



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: Diego Calvo Barreno

Barcelona, 5 de febrero de 2024

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	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
Nombre del autor	Eduardo Mora González
Nombre del Tutor	Diego Calvo Barreno
Fecha de entrega	01/2024
Titulación o programa	Máster Universitario en Ciencia de Datos
Área del Trabajo Final	Área 3
Idioma del trabajo	Español
Palabras clave	Cotización, Deep Learning, Noticias

Momentismo absoluto, vivo minuto a minuto
Al futuro lo he dejado atrás
El ayer me atormentaba, el mañana me asustaba
Sólo 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 Barreno* 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, sin ellos todo esto no hubiera sido lo mismo.

A todos ellos, muchas gracias.

Eduardo Mora González

Resumen

Este estudio se centra en entender y prever los movimientos del mercado de valores, conocido por ser volátil y estar influenciado por varios factores. Tomar decisiones en inversiones y estrategias empresariales, en este entorno, es complicado debido a su complejidad. Los métodos comunes de análisis financiero, a veces, no son suficientes y hay una creciente necesidad de herramientas más avanzadas para analizar datos financieros y predecir tendencias.

En concreto, nos enfocamos en las empresas eléctricas que forman parte del IBEX-35, donde las noticias juegan un papel crucial en sus cambios. Estas empresas eléctricas tienen una volatilidad única, los eventos y noticias específicos del sector tienen una influencia significativa, por lo que, necesitamos usar enfoques analíticos adaptados.

El objetivo principal de este estudio es aplicar técnicas avanzadas de análisis de datos para mejorar nuestra capacidad de prever y entender los cambios en el mercado financiero de las empresas eléctricas a través de la noticias.

Palabras clave: Mercado de valores, Volatilidad, Decisiones de inversión, Estrategias empresariales, Análisis financiero, Herramientas avanzadas, IBEX-35, Técnicas de análisis de datos, Empresas eléctricas, Impacto de noticias.

Abstract

This study focuses on understanding and predicting stock market movements, which are known for their volatility and susceptibility to various factors. Making decisions in investments and business strategies in this environment is challenging due to its complexity. Common methods of financial analysis are sometimes insufficient, and there is a growing need for more advanced tools to analyze financial data and predict trends.

Specifically, we are focusing on electric companies that are part of the IBEX-35, as news plays a crucial role in their changes. These electric companies have a unique volatility, and sector-specific events and news have a significant influence, requiring us to use tailored analytical approaches.

The main objective of this study is to apply advanced data analysis techniques to enhance our ability to predict and understand changes in the financial market of electric companies through news.

Keywords: Stock market, Volatility, Investment decisions, Business strategies, Financial analysis, Advanced tools, Data analysis techniques, IBEX-35, Electric companies, News impact.

Índice de figuras

2.1. Media móvil de 50-200 días (MA) de las acciones de Endesa	11
4.1. Diagrama de las fases del estudio	25
4.2. Crecimiento de las acciones de Endesa a lo largo de los años	28
4.3. Crecimiento de las acciones de Iberdrola a lo largo de los años	28
4.4. Crecimiento de las acciones de Acciona Energía a lo largo de los años	28
4.5. Crecimiento de las acciones de Naturgy a lo largo de los años	29
4.6. Crecimiento de las acciones de Solaria a lo largo de los años	29
4.7. Media móvil de 50-200 días (DMA) de las acciones de Endesa	30
4.8. 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)	30
4.9. Descuadre acciones de Endesa por año, semana y día	31
4.10. Noticias Endesa [12] y [13]	32
4.11. Descuadre acciones de Iberdrola por año, semana y día	33
4.12. Descuadre acciones de Acciona Energía por año y día	35
4.13. Descuadre acciones de Solaria por año, semana y día	36
4.14. Descuadre acciones de Naturgy por año, semana y día	38
4.15. Evolución de los dividendos Endesa	39
4.16. Comparación de la evolución mensual del precio de cierre de Endesa con y sin dividendos	40
4.17. Evolución de los dividendos Iberdrola	41

4.18. Comparación de la evolución mensual del precio de cierre de Iberdrola con y sin dividendos	41
4.19. Evolución de los dividendos de Solaria	42
4.20. Comparación de la evolución mensual del precio de cierre de Solaria con y sin dividendos	43
4.21. Evolución de los dividendos de Naturgy	44
4.22. Comparación de la evolución mensual del precio de cierre de Iberdrola con y sin dividendos	44
4.23. Correlaciones entre las empresas del sector eléctrico	45
4.24. Ejemplo salida ventana de Holt-Winters	47
4.25. Comparación de los hiperparámetros con MSE más bajo en el periodo de TEST de 2009-09-30 a 2010-08-31	48
4.26. Feature Importances con todas las variables	52
4.27. Feature Importances con las empresas agrupadas	53
4.28. Feature Importances de Iberdrola	53
4.29. Feature Importances de Acciona Energías	54
4.30. Feature Importances de Solaria	54
4.31. Feature Importances de Naturgy	55
4.32. Impacto de las noticias	56
4.33. Impacto de un tipo de noticia	57

Índice de cuadros

1.1. Cronograma de tareas	8
4.1. Top 5 días con mayor diferencia porcentual positiva en las acciones de Endesa . . .	32
4.2. Top 5 días con mayor diferencia porcentual negativa en las acciones de Endesa . . .	32
4.3. Top 5 días con mayor diferencia porcentual positiva en las acciones de Iberdrola . .	34
4.4. Top 5 días con mayor diferencia porcentual negativa en las acciones de Iberdrola . .	34
4.5. Top 5 días con mayor diferencia porcentual positiva en las acciones de Acciona Energía	34
4.6. Top 5 días con mayor diferencia porcentual negativa en las acciones de Acc. Energía	35
4.7. Top 5 días con mayor diferencia porcentual positiva en las acciones de Solaria . . .	37
4.8. Top 5 días con mayor diferencia porcentual negativa en las acciones de Solaria . . .	37
4.9. Top 5 días con mayor diferencia porcentual positiva en las acciones de Naturgy . . .	37
4.10. Top 5 días con mayor diferencia porcentual negativa en las acciones de Solaria . . .	38
4.11. Dividendos Endesa agrupados por año	39
4.12. Dividendos Iberdrola agrupados por año	40
4.13. Dividendos Solaria	42
4.14. Dividendos Naturgy agrupados por año	43
4.15. Matriz de correlación entre empresas del sector eléctrico.	45
4.16. MSE de las primeras 10 ventanas de Holt-Winters	46
4.17. Ejemplos resultados del modelo	47
4.18. Comparación MSE entre XGBoost y Holt-Winters	48
4.19. Comparación de errores MSE sin y con Noticias	51
4.20. Datos de noticias de Endesa	55

Índice general

Resumen	v
Abstract	vi
Índice de figuras	vii
Índice de cuadros	ix
Índice general	x
1 Introducción	2
1.1. Contexto y justificación del trabajo	2
1.2. Descripción general del problema	4
1.3. Motivación personal	4
1.4. Definición de los objetivos	5
1.4.1. Objetivo General	5
1.4.2. Objetivos Parciales	6
1.5. Descripción de la metodología	6
1.6. Planificación	7
2 Conceptos Teóricos	9
2.1. Fundamentos financieros	9
2.1.1. Mercado de valores	9
2.1.2. Trading	10

2.1.3.	Medias Móvil (Moving Average)	10
2.1.4.	Golden Cross y Death Cross	11
2.1.5.	Dividendos	11
2.2.	Fundamentos de minería de datos	12
2.2.1.	Deep Learning y las Redes Neuronales	12
2.2.2.	Procesamiento del lenguaje natural con la API de GPT	12
2.2.3.	XGBoost	13
2.2.4.	Holt-Winters	14
3	Estado del arte	16
3.1.	Análisis de sentimientos en la predicción del mercado de valores	16
3.1.1.	Aplicaciones del análisis de sentimientos	16
3.1.2.	Métodos de análisis de sentimientos	17
3.1.3.	Limitaciones del análisis de sentimientos	18
3.1.4.	Principales artículos	19
3.2.	Deep Learning en la predicción del precio de las acciones	21
3.2.1.	Ventajas del uso de Deep Learning	21
3.2.2.	Limitaciones y desafíos	21
3.2.3.	Aplicaciones de <i>Deep Learning</i> en la Predicción de Acciones	22
3.2.4.	Principales artículos	22
4	Diseño e implementación	25
4.1.	Obtención de los datos	25
4.1.1.	Justificación del uso de la API de yfinance	26
4.1.2.	Datos de cotización y dividendos	26
4.2.	Análisis de los datos	27
4.2.1.	Análisis de los precios de apertura y cierre de las acciones	27
4.2.2.	Promedio Móvil (Moving Average), Golden Cross y Death Cross	29
4.2.3.	Discrepancias en los valores de cierre y apertura de acciones	30

4.2.4.	Análisis de Dividendos	39
4.2.5.	Relación Entre Empresas	45
4.3.	Holt-Winters Vs XGBoost	46
4.3.1.	Holt-Winters	46
4.3.2.	XGBoost con Cross Validation	46
4.3.3.	Comparación de los modelos	48
4.4.	Análisis de noticias	49
4.4.1.	Justificación del uso de Web Scraping para obtener noticias	49
4.4.2.	Web Scraping	50
4.4.3.	GPT para el análisis de noticias	50
4.5.	Retroalimentación de XGBoost	50
4.6.	Pruebas del nuevo modelo XGBoost	51
4.6.1.	Feature Importances con todas las variables	51
4.6.2.	Feature Importances con el sentido incorporado a los tipos de noticias y generalizando todas las empresas distintas a Endesa	52
4.7.	Extrapolación al resto de empresas	53
4.7.1.	Iberdrola	53
4.7.2.	Acciona Energías	54
4.7.3.	Solaria	54
4.7.4.	Naturgy	55
4.8.	Contrastación práctica del estudio	55
4.9.	Dinámica del precio de las acciones después de la publicación de noticias	56
5	Conclusiones y líneas de trabajo futuras	58
5.1.	Conclusión del estudio	58
5.2.	Trabajo futuro	59
	Bibliografía	61

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.

1.1. Contexto y justificación del trabajo

Este trabajo de investigación se sitúa en el contexto de los mercados financieros globales los cuales 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: 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, se ha contemplado 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), enfocada en el uso de tecnologías avanzadas, incluyendo análisis de datos y *Machine Learning* (Aprendizaje Automático), para comprender y predecir los movimientos del mercado.

Predecir con rigor 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. Una predicción minuciosa 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 fundamenta en:

- **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 aplicación de técnicas avanzadas de análisis de datos y *Machine Learning* representa un avance importante en la capacidad de análisis y predicción en el campo financiero.
- **Enfoque Específico en el Sector Eléctrico:** Nuestro análisis se concentra en el sector eléctrico, asegurando la relevancia y aplicabilidad de los resultados dentro de este contexto empresarial y económico particular.
- **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.

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 toma de 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 necesitan herramientas más avanzadas y precisas para analizar datos financieros y predecir tendencias del mercado con mayor confiabilidad.

Esta investigación se enfoca en la aplicación de técnicas avanzadas de análisis de datos y *Machine Learning* para abordar el desafío de predecir con mayor exactitud y eficacia los movimientos del mercado de valores, específicamente dentro del sector eléctrico del IBEX-35. La pregunta central que guía esta investigación es cómo aprovechar los avances tecnológicos y el acceso a datos masivos para mejorar la comprensión y anticipación de los cambios en el mercado financiero del sector eléctrico, teniendo en cuenta el impacto de las noticias.

1.3. Motivación personal

La decisión para embarcarme en esta investigación proviene de una combinación de factores personales y profesionales que me han llevado a abordar este desafiante tema. Algunas de las razones que impulsan mi interés en este proyecto son:

- **Entusiasmo por las Finanzas y la Tecnología:** Siempre he sentido un profundo interés por el ámbito financiero y la innovación. Observar la transformación en la toma de decisiones financieras y la gestión de inversiones me ha fascinado. Esta investigación me permite fusionar estas dos áreas y explorar cómo enfoques avanzados 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 rigor 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. Esta investigación me otorga la oportunidad de aprender y aplicar las últimas técnicas y enfoques en este campo.
- **Impacto Potencial:** La idea de que los resultados de esta investigación puedan tener un impacto real y significativo en la toma de decisiones financieras y que las conclusiones puedan ayudar a los inversores y empresas a tomar decisiones más informadas y anticipar mejor los cambios en el mercado, compensaría todo el trabajo realizado.

1.4. Definición de los objetivos

En esta sección, se presenta el objetivo general que se busca alcanzar, así como los objetivos específicos que lo componen.

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 *Machine Learning* para 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 de empresas pertenecientes al sector eléctrico del IBEX-35.
- **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 *Generative Pre-Trained Transformer* (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

La metodología por la que se ha optado ha sido *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [17] que proporciona un marco estructurado y completo para la gestión de proyectos de minería de datos. Se divide en las siguientes etapas:

1. **Comprensión del Negocio:** Se definen los objetivos del proyecto y se evalúa su relevancia para el negocio en cuestión.

2. **Comprensión de los Datos:** 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 ofrece 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 [21].

1.6. Planificación

Este proyecto 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 el cronograma real con todas las actividades de cada módulo.

NOMBRE DE LA TAREA	FECHA DE INICIO	FECHA DE FINALIZACIÓN
MÓDULO 1	27.09.2023	10.10.2023
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
MÓDULO 2	11.10.2023	24.10.2023
Conceptos teóricos	11.10.2023	15.10.2023
Estado del arte	15.10.2023	24.10.2023
MÓDULO 3	25.10.2023	19.12.2023
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
MÓDULO 4	20.12.2023	16.01.2024
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
MODULO 5	17.01.2024	04.02.2024
Entrega documentación	18.01.2024	18.01.2024
Defensa	23.01.2024	23.01.2024

Cuadro 1.1: Cronograma de tareas

Capítulo 2

Conceptos Teóricos

El objetivo de este capítulo es 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 ámbito financiero, esencial en la economía global, se distingue por su dinamismo. Con elementos clave como indicadores empresariales, la dinámica bursátil y estrategias de *trading*, enfrenta una creciente competitividad. La capacidad para anticipar tendencias se convierte en una ventaja competitiva crucial. Esta sección introduce estos pilares fundamentales, sentando las bases para la exploración detallada en este estudio.

2.1.1. Mercado de valores

El mercado de valores es un conjunto de mercados e instituciones donde se negocian valores financieros como acciones, bonos y derivados. Cumple el propósito de facilitar el financiamiento para empresas y ofrecer oportunidades de inversión rentables y líquidas [1].

Regulado por autoridades como la Comisión Nacional del Mercado de Valores (CNMV) en España, el mercado busca transparencia, seguridad y eficiencia. Su importancia radica en su contribución al crecimiento, innovación y estabilidad financiera, permitiendo a empresas diversificar riesgos y obtener financiación alternativa [9]. Para inversores, ofrece oportunidades de renta, diversificación y participación en el desarrollo empresarial.

2.1.2. 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 [11].

Para hacer *trading* se necesita una plataforma electrónica que permita acceder al mercado y ejecutar las órdenes. Esta plataforma se denomina *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 [5].

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 [10].

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 [10].

2.1.3. Medias Móvil (Moving Average)

Las medias móviles (MA) son un componente esencial del análisis técnico y crucial para los *trader*. 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 [30].

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 muestra 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.

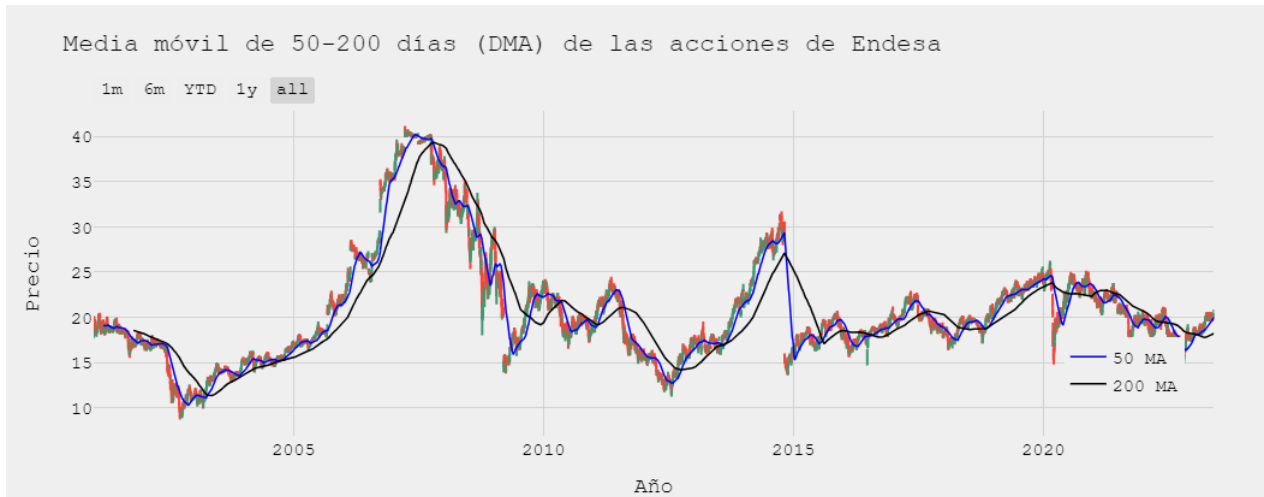


Figura 2.1: Media móvil de 50-200 días (MA) de las acciones de Endesa

2.1.4. 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 (se refiere al cruce de la media móvil de 50 días sobre la media móvil de 200 días). Es considerado como una señal alcista, indica un posible cambio positivo en la tendencia del mercado y sugiere un potencial aumento en los precios [14].

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, el cruce de la media móvil de 50 días está por debajo de la media móvil de 200 días. Es interpretado como una señal bajista que sugiere un posible cambio negativo en la tendencia del mercado y puede indicar una disminución en los precios [14].

2.1.5. Dividendos

Los dividendos [27] representan la porción de las ganancias que las empresas distribuyen a los accionistas. 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.

2.2. Fundamentos de minería de datos

La minería de datos consiste en explorar grandes conjuntos de datos para extraer conocimientos valiosos, utilizando técnicas estadísticas e inteligencia artificial [6]. Se sigue un ciclo de vida, como el CRISP-DM, que incluye comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue [22]. La minería de datos ofrece beneficios y aplicaciones amplias en diversos sectores [6].

2.2.1. Deep Learning y las Redes Neuronales

El *Deep Learning*, una rama del *Machine Learning*, utiliza Redes Neuronales artificiales para abordar problemas complejos de inteligencia artificial [7]. Incluye redes como las neuronales convolucionales (CNN) para imágenes, las recurrentes (RNN) para secuencias temporales y las profundas (DNN) con capas múltiples [16].

Este enfoque ha revolucionado la inteligencia artificial, posibilitando la resolución de desafíos como reconocimiento facial, traducción automática y conducción autónoma [3]. A pesar de sus éxitos, el *Deep Learning* enfrenta retos como la necesidad de grandes conjuntos de datos y la complejidad computacional [3].

2.2.2. Procesamiento del lenguaje natural con la API de GPT

El Procesamiento del Lenguaje Natural (PLN), una rama de la inteligencia artificial, se centra en la interacción entre las computadoras y el lenguaje humano, desarrollando algoritmos

para entender, interpretar y generar texto de manera efectiva. Incluye tareas como análisis de sentimientos, traducción automática y reconocimiento de voz [19].

La API de *Generative Pre-trained Transformer* (GPT) ha revolucionado el PLN al permitir el análisis contextual y la generación coherente de texto. Utilizada para tareas como resumen de texto, traducción automática y respuesta a preguntas, esta herramienta ofrece una comprensión más profunda de la información contenida en noticias recopiladas [25].

2.2.3. XGBoost

El *eXtreme Gradient Boosting* (XGBoost) es un algoritmo de aprendizaje automático supervisado que se utiliza para tareas de clasificación y regresión. Es una implementación 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 [33].

Las características clave de XGBoost son [26]:

- **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 [28].

El método de *Holt-Winters* tiene tres componentes principales para modelar series temporales [15]:

- **Nivel (*Level*):** Representa el valor medio de la serie temporal.
- **Tendencia (*Trend*):** Muestra la dirección general de la serie temporal.
- **Componente Estacional (*Seasonal*):** Identifica 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 [4]:

- **Suavizado Exponencial Simple (*Single Exponential Smoothing*):** Usa 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*):** Incorpora 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 combina finanzas y procesamiento del lenguaje natural (NLP) para evaluar las emociones y opiniones de inversores, analistas y del público sobre activos financieros. Esta disciplina, es esencial, para obtener datos valiosos sobre tendencias futuras de precios. Con avances tecnológicos y mayor disponibilidad de datos, el análisis de sentimientos se ha vuelto más avanzado y preciso [32].

3.1.1. 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. Las aplicaciones más destacadas son [18]:

1. **Predicción de Precios de 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:** 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:** 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 evaluar el sentimiento del mercado hacia empresas competidoras y comparar su rendimiento relativo.

3.1.2. 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 [32]. Algunos de los métodos más comunes incluyen [31]:

1. **Análisis de Polaridad:** 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:** 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:** 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:** Se utilizan para analizar grandes conjuntos de datos de texto y predecir el sentimiento del mercado. 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.

3.1.3. 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 [8]:

1. **Ruido en los Datos:** Los datos 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. Lo que 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.

3.1.4. Principales artículos

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

3.1.4.1. Twitter sentiment analysis for stock prediction

El artículo [29] 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 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 *Support Vector Machine* (SVM) 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 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. También proporciona información detallada sobre el proceso de preprocesamiento de datos y los niveles de precisión alcanzados con los diferentes algoritmos empleados.

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

El artículo [20] estudia 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 determina la relevancia de las palabras en un documento o una colección de documentos.

El estudio presenta 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, presenta 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. Ventajas del uso de Deep Learning

Las principales ventajas de usar técnicas de *Deep Learning* en la predicción de acciones son [23]:

- **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.

3.2.2. 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 [23]:

- **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.

3.2.3. Aplicaciones de *Deep Learning* en la Predicción de Acciones

Las diferentes aplicaciones del uso de las técnicas del *Deep Learning* son [18]:

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 se basan en modelos de *Deep Learning* para evaluar y gestionar mejor los riesgos asociados con la inversión en acciones.
5. **Simulación de Escenarios:** Las GAN sirven para simular escenarios y evaluar estrategias de inversión en diferentes condiciones del mercado.

3.2.4. 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

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

El artículo [34] 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 planteamiento se basa en 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 estudio también presenta un 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.

Los resultados de la investigación 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.

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

El artículo [24] 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 recopilaron 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).

Se caracteriza por su resistencia frente a valores atípicos y su capacidad para gestionar eficazmente la contaminación de datos en el proceso de predicción. 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.

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 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.

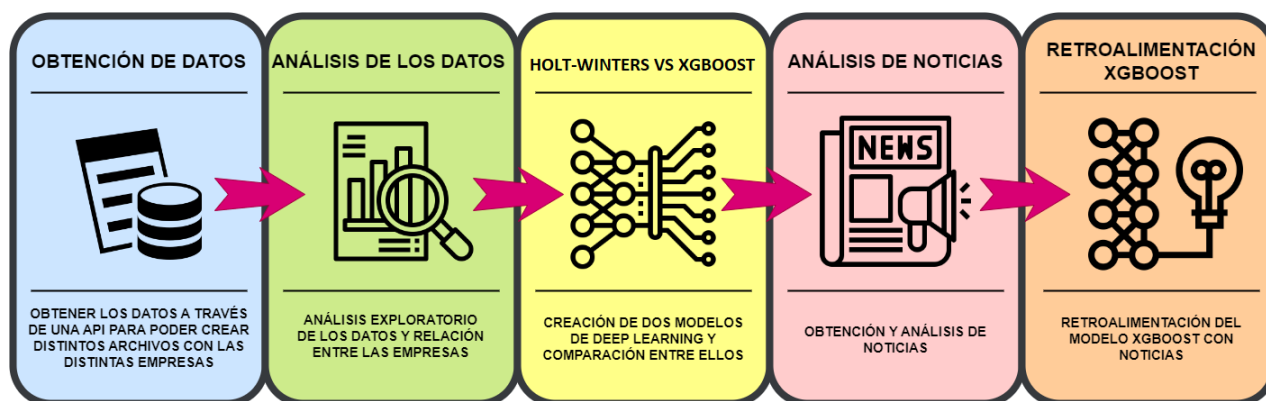


Figura 4.1: Diagrama de las fases del estudio

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 inicial 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.

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

La elección de utilizar la API *yfinance* en lugar de otros métodos, como el *web scraping*, se basa en consideraciones de fiabilidad y eficiencia. *Yfinance* ha demostrado ser una fuente confiable al proporcionar datos financieros precisos y consistentes, validados mediante pruebas comparativas. Su interfaz simple facilita la implementación y extracción de información relevante, evitando la complejidad y posibles restricciones legales asociadas al *web scraping*. La elección se fundamenta en la fiabilidad de los datos, la facilidad de uso y la preferencia por evitar la complejidad del *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 dividendos_Endesa = yf.Ticker('ELE.MC').dividends.loc[start:end]
4 dividendos_Endesa.to_csv('./data/dividendos_Endesa.csv')
```

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

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 diversos apartados con el objetivo de examinar aspectos clave.

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.

Como se puede analizar, 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

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

Tal y como se mencionó en el punto [2.1.3](#) 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 como se muestra en la figura [4.7](#) relativa a las acciones de Endesa.

Por otro lado, en la sección [2.1.4](#) se hace referencia al *Golden Cross* y *Death Cross* que aparecen cuando las líneas anteriores se cortan. 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 [4.8](#).

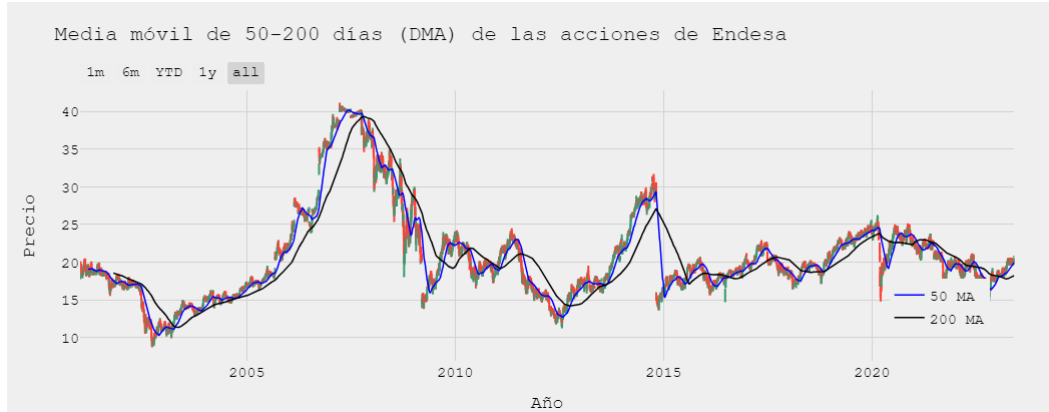


Figura 4.7: Media móvil de 50-200 días (DMA) de las acciones de Endesa

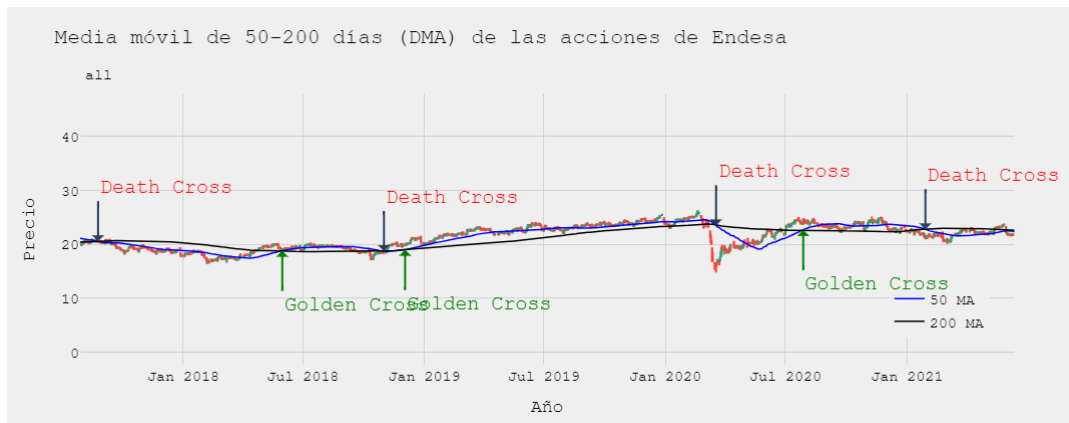


Figura 4.8: Media Móvil de 50-200 días (DMA) de las acciones de Endesa con marcadores Golden Cross y Death Cross (junio de 2018 - junio de 2021)

4.2.3. Discrepancias en los valores de cierre y apertura de acciones

Se realizó un análisis para detectar diferencias entre el cierre del día anterior y la apertura actual en el mercado de valores buscando desviaciones superiores al 2 %. Se observaron eventos destacados durante esos días para comprender mejor las causas, como información post-cierre, operaciones fuera de horario o transacciones internacionales. Se llevó a cabo un análisis manual inicial de las desviaciones más notables en los precios de las acciones, centrándose en explicar los eventos del día.

4.2.3.1. Discrepancias en los valores de Endesa

En el gráfico 4.9, 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.

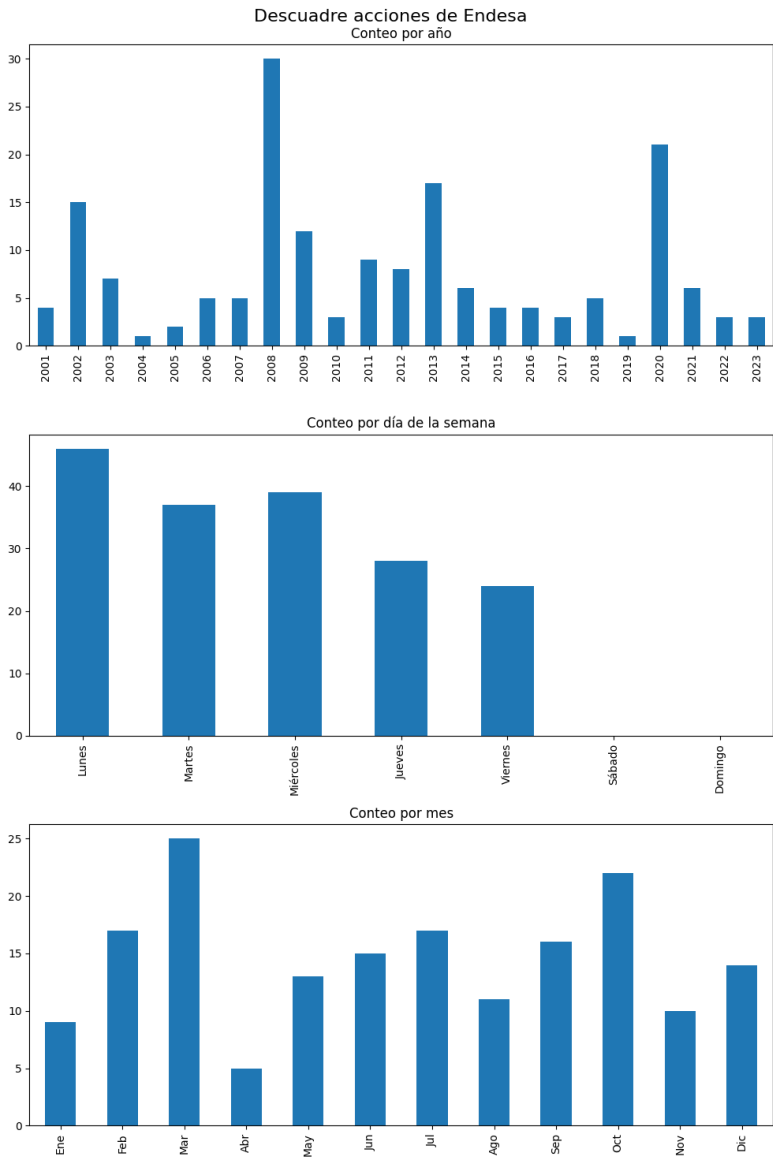


Figura 4.9: Descuadre acciones de Endesa por año, semana y día

En las tablas siguientes, se muestran las fechas en las que se observaron las mayores diferencias porcentuales, tanto positivas 4.1 como negativas 4.2, en las acciones de Endesa.

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

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

Se explora, a modo de ejemplo, dar una explicación para entender las razones detrás de estas variaciones positivas de la tabla 4.1.

El Mundo website screenshot showing news articles related to Endesa. The top article is 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". Below it is an article dated 2005-09-06 titled "El presidente de Endesa compró 50.000 acciones de la empresa el viernes". To the left, there is a section titled "Dividendo" with text about the 2019-2022 dividend policy. The bottom left shows a date "2020-05-05".

Figura 4.10: Noticias Endesa [12] y [13]

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 [4.10](#).

4.2.3.2. Discrepancias en los valores de Iberdrola

En el gráfico [4.11](#), 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.

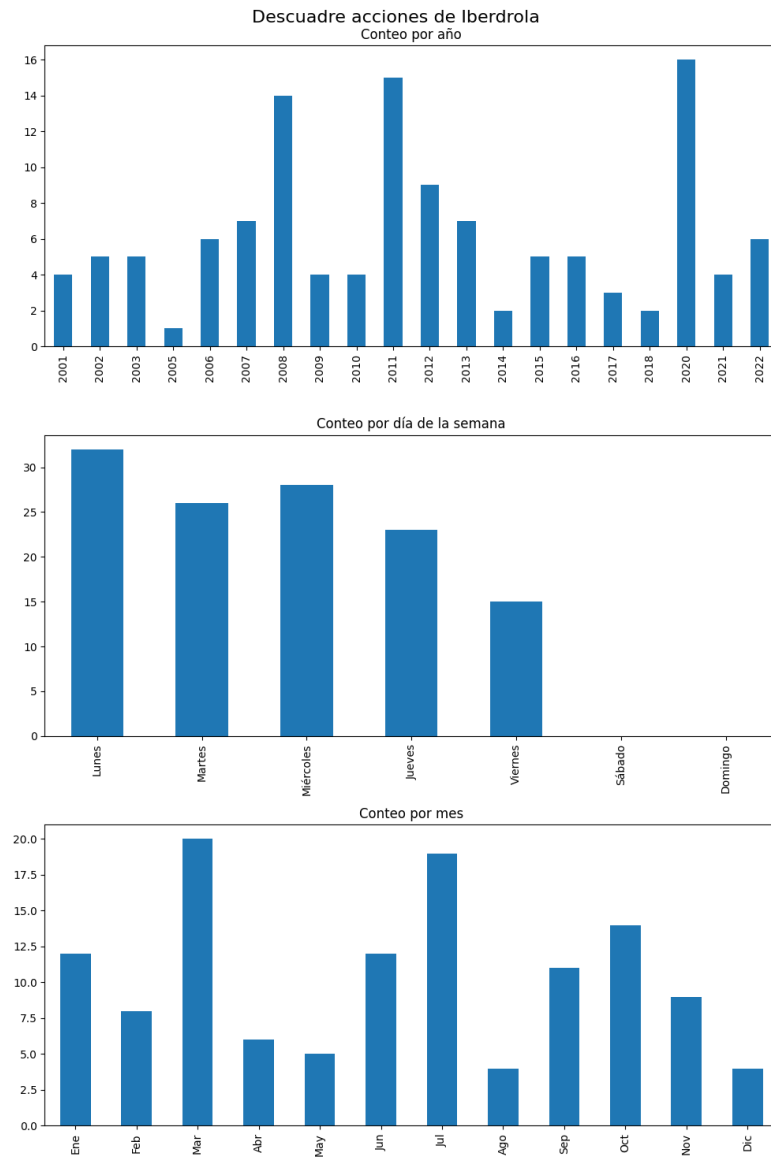


Figura 4.11: Descuadre acciones de Iberdrola por año, semana y día

En las tablas siguientes, se muestran 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

4.2.3.3. Discrepancias en los valores de Acciona Energía

En el gráfico 4.12, 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.

En las tablas siguientes, se muestran 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.

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

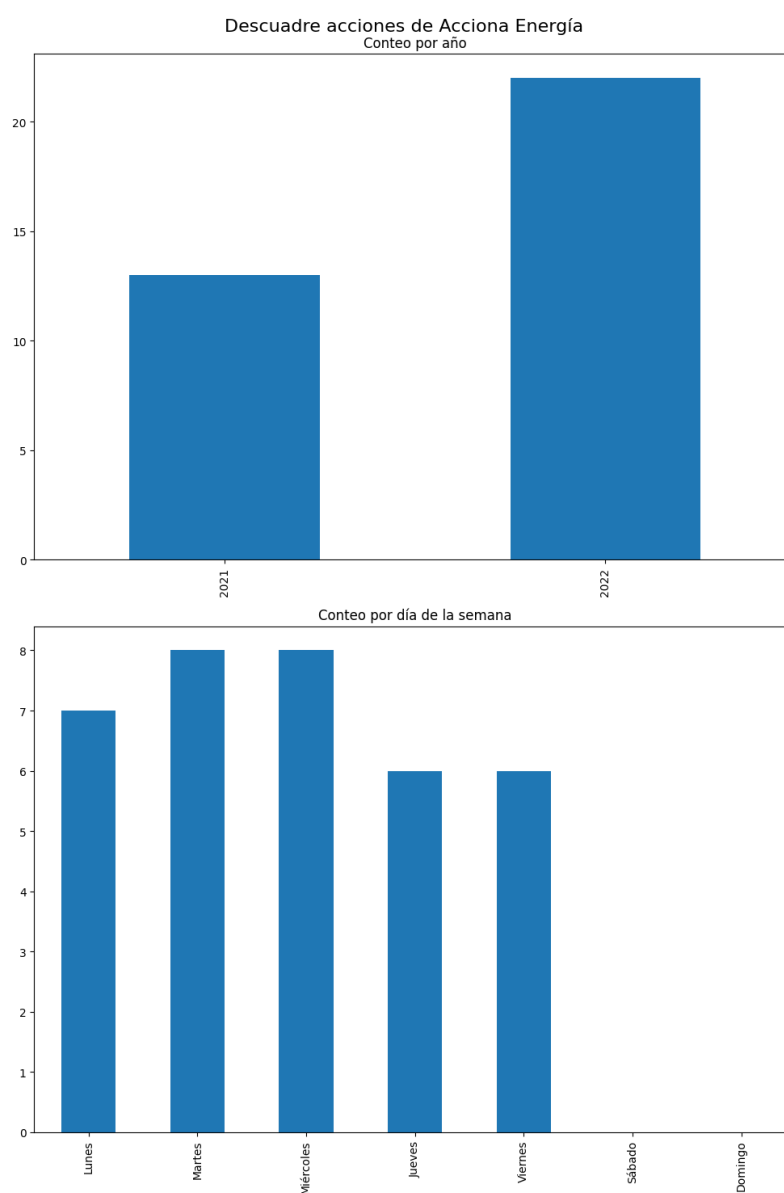


Figura 4.12: Descuadre acciones de Acciona Energía por año y día

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 Acc. Energía

4.2.3.4. Discrepancias en los valores de Solaria

En el gráfico 4.13, 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.

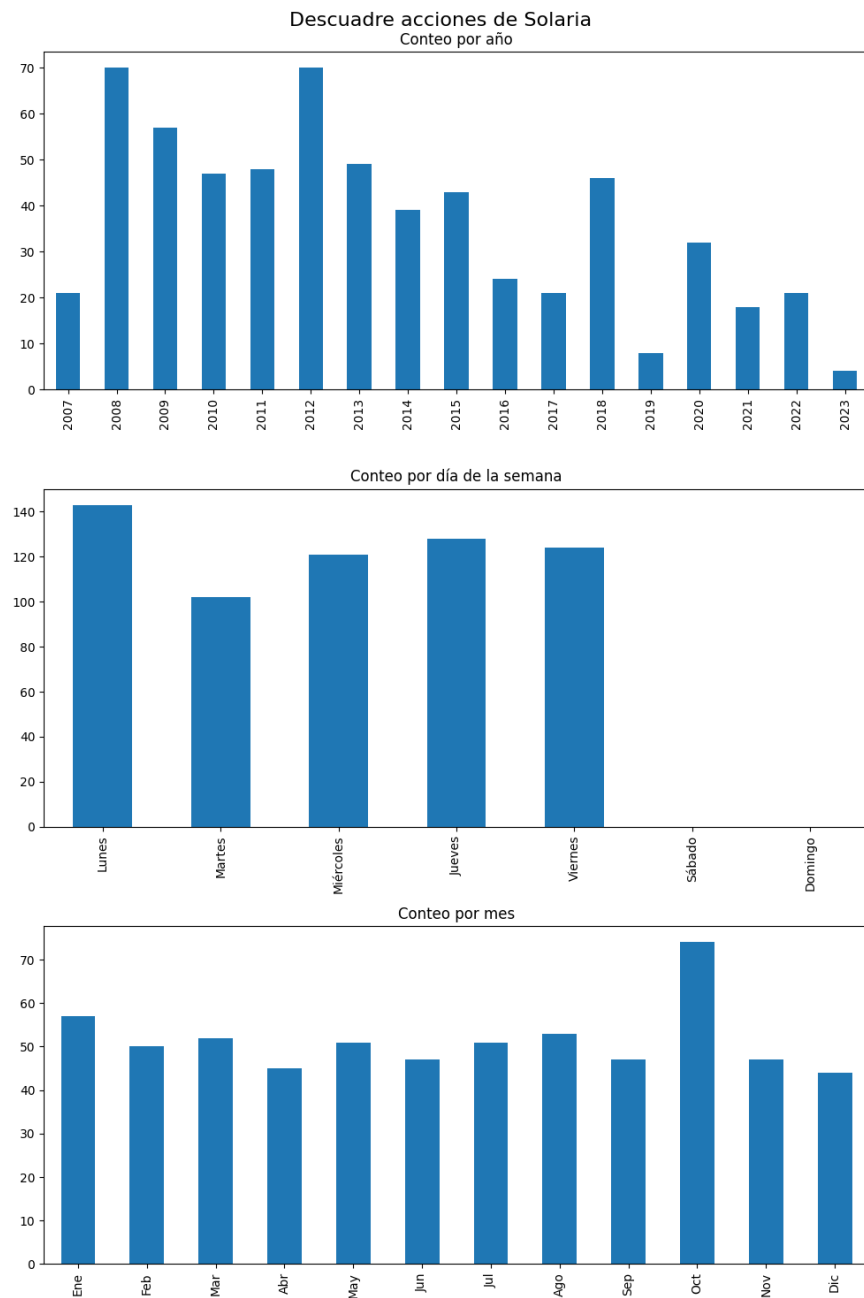


Figura 4.13: Descuadre acciones de Solaria por año, semana y día

En las tablas siguientes, se muestran 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

4.2.3.5. Discrepancias en los valores de Naturgy

En el gráfico 4.14, 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.

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

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

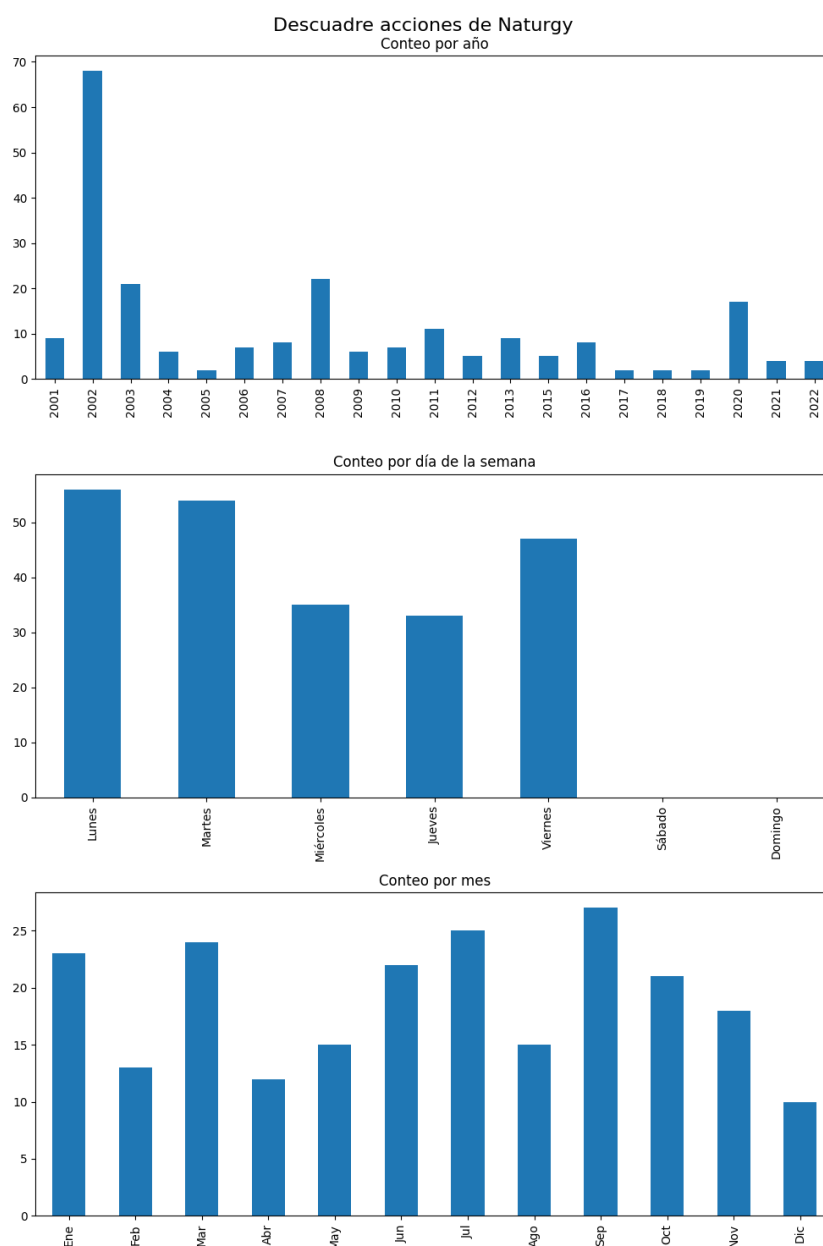


Figura 4.14: Descuadre acciones de Naturgy por año, semana y día

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

4.2.4. Análisis de Dividendos

Tal y como se mencionó en el punto 2.1.5 un dividendo de una empresa es una distribución de las ganancias de esa empresa a sus accionistas. Estos pagos se realizan periódicamente, generalmente trimestral o anualmente.

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.

Año	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Precio	0.325	0.341	0.341	0.351	0.369	0.786	0.820	0.766	5.897	0.514	0.509
Año	2012	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Precio	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 se registra en la figura 4.15.

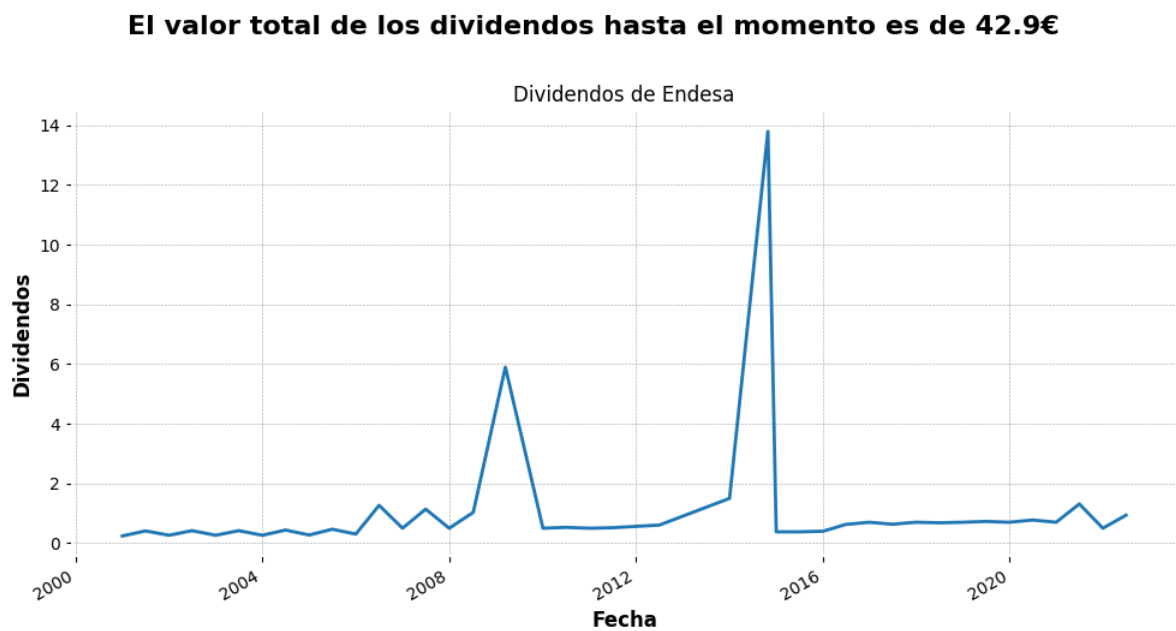


Figura 4.15: Evolución de los dividendos Endesa

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 4.16 se muestra el precio de la acción con y sin dividendos.

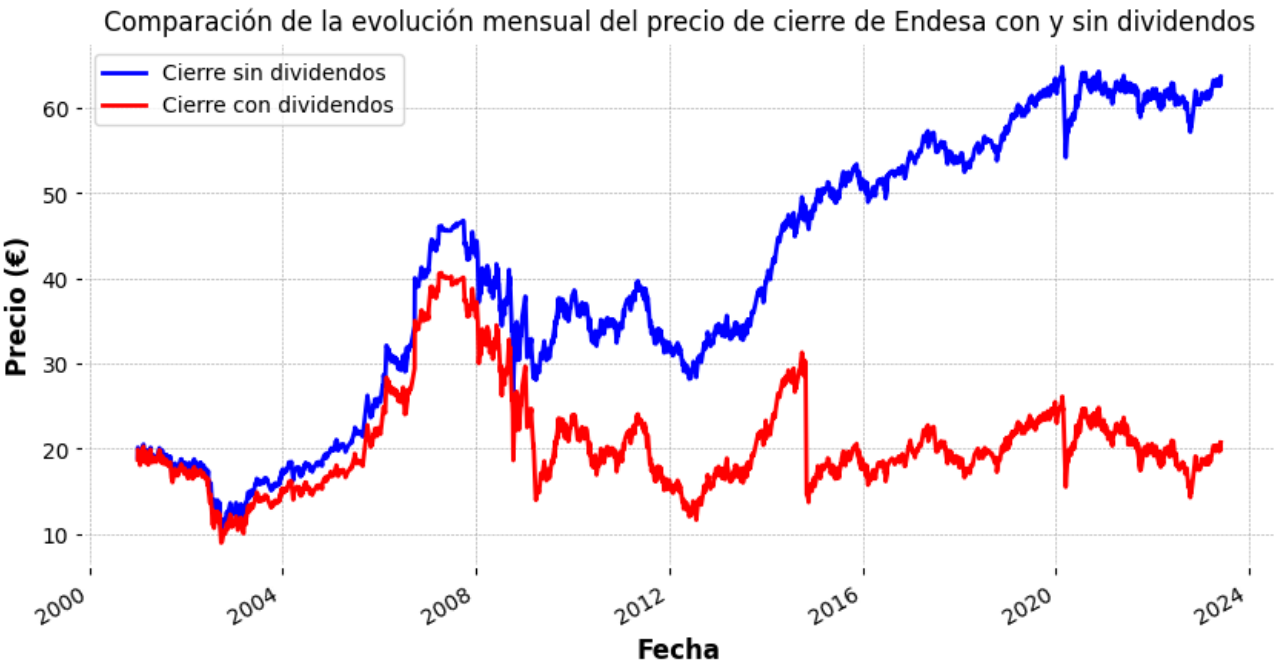


Figura 4.16: Comparación de la evolución mensual del precio de cierre de Endesa con y sin dividendos

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.

Año	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Precio	0.069	0.074	0.076	0.084	0.096	0.111	0.130	0.136	0.164	0.172	0.181
Año	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Precio	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 se registra en la figura 4.19.

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

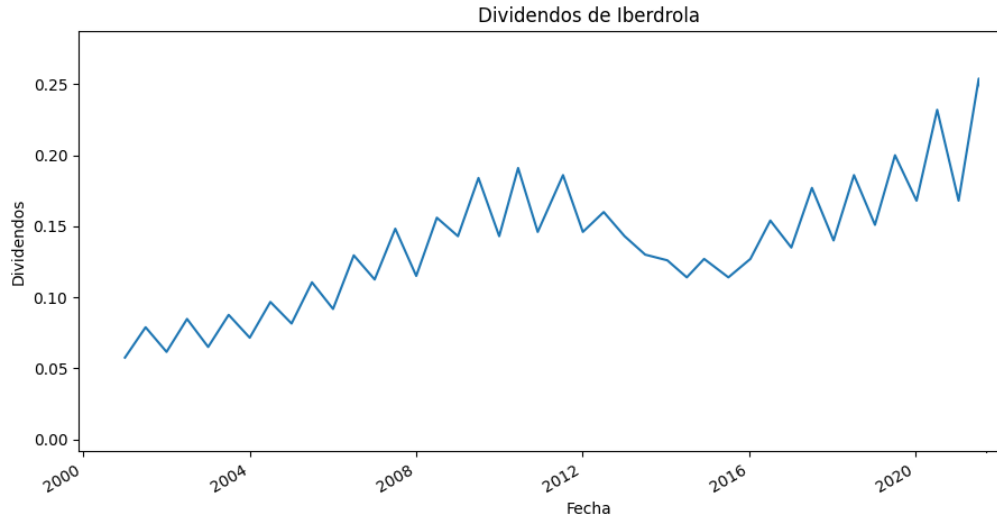


Figura 4.17: Evolución de los dividendos Iberdrola

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 4.20 se muestra el precio de la acción con y sin dividendos.

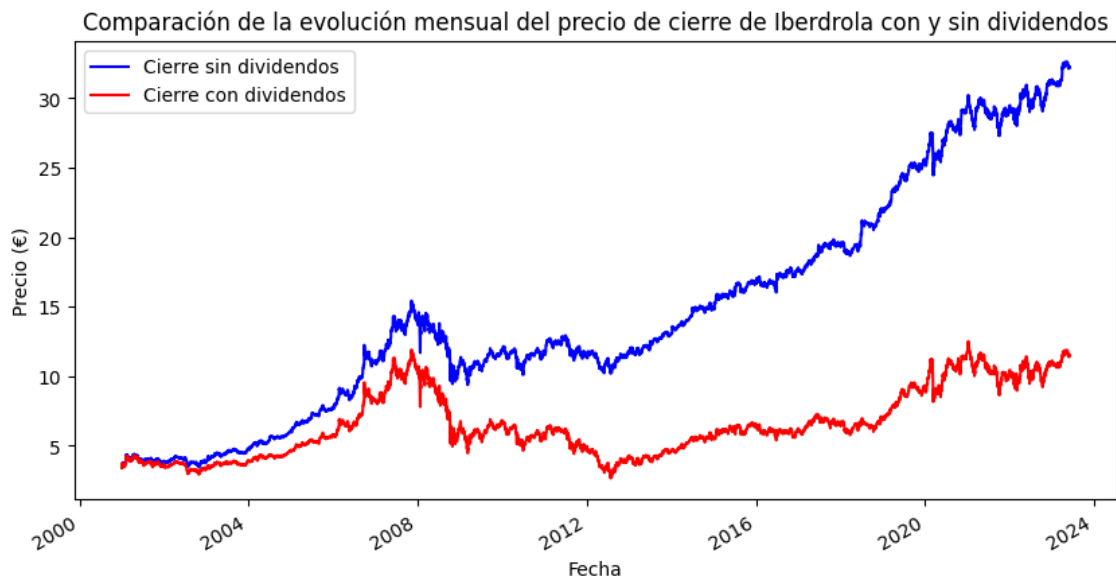


Figura 4.18: Comparación de la evolución mensual del precio de cierre de Iberdrola con y sin dividendos

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 se registra en la figura [4.19](#).

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

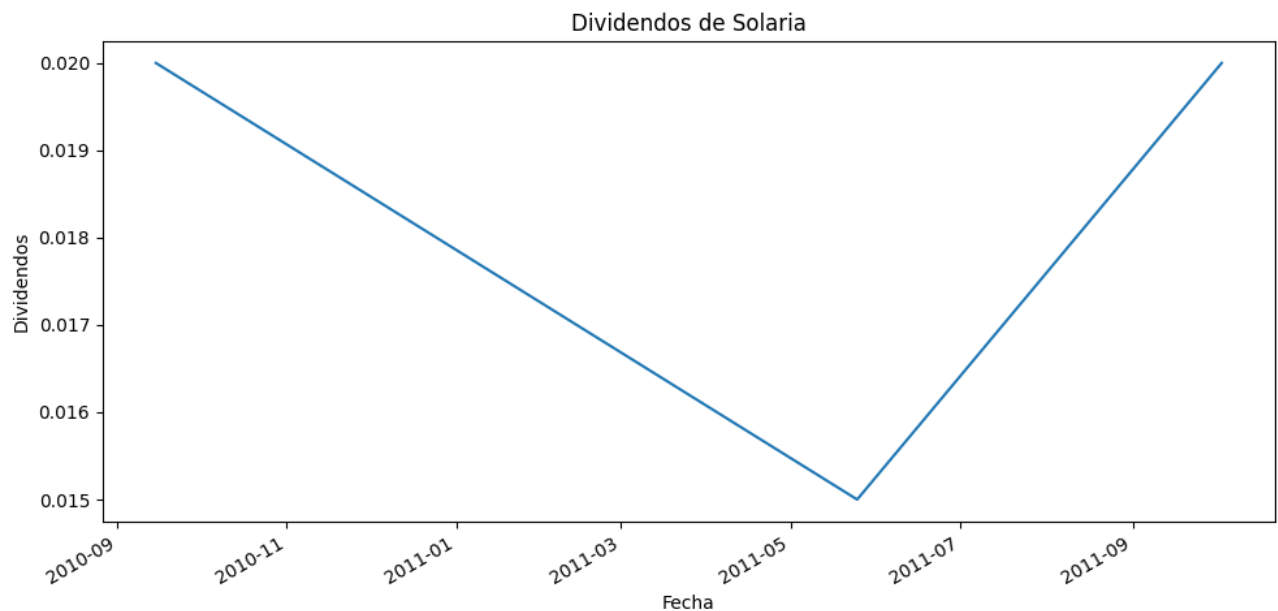


Figura 4.19: Evolución de los dividendos de Solaria

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 4.20 se muestra el precio de la acción con y sin dividendos.

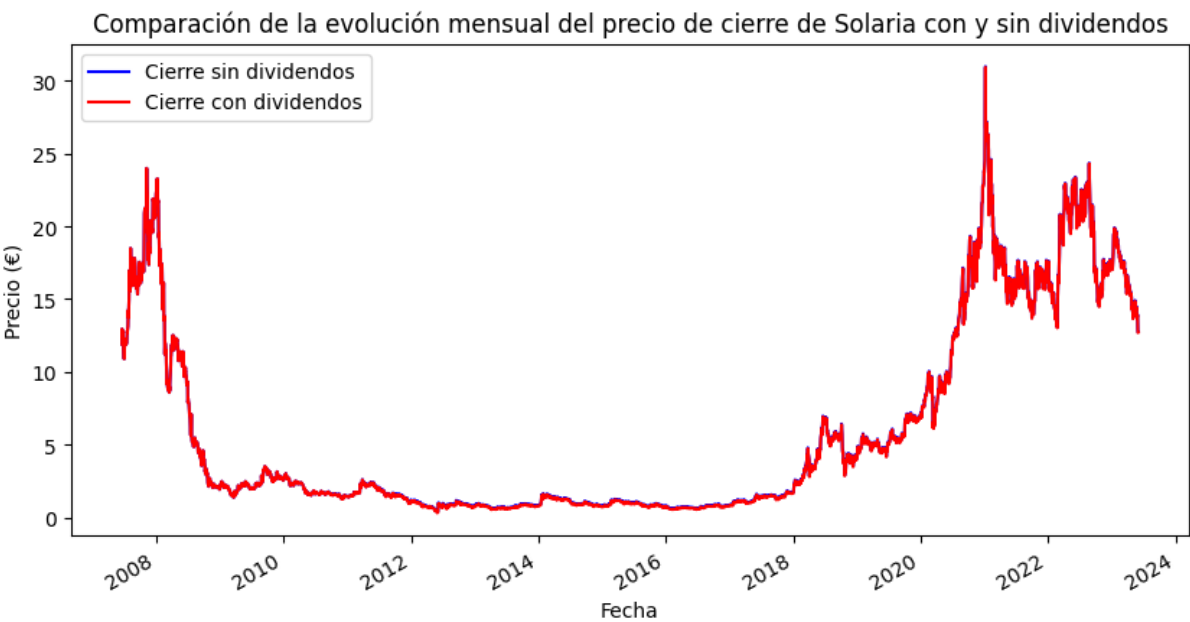


Figura 4.20: Comparación de la evolución mensual del precio de cierre de Solaria con y sin dividendos

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 se registra en la figura 4.21.

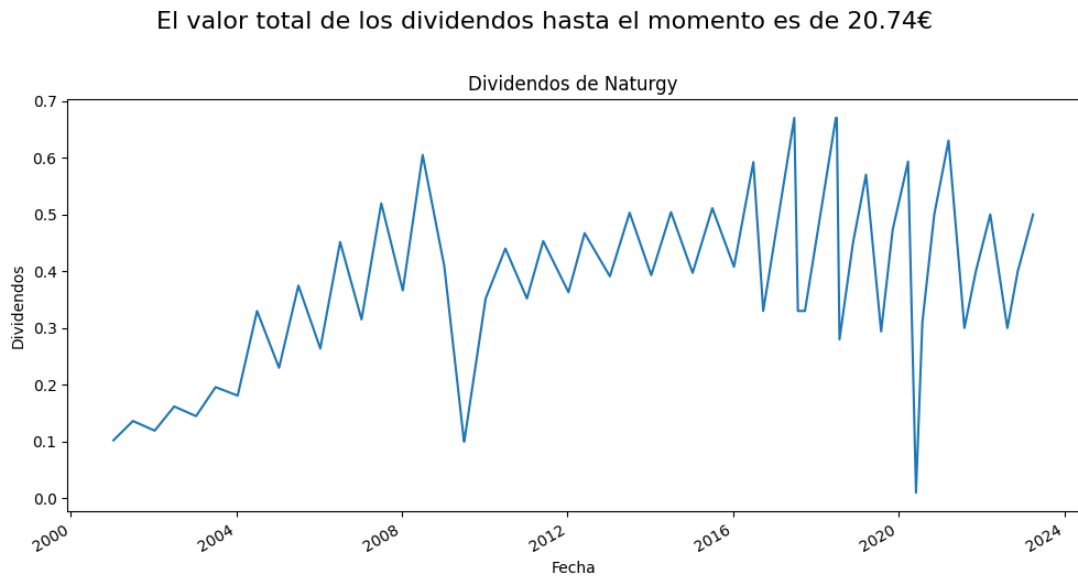


Figura 4.21: Evolución de los dividendos de Naturgy

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 4.22 se muestra el precio de la acción con y sin dividendos.

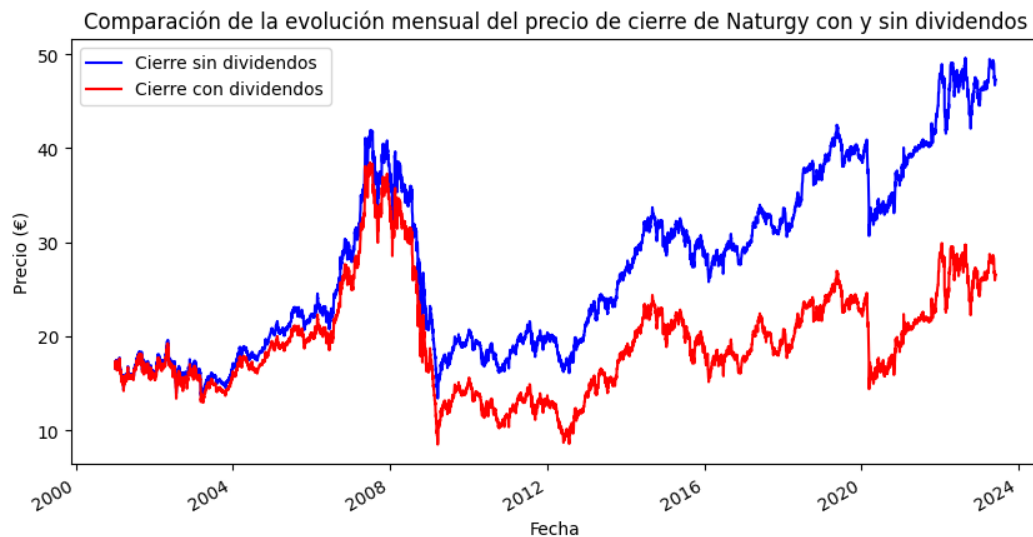


Figura 4.22: Comparación de la evolución mensual del precio de cierre de Iberdrola con y sin dividendos

4.2.5. Relación Entre Empresas

La matriz de correlación de las columnas *Close* de las empresas se presenta en la Tabla 4.15, mientras que la Figura 4.23 ofrece una representación visual de cómo las empresas se correlacionan entre sí en términos de rendimientos o precios.

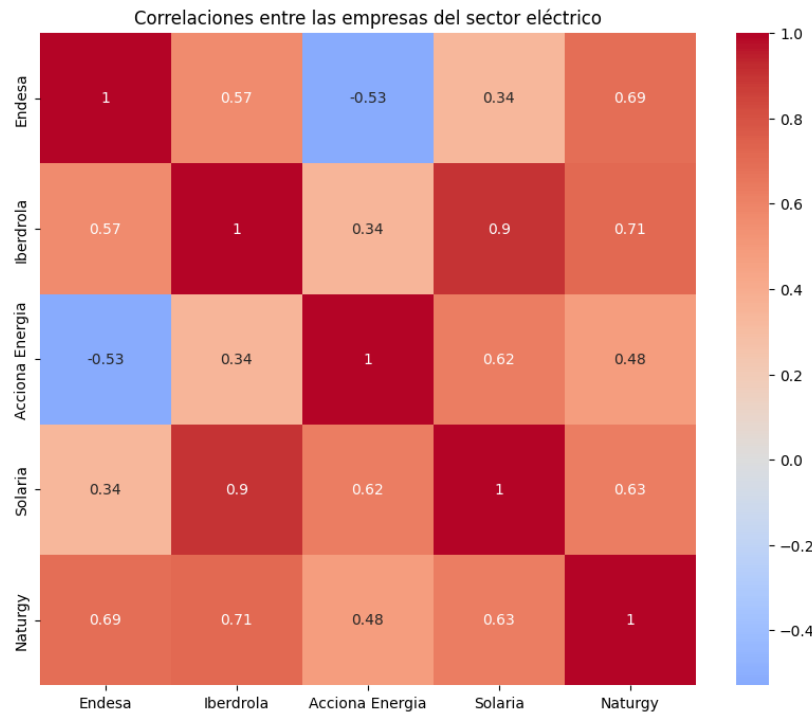


Figura 4.23: Correlaciones entre las empresas del sector eléctrico

	Endesa	Iberdrola	Acciona Energía	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 Energía	-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

Una vez explorados de los datos, se desarrollan y comparan 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*, tal y como se menciona en el apartado 2.2.4, es una técnica de pronóstico empleada en el análisis de series temporales para anticipar patrones y tendencias futuras. El procedimiento estructurado seguido para aplicar este algoritmo es:

1. Se han sumando los dividendos al precio de cierre de las acciones.
2. Se crean ventanas temporales para la validación cruzada, garantizando así una evaluación robusta del rendimiento del algoritmo.
3. Con el algoritmo compilado y entrenado, se generan predicciones para cada ventana temporal. En la Tabla 4.16 se detallan los resultados obtenidos para las 10 primeras ventanas.
4. Se visualizan la predicción asociada a cada ventana, como se muestra, a modo de ejemplo, en la Figura 4.24.

Ventana	0	1	2	3	4	5	6	7	8	9	10
MSE	2.14	14.15	92.80	26.21	51.78	17.22	6.31	25.77	2.80	9.19	50.76

Cuadro 4.16: MSE de las primeras 10 ventanas de Holt-Winters

4.3.2. XGBoost con Cross Validation

XGBoost, mencionado en la sección 2.2.3, es un algoritmo de aprendizaje automático utilizado para problemas de regresión y clasificación. La *Cross Validation* (validación Cruzada) es una técnica esencial para evaluar de manera robusta el rendimiento de un modelo. Divide el conjunto de datos en múltiples subconjuntos, entrena el modelo en algunos de ellos y evalúa el

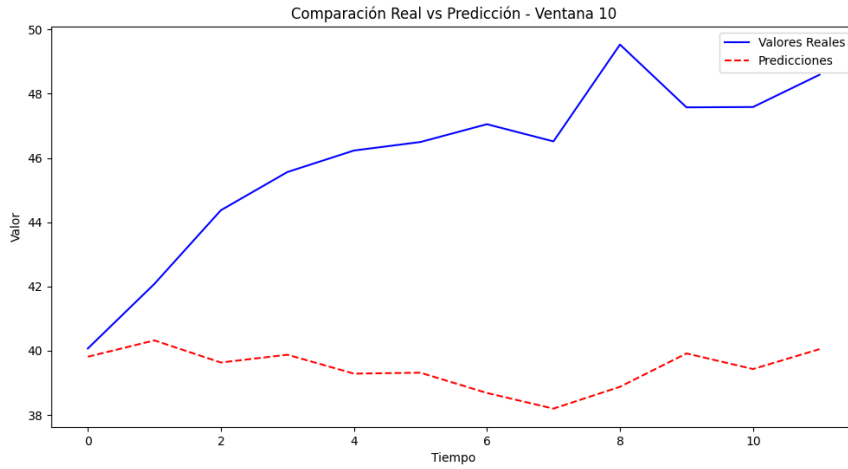


Figura 4.24: Ejemplo salida ventana de Holt-Winters

rendimiento en el subconjunto restante, repitiendo este proceso varias veces. El procedimiento estructurado seguido para aplicar este algoritmo:

1. Se realiza la preparación del dataframe, incorporando *lags* o pasos hacia atrás. La columna *Close* se mantuvo como la variable independiente central en este proceso.
2. Se procede a la creación de ventanas para la validación cruzada, garantizando así una evaluación robusta del rendimiento del algoritmo.
3. Se buscan de hiperparámetros, afinando y seleccionando las configuraciones óptimas para asegurar un desempeño óptimo del modelo. La Tabla 4.18 presenta un ejemplo de los resultados generados por el modelo en la búsqueda de hiperparámetros.
4. Se visualizaron de manera gráfica los hiperparámetros óptimos para cada intervalo de fechas. Para el periodo comprendido entre el 30 de septiembre de 2009 al 31 de agosto de 2010, los resultados a nivel gráfico se muestran en la figura 4.25.

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

Cuadro 4.17: Ejemplos resultados del modelo

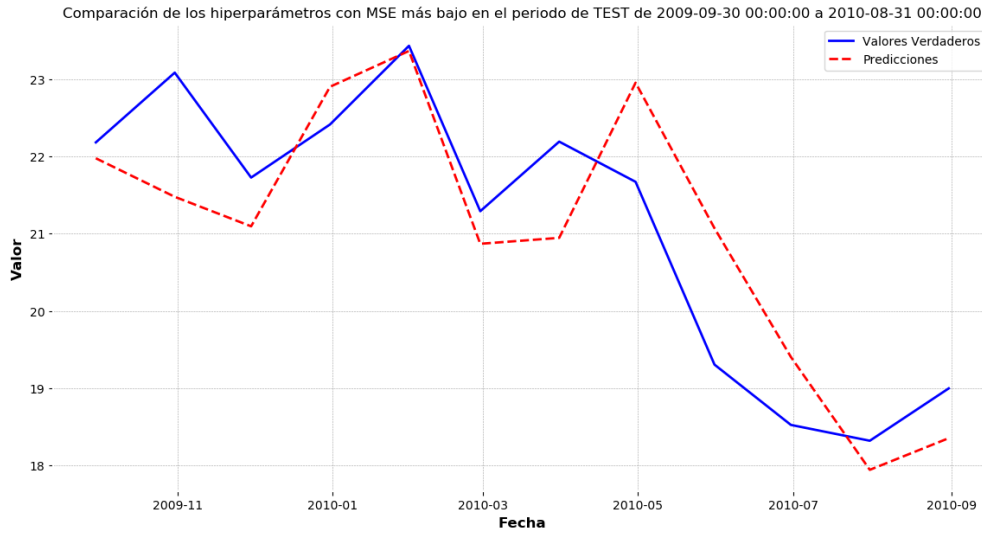


Figura 4.25: Comparación de los hiperparámetros con MSE más bajo en el periodo de TEST de 2009-09-30 a 2010-08-31

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

Una vez analizados los resultados, observamos 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

El enfoque se centra en la recopilación y evaluación de noticias para identificar con precisión aquellas que tienen un impacto significativo en las fluctuaciones de los precios de las acciones.

Tras aplicar diversas estrategias de aprendizaje profundo para prever el valor de las acciones, se concluye que es crucial incorporar información más allá del análisis histórico de precios. La necesidad de datos adicionales lleva a considerar las noticias como una fuente relevante.

En el análisis detallado, se observa que ciertas variaciones en los precios de las acciones están directamente influenciadas por eventos noticiosos significativos.

4.4.1. Justificación del uso de Web Scraping para obtener noticias

La elección de utilizar *Web Scraping* para obtener noticias en el ámbito financiero se basa en la necesidad de acceder a información relevante y oportuna que influye directamente en las dinámicas del mercado. Tras explorar y evaluar diversas API's, incluyendo servicios noticiosos como *GNews*, se observa que estas no proporcionan de manera eficiente la amplitud y velocidad requeridas para recopilar noticias financieras de manera integral. Las API's, a pesar de ofrecer datos estructurados, presentan limitaciones en términos de número de solicitudes diarias y en la actualización en tiempo real, comprometiendo la efectividad del análisis. Unidos a los cambios en las políticas de acceso gratuito y la implementación de tarifas en API's como *Twitter* que afectan a la accesibilidad de datos en tiempo real. En este contexto, el *Web Scraping* se presenta como una solución práctica y eficaz, que supera las limitaciones al extraer datos de diversas fuentes de manera flexible y sin restricciones impuestas por suscripciones o límites diarios de solicitudes.

4.4.2. Web Scraping

Se ha elegido la sección de noticias del sector eléctrico del periódico El País [2] como fuente para recopilar información. El resultado es un archivo que contiene todas las noticias obtenidas, listas para su análisis posterior. En total, se recopilaron 8305 noticias.

4.4.3. GPT para el análisis de noticias

Con GPT se han analizado las noticias, obteniendo como salida un fichero como el siguiente:

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

Listing 4.2: Salida de GPT

Donde:

- **Empresa:** Puede ser una de las siguientes opciones: Sector Energético, Iberdrola, Endesa, Naturgy, Acciona.
- **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 $[-1, 1]$, reflejando la dirección positiva o negativa de la noticia.

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

4.5. Retroalimentación de XGBoost

En esta sección, se reproduce el procedimiento previamente explicado, pero se incorpora el análisis de noticias.

Antes de integrar las noticias al nuevo modelo, se observa la presencia de diversas instancias de noticias registradas en distintas fechas. Para abordar esta variabilidad, se aplica un codificador *OneHot* a las columnas *empresa* y *tipo*.

Los resultados posteriores a la incorporación de las noticias se muestran en la tabla 4.19.

	Mediana MSE	Media MSE
Sin Noticias	3.4718	17.1820
Con Noticias	1.5926	3.85137

Cuadro 4.19: Comparación de errores MSE sin y con Noticias

Estos nuevos 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. La media del error MSE disminuye de 17.1820 a 3.85137, indicando una mejora promedio en la precisión del modelo. La introducción de noticias contribuye positivamente a la capacidad del modelo para generar respuestas más precisas.

4.6. Pruebas del nuevo modelo XGBoost

En esta sección, se comentan y analizan las conclusiones obtenidas después del empleo del nuevo modelo XGBoost, al que se le añadieron las noticias. Con este fin, se lleva a cabo el reentrenamiento de dicho modelo para extraer las *Feature Importances*.

4.6.1. Feature Importances con todas las variables

Se reentrena el nuevo modelo utilizando la totalidad de los datos de entrenamiento a excepción del último período (de mayo de 2022 a abril de 2023), el cual se reserva para la validación.

Posteriormente, se extrae las *Feature Importances* con el objetivo de comprender cuáles tipos de noticias, *lags* u otras variables tienen una mayor relevancia en el proceso de predicción. Se utilizan los hiperparámetros óptimos previamente obtenidos. En la imagen 4.26 se muestra el *Feature Importances*.

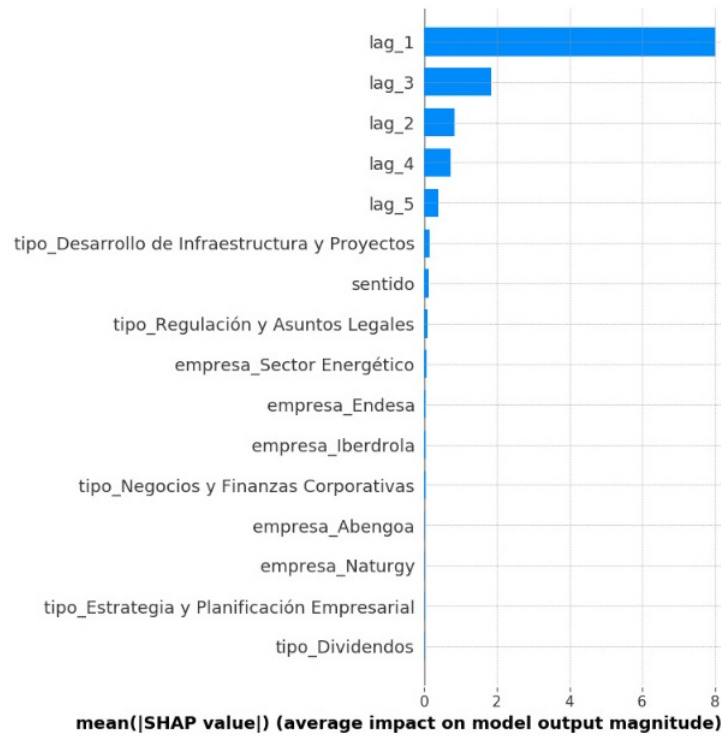


Figura 4.26: Feature Importances con todas las variables

A simple vista, es evidente que los *lags* juegan un papel fundamental. El precio de una acción no experimenta movimientos aleatorios, sino que se basa, en gran medida, en el precio de cierre previo.

4.6.2. Feature Importances con el sentido incorporado a los tipos de noticias y generalizando todas las empresas distintas a Endesa

Se reentrena el nuevo modelo y se incorpora el sentido dentro de las noticias. Las noticias relacionadas con empresas distintas de Endesa se agrupan en una única variable. La finalidad es simplificar la representación de noticias relacionadas con otras empresas, buscando mayor claridad y eficiencia en el análisis. El *Feature Importances* es la que muestra la imagen 4.27.

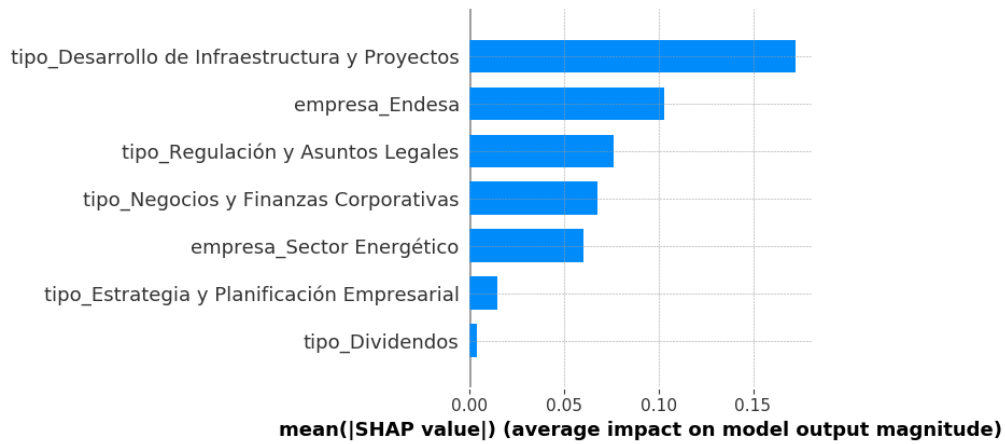


Figura 4.27: Feature Importances con las empresas agrupadas

4.7. Extrapolación al resto de empresas

Una vez efectuadas las pruebas y obtenidas las conclusiones, se realizan nuevos ensayos. En esta ocasión, se cargan los datos correspondientes al resto de empresas lo que permite identificar aquellas noticias que afectan en mayor o menor medida al precio de las acciones.

4.7.1. Iberdrola

Se ingresan los datos históricos de las acciones de Iberdrola y se han incorporan los correspondientes *lags*. Se sigue el mismo proceso de normalización obteniendo el *Feature Importances* de la imagen 4.28.

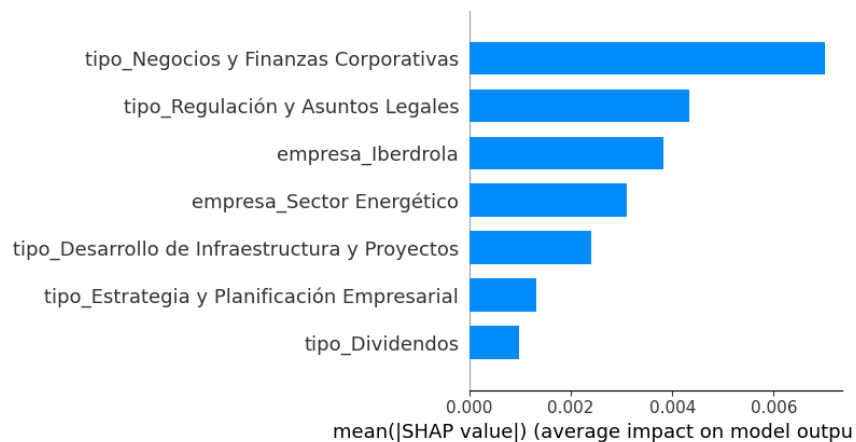


Figura 4.28: Feature Importances de Iberdrola

4.7.2. Acciona Energías

Se replica el mismo procedimiento mencionado obteniendo el *Feature Importances* de la imagen 4.29.

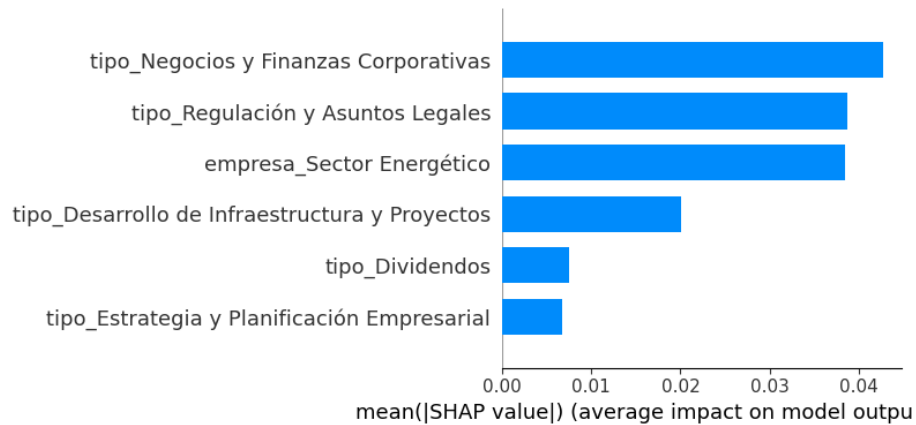


Figura 4.29: Feature Importances de Acciona Energías

4.7.3. Solaria

Se replica el mismo procedimiento mencionado obteniendo el *Feature Importances* de la imagen 4.30.

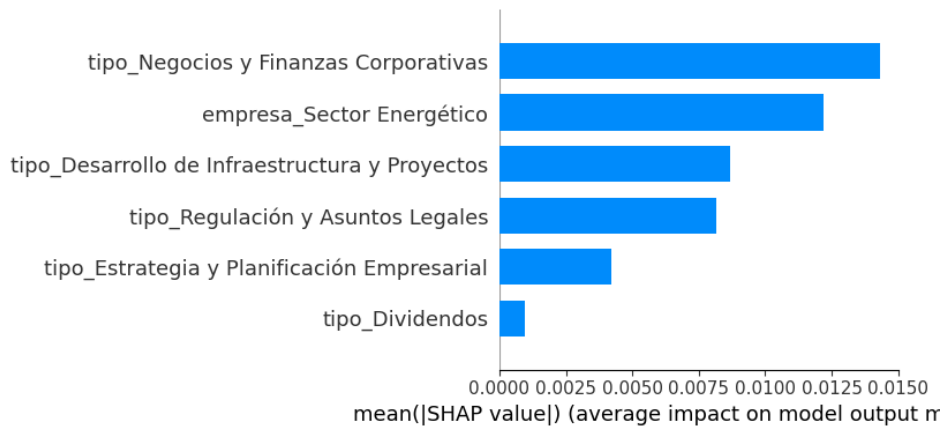


Figura 4.30: Feature Importances de Solaria

4.7.4. Naturgy

Se replica el mismo procedimiento mencionado obteniendo el *Feature Importances* de la imagen 4.31.

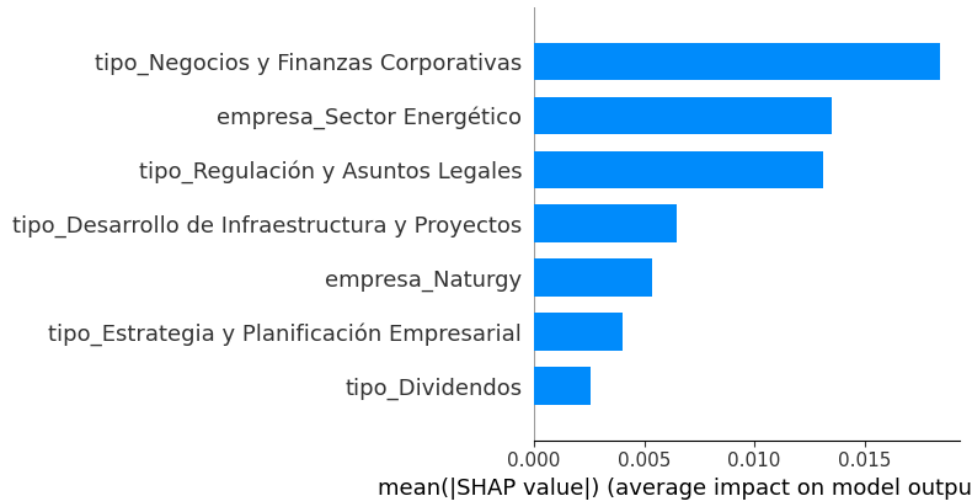


Figura 4.31: Feature Importances de Naturgy

4.8. Contrastación práctica del estudio

La validez del estudio se confirma mediante una verificación práctica. Se eligió un día específico en el cual se registraron más de dos noticias relevantes. Se examinó el tono de cada noticia (positiva o negativa) y se realizó un análisis detallado de los datos, incluyendo visualizaciones, para observar la reacción del precio de la acción ante estas noticias.

Después de filtrar las noticias relacionadas con la empresa Endesa y agruparlas por fecha, se decidió seleccionar un día al azar para llevar a cabo el análisis. Los resultados se presentan de manera detallada en la Tabla 4.20.

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 4.20: Datos de noticias de Endesa

Después de recopilar las noticias, se nota que estas reflejan una perspectiva sumamente positiva. Con el propósito de evaluar su impacto en Endesa, procederemos a examinar cómo se comportaron las acciones durante esos días en cuanto a su cotización, como se ilustra en la Figura 4.32.

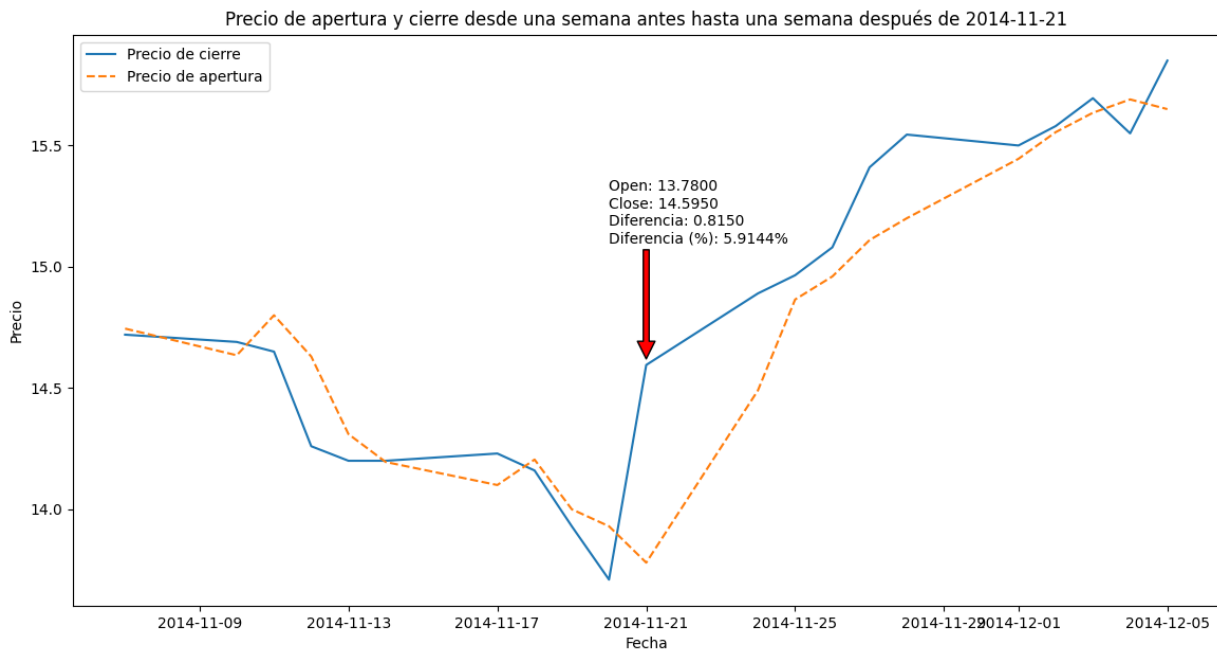


Figura 4.32: Impacto de las noticias

4.9. Dinámica del precio de las acciones después de la publicación de noticias

En nuestro estudio, hemos llevado a cabo un análisis exhaustivo explorando diversos modelos para comprender qué noticias influyen más en cada empresa. Después de identificar estas categorías clave, hemos avanzado a la siguiente fase: calcular de manera precisa el impacto de estas noticias en los precios de las acciones.

Para lograr esto, se ha implementado funciones que calculan el rendimiento de las acciones en periodos de 1, 2 y 5 días. Posteriormente, se ha realizado un filtrado por tipos de noticias, su orientación, y las empresas involucradas. Un ejemplo de salida de estas funciones se presenta en la Figura 4.33.

```
Una noticia del tipo Desarrollo de Infraestructura y Proyectos Positiva Endesa :

Hará subir las acciones un 68.75% de las veces en los próximos 1 días desde que se dio la noticia. Media de variación en subidas: 1.28%
Hará bajar las acciones un 25.00% de las veces en los próximos 1 días desde que se dio la noticia. Media de variación en bajadas: -0.70%
El precio se mantendrá igual un 6.25% de las veces en los próximos 1 días desde que se dio la noticia.
*****

Hará subir las acciones un 68.75% de las veces en los próximos 2 días desde que se dio la noticia. Media de variación en subidas: 1.48%
Hará bajar las acciones un 25.00% de las veces en los próximos 2 días desde que se dio la noticia. Media de variación en bajadas: -0.49%
El precio se mantendrá igual un 6.25% de las veces en los próximos 2 días desde que se dio la noticia.
*****

Hará subir las acciones un 68.75% de las veces en los próximos 5 días desde que se dio la noticia. Media de variación en subidas: 1.58%
Hará bajar las acciones un 25.00% de las veces en los próximos 5 días desde que se dio la noticia. Media de variación en bajadas: -2.12%
El precio se mantendrá igual un 6.25% de las veces en los próximos 5 días desde que se dio la noticia.
*****

Hará subir las acciones un 68.75% de las veces en los próximos 7 días desde que se dio la noticia. Media de variación en subidas: 2.09%
Hará bajar las acciones un 25.00% de las veces en los próximos 7 días desde que se dio la noticia. Media de variación en bajadas: -1.80%
El precio se mantendrá igual un 6.25% de las veces en los próximos 7 días desde que se dio la noticia.
*****
```

Figura 4.33: Impacto de un tipo de noticia

Capítulo 5

Conclusiones y líneas de trabajo futuras

En esta sección, se presenta las conclusiones de este estudio. Se analizan los resultados obtenidos sobre el impacto de las noticias en el mercado financiero.

Se determina 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 sugiere 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 observa que el precio de cotización se encuentra intrínsecamente ligado al precio de cierre del día anterior, como se evidenció en los *Feature Importances*.
- **Impacto Diferenciado en Endesa:** Las noticias sobre el desarrollo de infraestructura y proyectos tienen un impacto considerable en las acciones de Endesa. Las noticias positivas en este ámbito aumentan la probabilidad de que las acciones suban en los próximos 5 días en un 73.33 %, con un aumento promedio del 1.28 %. La probabilidad de que bajen es del

20.00 %, con una disminución promedio del -0.67 %, y hay un 6.67 % de probabilidad de que el precio se mantenga igual.

- **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.
- **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.

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. Trabajo futuro

En esta sección, se exploran las perspectivas y futuras líneas de trabajo, proporcionando un esbozo de las ideas a seguir. 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 desempeño 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] Cinco días: economía y mercados. <https://cincodias.elpais.com/>, 2023.
- [3] Blogthinkbig. ¿sabes en qué se diferencian las redes neuronales del deep learning? <https://blogthinkbig.com/redes-neuronales-deep-learning>, 2017.
- [4] Peter J Brockwell and Richard A Davis. *Introduction to Time Series and Forecasting*. Springer, 2016.
- [5] Novatos Trading Club. ¿qué es el trading? los 7 pasos para empezar. <https://www.novatostradingclub.com/que-es-el-trading/>, 2023.
- [6] Maria Coppola. Qué es la minería de datos: conceptos, técnicas y ejemplos. <https://blog.hubspot.es/marketing/mineria-datos>, 2023.
- [7] 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.
- [8] 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>, 2022.

-
- [9] Nerea Díaz. Mercado de valores: ¿qué es? definición y concepto. <https://economipedia.com/definiciones/mercado-de-valores-2.html>, 2023.
- [10] Economiasimple.net. Trading - qué es, definición y concepto. <https://economiasimple.net/glosario/trading>, 2023.
- [11] Economipedia. Qué es el trading y cómo funciona. <https://economipedia.com/definiciones/trading.html>, 2023.
- [12] EFE. Noticias economía de el mundo. <https://www.elmundo.es>, 2007.
- [13] Endesa. Informe anual 2020. <https://www.endesa.com/content/dam/enel-es/home/inversores/informe2020/ia-2020.pdf>, 2020.
- [14] 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.
- [15] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.
- [16] IAT. Deep learning, aprendizaje profundo y redes neuronales. <https://iat.es/tecnologias/inteligencia-artificial/deep-learning/>, 2018.
- [17] IBM. Ibm spss modeler - crisp-dm help overview. <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>, 2023.
- [18] 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.
- [19] Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Pearson, 2020.

-
- [20] 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.
- [21] Patricio Letelier. Metodologías ágiles para el desarrollo de software: Extreme programming. <http://www.cyta.com.ar/ta0502/v5n2a1.htm>, 2005.
- [22] Logicalis. Minería de datos: cómo funciona, elementos y requisitos. <https://www.logicalis.com/la/que-hacemos/data-and-ai/mineria-de-datos/>, 2023.
- [23] Macarena Salvador Maceira. Machine learning aplicado al trading. <https://repositorio.comillas.edu/rest/bitstreams/295750/retrieve>, 2019.
- [24] Babita Majhi, C.M. Anish, and Ritanjali Majhi. On development of novel hybrid and robust adaptive models for net asset value prediction. 2018.
- [25] OpenAI. Openai platform documentation. <https://platform.openai.com/docs/introduction>, 2023.
- [26] ArcGIS Pro. Cómo funciona xgboost. <https://pro.arcgis.com/es/pro-app/latest/tool-reference/geoai/how-xgboost-works.htm>.
- [27] Equipo Santander. ¿qué son los dividendos? <https://www.santander.com/es/stories/que-son-dividendos>, 2022.
- [28] Robert H Shumway and David S Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, 2017.
- [29] 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.

-
- [30] Equipo Skilling. Promedios móviles: ¿qué son en análisis técnico? <https://skilling.com/eu/es/blog/trading-articles/what-are-moving-averages/>, 2022.
- [31] Juan Chamorro Soler. Métodos para realizar un análisis de sentimiento. <https://isdfundacion.org/2023/02/03/metodos-para-realizar-un-analisis-de-sentimiento/>, 2023.
- [32] 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.
- [33] XGBoost. Xgboost documentation. <https://xgboost.readthedocs.io/en/stable/>.
- [34] 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.