



Facultad de Ciencias Económicas y Empresariales  
ICADE

**Evaluación del potencial de la  
Inteligencia Artificial en la predicción  
bursátil del sector energético del IBEX  
35**

Autor: Luis Albizu Arias  
Directora: Esther Vaquero Lafuente

## RESUMEN

El presente Trabajo de Fin de Grado analiza la capacidad y el potencial de diferentes modelos de Inteligencia Artificial en la predicción del comportamiento bursátil de las empresas del sector energético del IBEX 35 en los cinco días posteriores a la publicación de resultados trimestrales. El objetivo principal es concluir si diferentes técnicas de machine learning son capaces de anticipar movimientos en la cotización de las acciones e identificar las variables con mayor capacidad predictiva.

La metodología consiste en la construcción de una base de datos compuesta por 136 publicaciones de resultados, correspondientes a las compañías energéticas del IBEX 35 durante el período 2022-2026. Para ello se han considerado variables bursátiles, financieras, macroeconómicas y sectoriales, junto a eventos relevantes del sector energético. Posteriormente, se han desarrollado modelos predictivos mediante Orange Data Mining, utilizando algoritmos de Regresión Logística, Gradient Boosting y Random Forest.

Los resultados muestran que el modelo Random Forest, con una precisión del 66,9%, fue el modelo con mayor capacidad predictiva, claramente superior a la predicción aleatoria. Además, los modelos basados en árboles de decisión superaron a los modelos lineales, evidenciando la existencia de relaciones no lineales en los mercados financieros. Las variables más relevantes fueron la inflación, la sorpresa del beneficio por acción, la distancia a la media móvil de 200 sesiones, el volumen relativo y el momentum.

Las conclusiones reflejan el potencial de la Inteligencia Artificial como herramienta de apoyo en la predicción bursátil, pero también la necesidad de emplear variables relevantes y periodos temporales amplios para una mayor robustez de los resultados.

**PALABRAS CLAVE:** Inteligencia Artificial, Machine Learning, Predicción Bursátil, Sector Energético, IBEX 35, Publicación de Resultados, Random Forest.

## ABSTRACT

This Bachelor's Thesis analyses the capacity and potential of different Artificial Intelligence models in predicting the stock market behaviour of energy sector companies in the IBEX 35 within the five days following the publication of quarterly earnings. The main objective is to conclude whether different machine learning techniques can anticipate movements in stock prices and to identify the variables with the highest predictive capacity. The methodology consists of building a database comprising 136 earnings announcements corresponding to the energy companies of the IBEX 35 during the 2022-2026 period. To this end, stock market, financial, macroeconomic, and sectoral variables were considered, along with relevant events in the energy sector. Subsequently, predictive models were developed using Orange Data Mining, employing Logistic Regression, Gradient Boosting, and Random Forest algorithms.

The results show that the Random Forest model, with an accuracy of 66.9%, was the model with the highest predictive capacity, clearly outperforming random prediction. Furthermore, decision tree-based models outperformed linear models, evidencing the existence of non-linear relationships in financial markets. The most relevant variables were inflation, earnings per share (EPS) surprise, distance to the 200-day moving average, relative volume, and momentum. The conclusions reflect the potential of Artificial Intelligence as a support tool in stock market prediction but also highlight the need to employ relevant variables and broad timeframes to achieve greater robustness in the results.

**KEY WORDS:** Artificial Intelligence, Machine Learning, Stock Market Prediction, Energy Sector, IBEX 35, Earnings Announcements, Random Forest.

## ÍNDICE

<b>1. INTRODUCCIÓN.</b> .....	<b>4</b>
<b>2. MARCO TEÓRICO.</b> .....	<b>5</b>
2.1 Evolución de la Inteligencia Artificial en el sector financiero. ....	5
2.2 Modelos de machine learning empleados en el análisis de renta variable. ....	7
2.3 Revisión del sector energético en el IBEX 35. ....	9
<b>3. METODOLOGÍA.</b> .....	<b>11</b>
3.1 Recopilación de datos. ....	11
3.2 Justificación de la herramienta. ....	15
3.3 Transformación de datos. ....	16
3.4 Desarrollo del modelo predictivo. ....	18
<b>4. INTERPRETACIÓN DE RESULTADOS.</b> .....	<b>20</b>
4.1 Análisis exploratorio de los datos. ....	20
4.2 Interpretación del modelo. ....	26
<b>5. CONCLUSIONES.</b> .....	<b>28</b>
<b>6. DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA EN TRABAJOS DE FIN DE GRADO.</b> .....	<b>29</b>
<b>7. BIBLOGRAFÍA.</b> .....	<b>30</b>
<b>8. ANEXO.</b> .....	<b>37</b>
Anexo I. Desarrollo del modelo mediante Orange Data Mining .....	37
Anexo II. Configuración de los algoritmos de aprendizaje supervisado. ....	38
Anexo III. Análisis exploratorio de los datos. ....	40

## 1. INTRODUCCIÓN.

El presente Trabajo de Fin de Grado tiene por objetivo evaluar la capacidad de diferentes modelos de Inteligencia Artificial en la predicción del comportamiento de las acciones en los cinco días posteriores a la publicación de resultados. Adicionalmente, se pretende analizar la relevancia de distintas variables relacionadas, entre las que se incluyen la variación entre las expectativas y la realidad en la publicación de resultados y de beneficios por acción, rentabilidad del índice de referencia, variables macroeconómicas y el impacto de las noticias publicadas en el periodo analizado relativas al sector energético. Este trabajo aporta la capacidad predictiva de herramientas de Inteligencia Artificial en la predicción bursátil de un sector específico, evaluando sus limitaciones y potencial en la toma de decisiones de los inversores, ámbito en el que el aprendizaje automático ha ganado relevancia en la literatura reciente (Seow, 2025).

Aunque la Inteligencia Artificial no es un descubrimiento reciente, el aumento de la disponibilidad de datos y de nuevas técnicas de análisis han intensificado su uso en ámbitos económicos y empresariales. El sector financiero es uno de los principales campos de aplicación gracias a la capacidad de los modelos de aprendizaje automático de aprender patrones no lineales y relaciones entre variables difíciles de capturar mediante métodos tradicionales (Gu, Kelly y Xiu, 2020). Algoritmos como las redes neuronales se han aplicado con éxito a la predicción del comportamiento bursátil (Krauss, Do y Huck, 2017).

No obstante, la eficacia de la Inteligencia Artificial para anticipar el comportamiento de los mercados financieros debería ser limitada. Según la hipótesis de los mercados eficientes (Fama, 1970), los precios de los activos incorporan toda la información disponible, lo que impide obtener rentabilidades de manera sistemática. Es por ello que la capacidad de los modelos debería ser limitada, especialmente en intervalos temporales reducidos y ante eventos como la publicación de resultados.

A pesar de esto, a corto plazo pueden producirse ineficiencias en la interpretación de la información, o por la existencia de sesgos conductuales de los inversores (Kahneman y Tversky, 1979; Shiller, 2003). De esta manera, surge la siguiente pregunta: “¿Es capaz la Inteligencia Artificial de predecir el movimiento bursátil de las acciones?”

El análisis se centrará en el sector energético del IBEX 35. El motivo de la elección de este sector es, además de su relevancia en el mercado bursátil español<sup>1</sup>, su elevada volatilidad en los últimos años<sup>2</sup> debido a cuestiones macroeconómicas y regulatorias.

El periodo temporal del análisis abarca desde el 1 de enero de 2022 hasta el 16 de marzo de 2026. Se ha elegido este intervalo temporal debido al inicio del conflicto entre Rusia y Ucrania en febrero de 2022, el cual tuvo un gran impacto en la valoración de las acciones del sector energético español. El análisis finaliza en el primer trimestre de 2026 al ser la última fecha en la que se dispone de información sobre los precios de los activos.

## **2. MARCO TEÓRICO.**

### **2.1 Evolución de la Inteligencia Artificial en el sector financiero.**

La Inteligencia Artificial comienza formalmente en 1956, año en el que se celebró el Dartmouth Summer Research Project on Artificial Intelligence. El evento fue organizado por John McCarthy, que fue quien acuñó el término de Artificial Intelligence (Dartmouth College, 2026). En esa misma fecha, dos de los participantes de la conferencia, Allen Newell y Herbert Simon, desarrollaron *The Logic Theory Machine*, considerado el primer programa informático capaz de realizar razonamiento lógico automatizado (Newell & Simon, 1956).

Las primeras aplicaciones de la Inteligencia Artificial en el sector financiero surgieron en los años ochenta. En 1982, la empresa Apex lanzó al mercado PlanPower, un programa de inteligencia artificial que ofrecía asesoramiento fiscal y financiero a clientes con ingresos superiores a 75,000\$ (Ubuntu, 2023). De forma paralela, los mercados financieros comenzaron a implementar estrategias de arbitraje, con el objetivo de detectar discrepancias de precios. Inicialmente, estas estrategias estaban restringidas a los hedge funds y a las mesas de trading en los bancos de inversión. Poco después empezaron a introducirse sistemas de detección de fraude y algoritmos de machine learning en modelos de credit scoring, permitiendo la automatización del riesgo crediticio (Loanworks, s.f.).

---

<sup>1</sup> A 16 de marzo de 2026, las ocho compañías analizadas sumaban una capitalización de 245.572 millones de euros, equivalente al 30,6 % de la capitalización total del IBEX 35 (801.450 millones de euros)

<sup>2</sup> La volatilidad media anualizada de las cotizaciones de las compañías analizadas alcanzó el 25,2% durante el periodo de estudio

El comienzo del siglo XXI marcó el auge del aprendizaje automático. La expansión de Internet y el aumento de la capacidad computacional facilitaron la disponibilidad de grandes volúmenes de datos, permitiendo entrenar modelos de machine learning más robustos y precisos siendo empleados por bancos en la segmentación de clientes y la gestión del riesgo (Chicago Booth Review, 2024). Startups como Mint.com empezaron a ofrecer asesoramiento financiero personalizado usando herramientas de Inteligencia Artificial y machine learning (Loanworks, s.f.).

La adopción del deep learning supuso un hito significativo para el sector financiero en la década de 2010 (Loanworks, s.f.). El uso de este tipo de redes neuronales, junto al incremento del poder computacional y la disponibilidad de conjuntos de datos masivos permitió identificar patrones complejos en información financiera (Chicago Booth Review, 2024). Su aplicación en la predicción de series temporales o la detección de fraude sufrió un fuerte aumento (Loanworks, s.f.). Paralelamente, el auge de técnicas como los árboles de decisión, las redes neuronales y modelos híbridos han transformado la forma de gestión de carteras gracias a su capacidad predictiva y eficiencia en la selección de activos (Loanworks, s.f.).

Durante esos años, de forma paralela, surgieron los robo-advisors con plataformas como Betterment o Wealthfront (Redress Compliance, 2025). Estas plataformas eran capaces de construir y gestionar carteras automatizadas, ajustadas dinámicamente en función del horizonte temporal y aversión al riesgo del inversor. Su irrupción representó un punto de inflexión en la democratización de la inversión, anteriormente limitada a la actuación de los gestores tradicionales (Redress Compliance, 2025). De manera simultánea, las instituciones financieras comenzaron a integrar chatbots o automatización de procesos para aumentar su eficacia y reducir costes (Loanworks, s.f.).

El año 2017 supuso un punto de inflexión en el uso de la IA en el sector financiero. El lanzamiento del modelo Transformer por parte de Google permitió la integración del Natural Language Processing (NLP) en las finanzas (Sharkey & Treleaven, 2024). Desde entonces, modelos como BERT o GPT han permitido analizar grandes volúmenes de texto, como informes trimestrales o publicaciones en redes sociales, para entender el sentimiento del mercado y señales predictivas (Sharkey & Treleaven, 2024). Según la Chicago Booth Review, la NLP se considera como un subcampo clave en el mundo de

las finanzas, permitiendo a los analistas identificar patrones en textos que anteriormente requerían interpretación humana exhaustiva.

Durante la década de 2020, el crecimiento de la Inteligencia Artificial generativa ha permitido ampliar notablemente su aplicación financiera, siendo empleada en detección de fraude, gestión de carteras, análisis de riesgos sistémicos o cumplimiento normativo, entre otros (IMF, 2023). No obstante, la integración de estas herramientas también plantea riesgos regulatorios relacionados con la transparencia e interpretabilidad de los modelos (OECD, 2025).

En la actualidad, la expansión de la Inteligencia Artificial en el sector financiero está reflejada en el notable crecimiento del gasto global en estas tecnologías. Según datos del Fondo Monetario Internacional, la inversión mundial en software, hardware y sistemas de IA superará los 400.000 millones de dólares en 2027 (Kearns, 2023). En lo que al sector financiero respecta, su inversión en esa fecha se habrá duplicado, alcanzando los 97.000 millones de dólares (Kearns, 2023). Gestoras de activos como Amundi utilizan herramientas de Inteligencia Artificial para gestionar las carteras de algunos de sus más 100 millones de clientes (Kearns, 2023). Estas iniciativas muestran cómo la Inteligencia Artificial se ha convertido en un elemento estratégico dentro de los modelos de negocio financieros actuales (Kearns, 2023).

## **2.2 Modelos de machine learning empleados en el análisis de renta variable.**

El uso de técnicas de machine learning (ML) en el análisis predictivo de renta variable ha aumentado en los últimos años debido a una mayor disponibilidad de grandes volúmenes de datos financieros y avances computacionales que los permiten ser procesados de una manera más eficiente.

Tradicionalmente, el análisis bursátil se apoyaba principalmente en métodos estadísticos clásicos y en la interpretación humana respecto a los patrones de los precios. Sin embargo, estas aproximaciones representan importantes limitaciones en mercados caracterizados por relaciones no lineales entre valores, factores macroeconómicos que derivan en alta volatilidad y múltiples factores que, a pesar de no ser considerados, afectan a los precios de los activos (Saberironaghi, 2025). Incluso modelos multifactoriales como Fama-French o Carhart se basan en relaciones lineales, resultando inconsistentes para capturar tendencias complejas en los mercados actuales (Kelly & Xiu, 2023).

De esta manera, el machine learning sirve como herramienta para mejorar la capacidad predictiva en los precios respecto a los enfoques tradicionales. Estos modelos permiten aprender patrones ocultos en variables de mercado sin necesidad de imponer una estructura funcional fija. Estudios recientes demuestran que estos modelos son capaces de superar a los tradicionales en predicción de precios o volatilidad, especialmente en aquellos activos con comportamientos no lineales (El Amine El Amani et al, 2026).

Dentro de los enfoques supervisados clásicos aplicados a la predicción bursátil, la regresión lineal ha sido una de las más utilizadas gracias a que permite relacionar variables financieras mediante una estructura simple. Sin embargo, su capacidad para capturar patrones no lineales es limitada, por lo que se han desarrollado variantes regularizadas, como Ridge y Lasso, que mejoran la estabilidad del modelo ante problemas de multicolinealidad y sobreajuste, especialmente frecuentes en datos financieros con alta correlación entre variables (BSIC, 2025).

Otra técnica ampliamente utilizada son los árboles de decisión, que dividen el espacio en reglas jerárquicas, permitiendo capturar relaciones no lineales entre variables. A pesar de poder resultar inestables, técnicas de ensamble como Random Forest, que combina múltiples árboles entrenados sobre subconjuntos aleatorios de datos y variables, permiten aumentar la robustez y reducir la varianza del modelo. Estudios afirman que este último ofrece un rendimiento superior a modelos individuales en tareas de clasificación y regresión financiera (Yadav & Shailaja, 2024).

Otros métodos como Gradient Boosting, y especialmente XGBoost, permiten una elevada precisión en contextos con estructuras complejas. XGBoost construye árboles secuencialmente, corrigiendo los errores de modelos previos, y demostrando resultados muy positivos en predicción de precios y rentabilidades en comparación con otros algoritmos supervisados (Qian, 2025).

Recientemente, la aplicación de modelos basados en Transformers, especialmente en tareas relacionadas con el análisis de texto, se ha incrementado en el análisis financiero. Estos modelos permiten extraer información de noticias o informes, con el objetivo de analizar indicadores de sentimiento que ayuden en la predicción bursátil, obteniendo factores cualitativos que influyen en los precios. Entre los modelos más destacados se encuentran GPT-4 o BERT, cuya variante especializada en finanzas, FinBERT, ha

demostrado una alta precisión a la hora de clasificar el sentimiento de textos financieros e identificar patrones lingüísticos comunes en el mundo financiero (Kang & Choi, 2025).

A pesar de su auge en los mercados financieros, los modelos de machine learning siguen presentando limitaciones en su aplicación a la renta variable. Una de ellas es el sobreajuste, que se produce cuando el modelo aprende patrones del pasado que no se mantienen en datos nuevos, reduciendo su eficacia predictiva (Saberironaghi et al., 2025). Diversos estudios recientes destacan que, incluso modelos como Random Forest o XGBoost, pueden perder efectividad cuando analizan datos con relaciones no lineales que son inconsistentes (Yadav & Shailaja, 2024; Qian, 2025). Otra limitación es la interpretabilidad, ya que los modelos mencionados, que a su vez suelen ser los más precisos, fallan a la hora de identificar las variables que explican la predicción.

### **2.3 Revisión del sector energético en el IBEX 35.**

El sector energético cuenta con un papel central en la economía española actual y en el comportamiento del IBEX 35. Aunque BME no publique la ponderación por sectores, la importancia del sector es evidente. Iberdrola, que en 2026 alcanzó los 125.000 millones de capitalización bursátil, únicamente es superada por Inditex y Santander dentro del índice (Estrategias de Inversión, 2026). Este peso explica en gran parte la relevancia en el IBEX 35 del sector, compuesto por otras empresas como Endesa, Naturgy o Repsol entre otras, todas ellas con presencia estable. Adicionalmente, el índice sectorial IBEX 35 Energy ha ofrecido una revalorización del 53,83% desde comienzos de 2025, mostrando el potencial de crecimiento que ofrece el sector (FactSet, 2026).

Gráfico 1. Evolución del Ibex 35 Energy (01/01/2025 – 20/02/2026)



Fuente: Factset (2026)

El sector energético del IBEX 35, a pesar de que las empresas que lo componen se han mantenido en el periodo analizado, es extremadamente sensible a factores externos como los precios de las materias primas, la regulación y el entorno macroeconómico. Las principales compañías del sector (Iberdrola, Naturgy, Endesa o Repsol) desarrollan actividades vinculadas a la generación, distribución o comercialización de la energía, lo que hace que sus resultados financieros dependan directamente de la evolución de los precios de estas materias primas y de la obtención de márgenes. En los últimos años, el sector energético español ha intensificado su apuesta por las energías renovables, hasta el punto de que en 2024 más del 50% de la electricidad generada a nivel nacional procedió de fuentes renovables, principalmente eólica y solar (OBS Business School, 2025). La transición energética y la incertidumbre hacia los precios de la energía han hecho incrementar la atención del mercado hacia el sector, teniendo sus planes estratégicos y de inversión un fuerte impacto en la cotización bursátil de estas empresas (Estrategias de Inversión, 2024). Además, la presentación de resultados trimestrales es especialmente importante al servir como medidor de capacidad de adaptación a las nuevas tendencias. De esta manera, el sector supone un contexto idóneo para analizar el sentimiento del mercado y la información financiera mediante técnicas de machine learning.

En términos de composición bursátil, el sector energético del IBEX 35 está compuesto por compañías con modelos de negocio heterogéneos, pero ampliamente expuestos a cambios regulatorios, decisiones gubernamentales y el entorno macroeconómico.

Además, cuentan con perfiles distintos de riesgos: mientras Iberdrola o Endesa operan como utilities reguladas y tienen ingresos más estables, Repsol depende directamente del petróleo y del gas, llevando a una mayor volatilidad de sus resultados (elEconomista, 2025). Estas diferencias entre compañías hacen del sector un escenario idóneo para valorar la capacidad predictiva de los modelos de machine learning, al estar sujetas sus cotizaciones tanto a factores exógenos como a factores endógenos.

En consecuencia, el sector energético del IBEX 35 cuenta con una combinación única de relevancia económica, sensibilidad a factores externos y abundancia de información financiera pública. Estos factores justifican su elección para este Trabajo de Fin de Grado, con el objetivo de analizar si los modelos de machine learning son capaces de anticipar movimientos bursátiles tras presentación de resultados trimestrales y valorar el sentimiento del mercado.

### **3. METODOLOGÍA.**

#### **3.1 Recopilación de datos.**

Para la elaboración de este Trabajo de Fin de Grado se ha recopilado información relativa a las compañías del sector energético que han formado parte del IBEX 35 durante el período de estudio (Bolsas y Mercados Españoles, 2026). Las compañías analizadas son:

- Acciona, S.A.
- Endesa, S.A.
- Enagás, S.A.
- Iberdrola, S.A.
- Naturgy Energy Group, S.A.
- Redeia Corporación, S.A.
- Repsol, S.A.
- Solaria Energía y Medio Ambiente, S.A.

La incorporación de Acciona, S.A. al conjunto analizado requiere una justificación específica, al ser una compañía que, además de operar en el sector energético, también lo hace en los negocios de las infraestructuras, el agua y el sector inmobiliario. Su inclusión se fundamenta en su posición de liderazgo en el sector de las energías renovables, siendo el grupo Acciona Energía una de las mayores compañías de generación de energía

renovable a nivel internacional (Acciona Energía, 2024). No obstante, la variación en su cotización puede estar sujeta a sus negocios ajenos a la energía.

La selección de estas compañías se debe a tres criterios principales. El primero es su permanencia ininterrumpida en el IBEX 35 durante el periodo temporal analizado. El segundo es su elevada liquidez en el mercado, reduciendo el potencial ruido que afecta especialmente a operaciones de volumen más reducido (Amihud, 2002). Por último, la disponibilidad pública de su información trimestral es una condición necesaria para la construcción de las variables del modelo predictivo.

### **Datos bursátiles diarios**

La cotización diaria de las acciones ha sido recopilada desde el 1 de enero de 2022 hasta el 16 de marzo de 2026. El rango temporal cubre los ejercicios de 2022, 2023, 2024, 2025 y el primer trimestre de 2026. Este periodo permite disponer de un número razonable de presentaciones de resultados (17 por compañía, 136 eventos<sup>3</sup>). Este periodo se caracteriza por una alta volatilidad del sector energético, resultante de factores macroeconómicos como la crisis energética europea de 2022, fluctuaciones en los precios del petróleo y el gas, y una posterior normalización del mercado energético desde 2022 hasta 2026.

El conjunto de datos contiene las siguientes variables diarias:

- Date: fecha de cotización.
- Empresa: denominación social de la compañía.
- Código: variable obtenida de concatenar las columnas de empresa y date, cuyo objetivo es la identificación de los días para cada empresa correspondientes a la presentación de resultados trimestrales.
- Price: precio de cierre de la acción, medido en euros al ser la moneda local.
- Change: variación absoluta del precio de cierre respecto al día anterior.
- %Change: variación del precio respecto al día anterior, expresado en porcentaje.
- CVol: número de títulos negociados durante el día.
- Open: precio de apertura.
- High: precio máximo alcanzado durante la sesión.

---

<sup>3</sup> Se define como evento el día oficial de la publicación de los resultados trimestrales de las empresas de la muestra

- Low: precio mínimo negociado durante la sesión.

Estos indicadores permiten ver tanto la dinámica intradía del precio de las acciones, como sus tendencias a corto y medio plazo. La fuente empleada para la obtención de los datos ha sido Factset.

### **Datos financieros trimestrales**

Por otra parte, la información asociada a los datos financieros trimestrales se ha extraído del *Earnings Calendar* de Investing.com. Esta herramienta aporta información relativa al Beneficio por Acción (BPA) e ingresos, tanto reales como esperados según el consenso de analistas de ambas métricas. La extracción de la información se realizó según el siguiente procedimiento:

1. Selección del país: España
2. Selección de importancia: Alta, para filtrar únicamente por empresas incluidas en el IBEX 35.
3. Sector: Se filtró por compañías pertenecientes a los sectores de Energía y Suministros, para conseguir las compañías previamente acordadas.
4. Rango temporal: Rango personalizado que comprendía desde el 1 de enero de 2022 hasta el 16 de marzo de 2026, ambos incluidos.

El resultado de este procedimiento se encuentra en la pestaña “Resultados trimestrales”. A partir de esta pestaña se identifican, dentro de la pestaña “PRICE HISTORY IBEX 35”, las fechas en las que cada compañía publicó resultados trimestrales, así como variaciones en el precio de la acción en los días 1, 3 y 5 posteriores a la publicación.

### **Variables de mercado, macroeconómicas y sectoriales**

Para que el modelo pueda distinguir entre reacciones del precio atribuibles a la presentación de resultados o por movimientos del mercado o del sector, el conjunto de datos final incluye cuatro nuevas categorías de variables: un benchmark sectorial, variables macroeconómicas, materias primas relacionadas y un registro de eventos sectoriales.

**Benchmark sectorial:** Se ha seleccionado la cotización diaria del índice IBEX 35 Energy (IBEXIE), compuesto por los precios de las empresas del sector energético del IBEX 35.

El motivo de su incorporación es disponer de una referencia sectorial que permita comparar el comportamiento de los modelos desarrollados. También se busca obtener si la capacidad predictiva responde a patrones de las compañías analizadas o a una tendencia del sector. La serie se ha obtenido de Investing.

**Rentabilidad del bono español a 10 años:** Supone la primera de las variables macroeconómicas, y el motivo de su inclusión ha sido representar las condiciones financieras del mercado. Se busca incluir el efecto de los tipos de interés y la percepción del riesgo del mercado. Además, la rentabilidad libre de riesgo también juega un papel en la valoración de los activos. La información se ha obtenido de Factset.

**Euribor a doce meses:** se ha incorporado complementando la rentabilidad del bono español a 10 años, como indicador del coste de financiación. La serie diaria se ha obtenido del Banco de España.

**Inflación interanual de España:** se ha incluido la variación del Índice de Precios de Consumo (IPC), publicada mensualmente por el Instituto Nacional de Estadística (INE).

**Precio de materias primas relacionadas:** A pesar de que todas las compañías pertenecen al sector, la heterogeneidad entre ellas hace que la cotización de cada compañía se mueva por diferentes circunstancias. Es por ello que se han incluido las siguientes materias primas:

- Petróleo Brent, medido en dólares por barril, sirve como índice global de referencia del precio del crudo. Su variación tiene un impacto en la cotización de Repsol.
- Gas natural TTF, medido en euros por megavatio-hora, es el principal indicador del precio del gas natural. Tiene una gran influencia en el precio de las acciones de Naturgy y Enagás.
- Precio del mercado eléctrico mayorista español (OMIE), en euros por megavatio-hora, corresponde al precio medio diario del pool eléctrico ibérico. Esta variable es relevante en los precios de Iberdrola, Endesa, Acciona y Solaria.

Las series del petróleo Brent y el gas natural TTF se han obtenido de Factset, mientras que el precio del mercado eléctrico mayorista español se ha obtenido de la información histórica publicada por el Operador del Mercado Ibérico de Energía (OMIE).

**Eventos sectoriales:** con el objetivo de capturar sucesos relevantes en la cotización del sector energético que afectaban de manera heterogénea a las empresas, se han registrado los principales acontecimientos en el intervalo analizado. Se han catalogado dichos sucesos como geopolíticos, fiscales, macroeconómicos, regulatorios o sectoriales, y se ha mencionado tanto fecha de inicio como compañías principalmente afectadas. La identificación de estos sucesos se ha hecho a través de fuentes como Cinco Días, Expansión o Bolsamanía.

### **3.2 Justificación de la herramienta.**

Tras la recopilación de datos, se ha decidido que la herramienta utilizada para la predicción de la variación bursátil será la plataforma Orange Data Mining. El motivo del uso de esta aplicación responde a tres principales criterios.

En primer lugar, destaca la transparencia metodológica que ofrece. El funcionamiento de la aplicación es un flujo de trabajo visual, en el que se conectan widgets de manera consecutiva. El resultado es un diagrama legible y fácilmente interpretable, sin necesidad de la revisión de un código fuente complejo, lo cual resulta valioso en un trabajo académico.

El segundo motivo es la facilidad visual a la hora de comparar modelos. Al poder ejecutarse varios algoritmos en un único entorno y sobre un conjunto único de datos, la comparación de resultados es mucho más sencilla gracias al widget “Test and Score”. De esta manera, los modelos se pueden comparar de manera más clara y objetiva, y la elección de un único modelo está más justificada.

Por último, a diferencia de un entorno de programación como Python, en el que la construcción de un modelo requiere de conocimiento de bibliotecas específicas y de codificación, Orange Data Mining permite desarrollar modelos de manera más simple e intuitiva. Dado que el objetivo del presente trabajo es evaluar el potencial de la Inteligencia Artificial en el análisis financiero y no demostrar conocimientos avanzados de programación, este enfoque responde mejor a la naturaleza del estudio.

No obstante, algunas visualizaciones complementarias al análisis exploratorio de los datos se han elaborado con herramientas externas, al no estar disponibles en la aplicación.

### 3.3 Transformación de datos.

El conjunto de datos recogidos en el apartado 3.1, consolidado en un único archivo Excel, no puede utilizarse como entrada del modelo predictivo. La información se encuentra en frecuencias temporales distintas (trimestral para la presentación de resultados financieros, puntual para los eventos sectoriales y diaria para el resto de variables) y con formatos y cabeceras heterogéneos. Por este motivo, se han desarrollado dos archivos en formato CSV con diferentes objetivos. Se ha dividido el proceso en cuatro pasos:

1. El primer paso ha sido la homogeneización del contenido de las diferentes hojas. Lo primero ha sido ajustar las cabeceras y eliminar las filas iniciales con datos aportados por las distintas fuentes. Después, se ha unificado el formato de las fechas, siendo el resultante YYYY-MM-DD, y se han convertido a formato numérico las columnas que figuraban como texto.
2. Sobre la cotización diaria de las ocho compañías como tabla base se han incorporado, mediante operaciones de unión por fecha, las series del IBEX 35 Energy, las materias primas energéticas (Brent, TTF, OMIE) y las variables macroeconómicas (TIR del bono español a diez años, Euribor a doce meses e inflación interanual). Además, a las variables que no son macroeconómicas se ha calculado su retorno diario. Las jornadas en las que alguna serie externa no presenta cotización por festividad local se han completado mediante el criterio forward fill sobre el último valor disponible, conforme a la práctica habitual en estudios de microestructura de mercado, introduciendo el último valor disponible para evitar sesgo de anticipación. A esta tabla unificada se han añadido los siguientes indicadores técnicos:
  - Beta\_250d: mide cuánto se mueve la acción cuando lo hace su índice de referencia, en este caso, IBEX 35 Energy. Se ha calculado como la covarianza entre la acción y el benchmark entre la varianza del benchmark. Se exigió un mínimo de 60 sesiones para una estimación más robusta.
  - Momentum: se ha obtenido a 5, 20 y 60 días, e indica la rentabilidad acumulada de la acción en los días correspondientes anteriores. Es una variable importante al medir si la acción, en el momento del anuncio, se encuentra en tendencia alcista o bajista. Las primeras N sesiones de cada compañía salen vacías al no haber datos para comparar.
  - Vol\_20d: Mide cuánto ha fluctuado la acción en un periodo de 20 días, calculado como la desviación típica de los retornos diarios de las acciones en ese intervalo.

- Vol\_rel: Indica si la acción se ha negociado en mayor o menor medida en esa sesión respecto a la media de las 20 sesiones anteriores. Si es mayor que 1 se negoció más de lo habitual, y si es menor a 1 lo hizo en menor medida. Es importante para detectar interés por parte del mercado en torno a los anuncios.
- Dist\_ma\_200d: Calculado como el precio de la sesión entre la media de los precios de las 200 sesiones anteriores, indica cuánto se aleja el precio de su tendencia a largo plazo.
- Evento\_activo: Mide si hay algún evento activo de manera dicotómica, siendo 1 cuando existe este evento y 0 cuando no lo hay.
- Evento\_tipo: Variable categórica que distingue el tipo de evento activo, catalogándose como “ninguno” cuando no haya ningún evento en ese momento.
- Dia\_anuncio: Variable dicotómica, valiendo 1 si la compañía publicó resultados trimestrales ese día y 0 si no lo hizo.

El resultado de esta fase es el archivo panel\_diario\_orange.csv

3. A partir del panel diario y de la lista de los 136 anuncios trimestrales, se ha generado una tabla en la que cada fila representa un anuncio individual. Para cada uno se localiza el primer día bursátil posterior o coincidente con la fecha de publicación (denominado  $D_0$ ) y, sobre esa observación, se calculan dos bloques de variables: por un lado, los retornos del precio en las sesiones  $D+1$ ,  $D+3$  y  $D+5$ , junto con los retornos del benchmark sectorial en las mismas ventanas, que permiten construir las variables dependientes del modelo; por otro lado, el conjunto completo de indicadores técnicos, sectoriales y macroeconómicos calculados en  $D_0$ , que constituyen las variables explicativas. El resultado de esta fase es el archivo “eventos\_orange.csv”.
4. Las sorpresas porcentuales de beneficio por acción y de ingresos <sup>4</sup>se han calculado como la desviación entre el valor real y previsto medido en porcentaje, y se han acotado al intervalo  $[-3, +3]$  para evitar la influencia de un anuncio atípico (Solaria, 30 de septiembre de 2025, con una sorpresa de BPA del +1.346 %) que distorsionaba el modelo. Los valores ausentes en variables continuas (sorpresas no calculables por falta de previsión de consenso, betas iniciales con histórico insuficiente, etcétera) se han imputado mediante la mediana de la muestra. Las variables objetivo categóricas

---

<sup>4</sup> Se define como sorpresa porcentual la diferencia, en porcentaje, entre los rendimientos esperados por los analistas y los publicados por las empresas en sus resultados trimestrales

(signo del retorno) se han codificado con las etiquetas "sube" y "baja" para facilitar su lectura en Orange Data Mining.

El resultado final son dos archivos de Excel transformables a CSV y con tipos de datos homogéneos, directamente cargables en cualquier herramienta de análisis estadístico o de machine learning.

### **3.4 Desarrollo del modelo predictivo.**

Una vez construido el dataset de eventos (eventos\_orange.csv), se procede a la elaboración del modelo en Orange Data Mining (véase Anexo I). La plataforma se organiza mediante widgets conectados sucesivamente, de manera que cada uno procesa la información del anterior. El flujo desarrollado se estructura en cinco bloques principales:

1. Carga de los datos: Mediante el widget File se carga el archivo “eventos\_orange.csv”, que contiene 136 filas, correspondientes al número de presentaciones de resultados trimestrales de las ocho compañías del estudio. Dentro de su configuración, se asigna el tipo y el rol a cada una de las variables. Las variables “fecha\_anuncio” y “empresa” se definen como meta, para que su valor no influya en el entrenamiento, pero sí tengan funciones de trazabilidad. El conjunto de variables explicativas se define como feature y la variable que se busca predecir (sig\_ret\_d5\_cat) como target.  
Este widget se conecta con el widget de “Data Table”, con el objetivo de visualizar que la asignación de tipos y roles para las variables es correcta.
2. Al haber un elevado número de variables en comparación con el tamaño muestral de 136 observaciones, se conecta el widget “File” con el widget “Rank”. El objetivo de este componente es ordenar las variables en función de su capacidad predictiva sobre la variable objetivo, usando para ello las métricas de Information Gain Ratio y Gini Decrease. A partir de aquí, descartamos todas las variables salvo las cinco más relevantes según estos indicadores: “inflacion”, “bpa\_surprise”, “vol\_rel”, “dist\_ma200” y “mom\_5d”. Este proceso de reducción de dimensionalidad es importante para reducir el riesgo por sobreajuste, obteniendo después de este proceso modelos más precisos sobre el conjunto de validación.
3. Las variables seleccionadas, a través del widget “Select Columns” se conectan con tres algoritmos de aprendizaje supervisado con el fin de comparar su rendimiento y medir su capacidad predictiva.

- Gradient Boosting: equivalente al algoritmo XGBoost, construye los árboles de forma secuencial, corrigiendo cada árbol los errores de los anteriores en su conjunto. Se ha configurado con 200 árboles, una tasa de aprendizaje de 0,1 y una profundidad máxima de tres niveles. Mientras que la limitación en la profundidad cumple la misma función que en el algoritmo Random Forest, la tasa de aprendizaje hace que cada árbol aporte de forma moderada al modelo final (véase Anexo II.1).
- Random Forest: combina múltiples árboles de decisión entrenados con submuestras aleatorias de los datos, obteniendo un resultado según voto mayoritario del conjunto. Se ha configurado con un total de 200 árboles y una profundidad máxima de tres niveles, restringiendo la complejidad de cada árbol para evitar que aprendan el ruido de los datos de entrenamiento en vez de patrones generalizables (véase Anexo II.2).
- Logistic Regression: incluido como modelo lineal para comparar su rendimiento respecto a los árboles de decisión. Estima la probabilidad de que la observación pertenezca a cada clase aplicando una función logística sobre una combinación lineal de las variables explicativas (véase Anexo II.3).

4. La evaluación se realiza mediante el widget “Test and Score”, al que se conectan los algoritmos en calidad de learner y los datos del widget “Select Columns”, como conjunto al que entrenar y validar. El resultado es una tabla comparativa entre los tres algoritmos según seis métricas de evaluación:

- AUC (área bajo la curva ROC): mide la capacidad del modelo en cuestión para separar las dos clases con independencia del umbral de decisión.
- CA: precisión de clasificación, mide la proporción de predicciones acertadas sobre el total.
- F1: media armónica entre la precisión y la exhaustividad.
- Prec: de todas las observaciones clasificadas por los modelos como “sube”, en qué proporción acertaron.
- Recall: la proporción de subidas totales detectadas por cada modelo.
- MCC: coeficiente de correlación de Matthews, resume la matriz de confusión en un valor entre -1 y 1.

5. Finalmente, al widget “Test and Score” se conectan tres componentes de evaluación.

- Confusion Matrix: tabla que muestra los aciertos y errores desglosados por clase (verdadero positivo, verdadero negativo, falso positivo y falso negativo). Permite identificar si el modelo predice mejor las bajadas o las subidas.
- ROC Analysis: representa gráficamente la curva ROC de cada modelo y permite comparar su capacidad discriminatoria, es decir, la capacidad de los modelos de acertar en mayor o menor medida que el azar.
- Predictions: muestra, anuncio a anuncio, la predicción asignada para cada modelo, facilitando el análisis individual de los datos.

## 4. INTERPRETACIÓN DE RESULTADOS.

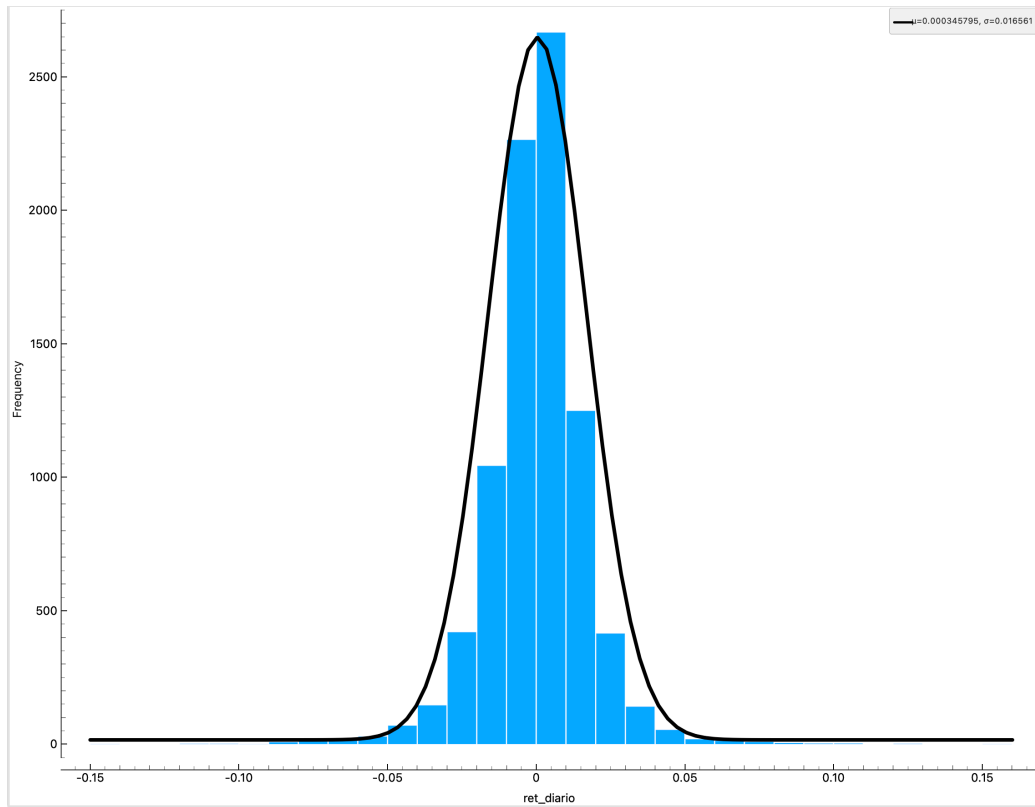
### 4.1 Análisis exploratorio de los datos.

El análisis exploratorio se realiza sobre el panel diario “panel\_diario\_orange.csv”, compuesto por 8.600 sesiones correspondientes a las ocho compañías analizadas en el periodo comprendido entre el 3 de enero de 2022 y el 16 de marzo de 2026. El análisis exploratorio se realiza sobre este fichero ya que de este resulta el fichero “eventos\_orange.csv”, el cual será el utilizado para el desarrollo de los modelos predictivos (véase Anexo III).

Analizando los datos mediante el widget “Data Table”, lo más destacable a primera vista es cómo Acciona es la compañía con menor volumen en la negociación de sus acciones, correspondiendo los primeros 345 días de menor negociación a esta compañía. Esto seguramente se deba al ser la empresa cuyas acciones tienen un mayor valor por bastante diferencia (100,30€ el precio más barato de Acciona respecto a 35,32€ de Endesa, la cotización más alta de las otras siete empresas en conjunto durante el periodo).

A través del widget “Column Statistics” vemos que el retorno diario de las compañías tiene una media (0,035%) y mediana (0,076%) muy próximas a 0, y una desviación típica del 1,66%. Los datos tienen una curtosis muy elevada (7,02), lo que indica la existencia de colas anchas, con movimientos extremos mucho más frecuentes de lo esperado en una distribución normal. Esto es resultado de un rango de retornos diarios comprendido entre -14,96% y 15,72%.

Gráfico 2: Histograma de los retornos diarios.



Fuente: Elaboración propia a partir de los resultados de Orange Data Mining

El comportamiento de las compañías es el siguiente:

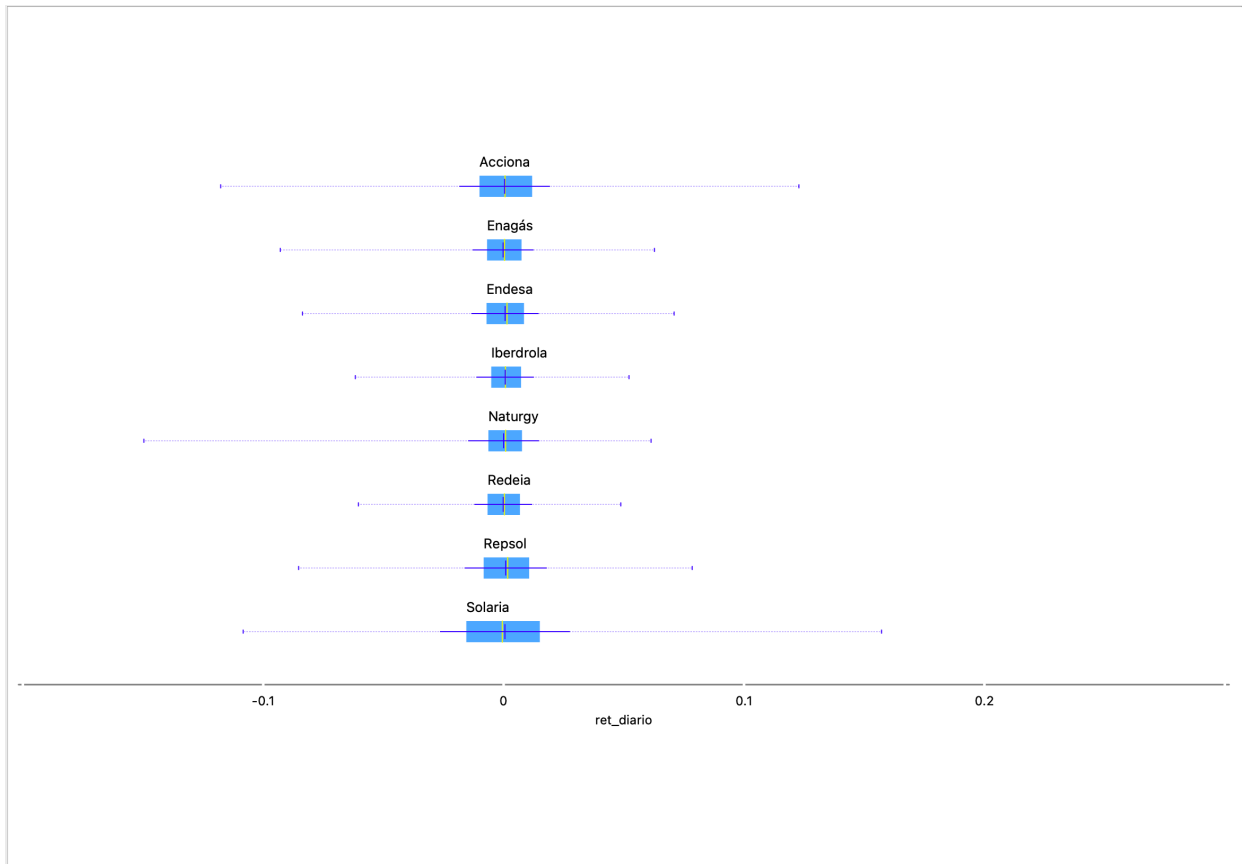
Tabla 1: Rentabilidad acumulada y volatilidad anualizada de las compañías.

Compañía	Rentabilidad acumulada	Volatilidad anualizada
Repsol	+119,0%	26,8%
Iberdrola	+88,9%	18,7%
Endesa	+75,0%	21,9%
Acciona	+27,8%	29,6%
Solaria	+24,2%	42,7%
Naturgy	-10,5%	23,1%
Redeia	-21,4%	18,8%
Enagás	-25,8%	19,9%

Fuente: Elaboración propia.

Mientras que las compañías con negocios relativos a la generación de energía y el petróleo cuentan con rentabilidades positivas en el periodo, las utilities reguladas de redes presentan rendimientos negativos. Esto resulta coherente en un periodo de subida de tipos de interés, al comercializar productos sensibles al tipo de financiación.

Gráfico 3: Boxplot de los retornos diarios de las compañías.

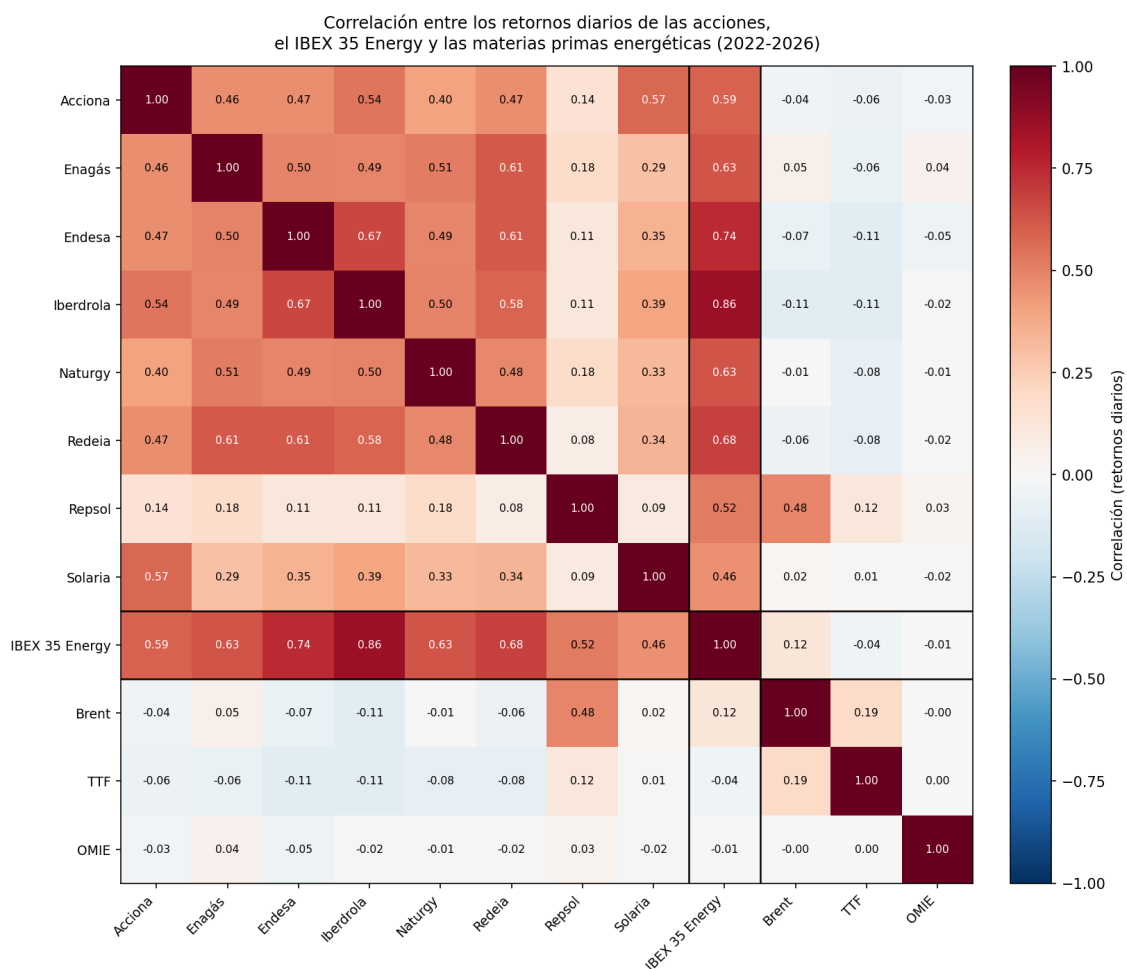


Fuente: Elaboración propia a partir de los resultados de Orange Data Mining

El gráfico de cajas confirma la heterogeneidad de las compañías: mientras que Acciona y Solaria cuentan con las cajas más anchas y mayor rango en sus retornos diarios, compañías como Iberdrola, Redeia o Endesa muestran distribuciones más concentradas y un rango de precios menor.

El análisis de correlación entre el retorno diario de cada acción con las variables externas es el siguiente:

Gráfico 4: Matriz de correlaciones entre retornos de las acciones, el benchmark y materias primas.



Fuente: Elaboración propia.

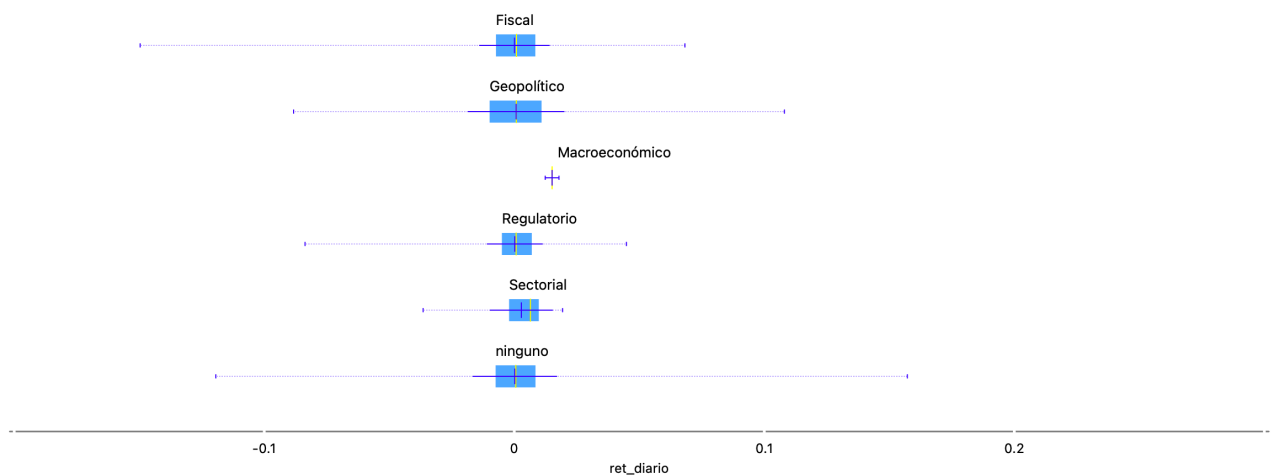
El benchmark IBEX 35 Energy es, con diferencia, el principal factor común al tener todas correlaciones positivas fuertes con él, siendo la menor de 0,461 (Solaria). También vemos que los precios de las acciones tiene una fuerte correlación entre sí, salvo Solaria y, en mayor medida, Repsol. Esto se puede deber al ser la única compañía cuya principal actividad de explotación está relacionada con el petróleo y no con otras fuentes de energía.

En cuanto a las materias primas, a diferencia de lo esperado, la única correlación fuerte se produce entre Brent y Repsol (0,485). A pesar de que esto es coherente con el modelo de negocio de la compañía, otras materias primas como gas natural (TTF) o electricidad (OMIE) muestran correlación prácticamente nula con las compañías cuyo modelo de negocio se basa en el precio de estos bienes.

En 4.480 observaciones de las 8.600 totales (52,09%) había presente algún evento sectorial influyente en la compañía, siendo el más frecuente los eventos geopolíticos

(2.283 observaciones), seguidos por los fiscales (1.402) y regulatorios (769), mientras que los de tipo sectorial (24) y macroeconómico (2) fueron los menos repetidos.

Gráfico 5: Boxplot de la distribución de eventos sectoriales.



Fuente: Elaboración propia a partir de los datos de Orange Data Mining.

Este gráfico, obtenido mediante el widget “Box Plot” conectado directamente a “File”, muestra la presencia de los eventos sectoriales, clasificados por tipo.

De esta manera, podemos obtener tres principales ideas anteriores al desarrollo del modelo. La primera es un conjunto de datos con retornos con colas anchas, lo que implica movimientos más extremos de lo que anticiparía una distribución normal. El segundo es la fuerte relación de los precios de los activos con su benchmark IBEX 35 Energy, pero la prácticamente nula correlación con las materias primas, exceptuando el caso Brent-Repsol. Por último, los eventos sectoriales no suponen un factor diferencial sobre los retornos diarios.

## 4.2 Interpretación del modelo.

La tabla obtenida en el widget Test and Score nos permite comparar el rendimiento de los tres algoritmos sobre la variable objetivo “sig\_ret\_d5\_cat”. Los resultados se resumen a continuación:

Tabla 2: Resultados de los modelos predictivos.

<b>Modelo</b>	<b>AUC</b>	<b>CA</b>	<b>F1</b>	<b>Prec</b>	<b>Recall</b>	<b>MCC</b>
Random Forest	0,684	<b>0,669</b>	0,666	0,676	0,669	<b>0,345</b>
Gradient Boosting	<b>0,697</b>	0,647	0,645	0,650	0,647	0,297
Logistic Regression	0,642	0,588	0,573	0,603	0,588	0,191

Fuente: Elaboración propia a partir de los resultados de Orange Data Mining

El modelo Random Forest supone el modelo más equilibrado, al tener la mejor precisión de clasificación (CA = 0,669) y los mejores F1 y coeficiente de Matthews (MCC = 0,345). Este resultado es coherente con la evidencia existente, que sitúa al Random Forest como el algoritmo con mayor capacidad predictiva en aplicaciones financieras y de sostenibilidad (Seow, 2025). Aunque el Gradient Boosting tiene un AUC ligeramente superior (0,697 frente a 0,684), de manera global tiene menor precisión. La Regresión Logística queda claramente por detrás en todas las métricas, lo que confirma que la dirección del precio de las acciones no es lineal y se puede predecir mejor mediante árboles de decisión.

La muestra está perfectamente balanceada con 68 subidas y 68 bajadas (baseline del 50%), lo que significa que el 66,9% de precisión del Random Forest mejora en casi 17 puntos porcentuales a la predicción aleatoria. El coeficiente de correlación de Matthews de 0,345, claramente positivo, confirma que el modelo acierta de manera consistente en predicciones tanto de subida como de bajada. El AUC de 0,684 indica una capacidad de discriminación moderada.

La matriz de confusión del Random Forest muestra que, de los 136 anuncios, el modelo clasifica bien 91 (52 bajadas y 39 subidas), suponiendo una precisión del 66,9%. No obstante, el comportamiento no es simétrico, identificando correctamente el 76,5% de las

caídas reales (52 de 68) y el 57,4% de las subidas reales (39 de 68). El principal error del modelo fueron los falsos negativos, clasificando 29 de las subidas reales como caídas, lo que evidencia un sesgo conservador. Sin embargo, cuando predice una subida acierta el 70,9% de las veces, siendo la señal más fiable del modelo.

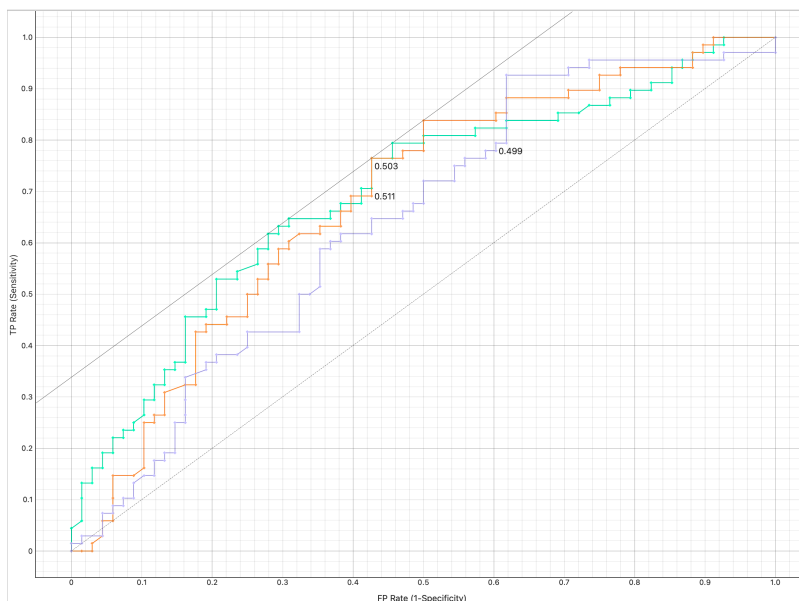
Gráfico 6: Matriz de confusión del modelo Random Forest.

		Predicted		$\Sigma$
		Baja	Sube	
Actual	Baja	52	16	<b>68</b>
	Sube	29	39	<b>68</b>
$\Sigma$		<b>81</b>	<b>55</b>	<b>136</b>

Fuente: Elaboración propia a partir de los resultados de Orange Data Mining

El análisis de la curva ROC confirma gráficamente que las tres curvas se sitúan por encima del clasificador aleatorio, situándose las de los modelos Random Forest y Gradient Boosting más separadas de esta curva que la de Regresión Logística, concordando con sus valores mayores de AUC.

Gráfico 7: Análisis de la curva ROC de los modelos predictivos.



Fuente: Elaboración propia a partir de los resultados de Orange Data Mining

No obstante, los resultados deben interpretarse considerando el tamaño muestral de 136 eventos. Un mayor rango del periodo y la incorporación de variables como EBITDA real y previsto o guidance comunicado por las compañías, que no se han podido obtener de manera homogénea en el intervalo temporal analizado, mejoraría la capacidad predictiva de los algoritmos en futuras extensiones.

## **5. CONCLUSIONES.**

El presente Trabajo Fin de Grado se planteó con el objetivo de evaluar la capacidad de diferentes modelos de Inteligencia Artificial en la predicción del comportamiento bursátil de las empresas del sector energético del IBEX 35 en los cinco días posteriores a la publicación de resultados trimestrales, junto a la identificación de las variables más relevantes en este aspecto. De esta manera, se pueden extraer las siguientes conclusiones:

Los modelos permiten responder a la pregunta de si la Inteligencia Artificial es capaz de predecir el comportamiento bursátil en los cinco días posteriores a la publicación de resultados trimestrales. El modelo más preciso, el Random Forest, obtuvo una precisión de clasificación del 66,9% sobre una muestra perfectamente balanceada, acertando casi 17 puntos porcentuales por encima del azar. Esto, junto a un coeficiente de correlación de Matthews de 0,345 y un AUC de 0,684, concluye que la Inteligencia Artificial tiene una capacidad predictiva real y moderada.

El resultado es coherente con la hipótesis de los mercados eficientes (Fama, 1970), ya que, si los mercados incorporasen toda la información disponible, la capacidad predictiva debería ser nula. La existencia de ineficiencias en el corto plazo es lo que hace que los modelos elaborados superen consistentemente al azar. No obstante, con las variables e intervalo temporal planteados, la Inteligencia Artificial no es capaz de batir de manera consistente al mercado.

Respecto a la metodología usada, se han usado diferentes modelos para dar respuesta al objeto del Trabajo de Fin de Grado. La superioridad de los árboles de decisión, Gradient Boosting y Random Forest, frente a la Regresión Logística, confirma la teoría de que el precio no es lineal y que se captura mejor mediante algoritmos que trabajen con interacciones complejas. Además, que el Random Forest sea el modelo más preciso coincide con la evidencia previa que lo posiciona como el algoritmo con mejor capacidad predictiva en aplicaciones financieras y de sostenibilidad (Seow, 2025). La identificación

en un mayor porcentaje de las bajadas (76,5%) frente a las subidas (57,4%) revela que las reacciones negativas sean más predecibles que las positivas.

De entre las variables usadas en el análisis, las que resultaron ser las más relevantes fueron la inflación, la sorpresa del beneficio por acción, el volumen relativo, la distancia a la media móvil de 200 sesiones y el momentum a cinco días. La sorpresa de resultados, a pesar de ser uno de los datos más relevantes en la presentación de resultados, no fue una de las variables más determinantes. Esto sugiere que el mercado, antes de la publicación de resultados, lo descuenta del precio de las acciones, haciendo que estas suban o bajen en fechas anteriores al evento.

A la hora de interpretar resultados hay que tener en cuenta ciertas limitaciones del modelo. El tamaño muestral de 136 anuncios obliga a, además de tener en cuenta las condiciones del sector en el intervalo seleccionado, restringir el número de variables seleccionadas para evitar un sobreajuste de los modelos. Otra limitación del modelo es la dificultad de obtener de manera homogénea variables relevantes, como el EBITDA previsto frente al real o la orientación publicada por las compañías, cuya inclusión podría haber resultado en una mayor predicción por parte de los modelos. No obstante, estas limitaciones pueden suponer puntos de partida para mejores predicciones de los modelos.

En conclusión, este trabajo evidencia el potencial real, pero limitado, de la Inteligencia Artificial en la predicción bursátil del sector energético del IBEX 35 tras la publicación de resultados. Su valor reside en servir como herramienta de apoyo en la toma de decisiones de los inversores, especialmente útil en la anticipación de bajada de los precios en mercados alcistas y en la identificación de variables más relevantes a considerar en la predicción.

## **6. DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA EN TRABAJOS DE FIN DE GRADO.**

Por la presente, yo, Luis Albizu Arias, estudiante de Administración y Dirección de Empresas de la Universidad Pontificia Comillas, al presentar mi Trabajo de Fin de Grado titulado “Evaluación del potencial de la Inteligencia Artificial en la predicción bursátil del sector energético del IBEX 35”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código solo en el contexto de las actividades descritas a continuación:

1. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
2. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
3. **Interpretador de código:** Para realizar análisis de datos preliminares.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 27 de mayo de 2026

Firma: Luis Albizu Arias

## 7. BIBLIOGRAFÍA.

Acciona, S.A. (s.f.). *Acciona — Áreas de negocio*. <https://www.acciona.com/es>

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31–56. [https://doi.org/10.1016/S1386-4181\(01\)00024-6](https://doi.org/10.1016/S1386-4181(01)00024-6)

Banco de España. (2023, 14 de septiembre). *El BCE sube los tipos 25 puntos básicos en septiembre*. <https://www.bde.es/wbe/en/noticias-eventos/actualidad-bce/decisiones-politica-monetaria/el-bce-sube-los-tipos-25-puntos-basicos-en-septiembre.html>

Bocconi Students Investment Club (BSIC). (2025). *Regularization techniques in financial machine learning: Ridge and Lasso regressions*. BSIC Quantitative Research. <https://bsic.it/regularization-techniques-in-financial-machine-learning/>

- Boletín Oficial del Estado (BOE). (2022, 13 de mayo). *Real Decreto-ley 10/2022, de 13 de mayo* [Excepción ibérica]. [https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2022-7843](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-7843)
- Boletín Oficial del Estado (BOE). (2022, 27 de diciembre). *Ley 38/2022, de 27 de diciembre, para el establecimiento de gravámenes temporales energético y de entidades de crédito*. <https://www.boe.es/buscar/act.php?id=BOE-A-2022-22684>
- Bolsamanía. (2022, 25 de febrero). *Endesa, Iberdrola, Naturgy, Acciona y Solaria se disparan en el Ibex 35*. <https://www.bolsamania.com/noticias/empresas/endesa-iberdrola-naturgy-acciona-y-solaria-se-disparan-en-el-ibex-35--9292621.html>
- Bolsamanía. (2023, 9 de octubre). *Cierre bajista en el Ibex 35 y Europa tras el estallido de la guerra en Israel*. <https://www.bolsamania.com/noticias/cronica-ibex/lunes-9-octubre-2023-caidas-ataque-hamas-israel--14903710.html>
- Bolsamanía. (2025). *Redeia sufre en bolsa: las eléctricas podrían reclamar indemnizaciones por el apagón*. <https://www.bolsamania.com/noticias/empresas/redeia-sufre-en-bolsa-las-electricas-podrian-reclamar-indemnizaciones-apagon--19501082.html>
- Bolsas y Mercados Españoles (BME). (2026). *IBEX 35: composición y datos del índice*. BME Exchange. <https://www.bolsasymercados.es/es/bme-exchange/mercados-y-cotizaciones/acciones/ibex-35-es0si0000005.html>
- Chicago Booth Review. (2024). *The evolution of AI in finance*. University of Chicago Booth School of Business. <https://www.chicagobooth.edu/review/evolution-ai-finance>
- Dartmouth College. (2026). *Artificial Intelligence (AI) coined at Dartmouth*. <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth>
- El Amine El Amani, A., et al. (2026). *Machine learning and deep learning in computational finance: A systematic review* (Preprint arXiv:2511.21588). arXiv. <https://arxiv.org/abs/2511.21588>

- El Cronista. (2026, marzo). *La guerra en Irán hunde al Ibex 35: Repsol logra esquivarlo con una subida del 6,55%*.  
<https://www.cronista.com/espana/economia-finanzas/la-guerra-en-iran-hunde-al-ibex-35-repsol-logra-esquivarlo-con-una-subida-del-655/>
- elDiario.es. (2022, 22 de febrero). *Iberdrola gana 4.339 millones en 2022 y pulveriza su quinto récord de beneficios consecutivo*.  
[https://www.eldiario.es/economia/iberdrola-gana-4-339-millones-2022-pulveriza-quinto-record-beneficios-consecutivo\\_1\\_9973626.html](https://www.eldiario.es/economia/iberdrola-gana-4-339-millones-2022-pulveriza-quinto-record-beneficios-consecutivo_1_9973626.html)
- elEconomista. (2024, 21 de marzo). *Iberdrola invertirá 41.000 millones hasta 2026 para crecer en redes y eólica marina*.  
<https://www.eleconomista.es/energia/noticias/12733202/03/24/iberdrola-invertira-41000-millones-hasta-2026-para-crecer-en-redes-y-eolica-marina.html>
- elEconomista. (2025, 27 de abril). *Las energéticas presentarán un resultado conjunto de 2.949 millones en plena incertidumbre geopolítica*.  
<https://www.eleconomista.es/energia/noticias/13336038/04/25/las-energeticas-presentaran-un-resultado-conjunto-de-2949-millones-en-plena-incertidumbre-geopolitica.html>
- El Español – Invertia. (2023, 4 de abril). *El recorte de crudo de la OPEP y el embargo al petróleo ruso abocará a la estancación y a tipos altos en Europa*.  
[https://www.lespanol.com/invertia/empresas/energia/20230404/recorte-ojep-embargo-petroleo-abocara-estanflacion-europa/753424922\\_0.html](https://www.lespanol.com/invertia/empresas/energia/20230404/recorte-ojep-embargo-petroleo-abocara-estanflacion-europa/753424922_0.html)
- El Español – Invertia. (2024, 6 de noviembre). *El Ibex 35 vive su peor sesión desde marzo de 2023 al desplomarse un 3% tras la victoria de Trump*.  
[https://www.lespanol.com/invertia/mercados/20241106/ibex-bolsa-directo-hoy-cotizaciones-noticias-tiempo-real/899260066\\_10.html](https://www.lespanol.com/invertia/mercados/20241106/ibex-bolsa-directo-hoy-cotizaciones-noticias-tiempo-real/899260066_10.html)
- El Independiente. (2024, 17 de abril). *El fondo de Emiratos Árabes TAQA lanzará una OPA para controlar el 100% de Naturgy*.  
<https://www.elindependiente.com/economia/2024/04/17/el-fondo-de-emiratos-arabes-taqa-lanzara-una-opa-para-controlar-el-100-de-naturgy/>

- El Periódico de la Energía. (2022). *El impuesto sobre las energéticas entra en vigor hoy y gravarán ya 2022*. <https://elperiodicodelaenergia.com/impuesto-energeticas-vigor-hoy-gravaran-2022/>
- El Periódico de la Energía. (2024). *Iberdrola presenta su plan estratégico para el periodo 2024-2026*. <https://elperiodicodelaenergia.com/iberdrola-presenta-plan-estrategico-periodo-2024-2026/>
- El Progreso. (2025, 23 de julio). *El sector eléctrico se desploma en bolsa tras el rechazo al decreto antiapagones*. <https://www.elprogreso.es/articulo/economia/sector-electrico-sufre-caidas-lidera-descensos-ibex-reves-decreto-antiapagones/202507231302211896797.html>
- Energías Renovables. (2022, 13 de mayo). *De la excepción ibérica y el tope al precio del gas*. <https://www.energias-renovables.com/panorama/de-la-excepcion-iberica-y-el-tope-20220513>
- Estrategias de Inversión. (2024, 27 de diciembre). *El sector energético español en el IBEX 35: cómo analizarlo*. <https://www.estrategiasdeinversion.com/actualidad/noticias/bolsa-espana/el-sector-energetico-espanol-en-el-ibex-35-como-n-773787>
- Estrategias de Inversión. (2025, 4 de abril). *Así afectan los aranceles a los valores de la bolsa española y al Ibex 35*. <https://www.estrategiasdeinversion.com/actualidad/noticias/bolsa-espana/asi-afectan-los-aranceles-a-los-valores-de-la-n-801971>
- Estrategias de Inversión. (2026, 1 de junio). *Iberdrola vuelve sobre máximos históricos*. <https://www.estrategiasdeinversion.com/actualidad/noticias/bolsa-espana/iberdrola-vuelve-sobre-maximos-historicos-en-un-n-877145>
- FactSet Research Systems. (2026). *Cotizaciones históricas de acciones, materias primas energéticas, Euríbor 12M e índices sectoriales* [Base de datos financiera]. <https://my.apps.factset.com>

- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.  
<https://academic.oup.com/rfs/article/33/5/2223/5758276>
- Iberdrola, S.A. (2022). *Las inversiones récord (10.730 millones) nos permiten alcanzar un beneficio neto de 4.339 millones de euros*. Sala de comunicación.  
<https://www.iberdrola.com/press-room/news/detail/las-inversiones-record-10730-millones-nos-permiten-alcanzar-beneficio-neto-4339-millones-euros>
- Iberdrola, S.A. (2025). *Iberdrola invertirá 58.000 millones hasta 2028 (+30%) para crecer más en redes en los Estados Unidos y el Reino Unido*. Sala de comunicación. <https://www.iberdrola.com/sala-comunicacion/noticias/detalle/iberdrola-invertira-58000-millones-hasta-2028-crecer-mas-redes-estados-unidos-reino-unido>
- Instituto Nacional de Estadística (INE). (2026). *Índice de Precios de Consumo (IPC). Variación interanual*. [https://www.ine.es/prensa/ipc\\_tabla.htm](https://www.ine.es/prensa/ipc_tabla.htm)
- International Monetary Fund (IMF). (2023). *Generative artificial intelligence in finance: Risk considerations* (IMF FinTech Note 2023/006).  
<https://www.imf.org/en/Publications/ftn/Issues/2023/08/23/Generative-Artificial-Intelligence-in-Finance-Risk-Considerations-537446>
- Investing.com. (2026). *Earnings Calendar*. <https://www.investing.com/earnings-calendar>
- Investing.com. (2026). *Datos históricos del IBEX 35 Energy*.  
<https://es.investing.com/indices/ibex-35-energy-historical-data>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.

- Kang, J.-W., & Choi, S.-Y. (2025). Comparative investigation of GPT and FinBERT's sentiment analysis performance in news across different sectors. *Electronics*, 14(6), Artículo 1090. <https://doi.org/10.3390/electronics14061090>
- Kearns, J. (2023). AI's reverberations across finance. *Finance & Development*, 60(4). International Monetary Fund. <https://www.imf.org/en/Publications/fandd/issues/2023/12/AI-reverberations-across-finance-Kearns>
- Kelly, B., & Xiu, D. (2023). *Financial machine learning* (Working Paper n.º 2023-100). Becker Friedman Institute, University of Chicago.
- Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689–702. <https://www.sciencedirect.com/science/article/abs/pii/S0377221716308657>
- LoanWorks. (s.f.). *Artificial intelligence in financial services: A historical review*. <https://loanworks.ai/artificial-intelligence-in-financial-services-a-historical-review/>
- Newell, A., & Simon, H. A. (1956). *The Logic Theory Machine: A complex information processing system* (Paper P-868). RAND Corporation. <https://www.rand.org/pubs/papers/P868.html>
- Newtral. (2022, 22 de diciembre). *Cómo afecta a España el tope europeo al precio del gas*. <https://www.newtral.es/tope-gas-europeo-espana-precio-union-europea/20221222/>
- OBS Business School. (2025). *Informe OBS: El sector energético en España, hacia una descarbonización sostenible*. OBS Business School. <https://www.obsbusiness.school/actualidad/informes-de-investigacion/informe-obs-el-sector-energetico-en-espana-hacia-una-descarbonizacion-sostenible>

- OMIE — Operador del Mercado Ibérico de Energía. (2026). *Precios diarios del mercado eléctrico mayorista español*. <https://www.omie.es/es/market-results/annual/daily-market/daily-prices>
- Organisation for Economic Co-operation and Development (OECD). (2025). *AI in finance: Balancing innovation and stability*. OECD.AI Policy Observatory. <https://oecd.ai/en/wonk/ai-in-finance-balancing-innovation-and-stability>
- Qian, Y. (2025). *Research on stock price prediction based on random forest & XGBoost* [Conference proceedings]. ICIAAI-25. Atlantis Press. <https://www.atlantispress.com/proceedings/iciaai-25/126015318>
- Redress Compliance. (2025). *The history of robo-advisors and automated investing*. <https://redresscompliance.com/the-history-of-robo-advisors-and-automated-investing/>
- Saberironaghi, M., et al. (2025). Stock market prediction using machine learning and deep learning techniques: A review. *FinTech*, 5(3), Artículo 76. <https://www.mdpi.com/2673-9909/5/3/76>
- Seow, R. Y. C. (2025). Transforming ESG analytics with machine learning: A systematic literature review using TCCM framework. *Corporate Social Responsibility and Environmental Management*, 32(6), 7358–7389.
- Sharkey, E., & Treleaven, P. (2024). *BERT vs GPT for financial engineering* [Preprint]. arXiv. <https://arxiv.org/abs/2405.12990>
- Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1), 83–104.
- Ubuntu (Canonical). (2023). *Machine learning in finance: History, technologies, and outlook*. <https://ubuntu.com/blog/machine-learning-in-finance-history-technologies-and-outlook>
- Yadav, V., & Shailaja, K. P. (2024). *Prediction of stock price using XGBoost: A machine learning technique*. International Journal for Research in Applied Science & Engineering Technology (IJRASET).

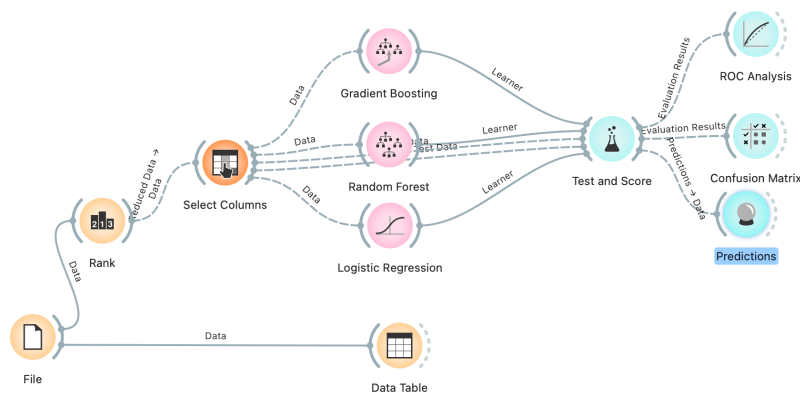
## 8. ANEXO.

Este apartado recoge la documentación visual complementaria del trabajo realizado en Orange Data Mining. Se incluyen el análisis exploratorio, el flujo de trabajo y las configuraciones de los algoritmos empleados.

### Anexo I. Desarrollo del modelo mediante Orange Data Mining

El presente anexo recoge la captura completa del flujo de trabajo en Orange Data Mining para la predicción del signo a cinco días posteriores a la publicación de resultados trimestrales. El diagrama refleja el proceso descrito en el apartado 3.4: la carga del archivo eventos\_orange.csv (File), la inspección (Data Table), la selección de variables más influyentes (Rank y Select Columns), el entrenamiento de los modelos utilizados (Gradient Boosting, Random Forest y Logistic Regression), la evaluación de los resultados (Test and Score) y la inspección de los resultados (ROC Analysis, Confusion Matrix y Predictions).

Figura 1. Flujo de trabajo desarrollado en Orange Data Mining para la predicción a cinco días



Fuente: Elaboración propia

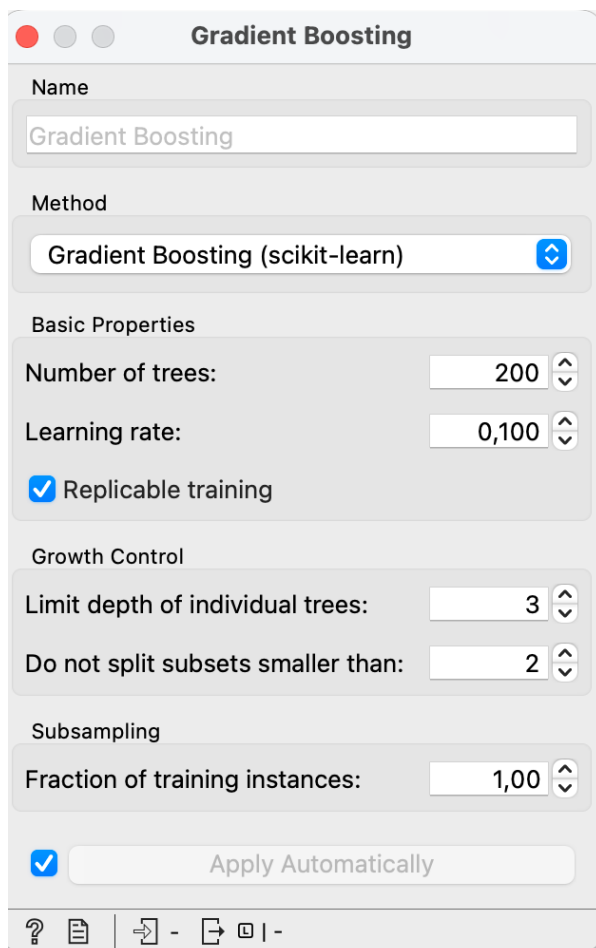
## Anexo II. Configuración de los algoritmos de aprendizaje supervisado

Este anexo recoge las capturas de las configuraciones de los tres algoritmos de aprendizaje supervisado: Gradient Boosting, Random Forest y Logistic Regression, descritos en el apartado 3.4.

### II.1. Gradient Boosting

Para el algoritmo Gradient Boosting, se han configurado 200 árboles con una tasa de aprendizaje de 0,1 y entrenamiento replicable. El control de crecimiento se estableció en una profundidad máxima de tres niveles con un tamaño de división mínimo de 2, manteniendo la fracción de muestreo en 1,00. La combinación de una tasa de aprendizaje moderada y profundidad limitada actúa como mecanismo de regularización que reduce el riesgo de sobreajuste sobre la muestra disponible.

Figura 2. Configuración del widget Gradient Boosting en Orange Data Mining

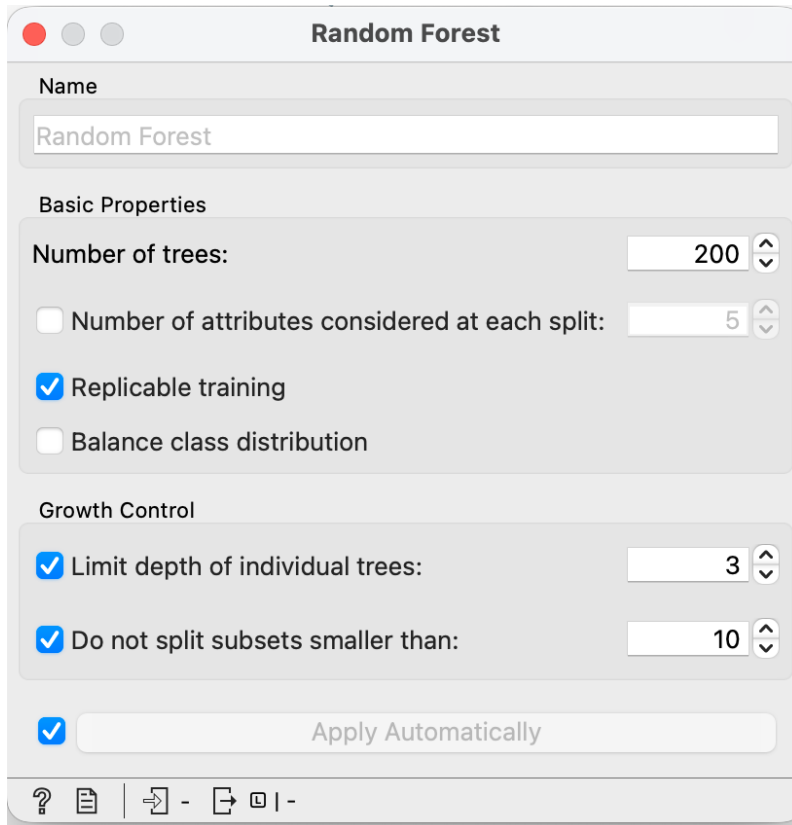


Fuente: Elaboración propia.

### II.2. Random Forest

En el algoritmo Random Forest se configuraron 200 árboles con entrenamiento replicable activado para asegurar la reproducibilidad. Se limitó la profundidad en el control de crecimiento a un máximo de tres niveles y un tamaño mínimo de división de 10. Se mantuvieron desactivadas las opciones de distribución de balanceo de clases y la limitación de atributos por división.

Figura 3. Configuración del widget Random Forest en Orange Data Mining

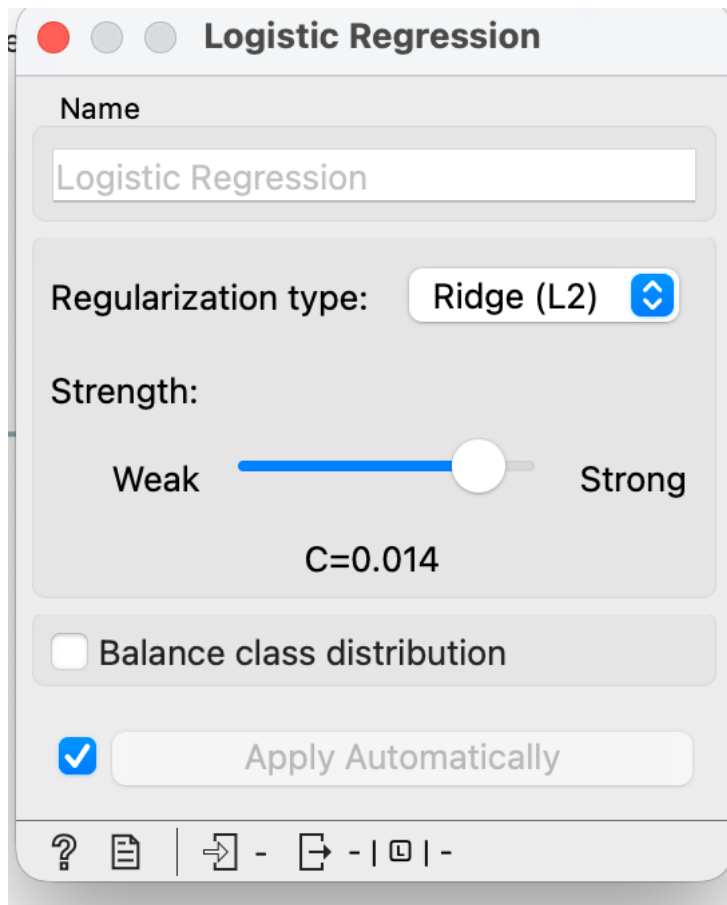


Fuente: Elaboración propia

### II.3. Regresión Logística

Por último, para el algoritmo de Regresión Logística se configuró una regularización Ridge (L2) y un parámetro de regularización  $C = 0,014$  para evitar el sobreajuste, así como una desactivación de la distribución del balanceo de clases.

Figura 4. Configuración del widget Logistic Regression en Orange Data Mining

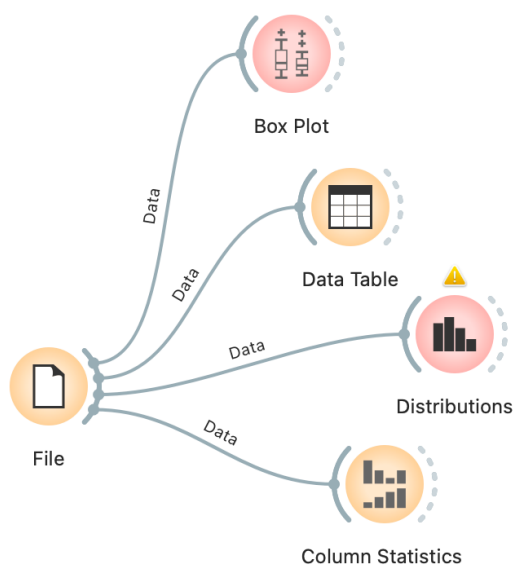


Fuente: Elaboración propia

### Anexo III. Análisis exploratorio de los datos

Este anexo contiene el flujo de trabajo en Orange Data Mining del análisis exploratorio de los datos descrito en el apartado 4.1. El widget File, que contiene el archivo “panel\_diario\_orange.csv” está conectado con otros cuatro widgets, cada uno con una función. El widget Data Table permite una revisión global de los datos. El widget Column Statistics aporta estadísticos descriptivos univariantes, como la media, mediana o desviación típica, de las variables continuas. El widget Distributions muestra el histograma de los retornos diarios ajustado a una distribución normal. Por último, el widget Box Plot representa la distribución de los retornos diarios por compañía y la distribución de los eventos activos por tipo, en función de la configuración escogida.

Figura 5. Flujo de trabajo desarrollado en Orange Data Mining para el análisis exploratorio de los datos



Fuente: Elaboración propia