticias afecta al comportamiento del precio del valor.

La incorporación del sentido a las noticias se llevó a cabo de la siguiente manera:

```
1 #Incorporamos el sentido a los valores de 'tipo de noticias'.
2 tipo_columns =
3     [col for col in cotizacion_gpt.columns if col.startswith('tipo')]
4
5 cotizacion_gpt[tipo_columns] =
6     cotizacion_gpt[tipo_columns].apply(lambda x: x * cotizacion_gpt['sentido'])
7
8 cotizacion_gpt = cotizacion_gpt.drop(columns=['sentido'])
```

Listing 4.13: Incorporación del sentido a los tipos de noticias

Los resultados numéricos de este entrenamiento son los siguientes:

Train desde 2002-09-30 hasta 2021-08-31

Test desde 2021-09-30 hasta 2022-08-31

Nº de lags: 5.

Hiperparámetros: {  
 'learning\_rate': 0.15,  
 'max\_depth': 5,  
 'min\_child\_weight': 2,  
 'n\_estimators': 350,  
 'subsample': 0.72  
}

MSE de: 1.0026513617846342

El *Feature Importances* es el que se muestra en la imagen [4.33](#).

En este caso, se acentúa aún más la disparidad entre las noticias relacionadas con el desarrollo de infraestructura y proyectos en comparación con otros tipos de noticias. Es notable que ahora las noticias específicas de Endesa, que no estaban clasificadas en categorías previas, han adquirido una relevancia más significativa. Además, se ha observado un aumento en la importancia de las noticias relacionadas con otras empresas.

Respecto al Error Cuadrático Medio (MSE), en esta ocasión se registra un valor de 1.002, en comparación con el valor anterior de 0.867. Este aumento en el MSE sugiere que, en el

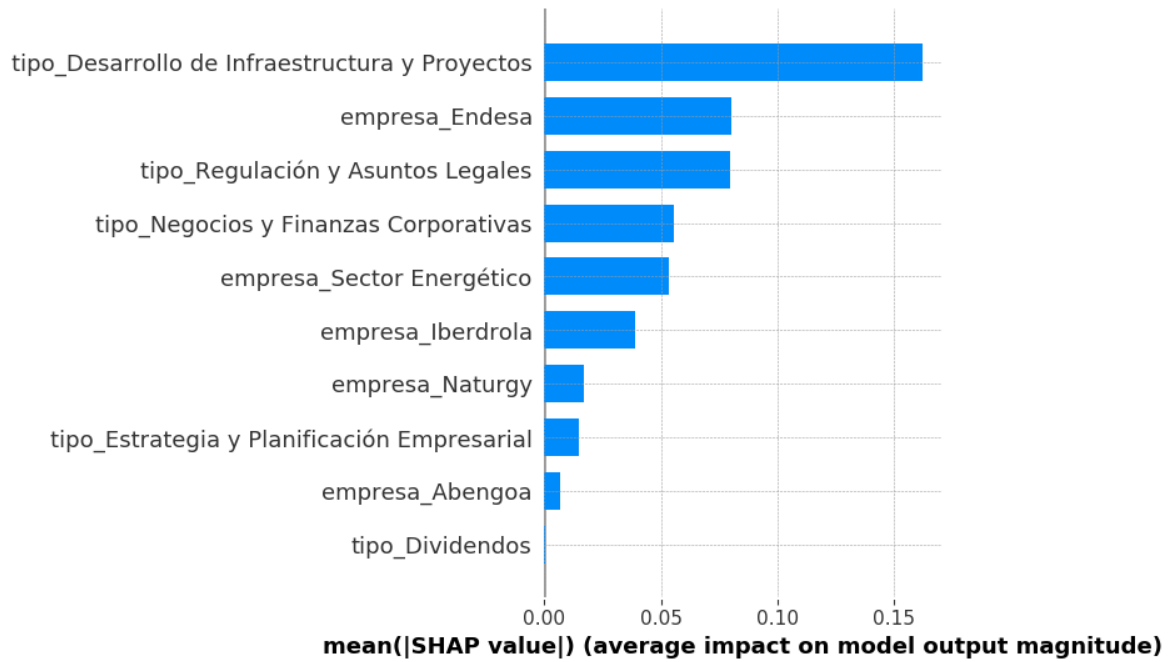


Figura 4.33: Feature Importances con el sentido incorporado en el Tipo de Noticia

nuevo modelo con el sentido incorporado a las noticias, la precisión del modelo podría haber disminuido ligeramente en relación con el modelo anterior.

#### 4.6.3. Feature Importances generalizando todas las empresas distintas a Endesa

Finalmente, se llevará a cabo un último proceso de reentrenamiento del modelo. En esta ocasión, todas las noticias referentes a empresas distintas de Endesa se agruparán en una única variable. Este enfoque tiene como objetivo simplificar la representación de noticias relacionadas con otras empresas, aportando una mayor claridad y eficiencia en el análisis. Con esta modificación, se busca optimizar la capacidad del modelo para identificar patrones y relaciones, facilitando así la interpretación de los resultados obtenidos.

Los resultados numéricos de este entrenamiento son los siguientes:

Train desde 2002-09-30 hasta 2021-08-31

Test desde 2021-09-30 hasta 2022-08-31

Nº de lags: 5.



```
Hiperparámetros: {  
    'learning_rate': 0.15,  
    'max_depth': 5,  
    'min_child_weight': 2,  
    'n_estimators': 350,  
    'subsample': 0.72  
}
```

MSE de: 1.3887587631245253

El *Feature Importances* es el que se muestra en la imagen 4.34.

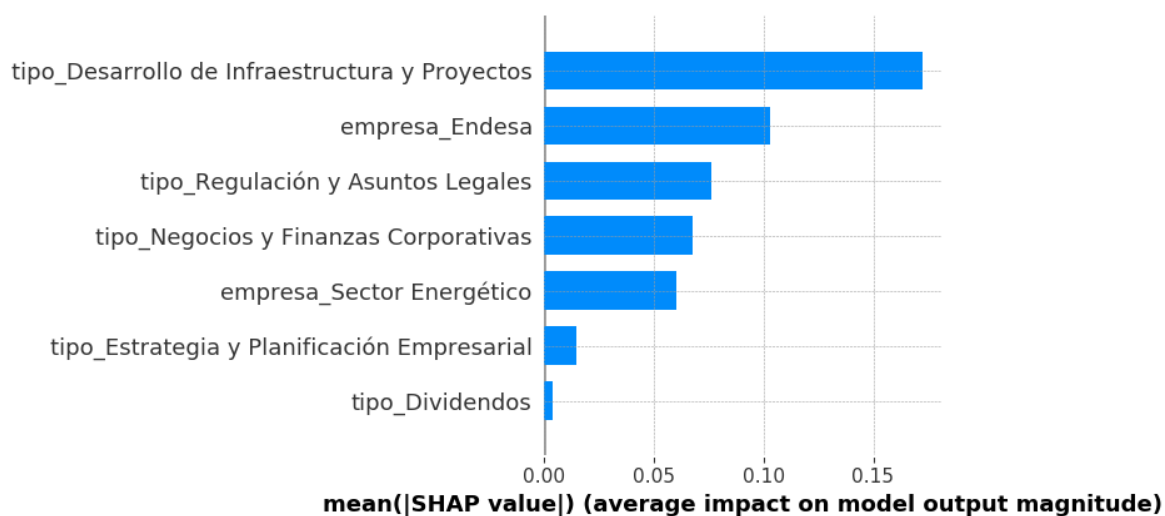


Figura 4.34: Feature Importances con las empresas agrupadas

En este último reentrenamiento, se aprecia que la importancia de las noticias apenas experimenta cambios significativos. Aunque el MSE ha aumentado a 1.388, es importante señalar que este pequeño incremento puede considerarse aceptable en comparación con las ventajas de consolidar las noticias relacionadas con otras empresas en una única variable. Este ajuste permite obtener una visión más clara y simplificada de la información, facilitando la interpretación de los resultados sin comprometer de manera significativa la calidad del modelo.

## 4.7. Extrapolación al resto de empresas

Después de realizar las pruebas y obtener conclusiones en la sección anterior, se llevarán a cabo nuevos ensayos. En esta ocasión, se cargarán los datos correspondientes al resto de empresas. Esto permitirá identificar qué noticias afectan en mayor o menor medida al precio de las acciones.

### 4.7.1. Iberdrola

Se han ingresado los datos históricos de las acciones de Iberdrola y se han incorporado los correspondientes lags. Además, se ha llevado a cabo la normalización de todas las empresas que no pertenecen a Iberdrola, siguiendo el mismo procedimiento utilizado previamente. Los resultados numéricos obtenidos son los siguientes:

Train desde 2001-09-09 hasta 2021-08-3

Test desde 2021-09-30 hasta 2022-08-31

Nº de lags: 5.

Hiperparámetros: {

    'learning\_rate': 0.15,

    'max\_depth': 5,

    'min\_child\_weight': 2,

    'n\_estimators': 350,

    'subsample': 0.72 }

MSE de: 0.3494

El *Feature Importances* es el que se muestra en la imagen [4.35](#).

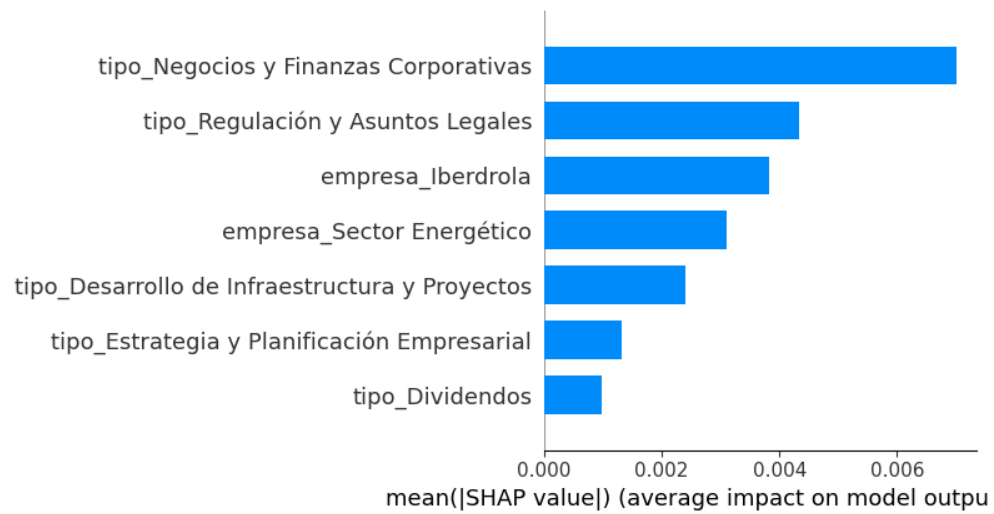


Figura 4.35: Feature Importances de Iberdrola

#### 4.7.2. Acciona Energías

Se ha replicado el procedimiento mencionado en la sección anterior pero ahora para las acciones de Acciona Energías. Los resultados numéricos obtenidos son los siguientes:

Train desde 2021-07-02 hasta 2022-08-31

Test desde 2022-09-31 hasta 2023-06-01

Nº de lags: 5.

Hiperparámetros: {

    'learning\_rate': 0.15,

    'max\_depth': 5,

    'min\_child\_weight': 2,

    'n\_estimators': 350,

    'subsample': 0.72

}

MSE de: 0.6034

El *Feature Importances* es el que se muestra en la imagen [4.36](#).

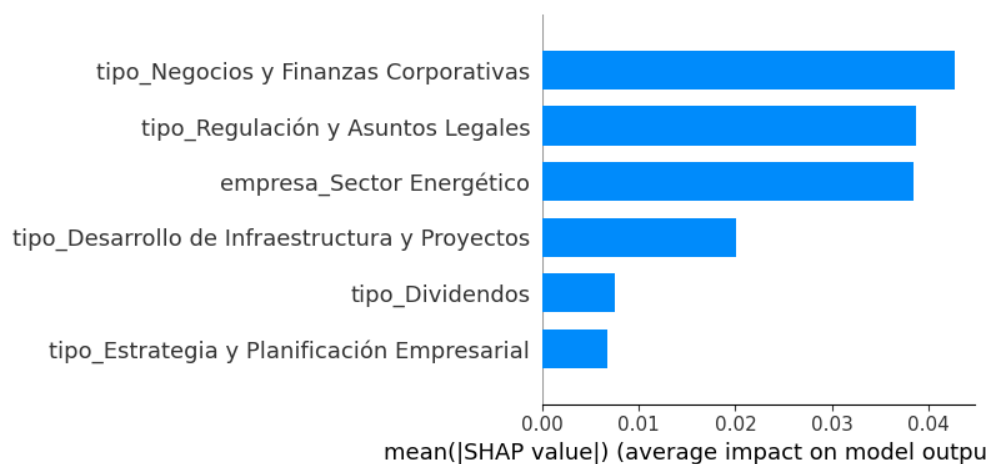


Figura 4.36: Feature Importances de Acciona Energías

### 4.7.3. Solaria

Se ha replicado el procedimiento mencionado en la sección anterior pero ahora para las acciones de Solaria. Los resultados numéricos obtenidos son los siguientes:

Train desde 2007-06-19 hasta 2021-08-31

Test desde 2021-09-30 hasta 2022-08-31

Nº de lags: 5.

Hiperparámetros: {

    'learning\_rate': 0.15,

    'max\_depth': 5,

    'min\_child\_weight': 2,

    'n\_estimators': 350,

    'subsample': 0.72

}

MSE de: 0.5576

El *Feature Importances* es el que se muestra en la imagen [4.37](#).

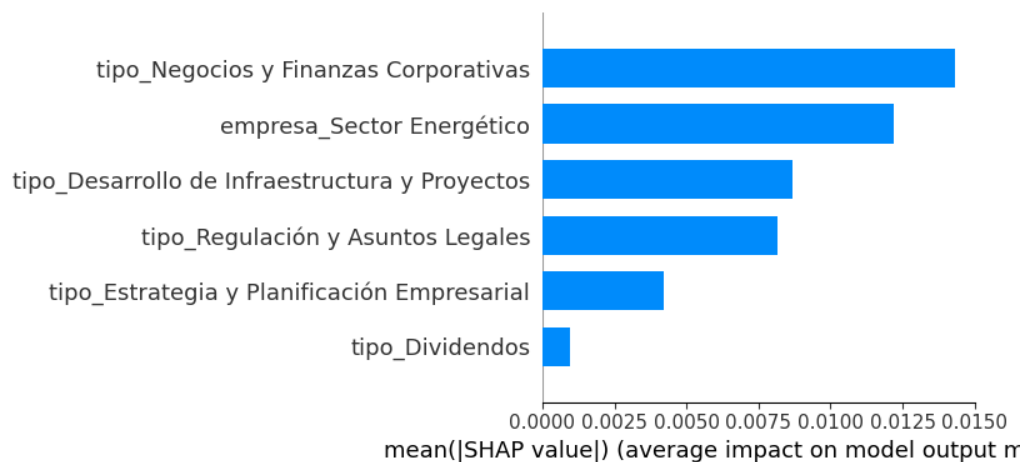


Figura 4.37: Feature Importances de Solaria

#### 4.7.4. Naturgy

Se ha replicado el procedimiento mencionado en la sección anterior pero ahora para las acciones de Naturgy. Los resultados numéricos obtenidos son los siguientes:

Train desde 2001-01-02 hasta 2021-08-31

Test desde 2021-09-30 hasta 2022-08-31

Nº de lags: 5.

Hiperparámetros: {

    'learning\_rate': 0.15,

    'max\_depth': 5,

    'min\_child\_weight': 2,

    'n\_estimators': 350,

    'subsample': 0.72

}

MSE de: 0.28921

El *Feature Importances* es el que se muestra en la imagen [4.38](#).



Figura 4.38: Feature Importances de Naturgy

## 4.8. Corroboración del estudio

Este último apartado tiene como objetivo validar el estudio mediante una comprobación práctica.

Para realizar la comprobación se ha seleccionado un día en el que se hayan registrado más de dos noticias relevantes. Se procederá a analizar el sentido de cada noticia (positivo o negativo) y se realizará un análisis a nivel de datos, incluyendo visualizaciones, para observar cómo ha reaccionado el precio de la acción en respuesta a estas noticias.

Después de filtrar las noticias relacionadas con la empresa Endesa y agruparlas según la fecha, se ha optado por seleccionar un día al azar para llevar a cabo el análisis. El proceso y resultado de se presenta a continuación:

```

1 # Leemos el fichero que contiene las salidas de GPT
2 gpt = pd.read_csv('./data/gpt.csv', index_col=0)
3
4 # Convierte la columna 'DATE' a datetime
5 gpt['Date'] = pd.to_datetime(gpt['Date'])
6 gpt['Date'] = gpt['Date'].dt.strftime('%Y-%m-%d')
7
8 # Filtra el DataFrame para incluir solo las filas donde la empresa es Endesa
9 endesa_news = gpt[gpt['empresa'] == 'Endesa']
10
11 # Agrupa por fecha y cuenta las entradas para cada día
12 news_counts = endesa_news.groupby('Date').size()
13

```

```

14 # Filtra los resultados para mostrar solo los días con mas de dos noticias
15 more_than_two_news = news_counts[news_counts > 2]
16
17 # Cogemos un valor aleatorio de los valores que cumplen la condicion
18 random_value = more_than_two_news.sample(1)
19
20 # Busca las filas que coinciden con la fecha '2014-11-21'
21 matching_rows = endesa_news.loc[endesa_news['Date'] == random_value[0]]

```

Listing 4.14: Obtención del caso de prueba final

El resultado es el que se observa en la Tabla 4.21.

ID	EMPRESA	TIPO	SENTIDO	FECHA
7144	Endesa	Negocios y Finanzas Corporativas	1	21/11/2014 9:02
5552	Endesa	Negocios y Finanzas Corporativas	0.8	21/11/2014 9:26
7070	Endesa	Negocios y Finanzas Corporativas	0.8	21/11/2014 18:50

Cuadro 4.21: Datos de noticias de Endesa

Luego de recopilar las noticias, hemos observado que estas expresan una perspectiva muy positiva. Con el objetivo de evaluar su impacto en Endesa, procederemos a analizar cómo reaccionaron las acciones durante esos días en términos de cotización.

Con el propósito de realizar dicho análisis, se han recopilado los datos de cotización del día anterior y del día posterior a la publicación de las noticias. Los resultados de esta recopilación se presentan en la Tabla 4.22.

Date	Open	High	Low	Close	Adj Close	Volume
2014-11-19	14.00	14.05	13.88	13.930	7.546004	1829649
2014-11-20	13.93	14.05	13.60	13.710	7.426827	2427887
2014-11-21	13.78	14.60	13.77	14.595	7.906241	256636414

Cuadro 4.22: Datos de cotización de Endesa entre el 2014-11-19 y el 2014-11-21

Si examinamos detenidamente el desarrollo de la acción a un nivel visual, observamos que estas noticias generaron un impacto sumamente positivo, como se ilustra gráficamente en la figura 4.39.

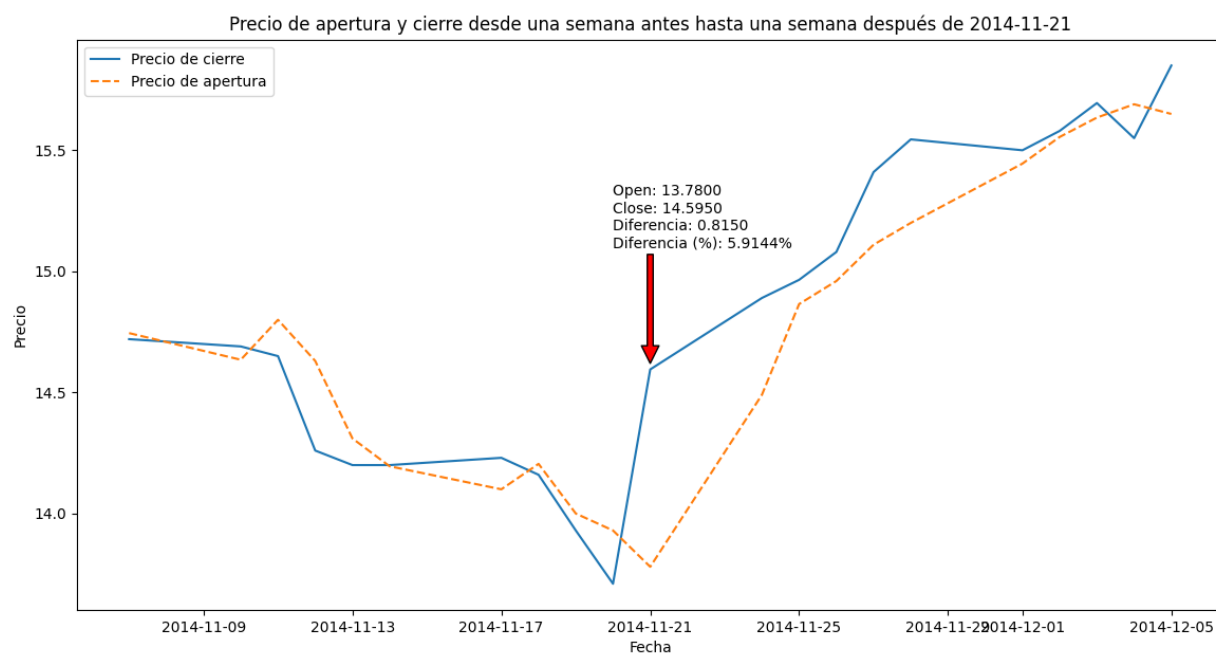


Figura 4.39: Impacto de las noticias





## Capítulo 5

# Conclusiones y Líneas de trabajo futuras

En esta sección, se presentarán las conclusiones del estudio, evaluando si se lograron los objetivos establecidos. Se analizarán los resultados obtenidos sobre el impacto de las noticias en el mercado financiero.

Se determinará si la investigación cumplió con su propósito inicial, proporcionando información clave sobre la relación entre las noticias analizadas y las variaciones en el precio de las acciones.

Finalmente, se sugerirá una dirección futura para la investigación, considerando la posibilidad de ampliar el análisis a nuevas variables o sectores, y explorar metodologías adicionales para una comprensión más detallada de las influencias en el mercado financiero.

### 5.1. Conclusión del estudio

En esta sección, se presentan las conclusiones derivadas del estudio realizado:

- **Dependencia del Precio de las acciones:** Se ha observado que el precio de cotización se encuentra intrínsecamente ligado al precio de cierre del día anterior, como se evidenció en los Features Importances.
- **Impacto Diferenciado en Endesa:** En el caso de Endesa, se destaca que el tipo de noticias relacionadas con el desarrollo de infraestructura y proyectos tiene una influencia

significativamente mayor en la fluctuación del precio de sus acciones en comparación con otros tipos de noticias.

- **Variedad de Factores Influyentes:** Para empresas como Iberdrola, Accionas Energías, Solaria y Naturgy, se observa que el tipo de noticias vinculadas a negocios y finanzas corporativas ejerce la influencia más destacada en la variación del precio de sus acciones. Esto sugiere que diferentes compañías pueden estar sujetas a factores específicos que afectan su rendimiento en el mercado.
- **Relevancia del Sentido y Evaluación de las Noticias:** Se concluye que la comprensión del sentido o evaluación de las noticias es crucial para determinar su influencia en el precio de las acciones. El análisis de este aspecto, junto con el porcentaje de importancia atribuido a cada tipo de noticias en relación con los demás, proporciona una base sólida para calcular la posible reacción de las acciones ante eventos informativos.

En resumen, estas conclusiones subrayan la importancia de considerar diversos factores, tales como el tipo de noticias, su evaluación y el contexto específico de cada empresa, al analizar el impacto de la información en el comportamiento del mercado de acciones. Este enfoque integral permite una comprensión más precisa y perspicaz de cómo las noticias pueden influir en las fluctuaciones de los precios de las acciones de empresas del sector energético.

## 5.2. Consecución de objetivos

En el desarrollo del presente estudio, se llevaron a cabo una serie de objetivos y subobjetivos con el fin de avanzar en el análisis financiero del sector eléctrico del IBEX-35, empleando técnicas de análisis de datos y aprendizaje automático. A continuación, se presenta un resumen de la consecución de cada objetivo:

El **Objetivo general** de este estudio fue desarrollar un enfoque avanzado de análisis financiero centrado en el sector eléctrico del IBEX-35, utilizando técnicas de análisis de datos y aprendizaje automático para predecir los movimientos específicos del mercado de valores, considerando la influencia de diversos tipos de noticias.

Referente a los **Objetivos Parciales**:

- **Recopilación de Datos Financieros:** Se logró obtener de manera exitosa el historial de cotización, los dividendos y otros indicadores relevantes de empresas pertenecientes al sector eléctrico del IBEX-35.
- **Análisis de Datos e Identificación de Noticias Influyentes:** Se realizó un análisis detallado de los datos financieros, identificando patrones y tendencias en el mercado de valores. Además, se integró el análisis de noticias mediante modelos de lenguaje como GPT para identificar factores influyentes basados en la información disponible en distintos tipos de noticias financieras.
- **Desarrollo de Modelos de *Machine Learning*:** Se logró desarrollar modelos de *machine learning* utilizando algoritmos avanzados que demostraron ser capaces de predecir con precisión los movimientos futuros de las acciones en el sector eléctrico del IBEX-35.
- **Evaluación de la Precisión de las Predicciones:** Se evaluó el rendimiento de los modelos de *machine learning*, considerando tanto los datos financieros como la información analizada de las noticias. Se utilizaron métricas de evaluación adecuadas, como el error cuadrático medio (MSE).
- **Documentación de Resultados:** Se documentaron y presentaron de manera clara y concisa los resultados de la investigación, incluyendo los factores identificados a través del análisis de datos financieros y noticias.

## 5.3. Trabajo futuro

En esta sección, se explorarán las perspectivas y futuras líneas de trabajo, proporcionando un esbozo de las ideas a seguir. A continuación, se detallan algunas sugerencias:

1. **Ampliación de Fuentes y Clasificación de Noticias:** Se contempla la posibilidad de expandir la variedad de fuentes de noticias para enriquecer la diversidad de datos.

Además, se propone explorar la inclusión de nuevos tipos de noticias en el proceso de clasificación, con el fin de mejorar la robustez del modelo.

2. **Exploración de Otros Modelos:** Como parte de la continua mejora y evolución, se considera la evaluación de otros modelos de aprendizaje automático. Esta comparación permitirá identificar posibles mejoras y optimizaciones, contribuyendo así a la eficacia general del estudio.
3. **Diversificación en Sectores de Aplicación:** Se plantea la posibilidad de ampliar el alcance del modelo a otros sectores. La exploración de cómo el modelo se adapta y comporta en sectores diversos puede proporcionar valiosas perspectivas.
4. **Implementación de Tecnologías Emergentes:** Se sugiere la evaluación e implementación de tecnologías emergentes en el campo del procesamiento del lenguaje natural y análisis de los datos.
5. **Optimización Continua y Retroalimentación:** Se propone establecer un ciclo de retroalimentación constante para optimizar el modelo en respuesta a los desafíos y oportunidades identificados durante su implementación. La mejora continua es esencial para mantener la relevancia y eficacia a lo largo del tiempo.

Estas ideas forman parte de una estrategia proactiva para avanzar en el desarrollo y empeño del proyecto en el futuro.

# Bibliografía

- [1] Sitio web de la comisión nacional del mercado de valores (cnmv). <https://www.cnmv.es/portal/home.aspx>.
- [2] Endesa gana un 5,3 [https://cincodias.elpais.com/cincodias/2002/10/29/empresas/1036061935\\_850215.html](https://cincodias.elpais.com/cincodias/2002/10/29/empresas/1036061935_850215.html), 10 2002.
- [3] Cinco días: economía y mercados. <https://cincodias.elpais.com/>, 2023.
- [4] Paulo Alburquenque and Luis Ángel Hernández. ¿cómo realizar un análisis fundamental de una empresa? <https://www.rankia.cl/blog/analisis-ipsa/3501534-como-realizar-analisis-fundamental-empresa>, 2022. Actualizado: 31/12/2022.
- [5] Blogthinkbig. ¿sabes en qué se diferencian las redes neuronales del deep learning? <https://blogthinkbig.com/redes-neuronales-deep-learning>, 2017.
- [6] Peter J Brockwell and Richard A Davis. *Introduction to Time Series and Forecasting*. Springer, 2016.
- [7] Novatos Trading Club. ¿qué es el trading? los 7 pasos para empezar. <https://www.novatostradingclub.com/que-es-el-trading/>, 2023.
- [8] Maria Coppola. Qué es la minería de datos: conceptos, técnicas y ejemplos. <https://blog.hubspot.es/marketing/mineria-datos>, 2023.
- [9] Datademia. ¿qué es deep learning y qué es una red neuronal? <https://datademia.es/blog/que-es-deep-learning-y-que-es-una-red-neuronal>, 2020.

- [10] Melanie de ComparaSoftware. Ventajas y desventajas de las herramientas de análisis de sentimiento. <https://blog.comparasoftware.com/analisis-de-sentimiento/#:~:text=Limitaciones%20del%20an%C3%A1lisis%20de%20sentimiento&text=No%20es%20capaz%20de%20reconocer,requerir%20tiempo%20de%20configuraci%C3%B3n%20extra.>, 2022.
- [11] Universidad Nacional Autónoma de México. Técnicas utilizadas en minería de datos. [https://www.cuaed.unam.mx/cursos/tecnologia/tecnicas\\_mineria\\_datos.pdf](https://www.cuaed.unam.mx/cursos/tecnologia/tecnicas_mineria_datos.pdf), 2023.
- [12] Nerea Díaz. Mercado de valores: ¿qué es? definición y concepto. <https://economipedia.com/definiciones/mercado-de-valores-2.html>, 2023.
- [13] easyMarkets. ¿cómo uso el análisis de sentimiento para operar? <https://www.easymarkets.com/eu/es/learn-centre/understanding-analysis/how-do-i-use-sentiment-analysis-to-trade/>, 2023.
- [14] Economiasimple.net. Trading - qué es, definición y concepto. <https://economiasimple.net/glosario/trading>, 2023.
- [15] Economipedia. Qué es el trading y cómo funciona. <https://economipedia.com/definiciones/trading.html>, 2023.
- [16] EFE. Noticias economía de el mundo. <https://www.elmundo.es>, 2007.
- [17] Endesa. Informe anual 2020. <https://www.endesa.com/content/dam/enel-es/home/inversores/informe2020/ia-2020.pdf>, 2020.
- [18] Fisher Investments España. El valor real de los indicadores de sentimiento. <https://www.estrategiasdeinversion.com/analisis/bolsa-y-mercados/informes/el-valor-real-de-los-indicadores-de-sentimiento-n-552607>, 2022.
- [19] Federico J. Caballero Ferrari. Sentimiento de mercado. <https://economipedia.com/definiciones/sentimiento-de-mercado.html>, 2023.

- [20] Javier Sánchez Galán. Empresa: qué es, características, tipos y ejemplos. <https://economipedia.com/definiciones/empresa.html>, 2020. Redactado por: Javier Sánchez Galán, Revisado por: José Francisco López, Actualizado el 1 de marzo de 2020.
- [21] Iván García. Principios de la organización empresarial. <https://emprendepyme.net/principios-de-la-organizacion-empresarial.html>, 2023. Actualizado: 23 de junio de 2023.
- [22] José Luis García. ¿cuál es el sentimiento de mercado? <https://www.estrategiasdeinversion.com/analisis/bolsa-y-mercados/el-experto-opina/cual-es-el-sentimiento-de-mercado-n-224554>, 2014.
- [23] Jose V. Gascó and Javier Lorenzo. Análisis por ratios: qué es y cómo se hace. <https://economipedia.com/definiciones/analisis-por-ratios.html>, 2023. Actualizado: 11/05/2023.
- [24] TIMOTHY LI GORDON SCOTT. Golden cross vs. death cross: What's the difference? <https://www.investopedia.com/ask/answers/121114/what-difference-between-golden-cross-and-death-cross-pattern.asp>, 2022.
- [25] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.
- [26] IAT. Deep learning, aprendizaje profundo y redes neuronales. <https://iat.es/tecnologias/inteligencia-artificial/deep-learning/>, 2018.
- [27] IBM. Ibm spss modeler - crisp-dm help overview. <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>, 2023.
- [28] ICHI.PRO. Una guía completa para el análisis de sentimientos y sus aplicaciones. <https://ichi.pro/es/una-guia-completa-para-el-analisis-de-sentimientos-y-sus-aplicaciones-1260903695>, 2022.



- [29] Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Pearson, 2020.
- [30] Ayman E. Khedr, S.E. Salama, and Nagwa Yaseen. Predicting stock market behavior using data mining technique and news sentiment analysis. *I.J. Intelligent Systems and Applications*, 7:22–30, July 2017.
- [31] Patricio Letelier. Metodologías ágiles para el desarrollo de software: Extreme programming. <http://www.cyta.com.ar/ta0502/v5n2a1.htm>, 2005.
- [32] Logicalis. Minería de datos: cómo funciona, elementos y requisitos. <https://www.logicalis.com/la/que-hacemos/data-and-ai/mineria-de-datos/>, 2023.
- [33] Macarena Salvador Maceira. Machine learning aplicado al trading. <https://repositorio.comillas.edu/rest/bitstreams/295750/retrieve>, 2019.
- [34] Babita Majhi, C.M. Anish, and Ritanjali Majhi. On development of novel hybrid and robust adaptive models for net asset value prediction. 2018.
- [35] Anukrati Mehta. A comprehensive guide to types of neural networks. <https://www.digitalvidya.com/blog/types-of-neural-networks/>, 2022.
- [36] OpenAI. Openai platform documentation. <https://platform.openai.com/docs/introduction>, 2023.
- [37] TIC Portal. Deep learning: ¿qué es, qué tipos hay y para qué sirve? <https://www.ticportal.es/glosario-tic/deep-learning-dl>, 2019.
- [38] ArcGIS Pro. Cómo funciona xgboost. <https://pro.arcgis.com/es/pro-app/latest/tool-reference/geoai/how-xgboost-works.htm>.
- [39] QuestionPro. Minería de datos: Qué es, ventajas, técnicas y cómo realizarla. <https://www.questionpro.com/blog/es/mineria-de-datos/>, 2023.
- [40] Equipo Santander. ¿qué son los dividendos? <https://www.santander.com/es/stories/que-son-dividendos>, 2022.

- [41] Data Science. A comprehensive guide to types of neural networks. <https://datascientest.com/es/que-es-el-transfer-learning>, 2023.
- [42] Robert H Shumway and David S Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, 2017.
- [43] Sanjam Singh and Amandeep Kaur. Twitter sentiment analysis for stock prediction. In *2nd International Conference on “Advancement in Electronics & Communication Engineering (AECE 2022)*, page 674, Ghaziabad, India, July 14-15 2022. Raj Kumar Goel Institute of Technology, Ghaziabad. Published on Page 674.
- [44] Equipo Skilling. Promedios móviles: ¿qué son en análisis técnico? <https://skilling.com/eu/es/blog/trading-articles/what-are-moving-averages/>, 2022.
- [45] Juan Chamorro Soler. Métodos para realizar un análisis de sentimiento. <https://isdfundacion.org/2023/02/03/metodos-para-realizar-un-analisis-de-sentimiento/>, 2023.
- [46] FXCM Research Team. Análisis de sentimiento: ¿qué es? ¿por qué lo necesita? <https://www.fxcm.com/es/insights/sentiment-analysis-what-is-it-why-do-you-need-it/>, 2023.
- [47] XGBoost. Xgboost documentation. <https://xgboost.readthedocs.io/en/stable/>.
- [48] Zoey Yi Zhao, Meng Xie, and Mike West. Dynamic dependence networks: Financial time series forecasting & portfolio decisions. *Applied Stochastic Models in Business and Industry*, 32:311–339, 2016. Original technical report, December 2015.