



Vigilada Mineducación

***Machine learning* como alternativa al modelo de Markowitz: Evaluación comparativa de rentabilidad y riesgo en el mercado bursátil español (2014-2024)**

Machine Learning as an Alternative to the Markowitz Model: A Comparative Evaluation of Profitability and Risk in the Spanish Stock Market (2014-2024)

Daniel Felipe Ricaurte Rodríguez

Trabajo de grado

Asesor, docente

Fernando Alonso Villamil

UNIVERSIDAD EAFIT

ESCUELA DE FINANZAS, ECONOMÍA Y GOBIERNO

MAESTRÍA EN ADMINISTRACIÓN FINANCIERA - MAF

MEDELLÍN

2025

Resumen

Este trabajo corresponde a la versión ajustada del trabajo de fin de máster presentado en la Universidad Pompeu Fabra-Barcelona School of Management (UPF-BSM), dentro del programa de doble titulación con la Universidad EAFIT. El estudio busca determinar si las estrategias basadas en modelos de *machine learning* (ML) constituyen una alternativa válida al modelo tradicional de Markowitz para la construcción de portafolios de inversión. Se analizaron precios diarios de 36 acciones del mercado bursátil español entre 2014 y 2024, utilizando modelos supervisados de clasificación (regresión logística, árboles de decisión, *random forest*, KNN y XGBoost) para predecir la dirección del precio y generar señales de compra. Los resultados muestran que, sin costos de transacción, Markowitz obtiene mayor rentabilidad; sin embargo, los modelos de ML presentan mejor control del riesgo y desempeño superior en escenarios con costos. Se concluye que su integración ofrece estrategias híbridas más eficientes bajo condiciones de mercado reales.

Palabras clave: *Machine learning*, modelo de Markowitz, portafolios de inversión, rentabilidad y riesgo.

Abstract

This work corresponds to the adjusted version of the Master's Thesis originally presented at Universitat Pompeu Fabra-Barcelona School of Management (UPF-BSM), within the double-degree program with Universidad EAFIT. The study aims to determine whether strategies based on Machine Learning models constitute a valid alternative to the traditional Markowitz model for investment portfolio construction. Daily prices of 36 stocks from the Spanish stock market between 2014 and 2024 were analyzed using supervised classification models (logistic regression, decision trees, random forest, KNN, and XGBoost) to predict price direction and generate buy signals. Results show that while the Markowitz model achieves higher returns in the absence of transaction costs, ML models provide better risk control and superior performance under cost-inclusive scenarios. It is concluded that integrating both approaches enables more efficient hybrid investment strategies under real market conditions.

Key words: Machine learning, Markowitz model, investment portfolios, return and risk.

Contenido

Lista de figuras	vi
Introducción	1
Motivación	1
Objetivos 2	
Objetivo general.....	2
Objetivos específicos	2
Pregunta de investigación	2
Marco teórico	2
Fundamentos del modelo de Markowitz.....	3
Fundamentos del aprendizaje automático <i>machine learning</i> (ML)	4
Revisión de la literatura.....	6
Aplicaciones de ML a la construcción de portafolios.....	6
Predicción de precios mediante ML e indicadores técnicos	7
Metodología	8
Selección de acciones.....	8
Periodo de estudio.....	10
Metodología para predecir las direcciones futuras de las acciones con ML	11
Problemas de pronósticos	11
Variable objetivo.....	12
Variables predictoras.....	13
Modelos de pronósticos.....	15
Entrenamiento de modelos	16
Métricas de evaluación: balanced accuracy	17
Validación cruzada.....	18
Predicción final.....	18
Estrategia de <i>trading</i> basada en ML.....	19
Generación de señales de trading	19
Reglas de negociación.....	21
Parámetros de negociación	22
Estrategias de inversión con el modelo de Markowitz	24
Métricas de desempeño	26
Resultados	28

Selección del mejor modelo.....	28
Resultados del análisis de sensibilidad para la inversión máxima.....	30
Selección de la ventana móvil óptima	31
Comparación de las estrategias de inversión	32
Rentabilidad.....	32
TCAC.....	34
Volatilidad.....	35
Max drawdown.....	36
Ratio de Sharpe.....	37
Conclusiones	38
Anexos....	40
Anexo 1. Distribución de clases de las acciones	40
Anexo 3. Evolución de las estrategias a los diferentes costes de transacción	43
Anexo 3. Gráfico max <i>drawdown</i>	45
Referencias	45

Lista de tablas

Tabla 1. <i>Acciones</i>	12
Tabla 2. <i>Ejemplo de definición de la variable objetivo</i>	15
Tabla 3. <i>Decisión de compra y venta para la acción ferroviaria</i>	23
Tabla 4. <i>Priorización de compra (millones de euros)</i>	27
Tabla 5. <i>Mejor modelo por acción</i>	32
Tabla 6. <i>Ventana móvil</i>	35
Tabla 7. <i>Rentabilidad total</i>	36
Tabla 8. <i>TCAC</i>	37
Tabla 9. <i>Volatilidad</i>	38
Tabla 10. <i>Max drawdown</i>	39
Tabla 11. <i>Ratio de Sharpe</i>	40

Lista de figuras

<i>Figura 1. Análisis de Sensibilidad del Ratio de Sharpe según el Porcentaje Máximo de Inversión y Costos de Transacción</i>	34
---	----

Introducción

La asignación de activos es un tema central en finanzas. Desde que Harry Markowitz presentó su modelo en 1952, su propuesta de portafolios de mínima varianza se volvió un referente para equilibrar rentabilidad y riesgo. No obstante, en los mercados reales los retornos no siempre son normales ni las covarianzas estables (Cont, 2001; Engle, 2002). En la última década, factores como la pandemia del covid-19, la inflación y los conflictos geopolíticos han desafiado la capacidad de los modelos clásicos para adaptarse. En este escenario, el *machine learning* (ML) surge como una alternativa capaz de aprender de los datos y responder a condiciones cambiantes.

Este trabajo compara ambos enfoques aplicados al mercado bursátil español entre 2014 y 2024, analizando cuál ofrece un mejor equilibrio entre rentabilidad y riesgo para un inversor conservador.

Motivación

Tomar decisiones sobre cómo distribuir los activos de una inversión es uno de los mayores retos para cualquier persona, especialmente para quienes buscan proteger su capital antes que asumir grandes riesgos. Durante décadas, el modelo de Markowitz ha sido una guía esencial para equilibrar rentabilidad y estabilidad, pero los mercados actuales son más complejos y volátiles. Factores como la pandemia, la globalización y la gran disponibilidad de datos han mostrado que sus supuestos teóricos no siempre se cumplen. En este contexto, el ML aparece como una herramienta capaz de adaptarse y aprender de los cambios del mercado. Este trabajo busca comprobar si esa inteligencia predictiva puede mejorar la toma de decisiones y ofrecer resultados más consistentes que el modelo tradicional, especialmente al considerar costos y limitaciones reales de inversión.

Objetivos

Objetivo General

Evaluar si las estrategias de inversión basadas en modelos de ML pueden ser una alternativa válida al modelo tradicional de Markowitz en la construcción de portafolios del mercado bursátil español entre 2014 y 2024, comparando su rentabilidad y nivel de riesgo desde la perspectiva de un inversor conservador.

Objetivos Específicos

1. Implementar modelos de ML para predecir la dirección diaria del precio de las acciones del mercado bursátil español.
2. Construir portafolios de inversión basados en las predicciones de los modelos de ML, aplicando reglas de gestión del capital.
3. Conformar portafolios utilizando el modelo tradicional de Markowitz de mínima varianza.
4. Comparar el desempeño de ambas estrategias en términos de rentabilidad y riesgo, con y sin considerar los costos de transacción, durante el periodo 2014-2024.

Pregunta de investigación

¿En qué medida las estrategias de inversión basadas en *machine learning* (ML) pueden ofrecer mejores resultados o complementar al modelo tradicional de Markowitz en la construcción de portafolios del mercado bursátil español entre 2014 y 2024, equilibrando la rentabilidad y el riesgo para un inversor conservador?

Marco teórico

En este capítulo se presentan las bases conceptuales y técnicas que dan sentido a este trabajo. Primero, se aborda el modelo clásico de Markowitz, que durante décadas ha sido un punto de partida para construir carteras eficientes. Se explican sus principales ideas, los supuestos que

lo sustentan y las limitaciones que enfrenta cuando se aplica en mercados reales, donde las condiciones cambian con rapidez y los comportamientos no siempre siguen la teoría.

Después, se introduce el enfoque de ML como una alternativa más flexible y adaptativa. Se explica de forma general cómo funciona el aprendizaje automático y se describen los cinco algoritmos elegidos para el estudio: regresión logística, árboles de decisión, *random forest*, KNN y XGBoost.

Por último, se presenta una revisión de la literatura reciente que permite entender cómo la inteligencia artificial ha ido ganando espacio en las finanzas y por qué estos modelos resultan relevantes para el análisis y la gestión de inversiones.

Fundamentos del modelo de Markowitz

El modelo de portafolios propuesto por Harry Markowitz (1952) sentó las bases de la teoría moderna de las finanzas al demostrar que la diversificación puede reducir el riesgo total sin disminuir necesariamente la rentabilidad esperada. Su principal contribución fue ofrecer un marco matemático que cuantifica la relación entre el riesgo y el retorno, permitiendo construir portafolios eficientes.

De acuerdo con Markowitz (1952), los inversionistas buscan maximizar la rentabilidad esperada para un nivel dado de riesgo o, de forma equivalente, minimizar el riesgo para una rentabilidad determinada. Este equilibrio da origen a la *frontera eficiente*, que reúne las combinaciones óptimas de activos.

- Rentabilidad esperada del portafolio:

$$E(R_p) = \sum_{i=1}^n w_i * E(R_i)$$

- Varianza del portafolio:

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i * w_j * \sigma_{ij}$$

donde w_i es el peso del activo i , $E(R_i)$ su rendimiento esperado y σ_{ij} la covarianza entre los activos i y j .

El problema de optimización se formula como:

$$\min_w \sigma_p^2 \text{ sujeto a } \sum_{i=1}^n w_i = 1 \text{ y } E(R_p) = R^*$$

Diversos autores han identificado limitaciones en la aplicación práctica del modelo. Michaud (1989) y Chopra y Ziemba (1993) advierten que pequeñas variaciones en las estimaciones de los parámetros pueden generar carteras inestables. Además, los rendimientos financieros rara vez siguen distribuciones normales, mostrando colas gruesas y asimetrías (Cont, 2001; Mandelbrot y Hudson, 2006), lo que subestima los riesgos extremos.

A pesar de estas limitaciones, el modelo de Markowitz continúa siendo una referencia fundamental en la gestión de portafolios y un punto de partida para métodos más flexibles, como los basados en ML, que buscan capturar la complejidad y la dinámica de los mercados financieros actuales.

Fundamentos del aprendizaje automático *machine learning* (ML)

El aprendizaje automático o *machine learning* (ML) es una rama de la inteligencia artificial que desarrolla algoritmos capaces de aprender patrones a partir de los datos y de mejorar su desempeño sin programación explícita (Mitchell, 1997). Su aplicación al análisis financiero ha

crecido por su capacidad para reconocer relaciones complejas, no lineales y dinámicas en los mercados.

A diferencia de los modelos estadísticos tradicionales, el ML no depende de supuestos teóricos rígidos, sino que extrae las relaciones entre variables desde la evidencia empírica. Esta flexibilidad lo hace especialmente útil en contextos donde la volatilidad, la heterocedasticidad y el ruido dificultan el uso de enfoques clásicos (Domingos, 2012).

En este estudio se emplea un enfoque de aprendizaje supervisado, donde los modelos se entrenan con datos históricos del mercado bursátil español para predecir la dirección diaria del precio de cada acción. Se seleccionan cinco algoritmos representativos por su complementariedad y su desempeño comprobado:

- **Regresión logística**, por su sencillez y su capacidad para estimar probabilidades direccionales.
- **Árboles de decisión**, que generan reglas interpretables.
- **Random forest**, que combina múltiples árboles para mejorar la estabilidad y reducir la varianza.
- **K-nearest neighbors (KNN)**, que clasifica según los vecinos más cercanos.
- **XGBoost**, un modelo de ensamble secuencial que optimiza errores y suele ofrecer alto rendimiento (Fernández *et al.*, 2014).

En el ámbito financiero, el ML se utiliza para predecir movimientos de precios, detectar tendencias y ajustar portafolios de forma dinámica. Su fortaleza radica en la capacidad de adaptarse a nuevas condiciones del mercado y modelar relaciones que los métodos

tradicionales no capturan con precisión. No obstante, requiere una adecuada selección de variables y validación rigurosa para evitar el sobreajuste.

Revisión de la literatura

En los últimos años, el uso de modelos de ML ha cobrado gran relevancia en el ámbito financiero, especialmente en tareas como la predicción de precios y la construcción de carteras de inversión. Esto se debe a su capacidad para identificar patrones complejos en grandes volúmenes de datos, adaptarse a condiciones cambiantes y superar algunas limitaciones de los enfoques tradicionales como el de Markowitz.

Aplicaciones de ML a la construcción de portafolios

Echeverri (2021) desarrolló un sistema automatizado para conformar portafolios mediante analítica financiera, evaluando distintos algoritmos de ML según su capacidad para generar señales de entrada y salida. Entre ellos, XGBoost obtuvo los mejores resultados en precisión y rentabilidad ajustada al riesgo, al identificar con mayor eficacia los momentos óptimos de inversión.

Por su parte, Albelda (2021) analizó la posibilidad de replicar el comportamiento del índice bursátil alemán DAX mediante carteras construidas con técnicas de ML. Al comparar modelos como *lasso regression*, *random forest* y *autoencoders*, encontró que *random forest* fue el más eficiente, al seleccionar activos y anticipar movimientos del mercado, alcanzando una rentabilidad superior y menor volatilidad respecto al índice de referencia.

En conjunto, estos estudios concluyen que los modelos predictivos, combinados con reglas claras de gestión del capital, permiten crear carteras dinámicas que, bajo ciertas condiciones de mercado, superan a las estrategias tradicionales.

Predicción de precios mediante ML e indicadores técnicos

Diversos estudios han explorado con éxito el uso del aprendizaje automático en la predicción de precios bursátiles, integrando indicadores técnicos y fuentes alternativas de información. Saetia y Yokrattanasak (2022) analizaron modelos de clasificación para anticipar la dirección del mercado accionario tailandés, combinando indicadores técnicos con datos de Google Trends sobre el comportamiento de búsqueda de los inversores. Entre los algoritmos probados, SVM, árboles de decisión, Naive Bayes y *random forest*, este último obtuvo la mayor precisión y estabilidad.

De forma similar, Chen *et al.* (2020) aplicaron *random forest* con análisis de sentimiento al mercado chino, demostrando que esta combinación mejoró la capacidad predictiva en contextos de alta incertidumbre, logrando *ratios* de Sharpe más altas y una menor frecuencia de pérdidas que los modelos tradicionales.

Finalmente, Echeverri (2021) comparó modelos complejos como XGBoost y redes neuronales con técnicas más clásicas como la regresión logística y KNN, concluyendo que estas últimas también ofrecen resultados sólidos y consistentes, especialmente para predecir movimientos direccionales en acciones del Colcap colombiano.

En conjunto, la literatura muestra que tanto los algoritmos avanzados como los más simples pueden generar valor cuando se entrenan con datos adecuados, reforzando la idea de que no existe un único modelo óptimo, sino enfoques complementarios que se adaptan a distintas condiciones de mercado.

Metodología

Este trabajo corresponde a una investigación aplicada que compara el desempeño de dos estrategias de inversión: una basada en ML y otra en el modelo de Markowitz con optimización cuadrática, aplicadas al mercado bursátil español entre 2014 y 2024.

Primero, se describe el proceso de entrenamiento y selección de los modelos predictivos para estimar la dirección diaria del precio de las acciones, a partir de los cuales se construyó una estrategia de *trading* automática. Paralelamente, se implementó el modelo de Markowitz con rebalanceo diario y límites de inversión, con el fin de reflejar condiciones realistas del mercado.

Finalmente, se presentan las métricas utilizadas para evaluar y comparar ambas estrategias, considerando no solo la rentabilidad, sino también el nivel de riesgo asumido durante el periodo analizado.

Selección de acciones

El estudio se centra en acciones del mercado bursátil español. Para aplicar los modelos predictivos de forma rigurosa y realizar simulaciones de portafolios fiables, fue necesario depurar el conjunto inicial de activos, garantizando la calidad y la consistencia de las series temporales.

Tras este proceso, se identificaron 26 acciones que cumplieron con los requisitos exigidos:

- Disponibilidad continua de datos diarios entre el 2 de enero de 2014 y el 31 de diciembre de 2024.
- Información completa sobre los precios, el volumen y las variables técnicas requeridas para el modelado.

Esta depuración asegura la estabilidad de las simulaciones y evita sesgos por series incompletas o interrupciones en la cotización. Las acciones seleccionadas se muestran a continuación:

Tabla 1. Acciones

Ticker	Empresa
ITX.MC	Inditex
IBE.MC	Iberdrola
SAN.MC	Banco Santander S. A.
BBVA.MC	Banco Bilbao Vizcaya Argentaria
CABK.MC	Caixabank S. A.
NTGY.MC	Naturgy Energy Group
TEF.MC	Telefónica
ELE.MC	Endesa
FER.MC	Ferrovial S. A.
REP.MC	Repsol YPF
ACS.MC	Actividades de Construcción y Servicios
SABE.MC	Banco Sabadell
REDE.MC	Red Eléctrica Corporación
MAP.MC	Mapfre S. A.
BKT.MC	Bankinter S. A.
ANA.MC	Acciona S. A.

GRLS.MC	Grifols
MTS.mc	ARCELORMITTAL S. A.
FLUI.MC	Fluidra S. A.
COL.MC	Inmobiliaria Colonial
ROVI.MC	Laboratorios Farmacéuticos Rovi S. A.
ENAG.MC	Enagas
IDR.MC	Indra Sistemas
SCYR.MC	Sacyr
ACX.MC	Acerinox
SLRS.MC	Solaria Energía y Medioambiente S. A.

Periodo de estudio

El periodo de estudio abarca desde el 2 de enero de 2014 hasta el 31 de diciembre de 2024, lo que representa una década de observaciones diarias del mercado español. Esta ventana temporal permite analizar diferentes fases económicas: expansión, corrección y recuperación, e incorporar eventos relevantes como el Brexit, la pandemia del covid-19, el conflicto en Ucrania y las decisiones del Banco Central Europeo sobre política monetaria.

La información proviene de la plataforma Refinitiv Eikon, que ofrece cobertura amplia y datos de alta calidad, garantizando la granularidad necesaria para entrenar los modelos predictivos y simular carteras con precisión.

Metodología para predecir las direcciones futuras de las acciones con ML

El objetivo principal de esta etapa es predecir, de forma diaria, la dirección futura del precio de cada acción, mediante la aplicación de modelos de ML. A partir de estas predicciones, se generan señales de compra que sirven como base para la construcción de carteras de inversión dinámicas, las cuales son evaluadas en términos de rentabilidad y riesgo.

Este enfoque se basa en la propuesta de Echeverri (2021), adaptada y ampliada para el contexto del mercado español. Las predicciones se realizan de forma individual para cada una de las 26 acciones seleccionadas, entrenando un modelo específico por activo y evaluando su rendimiento a partir de datos históricos diarios.

Problemas de pronósticos

En los últimos años, la aplicación de métodos computacionales en los mercados financieros ha aumentado de forma notable, impulsada por tecnologías que facilitan el procesamiento de grandes volúmenes de datos y la detección de patrones relevantes. Entre estas herramientas, los modelos de ML han ganado protagonismo, especialmente en tareas de predicción aplicadas a series temporales bursátiles. Esta tendencia ha sido documentada por Aloud (2020), quien evidenció el potencial de estas técnicas para anticipar movimientos del mercado a partir de datos históricos.

Entre los instrumentos financieros, las acciones se destacan por su potencial de rentabilidad, aunque su volatilidad implica un riesgo elevado. Esto ha impulsado el desarrollo de enfoques analíticos que buscan anticipar la evolución de los precios con base en información pasada. En la literatura se distinguen dos estrategias principales: una orientada a estimar el valor exacto del precio futuro y otra centrada en predecir su dirección, es decir, si el precio subirá o bajará en la siguiente jornada.

Este estudio adopta el segundo enfoque, formulando el problema como una tarea de clasificación binaria dentro del aprendizaje supervisado. El objetivo es predecir si el precio de cierre del día siguiente será mayor o menor que el actual. Esta elección combina la viabilidad técnica con el respaldo teórico, dado que trabajos como los de Chen *et al.* (2003) y Huang *et al.* (2005) demuestran que la predicción direccional suele ofrecer mejores resultados en términos de estabilidad y rentabilidad que la estimación del precio exacto.

Variable objetivo

La variable objetivo representa el fenómeno que se busca predecir: la dirección futura del precio de cada acción en el mercado bursátil español. Se construye a partir de los precios de cierre diarios y se define como una variable binaria que indica si el precio sube o baja en la siguiente sesión, de acuerdo con la siguiente expresión:




$$Y_t = \begin{cases} 1 & \text{si } P_{t+1} > P_t \text{ (Sube)} \\ 0 & \text{si } P_{t+1} < P_t \text{ (Baja)} \end{cases}$$



Donde:

- P_t : es el precio de cierre de la acción en el día t .

En la siguiente tabla se muestra un ejemplo de cómo se definió esta variable objetivo para el caso de Inditex.

Tabla 2: Ejemplo de definición de la variable objetivo

Ticker	Fecha	Precio cierre	Tendencia	Dirección
ITX.MC	41641	23,72	1	
ITX.MC	41642	23,78	0	
ITX.MC	41645	23,66	1	

ITX.MC	41646	23,89	0	
ITX.MC	41647	23,73	0	

Variables predictoras

El conjunto de variables predictoras empleadas para entrenar los modelos de ML integra tanto información de carácter microeconómico, específica de cada acción, como macroeconómico, relacionada con las condiciones generales del mercado. Esta combinación permite captar patrones temporales, señales técnicas y factores contextuales que pueden incidir en la evolución direccional del precio. Para su análisis, las variables se agrupan en tres categorías principales:

I. Variables de mercado específicas de cada acción:

Este grupo de variables captura el comportamiento reciente de cada activo a partir de sus precios y del volumen de negociación.

- Precio de cierre: P_t .
- Máximo diario: M_t .
- Mínimo diario: L_t .
- Volumen de negociación: V_t .

II. Variables de contexto de mercado (*benchmarks*)

Para incorporar información relevante sobre el entorno macroeconómico y las condiciones generales del mercado, se incluyen variables representativas tanto del ámbito global como del regional. Las variables son:

- Yield Bonos Españoles a 10 años: B_t como una variable proxy a la tasa de interés.
- Precio de cierre EXSA: ETF_t es un ETF que replica el STOXX Europe 600.
- Precio de cierre EUR/USD: Fx_t .

Con el propósito de capturar de forma más completa la dinámica temporal de los mercados financieros, se incorporaron no solo las variables actuales, sino también sus retardos de hasta quince días. En concreto, cada variable predictora, ya sea un precio, un volumen, un indicador técnico o una referencia externa de mercado (*benchmark*), se incluyó tanto en su valor actual como en sus versiones rezagadas desde $t - 1$ hasta el $t - 15$.

Esta estrategia se aplicó de manera uniforme, lo que permite a los modelos identificar patrones de corto plazo, como movimientos de *momentum*, cambios incipientes de tendencia o efectos diferidos que se manifiestan con varios días de desfase. Además, refuerza la capacidad del sistema para reconocer secuencias temporales para anticipar variaciones en la dirección del precio, como señalan Fischer y Krauss (2018).

III. Indicadores técnicos

Se incorporan indicadores ampliamente utilizados en el análisis técnico. Estos indicadores resumen aspectos clave del comportamiento del activo, como la tendencia, el *momentum* y la presión de compra o venta. Por lo tanto, los indicadores usados son:

- **Media móvil exponencial (EMA 50 y EMA10):** refleja la tendencia del precio, asignando mayor peso a los valores más recientes. Se calcula como:

$$EMA_t = \alpha * P_t + (1 - \alpha) * EMA_{t-1}, \text{ con } \alpha = \frac{2}{n + 1}$$

Donde P_t es el precio de cierre y n el número de días, por ejemplo EMA (10) o EMA (50).

- **Índice de fuerza relativa (RSI):** mide la relación entre las ganancias y las pérdidas recientes, identificando posibles condiciones de sobrecompra o sobreventa:

$$RSI_t = 100 - \left(\frac{100}{1 + RS_t} \right), RS_t = \frac{\text{Promedio de ganancias}}{\text{Promedio de pérdidas}}$$

Se utilizará con una ventana de 14 días, ya que es lo que comúnmente se usa.

- **MACD (*moving average convergence divergence*):** indica cambios en la fuerza y la dirección de la tendencia. Se define como:

$$MACD_t = EMA_{12,t} - EMA_{26,t}$$

Además, se acompaña de una línea de señal:

$$Señal_t = EMA_9(MACD_t)$$

- **Rate of change (ROC):** mide la variación porcentual del precio con respecto a su valor n días atrás (12 días), capturando el ritmo de cambio reciente. Se define como:

$$ROC_t = \left(\frac{P_t - P_{t-n}}{P_{t-n}} \right) * 100$$

Modelos de pronósticos

En este estudio se emplearán cinco modelos supervisados de clasificación, seleccionados tanto por su diversidad metodológica como por su uso probado en aplicaciones financieras. Cada modelo se entrena por separado con los datos históricos de cada acción, lo que permite capturar patrones particulares del comportamiento de cada activo.

Los modelos utilizados son los siguientes:

- Regresión logística.
- Árbol de decisión.
- *Random forest*.
- *K-nearest neighbors* (KNN).
- XGBoost (*extreme gradient boosting*).

Entrenamiento de modelos

El entrenamiento de los modelos de ML se realiza de forma individual para cada acción, utilizando las variables previamente descritas. Este enfoque permite adaptar los modelos a las particularidades de cada activo y capturar mejor sus patrones de comportamiento.

Dado que se trata de series temporales, la división entre entrenamiento y prueba se efectúa de manera secuencial, no aleatoria. El 80 % inicial del periodo se destina al entrenamiento y el 20 % restante a la evaluación, evitando fugas de información y garantizando que las predicciones se basen únicamente en datos disponibles hasta ese momento.

Antes de la partición, todas las variables predictoras se normalizan mediante la técnica de estandarización (*z-score*), que transforma cada valor en función de la media y la desviación estándar:

$$z_i = \frac{x_i - \mu_x}{\sigma_x}$$

donde x_i es el valor observado, μ_x la media y σ_x la desviación estándar de la variable x .

El procedimiento general seguido para el entrenamiento de los modelos incluye:

1. **Selección de variables:** Se construye la matriz de características para cada acción, incorporando los retardos del precio, el volumen, los indicadores técnicos y las variables de referencia.
2. **Normalización:** Las variables predictoras se estandarizan antes de dividir los datos, asegurando la coherencia entre los conjuntos de entrenamiento y prueba.
3. **Entrenamiento individual:** Para cada acción se ajustan cinco modelos (uno por algoritmo), empleando el mismo conjunto de variables.
4. **Predicción y evaluación:** Los modelos se prueban sobre el conjunto de evaluación y se comparan mediante la métrica de *balanced accuracy*, adecuada para clases desbalanceadas. El modelo con mejor desempeño se selecciona para las predicciones finales.

Métricas de evaluación: balanced accuracy

Para evaluar el desempeño de los modelos de clasificación binaria, se emplea la métrica *balanced accuracy*. Esta medida resulta adecuada en contextos como el presente, donde puede existir un desbalance entre clases (ver anexo 3), es decir, cuando una de las dos situaciones, la subida o la bajada del precio, ocurre con mayor frecuencia.

A diferencia de la precisión tradicional (*accuracy*), que puede verse afectada por la clase mayoritaria, la *balanced accuracy* corrige este sesgo al calcular el promedio entre la sensibilidad (capacidad de detectar correctamente las subidas) y la especificidad (capacidad de identificar correctamente las bajadas), según la siguiente expresión:

$$\text{Balanced accuracy} = \frac{TP/(TP + FN) + TN/(TN + FP)}{2} = \frac{\text{Sensibilidad} + \text{Especificidad}}{2}$$

Donde:

- TP (*true positives*): Subidas clasificadas correctamente.
- FN (*false negatives*): Subidas clasificadas como bajadas.
- TN (*true negatives*): Bajadas clasificadas correctamente.
- FP (*false positives*): Bajadas clasificadas como subidas.

El uso de esta métrica permite valorar de forma más justa la capacidad del modelo para distinguir ambos escenarios. Por ello, la *balanced accuracy* se adopta como criterio para seleccionar el mejor modelo por acción.

Validación cruzada

El rendimiento de un modelo depende tanto de la calidad de los datos como del ajuste adecuado de sus hiperparámetros. Para optimizarlos se aplicó validación cruzada, enfocada por eficiencia en las dos acciones con peor desempeño inicial mediante GridSearchCV. Los parámetros óptimos obtenidos se extrapolaron al resto, garantizando la coherencia y el ahorro computacional.

Se utilizó validación cruzada estratificada de tres particiones (k-fold) para mantener el equilibrio entre clases y evaluar el desempeño con la métrica de *balanced accuracy*. Los valores finales se aplicaron globalmente, buscando mejorar la precisión sin comprometer la viabilidad del proceso.

Predicción final

Tras seleccionar el modelo con mejor desempeño para cada acción, se reentrenó con todos los datos históricos disponibles, omitiendo la partición de validación, ya evaluada previamente. Esto permitió aprovechar toda la información y aumentar la precisión de las estimaciones.

Cada modelo genera luego una predicción diaria sobre la dirección del precio (sube o baja), que sirve como insumo principal para la estrategia de inversión y determina las decisiones de compra o venta del portafolio.

Estrategia de *trading* basada en ML

Con las predicciones diarias sobre la dirección del precio para cada una de las 26 acciones analizadas, se implementa una estrategia de *trading* que traduce dichas señales en decisiones concretas de inversión. Esta estrategia combina los resultados de los modelos de clasificación con un conjunto de reglas sobre la asignación de capital, la gestión del riesgo y la selección de activos para la cartera.

La lógica central es simple: se ejecutan órdenes de compra cuando la señal indica subida, y órdenes de venta cuando señala bajada o se alcanza la pérdida máxima permitida. Bajo esta premisa, se definen parámetros operativos orientados a aprovechar las señales más rentables, evitar concentraciones excesivas en un solo activo y proteger el portafolio frente a operaciones fallidas mediante mecanismos como el *stop-loss*.

En los apartados siguientes se detallan los principales componentes de la estrategia, comenzando por la generación de señales de compra y las reglas que guían las decisiones diarias de inversión.

Generación de señales de trading

La estrategia de inversión se construye a partir de las predicciones generadas por los modelos de ML. Estas predicciones indican, para cada acción y cada día, si se espera una subida (1) o una bajada (0) en el precio al cierre de la siguiente jornada. A partir de esta información, se generan señales operativas que permiten decidir si se debe comprar, mantener o vender una posición.

La lógica que rige la generación de señales se define mediante la siguiente regla:

$$S_t = \begin{cases} Pr_t & \text{si } t = 1 \\ Pr_t - Pr_{t-1} & \text{si } t \neq 1 \end{cases}$$

Donde:

- S_t es la señal generada para el día t , con los siguientes valores posibles:
 - $S_t = 1$: Señal de compra.
 - $S_t = 0$: Señal de mantener.
 - $S_t = -1$: Señal de venta.
- Pr_t es la predicción generada por el modelo para el día t .

A continuación, se muestra un ejemplo aplicado a la acción Ferrovial SE para ilustrar cómo se traduce esta lógica en decisiones operativas (tabla 3).

Tabla 3. Decisión de compra y venta para la acción ferroviaria

Ticker	t	Fecha	Predicción	Señal	Acción
FER.MC	1	2/01/2014	1	1	Compro
FER.MC	2	3/01/2014	0	-1	Vendo
FER.MC	3	6/01/2014	1	1	Compro
FER.MC	4	7/01/2014	1	0	Mantengo
FER.MC	5	8/01/2014	0	-1	Vendo

Reglas de negociación

Las decisiones dentro de la estrategia siguen un conjunto de reglas claras que definen cómo se gestionan las posiciones, el capital y la operativa diaria. Estas normas permiten evaluar el impacto real de las predicciones generadas por los modelos, eliminando cualquier intervención discrecional y asegurando la coherencia del sistema. Las principales reglas son las siguientes:

- **Una única posición activa por activo:** No se permiten compras adicionales sobre un mismo valor si ya existe una posición abierta.
- **Operaciones múltiples y evaluación diaria:** A lo largo del periodo de simulación (2014-2024) pueden ejecutarse tantas operaciones como lo permitan las señales y el capital disponible.
- **Reinversión de beneficios:** Los beneficios obtenidos por operaciones exitosas se reinvierten automáticamente, incrementando el capital disponible para futuras compras.
- **Capital inicial fijo y sin aportes adicionales:** La estrategia parte de un capital inicial definido y no contempla nuevas inyecciones de fondos durante la simulación.
- **Actualización diaria del portafolio:** La composición del portafolio se revisa diariamente. Esta actualización continua permite responder con agilidad a las señales generadas por los modelos.

En conjunto, estas reglas aseguran una ejecución disciplinada de la estrategia, permitiendo evaluar con transparencia y precisión hasta qué punto la capacidad predictiva de los modelos de ML se traduce en decisiones de inversión efectivas.

Parámetros de negociación

Para simular de forma realista la implementación de la estrategia de inversión basada en señales generadas por modelos de ML, se definen los siguientes parámetros de negociación. Estos condicionan la operativa diaria, establecen límites prudenciales y permiten evaluar la solidez de la estrategia en distintos contextos de mercado:

- **Capital inicial:** La simulación parte de un capital inicial de 1 000 000 de euros, sin aportes adicionales a lo largo del periodo.
- **Monto mínimo de inversión por operación:** Se establece un umbral mínimo de 50 000 euros por transacción. Esta regla busca evitar operaciones poco relevantes desde el punto de vista financiero.
- **Pérdida máxima permitida:** Para controlar el riesgo de cada operación, se aplica una pérdida máxima del 3 % respecto al precio de entrada. Si una posición alcanza esa pérdida latente, se liquida automáticamente, independientemente de la señal generada por el modelo.
- **Límite de inversión por operación:** No se permite invertir más del 40 % del capital disponible en una sola operación. Esta restricción favorece la diversificación del portafolio.
- **Precios de negociación:** Las compras y las ventas se ejecutan al precio de cierre de cada jornada.
- **Costes de transacción:** Para valorar el impacto de la fricción del mercado, se simulan cuatro escenarios con distintos niveles de coste por operación: 0, 1, 2 y 3 puntos básicos (pbs).

- **Priorización de compra:** En caso de recibir múltiples señales de compra en un mismo día, las acciones se ordenan según su nivel de capitalización bursátil al 31 de diciembre de 2024 (tabla 4). El capital se asigna primero a las compañías con mayor capitalización.

Tabla 4. Priorización de compra (millones de euros)

Empresa	Capitalización	Priorización de compra
Inditex	154 710,60 €	1
Iberdrola	84 644,50 €	2
Banco Santander S. A.	67 648,30 €	3
Banco Bilbao Vizcaya Argentaria	54 474,60 €	4
Caixabank S. A.	37 568,00 €	5
Naturgy Energy Group	22 669,60 €	6
Telefónica	22 323,40 €	7
Endesa	21 990,30 €	8
Ferrovial S. A.	19 183,60 €	9
Repsol YPF	13 763,80 €	10
Actividades de Construcción y Servicios	13 159,40 €	11
Banco Sabadell	10 211,30 €	12
Red Eléctrica Corporación	8 927,80 €	13
Mapfre S. A.	7 532,60 €	14
Bankinter S. A.	6 867,30 €	15

Acciona S. A.	5 962,30 €	16
Grifols	5 745,20 €	17
ARCELORMITTAL S. A.	5 727,40 €	18
Fluidra S. A.	4 518,90 €	19
Inmobiliaria Colonial	3 246,50 €	20
Laboratorios Farmacéuticos Rovi S. A.	3 225,30 €	21
Enagas	3 086,20 €	22
Indra Sistemas	3 017,30 €	23
Sacyr	2 481,70 €	24
Acerinox	2 356,20 €	25
Solaria Energía y Medioambiente S. A.	976,50 €	26

Fuente: Informe Anual de Mercados 2024 de la CNMV.

La estrategia de inversión desarrollada en este apartado combina modelos de predicción basados en ML con una política de gestión activa estructurada con base en reglas operativas claras y parámetros de control de riesgo bien definidos. Esta configuración permite evaluar de forma realista la viabilidad de tomar decisiones de inversión automatizadas en el contexto del mercado bursátil español. Además, establece una base sólida para comparar su desempeño con enfoques más tradicionales de asignación de activos, como el modelo de Markowitz, que se presenta a continuación.

Estrategias de inversión con el modelo de Markowitz

Como punto de comparación frente a la estrategia basada en modelos de ML, se implementa un enfoque clásico de optimización de carteras, desarrollado originalmente por Harry Markowitz. Esta estrategia se adapta a un entorno operativo realista en el mercado bursátil español, y se

centra en minimizar el riesgo del portafolio (medido como la varianza de los rendimientos) bajo un conjunto de restricciones prácticas y con rebalanceo diario.

Construcción y rebalanceo diario del portafolio

La simulación comienza el 1 de enero de 2014 con un capital inicial de un millón de euros. A partir de esa fecha y en cada jornada bursátil, se evalúa si hay suficiente información disponible para aplicar el modelo. Para ello, se utiliza una ventana móvil de 5 años (alrededor de 1 260 días), con la que se calculan los rendimientos diarios, la media esperada y la matriz de varianzas y covarianzas entre los activos.

Con estos datos, se plantea un problema de optimización cuadrática, cuyo objetivo es minimizar la varianza del portafolio, sujeto a las siguientes restricciones:

- El peso de cada activo debe situarse entre 0 y un máximo permitido (generalmente no superior al 40 %), definido en función de los resultados de la estrategia de ML.
- La suma total de los pesos debe ser igual a 1, lo que implica un portafolio completamente invertido.
- No se permiten posiciones cortas (todos los pesos deben ser no negativos).

Implementación operativa y restricciones de inversión:

Una vez determinados los pesos óptimos, se calcula el número de acciones a comprar o vender para ajustarse a esa distribución, utilizando el precio de cierre de la jornada. Para cada operación, se respeta la restricción de no invertir más del 40 % del capital disponible en un solo activo, lo que favorece la diversificación y evita la concentración excesiva del riesgo.

Asimismo, se contemplan diferentes escenarios de costes de transacción (0, 1, 2 y 3 puntos básicos), que se descuentan en cada operación para evaluar su impacto sobre la rentabilidad neta del portafolio.

El capital no invertido se mantiene en liquidez hasta el siguiente día de negociación. Este proceso se repite de forma diaria hasta el cierre del periodo analizado, lo que permite simular una estrategia de asignación dinámica que se ajusta constantemente al comportamiento histórico de los activos.

Métricas de desempeño

Las métricas de desempeño sirven para evaluar y comparar el rendimiento de los portafolios generados por ambos enfoques, el que se basa en modelos de ML y el que se basa en el modelo de Markowitz. Esto se hace mediante las siguientes métricas.

Rentabilidad total: Mide el crecimiento porcentual del valor del portafolio entre el inicio y el final del periodo analizado:

$$Rentabilidad\ total = \frac{V_f - V_i}{V_i}$$

Donde:

- V_f : Valor final del portafolio.
- V_i : Valor inicial del portafolio.

1. **Tasa de crecimiento anual compuesta (TCAC):** Refleja la rentabilidad promedio anual del portafolio, considerando el efecto del interés compuesto a lo largo del tiempo:

$$TCAC = \left(\frac{V_f}{V_i} \right)^{\frac{1}{T}} - 1$$

Donde:

- T : número de años, calculado como el número total de días dividido por 252.
2. **Volatilidad anualizada:** Representa la variabilidad de los retornos diarios, escalada a una base anual. Se calcula como:

$$\text{Volatilidad anual} = \sigma_d * \sqrt{252}$$

Donde:

- σ_d : Desviación estándar de los retornos diarios.
3. **Ratio de Sharpe:** Evalúa la rentabilidad del portafolio ajustada al riesgo, tomando como referencia la tasa libre de riesgo. Cuanto mayor es la *ratio*, mejor ha sido la rentabilidad en relación con la volatilidad asumida:

$$\text{Ratio de Sharpe} = \frac{\overline{r_p - r_f}}{\sigma_p} * \sqrt{252}$$

Donde

- σ_p : Desviación estándar diaria del portafolio.
 - $\overline{r_p - r_f}$: Media de los excesos de los retornos diarios.
4. **Máximo drawdown (MDD):** Indica la mayor pérdida relativa sufrida por el portafolio desde un pico hasta un valle durante el periodo analizado. Es una métrica clave para valorar la exposición a caídas abruptas:

$$MDD = \max_{t \in [0, T]} \left(\frac{\max_{\tau \in [0, t]} V(\tau) - V(t)}{\max_{\tau \in [0, t]} V(\tau)} \right)$$

Donde:

- $\max_{\tau \in [0, t]} V(\tau)$: Valor máximo alcanzado por el portafolio hasta el tiempo t .
- $V(t)$: Valor del portafolio en el tiempo t .

Estas métricas permiten realizar una evaluación integral del comportamiento de los portafolios, teniendo en cuenta tanto la rentabilidad generada como la estabilidad y la eficiencia en la gestión del riesgo. Serán utilizadas para comparar los resultados de ambas estrategias bajo distintos escenarios de costes de transacción.

Resultados

En este capítulo se presentan los resultados obtenidos al aplicar las dos estrategias de inversión propuestas: una basada en modelos de ML y otra en el modelo clásico de Markowitz. El objetivo es analizar el comportamiento de cada enfoque durante el periodo 2014-2024 y evaluar su respuesta bajo diferentes escenarios, incluyendo la presencia de costos de transacción.

Selección del mejor modelo

Tras completar el proceso de entrenamiento y evaluación, se seleccionó el modelo con mejor desempeño para cada una de las 26 acciones analizadas. El criterio principal fue la métrica de *balanced accuracy*, por su capacidad para evaluar de forma equitativa el rendimiento en escenarios con clases desbalanceadas, como ocurre en este estudio.

La siguiente tabla presenta el modelo que obtuvo los mejores resultados para cada activo, es decir, aquel con mayor capacidad predictiva sobre la dirección del precio para el día siguiente.

Tabla 5. Mejor modelo por acción

Acción	Mejor modelo
Inditex	NeuralNetwork
Iberdrola	DecisionTree
Banco Santander S. A.	NeuralNetwork
Banco Bilbao Vizcaya Argentaria	NeuralNetwork
Caixabank S. A.	LogisticRegression
Ferrovial S. A.	DecisionTree
Telefónica	XGBoost
Repsol YPF	RandomForest
Actividades de Construcción y Servicios	KNN
Banco Sabadell	LogisticRegression
Red Eléctrica Corporación	DecisionTree
Endesa	KNN
Bankinter S. A.	NeuralNetwork
Acciona S. A.	XGBoost
Naturgy Energy Group	XGBoost
Mapfre S. A.	RandomForest
Grifols	RandomForest
ARCELORMITTAL S. A.	LogisticRegression

Enagas	RandomForest
Indra Sistemas	DecisionTree
Fluidra S. A.	DecisionTree
Sacyr	RandomForest
Acerinox	KNN
Laboratorios Farmacéuticos Rovi S. A.	DecisionTree
Inmobiliaria Colonial	RandomForest
Solaria Energía y Medioambiente S. A.	LogisticRegression

Resultados del análisis de sensibilidad para la inversión máxima

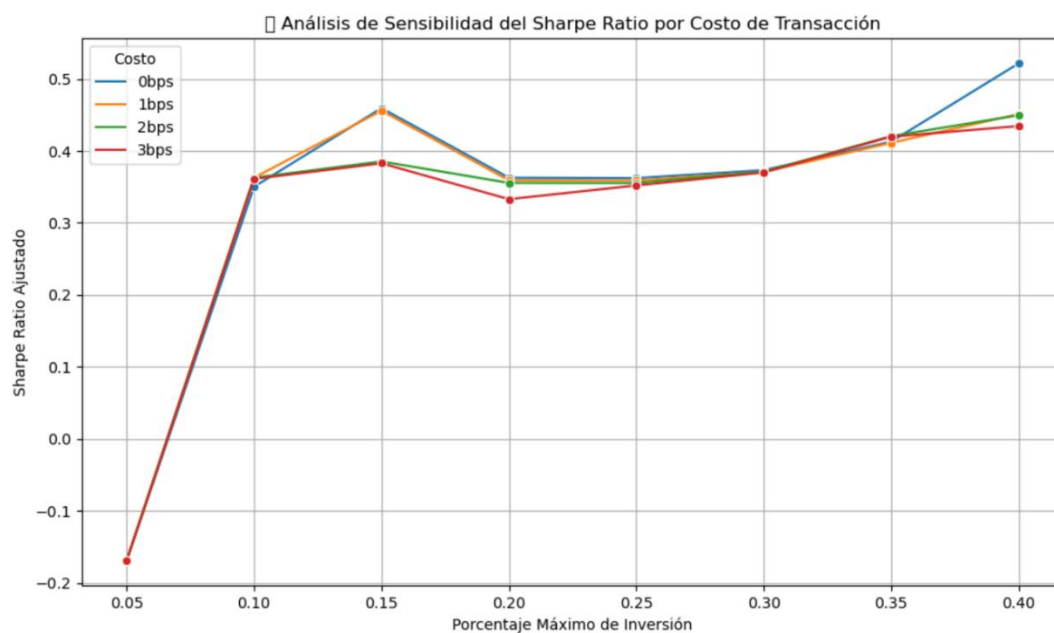
Los resultados indican que un límite muy bajo, como el 5 %, afecta negativamente el rendimiento: incluso sin costos, la *ratio* de Sharpe cae por debajo de cero. Esto evidencia que una diversificación excesiva reduce la capacidad del modelo para aprovechar oportunidades relevantes sin disminuir realmente el riesgo. A partir del 10 %, el desempeño mejora, destacándose el 15 % como nivel más eficiente sin costos, con un Sharpe superior a 0,45.

Al incorporar costos de transacción, el punto óptimo se desplaza hacia valores más altos. Con 1 pbs, el 15 % mantiene buen rendimiento; sin embargo, con 2 y 3 pbs los mejores resultados se ubican entre el 3 % y el 40 %. Esto sugiere que, ante las fricciones operativas, límites más amplios permiten conservar la eficiencia del portafolio.

En escenarios realistas con costos moderados, los valores entre 35 % y 40 % ofrecen el mejor equilibrio entre la diversificación, la rentabilidad y el riesgo. Por ello, se fija un límite máximo

del 40 %, que permite aprovechar las señales del modelo sin generar concentración excesiva y asegura un desempeño estable bajo condiciones de mercado reales.

Figura 1. Análisis de sensibilidad de la *ratio* de Sharpe según el porcentaje máximo de inversión y costos de transacción



Selección de la ventana móvil óptima

Con el fin de evaluar cómo influye el tamaño de la ventana móvil en el desempeño del portafolio optimizado con el modelo de Markowitz, se probaron cinco configuraciones que abarcaron entre 1 y 5 años de datos. Para cada caso se calcularon las métricas principales: rentabilidad total, TCAC, volatilidad y la *ratio* de Sharpe.

Según la tabla 6, la ventana de 5 años obtuvo el mejor desempeño general. Alcanzó la *ratio* de Sharpe más alta (0,55), igual a la de la ventana de 1 año, pero con una mayor rentabilidad: un TCAC de 9,22 % frente a 8,96 % y una rentabilidad acumulada del 167,84 % frente a 160,63 %. Además, su volatilidad fue ligeramente inferior (14,93 %), lo que refuerza un perfil riesgo-retorno más equilibrado.

Las ventanas intermedias de 2 a 4 años mostraron resultados más modestos, con las *ratios* de Sharpe entre 0,43 y 0,52 y menores tasas de crecimiento. En particular, la de 3 años fue la menos eficiente, con la *ratio* de Sharpe más baja (0,43) y una rentabilidad total de 118,69 %.

En conclusión, la ventana móvil de 5 años se consolida como la opción más consistente tanto técnica como prácticamente. Este horizonte captura de forma más precisa la relación riesgo-rentabilidad y suaviza los efectos de la volatilidad puntual, con lo cual resulta especialmente adecuado para inversores que buscan estabilidad y una visión de largo plazo.

Tabla 6. Ventana móvil

Ventana (años)	Rentabilidad	TCAC	Volatilidad	Ratio de Sharpe
1	160,63 %	8,96 %	14,30 %	0,55
2	150,78 %	8,58 %	14,67 %	0,52
3	118,69 %	7,26 %	14,94 %	0,43
4	144,58 %	8,34 %	14,99 %	0,49
5	167,84 %	9,22 %	14,93 %	0,55

Comparación de las estrategias de inversión

Rentabilidad

La rentabilidad acumulada entre 2014 y 2024 evidencia diferencias claras entre las tres estrategias analizadas (tabla 7). En un escenario sin costos de transacción, la estrategia basada en ML alcanza el mejor resultado, con una rentabilidad de 180,52 %, y refleja su capacidad para aprovechar señales del mercado de forma dinámica y eficiente.

Sin embargo, esta ventaja disminuye al incluir costos operativos: con 1 pbs la rentabilidad cae a 145,66 %, y con 3 pbs desciende a 138,00 %. Aun así, se mantiene como una alternativa altamente rentable incluso con operaciones frecuentes.

La estrategia diaria basada en el modelo de Markowitz parte de 168,01 % sin costos, pero muestra una caída más pronunciada con el aumento de fricciones: 158,54 % con 1 pbs, 152,95 % con 2 pbs y 143,07 % con 3 pbs. Aunque no supera a ML en condiciones ideales, sí lo hace cuando los costos son significativos, demostrando mayor eficiencia estructural en entornos donde operar resulta más costoso.

El IBEX, utilizado como referencia pasiva, presenta el peor desempeño: 74,34 % sin costos y apenas 11,77 % con 3 pbs, confirmando su baja eficiencia para estrategias orientadas hacia el rendimiento ajustado al riesgo.

En conclusión, la estrategia de ML se destaca en escenarios sin fricciones, mientras que la de Markowitz muestra una mayor resistencia cuando los costos de transacción afectan la rentabilidad. La elección final dependerá del entorno operativo y del perfil del inversor.

Tabla 7. Rentabilidad total

Estrategia	Rentabilidad	Rentabilidad	Rentabilidad	Rentabilidad
	total 0pbs	total 1pbs	total 2pbs	total 3pbs
Trading	180,52 %	145,66 %	144,80 %	138,00 %
Markowitz Diaria	168,01 %	158,54 %	152,95 %	143,07 %
IBEX35	74,34 %	50,75 %	29,99 %	11,77 %

TCAC

El análisis del TCAC permite observar el ritmo de crecimiento anual de cada portafolio durante el periodo 2014-2024, incorporando el efecto del interés compuesto. En un escenario sin costos de transacción, la estrategia basada en ML obtiene el mejor resultado, con un crecimiento anual compuesto del 9,68 %, ligeramente superior al 9,23 % alcanzado por la estrategia de Markowitz (tabla 8).

Al incluir costos, esta ventaja se reduce. El rendimiento de la estrategia de ML desciende gradualmente: 8,38 % con 1 pbs, 8,35 % con 2 pbs y 8,08 % con 3 pbs. En cambio, la estrategia de Markowitz muestra mayor resistencia, con valores de 8,88 %, 8,67 % y 8,28 %, respectivamente.

Este comportamiento evidencia un patrón claro: la estrategia de ML sobresale en condiciones ideales, pero es más sensible a los costos; mientras que Markowitz mantiene una trayectoria más estable, lo que la hace atractiva para inversores que priorizan la consistencia en el tiempo.

En contraste, el IBEX exhibe un crecimiento mucho más limitado: 3,55 % sin costos y solo 0,70 % con 3 pbs. Estos resultados confirman que, durante el periodo analizado, la inversión pasiva en el índice fue significativamente menos eficiente que las estrategias activas.

Tabla 8. TCAC

Estrategia	TCAC 0pbs	TCAC 1pbs	TCAC 2pbs	TCAC 3pbs
Trading	9,68 %	8,38 %	8,35 %	8,08 %
Markowitz Diaria	9,23 %	8,88 %	8,67 %	8,28 %
IBEX35	3,55 %	2,61 %	1,66 %	0,70 %

Volatilidad

La tabla 9 presenta los niveles de volatilidad anual de cada estrategia considerando distintos escenarios de costos de transacción. Esta métrica resulta esencial para evaluar el riesgo asumido por cada enfoque a lo largo del tiempo.

La estrategia construida con el modelo de Markowitz muestra una ventaja clara: su volatilidad se mantiene estable alrededor del 14,9 %, sin verse afectada por los costos. Esto refleja una cartera sólida, bien diversificada y con un comportamiento predecible, lo que evidencia un control eficaz del riesgo incluso en condiciones operativas exigentes.

Por su parte, la estrategia basada en ML presenta una volatilidad ligeramente mayor, cercana al 17,4 %, también con poca variación entre escenarios. Aunque mantiene buenos resultados en rentabilidad, este nivel superior de riesgo puede deberse a una rotación más alta o a concentraciones derivadas de las señales del modelo.

El IBEX exhibe el mayor riesgo de los tres, con una volatilidad de 19,03 % sin costos y 16,17 % con 3 pbs. Esta reducción no implica mayor estabilidad, sino el efecto de los costos sobre los resultados.

En conjunto, además de su buen desempeño en cuanto a la rentabilidad, la estrategia de Markowitz se destaca por mantener un control constante del riesgo, consolidándose como la alternativa más equilibrada en la relación rentabilidad-riesgo.

Tabla 9. Volatilidad

Estrategia	Volatilidad 0pbs	Volatilidad 1pbs	Volatilidad 2pbs	Volatilidad 3pbs
Trading	17,27 %	17,38 %	17,40 %	17,47 %

Markowitz Diaria	14,93 %	14,93 %	14,89 %	14,92 %
IBEX35	19,17 %	18,09 %	17,08 %	16,17 %

Max drawdown

El máximo *drawdown* mide la mayor pérdida que puede sufrir un portafolio desde su punto más alto hasta un mínimo posterior. Es una métrica clave para evaluar la capacidad de una estrategia para proteger el capital en periodos adversos.

Según la tabla 10, la estrategia basada en ML es la que mejor contiene estas caídas extremas. Su *drawdown* se mantiene entre -29,17 % (sin costos) y -30,39 % con 3 pbs, mostrando una estabilidad destacable incluso al incluir costos de transacción. Este comportamiento refleja una buena capacidad de reacción ante escenarios negativos, limitando las pérdidas y evitando las caídas prolongadas.

La estrategia de Markowitz presenta descensos algo mayores, entre -31,59 % y -31,74 %, lo que sugiere una menor capacidad de ajuste en momentos de tensión, posiblemente por su baja rotación y la permanencia en activos que tardan más en recuperarse.

El IBEX, como referencia pasiva, es el más afectado, con *drawdowns* superiores al 40 % en todos los escenarios y un mínimo de -41,65 % con 3 pbs, lo que confirma su alta vulnerabilidad y la ausencia de mecanismos activos de gestión del riesgo.

En conjunto, la estrategia de ML se destaca por su capacidad para limitar pérdidas severas, convirtiéndose en una alternativa atractiva para inversores conservadores o enfocados en preservar el capital ante los entornos de alta volatilidad.

Tabla 10. Max drawdowns

Estrategia	Max drawdowns 0pbs	Max drawdowns 1pbs	Max drawdowns 2pbs	Max drawdowns 3pbs
Trading	-29,17 %	-29,43 %	-29,29 %	-30,39 %
Markowitz Diaria	-31,59 %	-31,64 %	-31,70 %	-31,74 %
IBEX35	-38,88 %	-40,24 %	-40,24 %	-41,65 %

Ratio de Sharpe

La *ratio* de Sharpe permite evaluar qué tan bien justifica una estrategia el riesgo asumido con la rentabilidad obtenida. Los resultados de la tabla 11 evidencian diferencias claras entre las tres estrategias analizadas.

La estrategia basada en el modelo de Markowitz se consolida como la más eficiente. Incluso con costos de transacción, mantiene un desempeño sólido: parte de 0,520 sin costos y apenas desciende a 0,492 con 3 pbs. Esta estabilidad confirma que el modelo asigna los pesos del portafolio de manera eficaz, logrando una relación riesgo-rentabilidad consistente.

La estrategia de ML también presenta resultados positivos, con *ratios* entre 0,521 y 0,434. Aunque su eficiencia es ligeramente menor, sigue siendo competitiva. Además, su capacidad de reacción ante señales negativas reflejada en un menor *drawdown* la hace especialmente atractiva para inversores que priorizan el control de pérdidas sobre la rentabilidad máxima.

Por el contrario, el IBEX muestra el desempeño más débil: parte de 0,265 y cae hasta 0,035 con los mayores costos, evidenciando que no compensa adecuadamente el riesgo asumido.

En conclusión, el análisis de la *ratio* de Sharpe refuerza las tendencias observadas: la estrategia de Markowitz ofrece la mejor eficiencia general, mientras que la de ML se posiciona como una alternativa sólida y flexible, especialmente en entornos volátiles. El IBEX, en cambio, confirma su carácter pasivo y su limitado rendimiento ajustado al riesgo.

Tabla 11. *Ratio* de Sharpe

Estrategia	<i>Ratio</i> de Sharpe 0pbs	<i>Ratio</i> de Sharpe 1pbs	<i>Ratio</i> de Sharpe 2pbs	<i>Ratio</i> de Sharpe 3pbs
Trading	0,52	0,45	0,45	0,43
Markowitz Diaria	0,55	0,53	0,52	0,49
IBEX35	0,27	0,20	0,12	0,04

Conclusiones

Los resultados de esta investigación muestran que los modelos de ML aplicados al mercado bursátil español pueden generar decisiones de inversión fundamentadas. Aunque no superan de forma sistemática al modelo clásico de Markowitz en cuanto a la rentabilidad total, logran construir carteras competitivas y, en algunos aspectos, más consistentes.

La estrategia de mínima varianza de Markowitz alcanzó la mayor rentabilidad acumulada y el mejor crecimiento anual compuesto (CAGR) entre 2014 y 2024, pero evidenció la alta

sensibilidad frente a los costos de transacción, que reducen considerablemente su rendimiento. Esto indica que su efectividad práctica depende de que se opere en entornos con ejecución eficiente y bajos costos.

Por su parte, la estrategia basada en ML, pese a partir de una rentabilidad ligeramente menor, mostró mayor solidez frente a los costos, manteniendo un desempeño estable incluso en escenarios con fricciones elevadas. En términos de riesgo, Markowitz construyó carteras más estables, pero el ML ofreció mejor protección del capital, registrando los menores *drawdowns* en todos los casos.

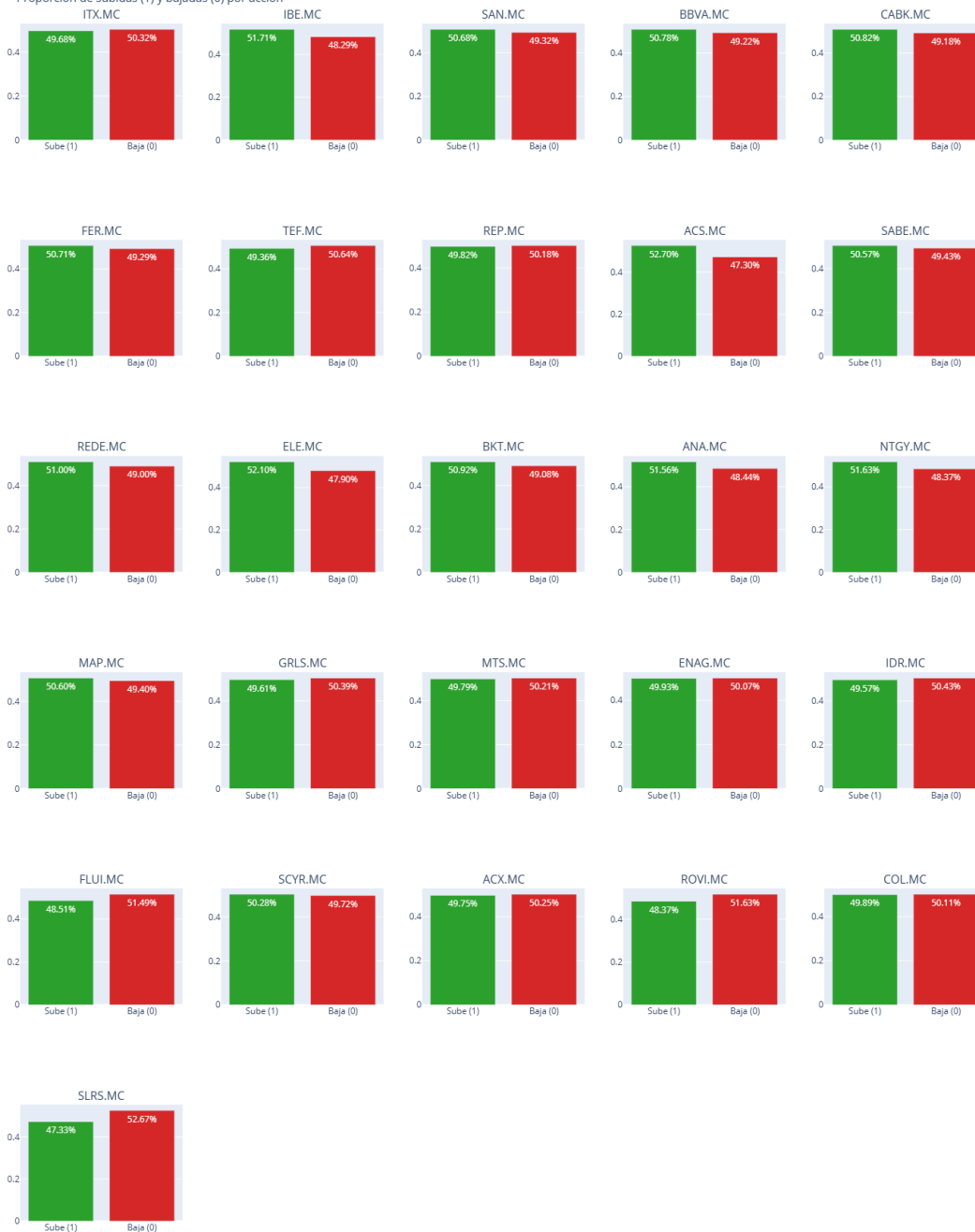
El análisis de la *ratio* de Sharpe refuerza esta conclusión: Markowitz es más eficiente en condiciones ideales, mientras que el ML mantiene una relación riesgo-retorno más equilibrada cuando se introducen los costos, consolidándose como una alternativa más robusta y operativamente eficiente.

En síntesis, para los inversores conservadores, Markowitz sigue siendo atractivo en entornos sin fricciones, mientras que el ML presenta ventajas claras en contextos reales con costos e incertidumbre. Una estrategia híbrida que combine la capacidad adaptativa del ML con la estructura de optimización de Markowitz podría representar una solución más completa. En conjunto, ambas estrategias superan al IBEX en rentabilidad y control del riesgo, diferenciándose según el perfil y las condiciones operativas del inversor.

Anexos

Anexo 1. Distribución de clases de las acciones

Proporción de subidas (1) y bajadas (0) por acción



Anexo 2. *Balanced accuracy score* de los modelos por acción

Acción	Logistic Regression	Random Forest	Decision Tree	KNN	Neural Network	XGBoost
ITX.MC	50,88 %	50,22 %	48,31 %	50,82 %	51,28 %	49,34 %
IBE.MC	51,17 %	50,30 %	51,79 %	49,53 %	50,14 %	50,50 %
SAN.MC	49,35 %	49,97 %	47,79 %	49,90 %	50,08 %	49,15 %
BBVA.MC	49,21 %	51,97 %	50,64 %	48,79 %	52,43 %	51,31 %
CABK.MC	52,84 %	50,92 %	48,00 %	51,48 %	50,23 %	50,24 %
FER.MC	50,48 %	50,84 %	52,01 %	51,84 %	50,77 %	50,69 %
TEF.MC	51,35 %	52,58 %	51,04 %	50,43 %	50,89 %	55,27 %
REP.MC	49,95 %	53,34 %	51,98 %	49,90 %	47,48 %	50,29 %
ACS.MC	50,48 %	50,09 %	48,81 %	50,59 %	49,93 %	49,39 %
SABE.MC	51,11 %	48,54 %	46,36 %	48,82 %	47,89 %	46,50 %
REDE.MC	50,18 %	49,75 %	50,86 %	42,83 %	46,94 %	49,73 %
ELE.MC	50,45 %	50,16 %	50,64 %	52,94 %	49,55 %	49,03 %
BKT.MC	51,88 %	52,74 %	51,52 %	51,67 %	55,71 %	50,48 %
ANA.MC	50,42 %	49,66 %	48,88 %	50,20 %	49,58 %	50,91 %
NTGY.MC	46,76 %	50,66 %	47,80 %	46,38 %	48,30 %	51,02 %
MAP.MC	51,90 %	51,99 %	50,38 %	48,50 %	50,80 %	50,81 %
GRLS.MC	48,66 %	51,28 %	49,75 %	49,43 %	49,08 %	51,03 %
MTS.MC	52,16 %	50,17 %	51,42 %	51,91 %	50,00 %	49,23 %

ENAG.MC	49,93 %	51,31 %	46,33 %	48,06 %	50,08 %	50,93 %
IDR.MC	50,33 %	50,92 %	51,69 %	50,31 %	47,66 %	49,86 %
FLUL.MC	50,95 %	50,56 %	52,24 %	51,85 %	50,73 %	50,06 %
SCYR.MC	47,07 %	51,27 %	50,19 %	49,00 %	50,97 %	50,19 %
ACX.MC	48,76 %	50,88 %	46,39 %	51,79 %	50,22 %	49,32 %
ROVL.MC	50,88 %	50,05 %	51,58 %	50,39 %	50,21 %	47,06 %
COL.MC	49,41 %	50,18 %	49,40 %	48,32 %	49,73 %	49,90 %
SLRS.MC	49,71 %	49,39 %	49,61 %	49,29 %	48,66 %	49,22 %

Anexo 3. Evolución de las estrategias a los diferentes costes de transacción

Evolución del Valor del Portafolio (0bps de Costo de Transacción)



Evolución del Valor del Portafolio (1bps de Costo de Transacción)



Evolución del Valor del Portafolio (2bps de Costo de Transacción)



Evolución del Valor del Portafolio (3bps de Costo de Transacción)



Anexo 3. Gráfico max *drawdown*



Referencias

- Albelda, A. (2021). Composición de cartera réplica del índice bursátil alemán DAX utilizando técnicas pertenecientes al *machine learning*: comparativa entre *lasso*, *random forest* y *autoencoder*. [Tesis de máster]. Universidad Complutense de Madrid. https://www.uv.es/bfc/TFM2021/Andres_Albelda.pdf.
- Chen, M., Zhang, Z., Shen, J., Deng, Z., He, J. y Huang, S. (2020). A quantitative investment model based on random forest and sentiment analysis. *Journal of Physics: Conference Series*, 1575. <https://doi.org/10.1088/1742-6596/1575/1/012083>.
- Chopra, V. K. y Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2), 6-11. <https://doi.org/10.3905/jpm.1993.409440>.
- Comisión Nacional del Mercado de Valores (2024). *Informe anual 2024*. https://www.cnmv.es/DocPortal/Publicaciones/Informes/Informe_Anual_ES_2024.pdf.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223-236. <https://doi.org/10.1080/713665670>.

- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87. <https://doi.org/10.1145/2347736.2347755>.
- Echeverri, L. C. (2021). *Conformación automática de portafolios de inversión usando analítica financiera*. Universidad Nacional de Colombia.
- Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3), 339-250. <https://doi.org/10.1198/073500102288618487>.
- Fischer, T. y Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669. <https://doi.org/10.1016/j.ejor.2017.11.054>.
- Mandelbrot, B. y Hudson, R. L. (2006). *The (mis)behavior of markets: A fractal view of financial turbulence*. Basic Books.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91. <https://doi.org/10.2307/2975974>.
- Michaud, R. O. (1989). The Markowitz optimization enigma: Is “optimized” optimal? *Financial Analysts Journal*, 45(1), 31-42. <https://doi.org/10.2469/faj.v45.n1.31>.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Saetia, K. y Yokrattanasak, J. (2022). Stock movement prediction using machine learning based on technical indicators and Google trend searches in Thailand. *International Journal of Financial Studies*, 11(1), 5. <https://doi.org/10.3390/ijfs11010005>.