

**PREDICCIÓN DE DIRECCIONES DE ACTIVOS FINANCIEROS BASADOS EN
LA VOLATILIDAD EN SERIES TEMPORALES UTILIZANDO MACHINE
LEARNING**

MATEO HOLGUIN CARVALHO

Proyecto de Grado

Asesor, docente:

Henry Giovanni Velasco Vera

UNIVERSIDAD EAFIT
ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA
MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA
MEDELLÍN

2025

CONTENIDO

INTRODUCCIÓN	5
PLANTEAMIENTO DEL PROBLEMA	7
JUSTIFICACIÓN.....	8
OBJETIVOS	9
GENERAL.....	9
ESPECÍFICOS	9
MARCO CONCEPTUAL	10
DISEÑO METODOLÓGICO.....	16
MATERIALES Y METODOS	17
COMPRESIÓN DE LOS DATOS	17
PREPARACIÓN DE LOS DATOS	18
DESCRIPCIÓN DE MODELOS	24
RESULTADOS	28
EVALUACIÓN	28
BACKTESTING	30
CONCLUSIONES	32
TRABAJOS FUTUROS	34
REFERENCIAS	34

RESUMEN

La identificación de señales efectivas para una estrategia de trading en activos financieros es un desafío que capta el interés de múltiples disciplinas, dado la naturaleza volátil y dinámica de los mercados. La complejidad del análisis de inversionistas radica en los numerosos factores que afectan los diferentes activos financieros, como variables macroeconómicas, decisiones corporativas y eventos inesperados, lo que dificulta la obtención de estimaciones precisas sobre los movimientos futuros. Esto es especialmente relevante para los inversores que buscan construir un portafolio que maximice sus ganancias. En este contexto, algunas variables muestran una mayor relación con los factores que influyen en el mercado, convirtiéndose en indicadores útiles para anticipar movimientos. No obstante, los avances recientes en computación y en técnicas de Machine Learning y Deep Learning han permitido desarrollar modelos más sofisticados que facilitan esta tarea. Este estudio propone comparar metodologías de aprendizaje automático basadas en series de tiempo, específicamente redes neuronales LSTM y modelos de árboles de decisión LightGBM, incorporando modelos de Heterocedasticidad Condicional Autorregresiva (GARCH) buscando una mejor aproximación para la clasificación de señales de compra y venta en instrumentos financieros, tomando en cuenta tanto sus patrones históricos como las variables externas que afectan su comportamiento. Los resultados obtenidos muestran que LightGBM obtuvo el mejor desempeño predictivo, alcanzando valores destacados como un F1 Score de 0.823 y un AUC-ROC de 0.923 en validación, mientras que LSTM fue el modelo con mejor desempeño financiero, logrando un retorno acumulado del 28.05% y un Sharpe Ratio de 0.70, superando ampliamente la estrategia de Buy & Hold. Estas evidencias sugieren que, aunque la predicción direccional diaria es una tarea compleja, los modelos avanzados de Machine Learning pueden transformar señales débiles en estrategias rentables.

Palabras clave: Predicción; Mercados; Volatilidad; Machine Learning; Deep Learning; Series de Tiempo; Heterocedasticidad; GARCH, LSTM, LightGBM; Red Neuronal; Arboles de Decisión

ABSTRACT

Identifying effective trading signals in financial assets is a challenge that draws attention across multiple disciplines due to the volatile and dynamic nature of financial markets. The complexity investors face stems from the wide range of factors that influence asset prices, including macroeconomic variables, corporate decisions, and unexpected events, making it difficult to obtain precise estimates of future movements. This is particularly relevant for investors seeking to build portfolios that maximize returns. In this context, some variables exhibit stronger relationships with market-driven factors, making them useful indicators for anticipating price direction. Nevertheless, recent advances in computing and in Machine Learning and Deep Learning techniques have enabled the development of more sophisticated models that facilitate this task. This study compares time-series-based machine learning methodologies, specifically LSTM neural networks and LightGBM decision-tree models, while incorporating Conditional Heteroskedasticity models (GARCH) to improve the classification of buy and sell signals in financial instruments, accounting for both historical patterns and external variables affecting asset behavior. The results show that LightGBM achieved the best predictive performance, with notable metrics such as an F1 Score of 0.823 and an AUC-ROC of 0.923 in validation, whereas LSTM delivered the best financial performance, reaching a cumulative return of 28.05% and a Sharpe Ratio of 0.70, clearly outperforming a Buy-and-Hold strategy. These findings suggest that although daily directional prediction is inherently complex, advanced Machine Learning models can transform weak signals into profitable trading strategies.

Keywords: Prediction; Markets; Volatility; Machine Learning; Deep Learning; Time Series; Heteroskedasticity; GARCH; LSTM; LightGBM; Neural Network; Boosting.

INTRODUCCIÓN

El mercado de capitales constituye uno de los pilares fundamentales del sistema financiero, al facilitar la compra y venta de activos de renta fija y renta variable entre inversionistas, intermediarios y empresas que buscan liquidez para financiar sus operaciones o impulsar su crecimiento. La emisión y negociación de acciones cobra especial relevancia dentro de este entorno, ya que permite a las compañías captar recursos directamente del mercado, mientras que los inversionistas participan al obtener la oportunidad de generar valor a corto y largo plazo mediante la valorización de dichas acciones. En este contexto, contar con herramientas capaces de anticipar movimientos de precios se convierte en un elemento estratégico tanto para la construcción como para la gestión eficiente de portafolios de inversión.

Las capacidades de predicción de precios de las acciones son cruciales para los inversionistas que buscan minimizar el riesgo en sus portafolios y maximizar sus ganancias, como lo menciona D. AL-Najjar (2016) en su análisis. No obstante, esta tarea resulta especialmente desafiante debido a la volatilidad inherente de los mercados y de los propios instrumentos financieros. Los precios están constantemente influenciados por factores como la evolución económica, los eventos políticos, las noticias diarias, situaciones inesperadas como desastres naturales o crisis internacionales, y las variaciones entre la oferta y la demanda. En conjunto, estas dinámicas generan un entorno altamente incierto, en el que anticipar el movimiento diario de un activo, incluso uno tan líquido y ampliamente seguido como Apple, puede asemejarse en muchos casos al comportamiento aleatorio de tirar una moneda al aire.

Tradicionalmente, los inversionistas han recurrido al análisis fundamental y al análisis técnico para orientar sus decisiones de compra o venta. El análisis fundamental busca estimar el valor intrínseco de una acción a partir de los estados financieros de la empresa y de su contexto económico, mientras que el análisis técnico se basa en la interpretación de patrones históricos de precios y volumen, bajo la premisa de que el mercado tiende a reflejar la información disponible y a repetir ciertos comportamientos (The Nippon Technical Analyst Association, 2018). Sin embargo, ambos enfoques presentan limitaciones importantes frente al creciente volumen de datos, la presencia de relaciones no lineales y la complejidad dinámica del mercado, factores que dificultan la capacidad de estas metodologías tradicionales para capturar de manera adecuada la naturaleza cambiante de los precios financieros.

El desarrollo de la computación y la expansión de metodologías avanzadas han permitido la integración de modelos estadísticos y de aprendizaje automático que ofrecen nuevas posibilidades en el análisis de series temporales financieras. Entre ellas, los modelos de Heterocedasticidad Condicional Autorregresiva (ARCH/GARCH) destacan por su capacidad para modelar la volatilidad dinámica, caracterizada por fenómenos como agrupamiento de volatilidad, efectos de

apalancamiento y leptocurtosis (Rusere & Kaseke, 2021). En particular, el modelo GARCH se ha convertido en una extensión fundamental al permitir que la varianza condicional no dependa exclusivamente de los errores pasados, sino también de su propia trayectoria, lo que introduce mayor flexibilidad y capacidad de ajuste frente a cambios persistentes en la volatilidad, tal como lo plantea Bollerslev (1986).

De manera complementaria, las redes neuronales recurrentes y, en particular, las arquitecturas LSTM, han demostrado un desempeño sobresaliente al capturar dependencias temporales de corto y largo plazo, evitando problemas como la pérdida de gradiente y permitiendo representar de forma más realista la naturaleza secuencial de los precios (Mutinda & Langat, 2024). A diferencia de las RNN tradicionales, LSTM introduce una estructura interna más compleja para gestionar el flujo de datos y evitar la pérdida de información relevante en secuencias largas.

En conjunto, una red Neuronal como LSTM y los modelos de Heterocedasticidad Condicional Autorregresiva, se han convertido en una herramienta práctica para el análisis de series de tiempo, ofreciendo a inversionistas y analistas una comprensión más profunda de las dinámicas de riesgo y rendimiento en los mercados de capitales.

En años recientes, los modelos basados en árboles de decisión, como LightGBM, también han ganado protagonismo debido a su capacidad para manejar grandes volúmenes de variables explicativas y capturar interacciones no lineales de manera eficiente. Su estructura de crecimiento leaf-wise y técnicas como GOSS o EFB permiten entrenar modelos altamente competitivos, rápidos y precisos, características que han impulsado su adopción en diversas aplicaciones financieras, incluyendo predicción de tendencias, clasificación de señales y gestión cuantitativa de portafolios (Hartanto et al., 2023).

Este estudio integra estas metodologías y las contrasta con el propósito de formular una estrategia de inversión basada en la predicción del movimiento diario de la acción de Apple, combinando información histórica, volatilidad condicional, variables macroeconómicas y técnicas avanzadas de aprendizaje automático. El objetivo es identificar cuál de estos enfoques ofrece una aproximación más precisa, estable y útil para apoyar en estrategias de compra o venta bajo un entorno real de mercado. A medida que avancemos en las siguientes secciones de este trabajo, examinaremos con mayor detalle el funcionamiento de estos modelos y su utilidad en la formulación de una estrategia de inversión óptima.

PLANTEAMIENTO DEL PROBLEMA

La toma de decisiones en los mercados financieros depende en gran medida de anticipar, con la mayor precisión posible, la dirección que tomará el precio de un activo en el corto plazo. Para inversionistas que gestionan portafolios activos, incluso pequeños errores en la anticipación de la dirección diaria del precio pueden acumularse y deteriorar de forma importante el desempeño financiero. Sumado a esto, la toma de decisiones se realiza en un entorno donde coexisten choques macroeconómicos, noticias corporativas, cambios en la política monetaria y estrategias de grandes participantes del mercado, lo que vuelve especialmente difícil extraer señales claras a partir de la información disponible.

Con el tiempo, se han desarrollado diversas técnicas para intentar predecir los precios de las acciones, pero muchas aún enfrentan limitaciones significativas. Sin embargo, los recientes avances en computación y en técnicas de aprendizaje automático, como Machine Learning y Deep Learning, han abierto nuevas posibilidades. Estos avances permiten crear modelos mucho más sofisticados que podrían mejorar la precisión en la predicción de precios.

Este estudio se centra en explorar y comparar diferentes enfoques modernos aplicados a series de tiempo financieras, con el propósito de identificar señales predictivas que puedan ser utilizadas en la formulación de una estrategia eficiente de trading. En concreto, se analizarán los modelos de Heterocedasticidad Condicional Autorregresiva, que son fundamentales para modelar la volatilidad del mercado, junto con la aplicación de una red neuronal recurrente como LSTM y modelos basados en árboles de decisión como LightGBM. El objetivo es evaluar la capacidad de estos métodos para orientar decisiones de inversión y determinar si generan señales útiles y consistentes para la predicción de la dirección de la acción de Apple bajo condiciones reales de mercado.

JUSTIFICACIÓN

La creciente complejidad y volatilidad de los mercados financieros exige herramientas más avanzadas para apoyar la toma de decisiones de inversión. En ese sentido, los modelos basados en redes neuronales, árboles de decisión y medidas de volatilidad condicional permiten capturar dinámicas no lineales y patrones temporales que los métodos tradicionales suelen pasar por alto. Estas técnicas facilitan una comprensión más precisa de los cambios en la dirección del mercado y de la estructura temporal de la volatilidad diaria.

Además, su aplicación en series de tiempo permite procesar grandes volúmenes de información y extraer conclusiones más robustas sobre las tendencias del mercado. Al identificar períodos de alta y baja volatilidad, los modelos de Heterocedasticidad Condicional Autorregresiva no solo generan estimaciones sobre la incertidumbre de los activos financieros, sino que también ofrecen elementos clave para la gestión del riesgo, especialmente en entornos donde eventos inesperados pueden afectar drásticamente los precios.

En este contexto, se considera la acción de Apple como variable de estudio debido a su solidez corporativa, alta liquidez y amplio historial de datos, cualidades que la convierten en un activo especialmente adecuado para el análisis de series temporales. Su comportamiento refleja dinámicas asociadas a inversionistas institucionales, rotación sectorial y eventos corporativos que generan fluctuaciones significativas en la volatilidad. Por estas razones, Apple resulta idónea para evaluar modelos predictivos avanzados y validar estrategias de trading basadas en Machine Learning, además de servir como un punto de partida sólido para replicar estas metodologías en otras acciones del mercado.

En esencia, la pregunta central que orienta este estudio es: ¿cuál de los modelos de Machine Learning ofrece una mejor aproximación para desarrollar una estrategia de trading efectiva y basada en evidencia? Los resultados pueden aportar información relevante para optimizar estrategias cuantitativas y para comprender cómo los modelos avanzados pueden mejorar la capacidad predictiva en mercados financieros altamente inciertos.

OBJETIVOS

GENERAL

Comparar la capacidad predictiva de diferentes enfoques modernos de análisis predictivo, incluyendo modelos de heterocedasticidad condicional autorregresiva (GARCH), redes neuronales recurrentes LSTM y algoritmos de árboles de decisión como LightGBM, para clasificar la dirección del movimiento diario de activos financieros, específicamente en la acción de Apple, y aportar elementos técnicos que respalden una estrategia de trading para inversionistas.

ESPECÍFICOS

- Preparar y depurar el conjunto de datos incorporando variables de mercado, transformaciones relevantes, rezagos temporales y mediciones de volatilidad que permitan estructurar adecuadamente las series para los modelos predictivos analizados.
- Implementar un modelo GARCH para estimar la volatilidad condicional diaria y evaluar su utilidad como variable explicativa en el proceso de clasificación de dirección de activos financieros.
- Entrenar una red neuronal LSTM orientada a la predicción secuencial, ajustando su arquitectura, hiperparámetros y estructura de ventanas temporales para capturar patrones no lineales y dependencias de largo plazo en las series de tiempo.
- Implementar un modelo de árboles de decisión mediante LightGBM con validación temporal, explorando su capacidad para manejar múltiples características, identificar interacciones no lineales y generar predicciones eficientes en contextos financieros.
- Comparar el desempeño de los modelos LSTM y LightGBM mediante métricas de clasificación como Accuracy, F1-score, MCC, Sensibilidad, Especificidad y AUC-ROC, determinando cuál enfoque ofrece mejores resultados en la predicción de la dirección diaria del mercado.
- Realizar un Backtesting con los modelos evaluados para analizar su desempeño bajo condiciones reales de mercado y determinar si permiten construir una estrategia que supere los resultados de la estrategia tradicional Buy & Hold.
- Proponer una metodología basada en los modelos previamente mencionados, que facilite la identificación de la estrategia más adecuada para optimizar resultados y mejorar la toma de decisiones.

MARCO CONCEPTUAL

Mercado de Capitales

Los mercados financieros son plataformas que agrupan la interacción de ahorradores e inversores a través de la compra y venta de diferentes activos de corto, mediano y largo plazo. Juegan un papel muy importante en la economía global, permitiendo que las empresas, públicas y privadas de diferentes sectores, puedan generar crecimiento a través de diferentes formas de financiamiento. Al mismo tiempo, brindan a individuos y organizaciones la oportunidad de invertir su dinero con la esperanza de obtener rendimientos. Como parte de los mercados financieros, sobresale los mercados de capitales donde se negocian acciones y bonos de deuda pública. De esta forma, las empresas obtienen recursos que les permiten financiar su crecimiento, dándole la oportunidad a inversionistas de participar con la expectativa de obtener beneficios con la valorización de los activos o con el pago de dividendos.

Por otro lado, los mercados de futuros ofrecen a los inversionistas la oportunidad de anticipar y gestionar el riesgo asociado a activos como commodities y acciones, proporcionando herramientas valiosas para protegerse ante la incertidumbre. Existen numerosas opciones para que las personas gestionen su dinero, y una que ha ganado relevancia en las últimas décadas es el mercado de criptomonedas. Este mercado ha captado la atención de millones, ofreciendo nuevas formas de inversión a través de activos digitales que operan de manera descentralizada.

Análisis Técnico y Fundamental

El análisis fundamental se enfoca en determinar el valor intrínseco de un activo financiero, como una acción, al examinar los factores económicos y financieros que influyen en una empresa emisora. Este enfoque implica realizar un análisis detallado de los estados financieros de la compañía, su posición dentro de la industria, el entorno macroeconómico y otros factores relevantes (Bentes & Navas (2013)). El objetivo es evaluar si el precio actual de mercado de un activo refleja realmente su valor intrínseco. Si el precio de mercado está por debajo del valor intrínseco, se considera que el activo está infravalorado, lo que puede representar una buena oportunidad de compra. En cambio, si el precio de mercado supera el valor intrínseco, se considera que el activo está sobrevalorado y podría ser momento de vender.

El análisis técnico, por otro lado, es una práctica que usa datos históricos y la acción del precio actual para predecir movimientos futuros del precio (Agarwal, 2022). Se centra en el estudio de la oferta y la demanda en el mercado, empleando herramientas como gráficos de velas (*Figura 1.*), líneas de tendencia y medias móviles. Estas herramientas ayudan a identificar patrones y tendencias que pueden predecir los movimientos futuros de los precios. La premisa fundamental de este

enfoque es que los precios reflejan las decisiones de inversión de individuos con información privilegiada, y que la historia tiende a repetirse (The First Step Guide to Technical Analysis, 2018). A diferencia del análisis fundamental, que se enfoca en determinar el valor intrínseco de un activo, el análisis técnico se ocupa de los cambios en los precios para identificar el momento más adecuado para comprar o vender.



Figura 1: Gráfico de Velas Acción de Apple (Yahoo Finance, 2025)

Aunque el análisis técnico no proporciona una evaluación directa del valor intrínseco de un activo, puede ser una herramienta muy eficaz para detectar oportunidades de inversión, especialmente cuando se combina con el análisis fundamental.

Heterocedasticidad Condicional Autorregresiva Generalizada

El modelo ARCH, introducido por Robert Engle (1982), constituyó un avance fundamental en la econometría financiera al permitir modelar la volatilidad cambiante en series de tiempo, superando la antigua suposición de varianza constante en los errores (AL-Najjar, 2016). Su enfoque se basa en que la volatilidad condicional depende de los errores pasados, lo que hace posible capturar el comportamiento dinámico de la varianza a lo largo del tiempo, una característica típica de los mercados financieros.

Como una extensión del modelo ARCH, Tim Bollerslev introdujo en 1986 el modelo GARCH, que busca abordar algunas de sus limitaciones. Este modelo se basa en el principio de que la varianza de los rendimientos de los activos financieros no es constante a lo largo del tiempo, sino que exhibe patrones de agrupamiento. Esto significa que, en períodos de alta volatilidad, es probable que la tendencia continúe alta en los siguientes períodos, al igual que en períodos de baja volatilidad. En resumen, el modelo GARCH asume que las varianzas del término de error en un periodo determinado no solo dependen de los errores pasados, sino también de las varianzas pasadas (Engle, 2001).

El modelamiento de GARCH está dado por la siguiente función:

$$\sigma_t^2 = \omega + \sum_{j=1}^q \alpha_j \epsilon_{t-j}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

donde: $i = 1$ hasta p representa la varianza condicional; ω , α y β son constantes no negativas en donde $\alpha_j + \beta_j$ debe ser menor que 1 pero a su vez debe estar cerca de la unidad para asegurar mayor precisión en el modelo; los residuos ϵ_{t-j}^2 representan la volatilidad condicional rezagada. Ambos modelos, ARCH y GARCH, tienen la suposición fundamental de que todos los choques sobre la volatilidad tienen una distribución simétrica.

Long Short-Term Memory – LSTM

Mutinda & Langat (2024) explican que LSTM son particularmente efectivas para modelar y predecir datos secuenciales, lo que las convierte en una herramienta valiosa para diversas aplicaciones, como el pronóstico de series de tiempo. Es un tipo de red neuronal recurrente (RNN) diseñada para abordar problemas de desvanecimiento de gradiente, lo cual dificulta la capacidad de una red neuronal tradicional para aprender y retener información sobre secuencias largas.

LSTM está compuesta por los siguientes componentes: Un estado de celda que funciona como la memoria de la red y tres puertas que permiten que se actualice, retenga o descarte información de manera selectiva, lo que permite gestionar adecuadamente las dependencias a largo plazo (*Figura 2*).

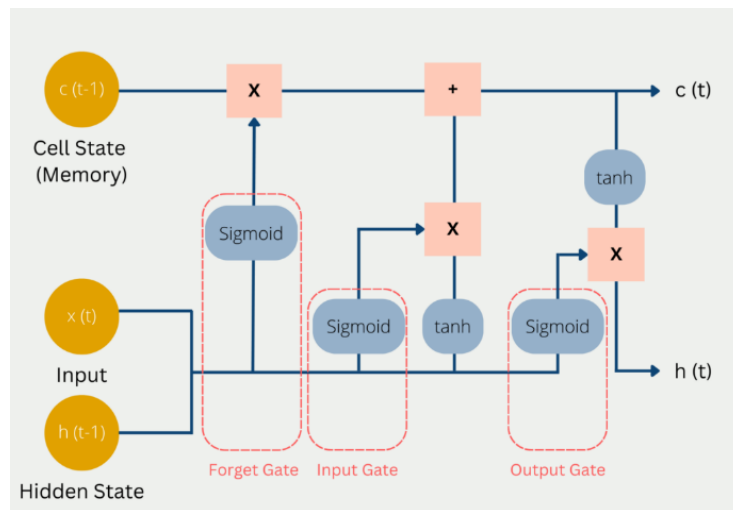


Figura 2. Esquema LSTM tomado de <https://cienciadedatos.net/documentos/py54-forecasting-con-deep-learning>

LightGBM

El algoritmo Light Gradient Boosting Machine (LightGBM) se ha posicionado como uno de los métodos de Boosting más eficientes para tareas de predicción en series de tiempo y problemas supervisados. Su principal fortaleza radica en la forma en que construye los árboles de decisión: en lugar de crecerlos nivel por nivel como los métodos tradicionales, LightGBM emplea una estrategia de crecimiento leaf-wise, lo que le permite profundizar en las divisiones más representativas y capturar relaciones no lineales con mayor precisión (ver *Figura 3*). De acuerdo con Padhi et al. (2021), esta arquitectura enfocada en la optimización del crecimiento del árbol contribuye a mejorar el desempeño predictivo en contextos financieros donde la complejidad y la no linealidad son dominantes, resaltando su capacidad para manejar grandes volúmenes de información sin degradación significativa del rendimiento.

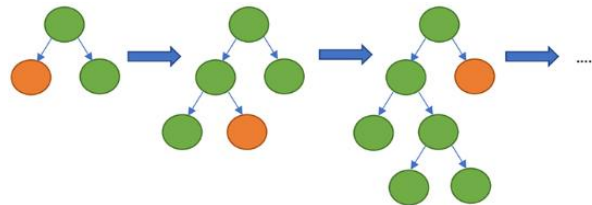


Figura 3. Crecimiento Leaf-wise

Adicionalmente, LightGBM integra técnicas avanzadas como Gradient based One Side Sampling (GOSS) y Exclusive Feature Bundling (EFB), que reducen el costo computacional sin obviar capacidad de generalización. Hartanto et al. (2023) evidencian que estas optimizaciones permiten que LightGBM no solo supere a otros modelos basados en boosting como XGBoost sino que también se convierta en una herramienta robusta para pronósticos financieros al minimizar riesgos de sobreajuste y mantener un rendimiento estable incluso en series con ruido, volatilidad o presencia de múltiples variables exógenas. Gracias a esta combinación de eficiencia y escalabilidad, LightGBM se consolida como un modelo particularmente adecuado para estudios de predicción de series de tiempo financieras.

Conjunto de Datos de Validación y Prueba

En el contexto del presente estudio, la correcta división del conjunto de datos es crucial para el entrenamiento y prueba de los modelos a implementar. La naturaleza de los datos utilizados exige una metodología que respete la secuencia de estos, evitando así la fuga de información que podría sesgar los datos.

Por lo tanto y como propuesta a desarrollar, se dividirán los datos de la siguiente manera teniendo en cuenta que se trabajara con series de precios diarias de los últimos 10 años:

La partición del conjunto de datos se realizará de manera estrictamente secuencial, respetando el orden temporal de las observaciones. Se asigna la mayor proporción de datos al conjunto de entrenamiento facilita que el modelo capture patrones históricos más robustos, mientras que un conjunto de validación independiente permite ajustar hiperparámetros, seleccionar el umbral de decisión y evaluar el rendimiento del modelo sin intervenir el proceso de aprendizaje. Por último, se define un conjunto de prueba para asegurar una medición objetiva del desempeño en datos no vistos, simulando un escenario optimo de predicción futura.

Métricas de Evaluación

Las métricas de evaluación se seleccionan con el objetivo de analizar el desempeño predictivo de los modelos de clasificación binaria empleados, teniendo en cuenta las particularidades del contexto financiero y la asimetría de los errores de predicción. Tal como señalan Kuhn & Johnson (2013), la elección adecuada de métricas es fundamental para evaluar de manera objetiva la calidad de los modelos y evitar interpretaciones sesgadas de su desempeño.

Se emplean métricas dependientes del umbral de clasificación, tales como *accuracy*, *precision*, *recall*, *specificity* y *F1-score*, las cuales permiten analizar distintos compromisos entre detección de movimientos positivos y control de errores. La matriz de confusión se utiliza como herramienta complementaria para descomponer el comportamiento del modelo en términos de verdaderos y falsos positivos y negativos. Adicionalmente, se incorpora el coeficiente de correlación de Matthews (MCC) como una métrica global, debido a su capacidad para resumir el desempeño del clasificador utilizando toda la información contenida en la matriz de confusión y su robustez frente a posibles desbalances de clase.

De manera complementaria, se utiliza el área bajo la curva ROC (AUC-ROC) como métrica independiente del umbral de decisión, lo que permite evaluar la capacidad discriminativa del modelo sin fijar una regla de clasificación específica. Esta métrica resulta particularmente útil para comparar modelos durante la fase de validación.

Sea N el tamaño de las observaciones, y_j el valor real, \hat{y}_j el valor predicho y \bar{y}_j la media de los valores reales:

$$\text{Accuracy:} \quad Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$\text{Precision:} \quad Precision = \frac{TP}{(TP + FP)}$$

Recall:

$$Recall = \frac{TP}{(TP + FN)}$$

Specificity:

$$Specificity = \frac{TN}{(TN + FP)}$$

F1-Score:

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

AUC-ROC:

$$AUC - ROC = \int Recall \times (1 - Specificity)$$

MCC:

$$1 = \frac{(TP \times TN - FP \times FN)}{\sqrt{((TP + FP)(TP + FN)(TN + FP)(TN + FN))}}$$

Igualmente, se utilizará el Backtesting como una técnica de validación que permite evaluar el desempeño del modelo a partir de datos históricos, con el propósito de analizar cómo habría respondido ante condiciones reales del mercado, permitiendo estimar su capacidad de generalización y su potencial utilidad en escenarios futuros.

DISEÑO METODOLÓGICO

Este estudio se desarrollará bajo la metodología CRISP-DM, la cual es ampliamente utilizada en proyectos de Machine Learning. Está compuesta por 6 fases secuenciales fundamentales y permite a su vez flexibilidad en su implementación (IBM, s/f). La *Figura 4* muestra la ilustración del ciclo de la metodología CRISP-DM. Estas fases comprenden:

1. **Entendimiento del negocio:** Esta etapa se centra en identificar las necesidades de los participantes del mercado financiero y comprender las metodologías empleadas para predecir precios de activos financieros. Incluye además la definición de objetivos específicos y criterios de éxito del proyecto.
2. **Comprensión de los datos:** Consiste en identificar las variables del mercado a utilizar, obtener los datos necesarios, y analizar la información proporcionada mediante la API de Yahoo Finance.
3. **Preparación de los datos:** En esta fase, se lleva a cabo el preprocesamiento, la transformación y la ingeniería de características para asegurar que los datos estén en condiciones óptimas para el modelado.
4. **Modelado:** Se implementan modelos de Regresión Logística y un modelo híbrido de Heteroscedasticidad Condicional Autorregresiva (ARCH/GARCH) combinado con redes neuronales LSTM, enfocados en la predicción de una variable categórica basada en múltiples variables independientes.
5. **Evaluación:** Se aplican métricas de desempeño adecuadas para cada modelo con el objetivo de evaluar su precisión y efectividad, utilizando las métricas previamente definidas.
6. **Despliegue:** Esta etapa busca implementar el modelo con mejor desempeño mediante una API, permitiendo la generación de clasificaciones en tiempo real.

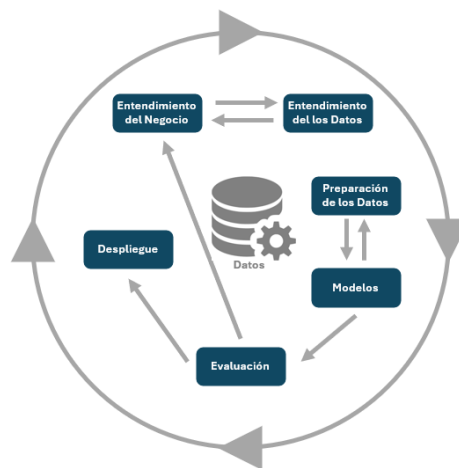


Figura 4. Ciclo de Metodología CRISP-DM.

MATERIALES Y METODOS

COMPRESIÓN DE LOS DATOS

El conjunto de datos empleado en este estudio fue construido a partir de series temporales financieras y variables macroeconómicas obtenidas de las bases de datos del Federal Reserve Bank of St. Louis (2025) y Yahoo Finance (2025), fuentes reconocidas por su rigor, alta confiabilidad y amplia cobertura histórica. La preparación de los datos constituye una etapa crítica para garantizar la calidad del modelo predictivo y la consistencia en el análisis de series temporales aplicadas al mercado financiero. Para tal fin, se recopilaron precios diarios y variables macroeconómicas desde enero de 2000 hasta septiembre de 2025, abarcando un total de 9.403 registros y 39 variables iniciales, lo que permite capturar distintos ciclos económicos, eventos de mercado y periodos de alta volatilidad a lo largo de más de dos décadas

El DataFrame original contiene variables estructuradas en formato OHLCV (Open, High, Low, Close, Volume), Todas las variables mantienen una frecuencia temporal diaria, lo que facilita la integración y análisis conjunto de los diferentes activos financieros. A continuación, se describen las principales variables consideradas:

Activos e índices bursátiles

- AAPL: Acciones de Apple Inc., una de las mayores empresas tecnológicas a nivel mundial, referente en innovación, computación personal y servicios digitales.
- S&P500: Índice estándar del mercado estadounidense compuesto por las 500 empresas más grandes por capitalización bursátil.
- Dow Jones Industrial Average: Índice compuesto por 30 grandes empresas industriales estadounidenses.
- Nasdaq Composite: Índice que agrupa más de 3.000 compañías, principalmente del sector tecnológico e innovación, incluyendo empresas de software, hardware, biotecnología y comunicaciones.
- VGT (Vanguard Information Technology ETF): Fondo cotizado en bolsa que replica el desempeño del sector tecnológico estadounidense.

Activo digital

- Bitcoin (BTC-USD): Criptomoneda descentralizada y el principal activo digital del mercado cripto, reconocido como reserva de valor alternativa y activo con alta volatilidad.

Variables Macroeconómicas

- CPI: Índice de precios al consumidor estadounidense, utilizado como medida de inflación y variaciones en el costo de vida..
- FEDFUNDS (Federal Funds Rate): Tasa de interés de referencia de la Reserva Federal de Estados Unidos.,
- INDPRO (Industrial Production Index): Índice de producción industrial que mide la actividad manufacturera, minera y de servicios públicos en Estados Unidos.
- DGS10 (10-Year Treasury Constant Maturity Rate): Rendimiento del bono del Tesoro estadounidense a 10 años.

PREPARACIÓN DE LOS DATOS

El conjunto de datos, como fue mencionado anteriormente, tiene valores faltantes para los días sábado, domingo y festivos, los cuales fueron completados utilizando el método de Fill Forward, el cual completa los valores faltantes en una serie de tiempo utilizando el último valor válido conocido hacia adelante.

La variable objetivo para determinar la dirección del rendimiento fue la acción de Apple, por lo cual, a partir del conjunto de datos obtenido, se generaron otras variables con el fin de resumir su comportamiento y poder obtener una mayor explicabilidad de esta variable:

- **Amplitud:** La diferencia entre el precio más alto y el precio más bajo en las operaciones intraday.
- **Valor Nominal:** Obtenida con el producto entre el volumen y el precio de cierre de cada variable.
- **Retorno Logarítmicos:** Calculara a través de la variabilidad diaria de los precios de cierre, asegurando así la estacionaridad de los datos:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

Donde R_t es el retorno logarítmico en el período t , P_t es el precio en el periodo actual t y P_{t-1} es el precio en el periodo $t - 1$.

- **Rezago:** de retornos, amplitud y volumen para capturar memoria de corto plazo en los mercados financieros. Adicionalmente, para capturar patrones temporales de mayor profundidad, se construyeron 30 rezagos del retorno

logarítmico de Apple, permitiendo así que el modelo identifique dependencias más complejas

- **Label:** Se define una variable objetivo binaria denominada *Label*, donde 1 representa un retorno positivo de la acción de Apple en el día siguiente, y 0 un retorno negativo o neutro.

Igualmente, se realizó un análisis de estacionaridad sobre la serie temporal entorno a su tendencia lineal específicamente para los precios de la acción de Apple (ver *Figura 5.*), considerando esta como la base de la variable objetivo. Se utilizan entonces las pruebas de Dickey-Fuller Aumentada (ADF) y Kwiatkowski-Phillips-Schmidt-Shin (KPSS) para determinar dicha estacionaridad (Amat Rodrigo & Escobar Ortiz (s/f)). Los resultados de las pruebas se observan a continuación en la *Tabla 1.* y dejan en evidencia de que la serie es estacionaria diferenciada una vez.

Serie evaluada	ADF Statistic	p-value ADF	KPSS Statistic	p-value KPSS
Serie original	-16.794	0.4418	43.306	0.01
Diferenciada (orden 1)	-332.964	0.0000	0.0383	0.10
Diferenciada (orden 2)	-120.615	0.0000	0.0799	0.10

Tabla 1. Test Estacionaridad Acción de Apple

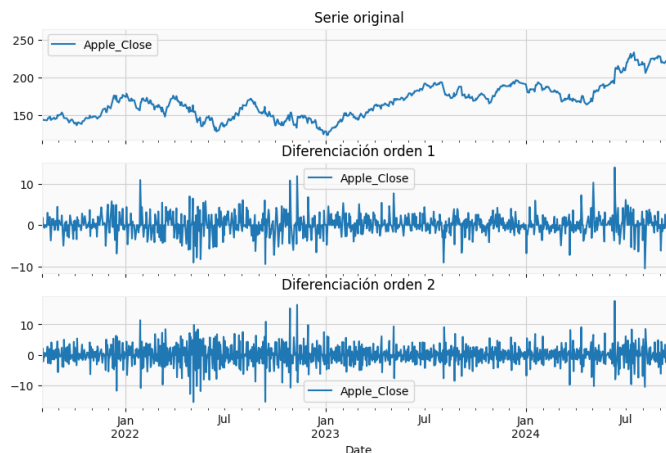
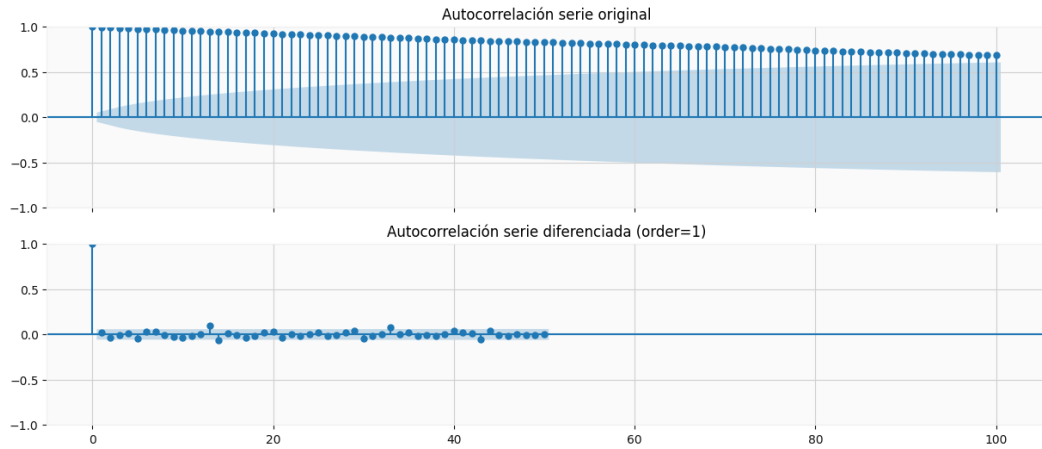


Figura 5. Test de Estacionaridad sobre Acción de Apple

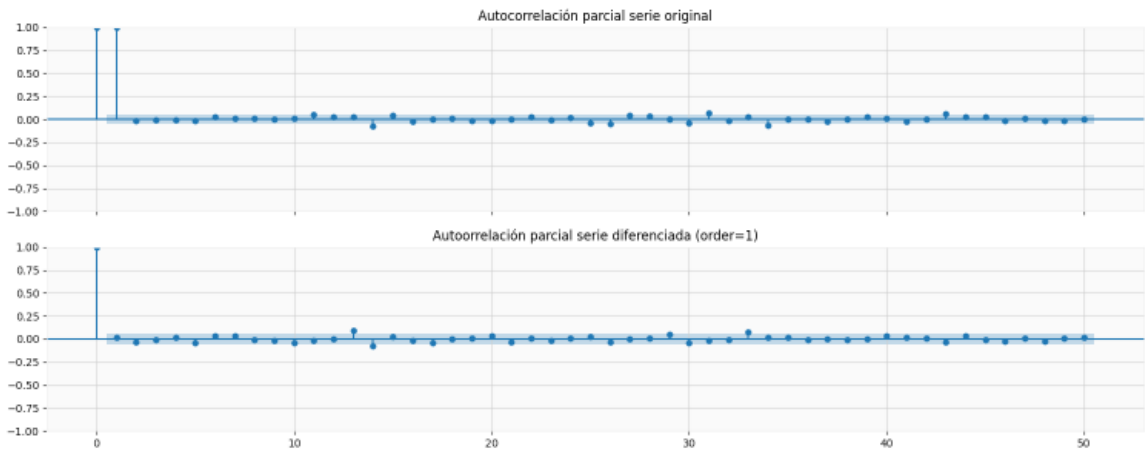
Como los p-values obtenidos en las pruebas son estadísticamente significativos de acuerdo al umbral de 0.05, se puede inferir que la serie es estacionaria en el primer orden de diferenciación, rechazando de esta forma la hipótesis nula que indica que la serie no es estacionaria.

La *Figura 5.* muestra el análisis de autocorrelación con las funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) para calcular la correlación

entre la serie temporal y sus valores rezagados Amat Rodrigo & Escobar Ortiz (2023), las cuales a su vez sugirieron rezagos de orden 1 para la serie evaluada.



a. ACF



b. PACF

Figura 5. Análisis de Autocorrelación de la Acción de Apple.

A continuación, se realizó un análisis de tendencias de los precios de cierre de cada activo los cuales se detallan en la *Figura 7.* y el calculó la correlación de Pearson para todas las variables del conjunto de datos (*Figura 8.*), dando como resultado una alta correlación entre los precios de cierre y la mayoría de los retornos de las variables. Las variables del BTC no tienen una relación lineal con los retornos de la acción de Apple. Además, respecto a la variable objetivo Label, las variables que mostraron una mayor relación con ella fueron los retornos de los índices del mercado bursátil.

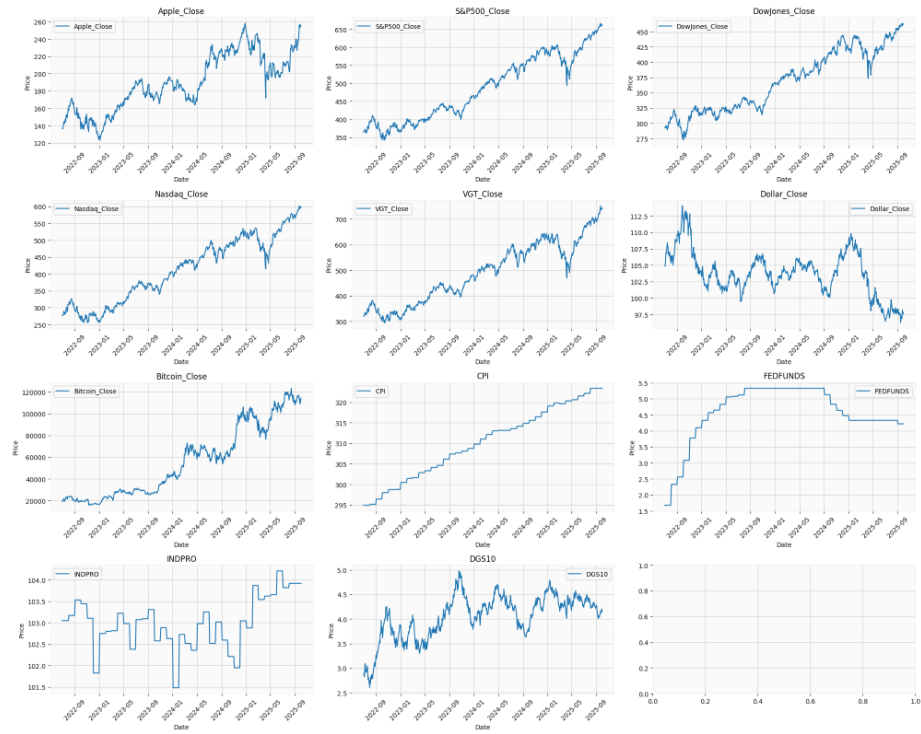


Figura 7. Análisis de tendencias de los precios de cierre de cada Activo

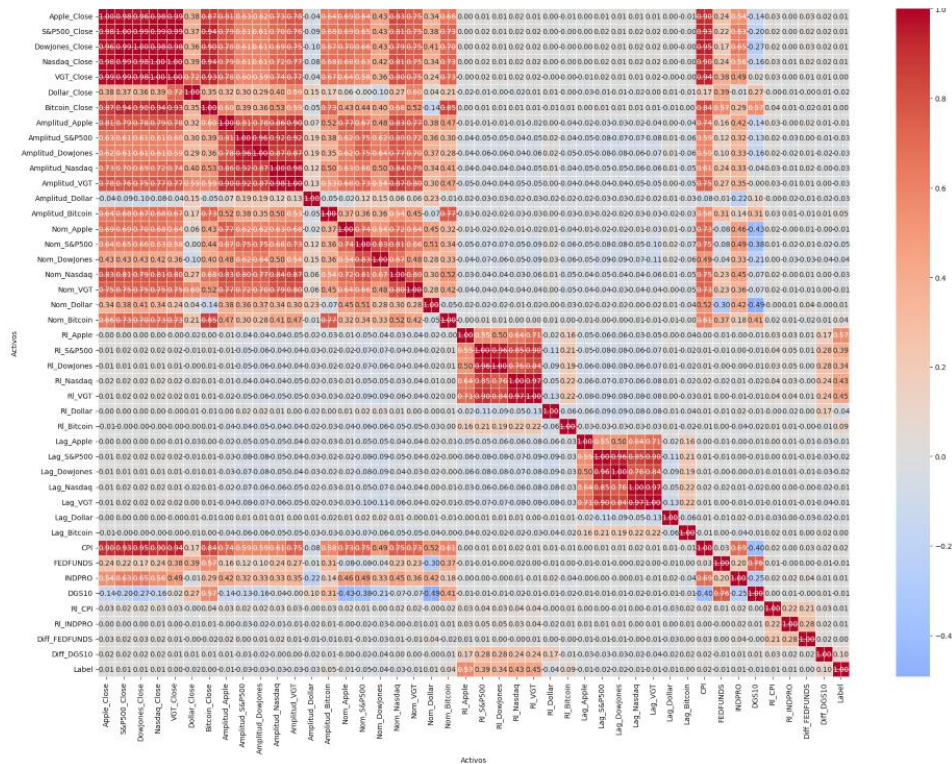


Figura 8. Análisis de Correlaciones entre las variables definidas

Finalmente, y a partir de los análisis de la matriz de correlaciones, se realiza el Test de Causalidad de Grenger para determinar si las variables de los retornos de los activos tienen relación causal con la variable objetivo. Los resultados arrojados por el test se muestran en la *Tabla 2*. e indican que las variables de los retornos tienen relación de Causalidad con la variable objetivo:

Test statistic	Critical value	p-value	df
4,140	1,939	0,000	(8, 32742)

Tabla 2. Test de Causalidad de Grenger: retornos de Apple vs Retornos de los demás activos

El mercado bursátil opera únicamente de lunes a viernes, por lo que las fechas correspondientes a sábados y domingos fueron eliminadas de la serie temporal. Esta decisión es importante dado que la inclusión de días sin negociación podría generar ruido y afectar el comportamiento temporal observado por los modelos, especialmente en esquemas sensibles a dependencias temporales como las redes LSTM. Tras este filtro, la base de datos se reduce de 9.403 a 6.717 observaciones, ajustándose únicamente a días hábiles del mercado.

Para la construcción del conjunto final de características se seleccionaron aquellas variables que resumen de manera más efectiva la dinámica del mercado, incluyendo el retorno logarítmico de Apple como variable base, los rezagos de retornos, amplitud y volumen de los activos considerados, así como indicadores macroeconómicos relevantes como CPI, INDPRO, FEDFUNDS y DGS10, junto con la variable objetivo (Label). Esta selección permite capturar tanto la memoria temporal del comportamiento del precio como el efecto de factores externos sobre el activo de interés. El número total de variables incluidas en el modelo depende del número de rezagos utilizados, oscilando aproximadamente entre 45 y 60 características, garantizando así una representación amplia y adecuada del comportamiento del mercado para fines predictivos.

Algunas series financieras y macroeconómicas utilizadas, en particular Bitcoin, el ETF tecnológico VGT y el índice del dólar, presentan valores faltantes derivados de diferencias en calendarios de negociación, fechas de inicio más recientes y discontinuidades en la disponibilidad histórica de la información. Para asegurar la coherencia temporal del conjunto de datos, estas series exógenas fueron alineadas al calendario del activo subyacente. En el caso de variables macroeconómicas y financieras de baja frecuencia, los valores ausentes se imputaron mediante arrastre temporal (*forward fill*), bajo el supuesto de persistencia de la información hasta la siguiente publicación oficial, preservando la causalidad temporal y evitando la introducción de información futura.

Adicionalmente, para aquellos periodos en los que activos como Bitcoin, VGT o el índice del dólar no contaban con observaciones disponibles, se aplicaron esquemas de imputación conservadores con el fin de garantizar la continuidad de las series durante el entrenamiento del modelo. Previo al proceso de estimación, las variables exógenas fueron escaladas para normalizar sus magnitudes y evitar que diferencias en unidades influyeran de forma desproporcionada en el aprendizaje. La normalización interna mediante *Layer Normalization* se utilizó exclusivamente como mecanismo de estabilización numérica. Estas decisiones de preprocesamiento, aunque necesarias desde el punto de vista operativo, incorporan supuestos que pueden incidir en la capacidad predictiva del modelo y se discuten de manera crítica en la sección de limitaciones.

Como parte del proceso de enriquecimiento del conjunto de variables, se incorporó una medida de volatilidad condicional estimada mediante un modelo GARCH(1,1) aplicado a los retornos de la acción de Apple. Este modelo permite capturar el comportamiento dinámico de la varianza a lo largo del tiempo, identificando periodos de alta y baja volatilidad característicos de los mercados financieros, lo cual se muestra en la *Figura 9*. La volatilidad condicional obtenida se desplazó un periodo hacia atrás para asegurar que su valor represente únicamente información disponible en cada momento del tiempo, evitando así fuga de información hacia el modelo predictivo. Esta variable se integró como un factor adicional para mejorar la capacidad del modelo de identificar cambios en la dirección del mercado, considerando que aumentos repentinos en la volatilidad suelen reflejar incertidumbre o eventos relevantes que pueden anticipar movimientos significativos en el precio del activo, tal como lo expone Engle (2001).

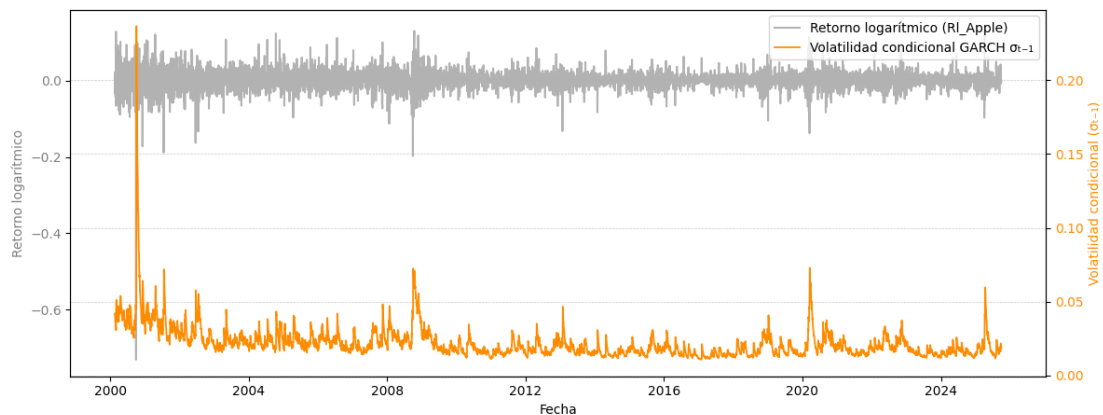


Figura 9. Evolución de los Retornos Vs Volatilidad GARCH(1,1) de Apple

Para garantizar la correcta secuencia temporal y evitar fuga de información, el conjunto de datos se dividió de forma cronológica en tres secciones: entrenamiento (enero de 2000 a mayo de 2023), validación (junio de 2023 a mayo de 2025) y

prueba (junio a septiembre de 2025). Esta estructura permite evaluar el modelo bajo un escenario realista, utilizando únicamente información disponible hasta cada fecha.

Para el modelo LSTM, los datos fueron transformados en tensores adecuados para aprendizaje secuencial, organizando los rezagos del retorno de Apple como una serie temporal y manteniendo el resto de las variables como características exógenas. En el caso de 30 rezagos, el conjunto final incluyó aproximadamente 6.077 muestras para entrenamiento, 522 para validación y 87 para prueba, conservando un balance adecuado entre clases y asegurando consistencia en las secuencias utilizadas por la red.

DESCRIPCIÓN DE MODELOS

LSTM - GARCH

Para este estudio se desarrolló un modelo híbrido diseñado para comprender tanto el comportamiento histórico del precio de la acción de Apple como el contexto financiero y macroeconómico que influye en su evolución. La elección de esta arquitectura responde a la necesidad de capturar patrones secuenciales presentes en series temporales y, al mismo tiempo, integrar información externa que aporta señales adicionales sobre el entorno del mercado.

El modelo está compuesto por dos ramas principales. La primera se encarga de procesar la secuencia de retornos logarítmicos de Apple, organizada como una serie de n valores pasados. Esta rama utiliza dos capas LSTM con 96 y 48 unidades, respectivamente, las cuales permiten identificar dependencias a corto y largo plazo, una característica clave en mercados donde los precios tienden a presentar continuidad o patrones inmediatos, alineado con la capacidad de las LSTM para capturar información de etapas pasadas y utilizarla para predecir comportamientos futuros como lo mencionan Moghar y Hamiche (2020). Para mejorar la estabilidad del aprendizaje y reducir el riesgo de sobreajuste, se incorporaron mecanismos de dropout durante el procesamiento secuencial.

La segunda rama recibe las variables exógenas, que incluyen indicadores macroeconómicos y datos de mercado complementarios. Este conjunto de información se ingresa como un vector de características y pasa primero por una capa de normalización, lo que contribuye a un entrenamiento más estable. Luego, dos capas densas de 128 y 64 neuronas procesan esta información mediante funciones de activación ReLU, logrando capturar relaciones no lineales entre estos factores y el comportamiento del activo.

Ambas ramas se integran en una capa que combina las señales temporales con las exógenas. A partir de esta fusión, el modelo continúa con una capa densa intermedia de 64 neuronas y, finalmente, una capa de salida con activación sigmoide, que devuelve la probabilidad de que el movimiento del precio sea positivo en la siguiente sesión (Ver *Figura 10.* y *Tabla 3.*).

Para entrenar el modelo se utilizó el optimizador Adam, junto con una función de pérdida basada en *Focal Loss* ponderada. Esta decisión busca responder de manera más efectiva a posibles desequilibrios entre clases, haciendo mayor énfasis en las predicciones difíciles. Además, se emplearon mecanismos de *early stopping* y ajuste dinámico de la tasa de aprendizaje para promover un entrenamiento eficiente y evitar sobreajuste.

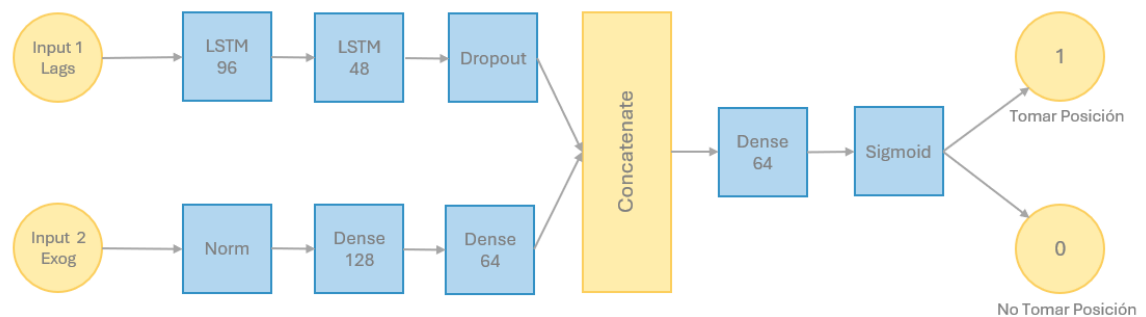


Figura 10. Arquitectura Modelo Híbrido LSTM

Componente	Tipo de Capa	Dimensión / Neuronas	Función / Propósito
Entrada secuencial	Input (serie temporal)	(30 pasos, 1 característica)	Secuencia de retornos logarítmicos rezagados de Apple
LSTM 1	LSTM	96 unidades	Captura dependencias temporales de largo plazo
LSTM 2	LSTM (sin return_sequences)	48 unidades	Captura patrones temporales de corto plazo
Dropout	Regularización	0.20	Reduce sobreajuste sobre características temporales
Entrada exógena	Input (features externas)	n exógenas (~25 variables)	Factores macroeconómicos y de mercado
Normalización	Layer Normalization	—	Estabiliza distribución de entrada exógena
Densa 1	Dense	128 neuronas (ReLU)	Extracción de representaciones no lineales
Dropout	Regularización	0.20	Previene sobreajuste exógeno
Densa 2	Dense	64 neuronas (ReLU)	Reducción dimensional y abstracción de señales
Concatenación	Fusion layer	—	Combina señales secuenciales + exógenas
Densa 3	Dense	64 neuronas (ReLU)	Integración final de patrones

Componente	Tipo de Capa	Dimensión / Neuronas	Función / Propósito
Capa de salida	Dense + Sigmoid	1 neurona	Probabilidad de subida del activo (clasificación binaria)
Pérdida	Focal Loss ponderada	$\gamma=2, \alpha=(0.70,0.30)$	Enfatiza errores relevantes (control de clase)
Optimizador	Adam	LR inicial = 0.0003	Optimización adaptativa del gradiente

Tabla 3. Arquitectura Modelo Híbrido LSTM

La elección de una arquitectura híbrida se fundamenta en la naturaleza del problema y en la dinámica propia de los mercados financieros. Mientras los modelos autoregresivos tradicionales capturan únicamente información histórica del activo, esta propuesta combina el aprendizaje secuencial de las LSTM con señales exógenas relevantes del entorno macroeconómico y del mercado, lo que permite al modelo comprender tanto patrones internos como factores externos que influyen en la dirección del precio. Este enfoque busca equilibrar complejidad y capacidad predictiva, ofreciendo una solución más flexible y robusta que modelos lineales o puramente secuenciales, y adaptándose mejor a escenarios reales de decisión financiera donde múltiples fuentes de información convergen simultáneamente.

Con el fin de garantizar la reproducibilidad del entrenamiento, el modelo se realizó utilizando un máximo de 300 épocas y un tamaño de lote de 64 observaciones, preservando estrictamente el orden temporal de los datos para evitar sesgos de anticipación. La convergencia del modelo se controló mediante un esquema de *early stopping* basado en el AUC-PR sobre el conjunto de validación, con una paciencia de 20 épocas y restauración automática de los mejores pesos. De manera complementaria, se empleó un mecanismo adaptativo de reducción de la tasa de aprendizaje, el cual disminuye el *learning rate* en un factor de 0.5 tras seis épocas sin mejora, estableciendo un valor mínimo de 10^{-6} para garantizar estabilidad numérica. Finalmente, la regla de decisión se definió a partir de una búsqueda exhaustiva del umbral de clasificación sobre el conjunto de validación, seleccionando aquel que maximiza el coeficiente de correlación de Matthews bajo una restricción mínima de especificidad del 40%, alineando así el proceso de entrenamiento con los objetivos operativos del problema financiero.

LightGBM + GARCH

El modelo comparativo basado en Light Gradient Boosting Machine (LightGBM) se implementó como una alternativa no secuencial para la predicción de la dirección diaria del precio de Apple. LightGBM es un algoritmo de gradient boosting que construye un conjunto de árboles de decisión optimizados para capturar relaciones no lineales y efectos de interacción entre las variables predictoras, incluyendo los rezagos del activo y los factores exógenos del mercado. Su arquitectura orientada a hojas permite un aprendizaje más profundo y eficiente, característica que ha sido

destacada en estudios recientes donde LightGBM ha mostrado un desempeño notable al modelar dinámicas financieras complejas bajo esquemas de ensamble avanzados (Padhi et al., 2021). Como punto de partida se utilizó un estimador base con parámetros moderadamente regulares (400 árboles, tasa de aprendizaje de 0.05, num_leaves=31, muestreo por filas y columnas), lo que asegura un buen equilibrio entre capacidad predictiva y control del sobreajuste.

Posteriormente se aplicó una búsqueda aleatoria de hiperparámetros (RandomizedSearchCV) con validación temporal mediante TimeSeriesSplit, optimizando la métrica AUC-ROC, que es independiente del umbral de clasificación. Los mejores parámetros se refinaron con un procedimiento de early stopping sobre el conjunto de validación para determinar la cantidad óptima de árboles. Finalmente, se evaluaron distintos umbrales de decisión, seleccionando el óptimo mediante la métrica MCC, y se realizó una poda de características basada en importancia de variables. La definición de los hiperparámetros del modelo se realizó de manera explícita y sistemática, permitiendo una descripción precisa del proceso de entrenamiento y facilitando la replicabilidad de los resultados obtenidos, lo cual se resume a continuación en *Tabla 4*.

Categoría	Hiperparámetro	Valor / Configuración	Descripción
Modelo base	n_estimators (base)	400	Número inicial de árboles antes de optimización
Optimización	learning_rate	{0.02, 0.03, 0.05, 0.08}	Tasa de aprendizaje explorada
Complejidad	num_leaves	{31, 63, 127}	Control de complejidad del árbol
Regularización	min_child_samples	{20, 40, 80, 120}	Mínimo de observaciones por hoja
Submuestreo	subsample	{0.6, 0.8, 1.0}	Fracción de muestras por árbol
Submuestreo	colsample_bytree	{0.4, 0.6, 0.8, 1.0}	Fracción de variables por árbol
Regularización	reg_alpha	{0.0, 0.1, 0.3, 0.5}	Penalización L1
Regularización	reg_lambda	{3, 5, 7}	Penalización L2
Desbalance	class_weight	balanced	Ajuste automático por clase
Validación	Tipo de CV	TimeSeriesSplit (3 folds)	Validación cruzada temporal
Selección	Métrica de búsqueda	ROC AUC	Métrica invariante al umbral
Optimización	Método de búsqueda	RandomizedSearchCV (16 iter.)	Exploración estocástica
Entrenamiento	n_estimators (máx.)	2000	Límite superior para early stopping
Optimización	Early stopping	150 rondas	Control de sobreajuste
Aleatoriedad	random_state	42	Reproducibilidad del experimento
Computación	n_jobs	-1	Paralelización completa

Categoría	Hiperparámetro	Valor / Configuración	Descripción
Decisión	Umbral final	Optimizado por MCC en VALID	Regla de clasificación

Tabla 4. Hiperparámetros del Modelo LihtGBM

RESULTADOS

EVALUACIÓN

Para evaluar el desempeño de los modelos anteriormente mencionados (LSTM y LightGBM), cada enfoque fue probado con diferentes ventanas de tiempo construidas a partir de múltiples rezagos. Aunque se exploraron varias alternativas dentro del conjunto de datos, se destacan aquí las ventanas que mostraron el mejor comportamiento predictivo: 30 y 35 rezagos, seleccionadas por ofrecer un equilibrio superior entre métricas de validación, prueba y desempeño en la búsqueda de la mejor estrategia de trading. A continuación, se resume en la *Tabla 5.* el rendimiento de cada modelo mediante las métricas empleadas para su evaluación.

Modelo	Conjunto	Accuracy	F1 Score	Precisión	Recall	Especificidad	Sensibilidad	AUC ROC
LSTM + 30 lags	Validación	0.540	0.485	0.585	0.414	0.679	0.414	0.531
	Prueba	0.529	0.610	0.533	0.711	0.333	0.711	0.522
LSTM + 35 lags	Validación	0.519	0.493	0.550	0.447	0.598	0.447	0.500
	Prueba	0.425	0.432	0.442	0.422	0.429	0.422	0.425
LightGBM + 30 lags	Validación	0.793	0.823	0.745	0.919	0.655	0.919	0.923
	Prueba	0.540	0.592	0.547	0.644	0.429	0.644	0.563
LightGBM + 35 lags	Validación	0.739	0.753	0.747	0.758	0.719	0.758	0.783
	Prueba	0.471	0.489	0.489	0.489	0.452	0.489	0.482

Tabla 5. Métricas de Evaluación Modelos LSTM y LightGBM

El análisis comparativo de las métricas de evaluación revela diferencias significativas en el comportamiento de los modelos probados. En términos predictivos, LightGBM con una ventana de 30 rezagos obtuvo los mejores

resultados en validación, alcanzando los valores más altos en *accuracy* (0.793), *F1-score* (0.823) y *AUC-ROC* (0.923). Mientras que el *AUC-ROC* evidencia una mayor capacidad discriminativa independiente del umbral de clasificación, las métricas dependientes del umbral reflejan un equilibrio favorable entre precisión y sensibilidad para esta configuración. Por su parte, los modelos LSTM mostraron un desempeño moderado en métricas dependientes del umbral, con una ligera ventaja para la configuración de 30 rezagos, aunque con valores inferiores a los obtenidos por LightGBM en la fase de validación. Las versiones con 35 rezagos, tanto en LSTM como en LightGBM, presentaron un deterioro consistente en su rendimiento, sugiriendo un incremento en la complejidad del modelo sin una mejora proporcional en su capacidad predictiva. En conjunto, las métricas indican que LightGBM con 30 rezagos fue el modelo más sólido desde una perspectiva predictiva durante el proceso de validación.

En el conjunto de prueba, tanto el modelo LSTM con 30 rezagos como LightGBM con 30 rezagos evidencian un comportamiento moderado, caracterizado por una tasa relevante de falsos positivos y falsos negativos, reflejando la dificultad inherente de predecir la dirección diaria del activo en periodos no vistos, lo cual se denota en la *Figura 10*. El modelo LSTM mostró una mayor sensibilidad, logrando identificar un mayor número de movimientos positivos correctos, mientras que LightGBM presentó un equilibrio ligeramente superior entre precisión y sensibilidad. Estas diferencias, junto con la degradación de algunas métricas respecto a la validación, refuerzan la naturaleza compleja y ruidosa del problema de predicción financiera y subrayan la importancia de interpretar los resultados más allá de una única métrica.

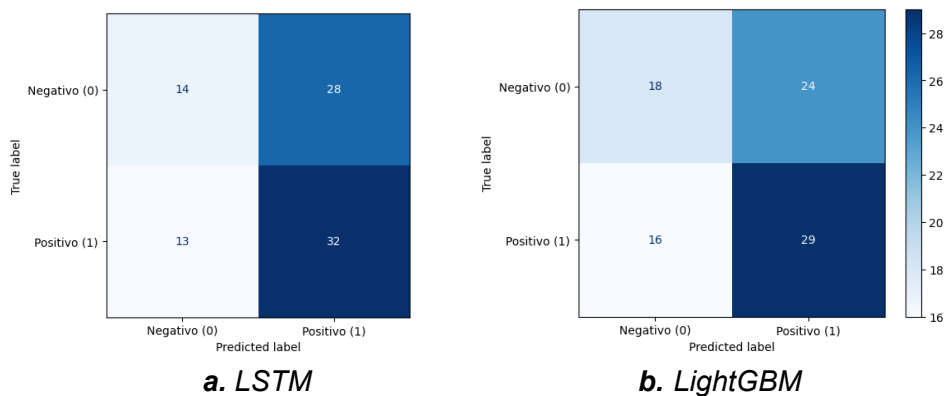


Figura 11. Matriz de confusión

BACKTESTING

El análisis del backtesting revela que el modelo LSTM con 30 rezagos obtuvo el mejor desempeño financiero entre todas las configuraciones evaluadas. Este modelo alcanzó un retorno acumulado del 28.05%, con un retorno anualizado del 13.90% y un Sharpe Ratio de 0.70, superando de forma notable tanto a los demás modelos como a la estrategia de Buy & Hold (*Figura 12.*), cuya rentabilidad acumulada fue del 14.45% con un Sharpe de 0.38. Además de presentar una relación riesgo–retorno más favorable, el LSTM mostró una volatilidad inferior a la del mercado de referencia, lo que refuerza su estabilidad en la generación de señales. Estos resultados sugieren que el modelo probablemente está identificando como positivos aquellos días donde la rentabilidad del activo es más marcada, lo que contribuye a mejorar la efectividad de la estrategia. En conjunto, la evidencia indica que, desde una perspectiva financiera, el LSTM con 30 rezagos es el modelo que mejor logra transformar su capacidad predictiva en una estrategia de trading rentable y con un perfil de riesgo más atractivo. A continuación, en la *Tabla 6.* se presentan los resultados del Backtesting de los modelos evaluados:

Modelo	Retorno Acumulado	Retorno Anualizado	Volatilidad	Sharpe Ratio
LSTM + 30 lags	28.05%	13.90%	19.95%	0.70
LSTM + 35 lags	15.49%	8.68%	18.82%	0.46
LightGBM + 30 lags	23.40%	13.25%	25.04%	0.53
LightGBM + 35 lags	19.67%	10.13%	16.97%	0.60
Buy & Hold (Referencia)	14.45%	10.10%	26.90%	0.38

Tabla 6. Backtesting Modelos Evaluados comparado con estrategia de Referencia Buy & Hold

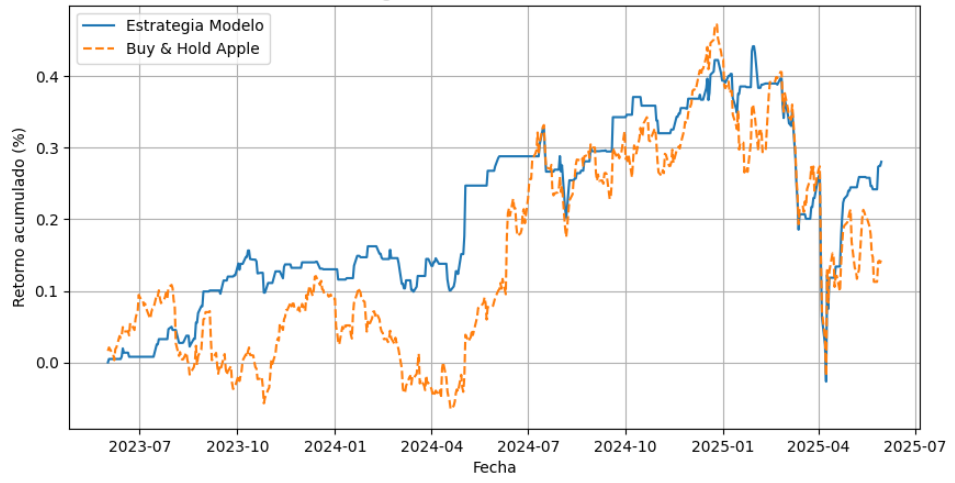


Figura 12. Backtesting – Retorno Acumulado LSTM + Lag 30

CONCLUSIONES

Los resultados obtenidos en este estudio permiten establecer varias conclusiones relevantes tanto desde la perspectiva predictiva como desde la construcción de una estrategia de trading basada en señales cuantitativas. En primer lugar, se evidencia que el modelo LSTM fue el que mejor desempeño financiero alcanzó, obteniendo el mayor retorno acumulado y una volatilidad más controlada en comparación con los demás modelos evaluados. Su comportamiento sugiere que esta arquitectura logró capturar patrones temporales que, aunque moderados en capacidad predictiva, fueron suficientes para generar señales más rentables dentro del backtesting.

Por otra parte, el análisis de las métricas de evaluación demuestra que LightGBM fue el modelo con mejor desempeño predictivo, destacándose en accuracy, F1 score y AUC-ROC durante la fase de validación. Esto confirma su capacidad para capturar relaciones no lineales y explotar de manera eficiente la estructura de los datos en términos de clasificación direccional, aunque su capacidad de generalización hacia el conjunto de prueba resultó más limitada que la del modelo LSTM.

Los resultados evidencian que, aunque los modelos presentan un desempeño competitivo, la predicción de la dirección diaria de la acción de Apple continúa siendo un desafío debido al carácter altamente ruidoso del mercado y a la influencia simultánea de factores externos como noticias corporativas, decisiones de política monetaria y dinámicas propias de los inversionistas institucionales. Esta naturaleza volátil, sumada al riesgo de sobreajuste asociado a la alta dimensionalidad del conjunto de características, limita la capacidad de generalización de los modelos y plantea oportunidades para perfeccionar las metodologías en investigaciones futuras.

Adicional a lo anterior, la predicción de movimientos en activos financieros enfrenta limitaciones estructurales derivadas de la hipótesis de eficiencia de los mercados, la naturaleza no estacionaria de las series y la presencia de ruido significativo. Incluso modelos no lineales de alta capacidad, como las redes LSTM y LightGBM, no garantizan una capacidad predictiva estable en el tiempo.

Además, la relación entre variables exógenas y retornos financieros puede cambiar abruptamente ante eventos estructurales, choques macroeconómicos o cambios regulatorios, lo que limita la capacidad de generalización del modelo fuera del periodo de entrenamiento. El uso de ventanas temporales fijas y la imputación de valores faltantes implican supuestos que, si bien necesarios para la implementación práctica, pueden introducir sesgos difíciles de cuantificar. Por ello, los resultados deben interpretarse como evidencia empírica condicionada al periodo analizado y no como relaciones causales permanentes.

Tanto en redes neuronales como en algoritmos de boosting, la selección adecuada de parámetros constituye un componente crítico para lograr niveles superiores de precisión y estabilidad. Por ello, una búsqueda más amplia y sistemática podría potenciar significativamente el rendimiento de los modelos evaluados y abrir la puerta a resultados aún más robustos.

El uso de volatilidad condicional mediante GARCH aportó estabilidad al conjunto de características y permitió incorporar información relevante sobre los periodos de aumento o disminución de incertidumbre en el mercado, su inclusión sigue siendo adecuada para estrategias de inversión que buscan capturar cambios de régimen y gestionar el riesgo de manera más informada.

Del mismo modo, el enfoque híbrido que combina análisis secuencial, factores exógenos y modelación de volatilidad constituye una base sólida para el desarrollo de estrategias de trading más avanzadas. Este enfoque puede potenciarse mediante la integración de modelos de amplitud para la definición de niveles de take profit y stop loss, la automatización en la selección de umbrales de decisión y la implementación de modelos de ensamblaje que combinen las fortalezas de cada técnica.

En conjunto, los resultados del estudio muestran que, aunque las señales direccionales diarias sean débiles y difíciles de anticipar, modelos como LSTM y LightGBM logran transformar información limitada en estrategias rentables. Este proyecto sienta las bases para continuar construyendo modelos más robustos y explorar enfoques adicionales que permitan obtener señales más consistentes dentro de un entorno altamente incierto como el mercado accionario.

TRABAJOS FUTUROS

Existen varias oportunidades de mejora y ampliación del presente estudio que podrían fortalecer la estrategia de trading propuesta y contribuir al desarrollo de modelos predictivos con mayor capacidad de generalización:

- Profundizar en técnicas de sampling y en la búsqueda de hiperparámetros: La exploración de métodos avanzados de muestreo, así como la implementación de búsqueda más exhaustiva y sistemática de hiperparámetros, podría mejorar el equilibrio entre clases y optimizar el rendimiento de los modelos evaluados.
- Modelar la amplitud de precios para definir reglas de take profit y stop loss: Incorporar la predicción de la amplitud intradía o de rangos de movimiento permitiría diseñar límites de pérdidas y objetivos de ganancia más precisos, lo cual podría mejorar las métricas actuales del modelo y evitar operaciones en días con movimientos marginales que aportan poco valor a la estrategia.
- Incluir modelos de referencia (Baselines) que sirvan como punto de partida para medir el desempeño de los modelos desarrollados, permitiendo contrastar sus resultados frente a enfoques más sencillos y justificar de forma clara el aporte adicional de las técnicas avanzadas de Machine Learning empleadas.
- Desplegar la estrategia en un entorno real, implementándola en un escenario de mercado con dinero de prueba para evaluar su desempeño operativo bajo condiciones efectivas de negociación. Este proceso permitiría validar la aplicabilidad práctica del modelo y analizar cómo responde la estrategia frente a la dinámica propia del mercado.
- Probar las metodologías desarrolladas en otros activos financieros con el fin de analizar su capacidad de generalización, comparar su rendimiento en diferentes entornos de volatilidad y validar la robustez de la estrategia más allá del caso específico de Apple.
- Implementar esquema de validación tipo Walk-Forward en los modelos LSTM lo cual permitiría evaluar de forma más realista la capacidad de generalización del modelo en distintos periodos del mercado. Este enfoque podría ayudar a reducir el riesgo de sobreajuste, mejorar la estabilidad de las métricas y obtener estimaciones más robustas del desempeño en escenarios fuera de muestra.

REFERENCIAS

- Agarwal, A. (2022). *Technical analysis: Getting started basic introduction and beginner's guide*. <https://www.babson.edu/finance-institute/guides/>
- AL-Najjar, D. (2016). Modelling and estimation of volatility using ARCH/GARCH models in Jordan's stock market. *Asian Journal of Finance & Accounting*, 8(1), 152–163. <https://doi.org/10.5296/ajfa.v8i1.9129>
- Amat Rodrigo, J., & Escobar Ortiz, J. (2023, septiembre). *Deep learning para la predicción de series temporales: Redes neuronales recurrentes (RNN), Gated Recurrent Unit (GRU) y Long Short-Term Memory (LSTM)*. <https://www.cienciadedatos.net/documentos/py51-modelos-arima-sarimax-python.html>
- Bentes, S. R., & Navas, R. (2013). The fundamental analysis: An overview. *International Journal of Latest Trends in Finance and Economic Sciences*, 3, 389–393. <https://www.researchgate.net/publication/340511850>
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327. [https://doi.org/https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/https://doi.org/10.1016/0304-4076(86)90063-1)
- Engle, R. (2001). GARCH 101: The use of ARCH GARCH in applied econometrics. *The Journal of Economic Perspectives*, 15, No 4, 157–168. <https://doi.org/https://www.aeaweb.org/articles?id=10.1257/jep.15.4.157>
- Federal Reserve Bank of St. Louis. (2025). *FRED Economic Data*. <https://fred.stlouisfed.org>.
- Hartanto, A. D., Kholik, Y. N., & Pristyanto, Y. (2023). Stock price time series data forecasting using the Light Gradient Boosting Machine (LightGBM) model. *International Journal on Informatics Visualization*, 7(4), 2270–2279. <https://joiv.org/index.php/joiv/article/view/1740>

- IBM. (s/f). *Guía de CRISP-DM de IBM SPSS Modeler*. <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. <https://doi.org/https://doi.org/10.1007/978-1-4614-6849-3>
- Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, 170, 1168–1173. <https://doi.org/10.1016/j.procs.2020.03.049>
- Mutinda, J. K., & Langat, A. K. (2024). Stock price prediction using combined GARCH-AI models. *Scientific African*, 26, e02374. <https://doi.org/10.1016/j.sciaf.2024.e02374>
- Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Ijaz, M. F. (2021). A fusion framework for forecasting financial market direction using enhanced ensemble models and technical indicators. *Mathematics*, 9(21), 2646. <https://doi.org/10.3390/math9212646>
- Rusere, W., & Kaseke, F. (2021). Modeling South African stock market volatility using univariate symmetric and asymmetric GARCH models. *Indian Journal of Finance and Banking*, 6(1), 1–15. <https://www.cribfb.com/journal/index.php/ijfb/article/view/1177/987>
- The Nippon Technical Analysts Association. (2018). *The first step guide to technical analysis*. <https://anyflip.com/meoij/yknz/basic>
- Yahoo Finance. (2025). *Historical market data*. <https://finance.yahoo.com>