

UNIVERSIDAD DE EL SALVADOR
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
ESCUELA DE POSGRADO



TRABAJO DE POSGRADO

**USO DE MODELOS SUPERVISADOS POR ANALISTAS FINANCIEROS EN LA
PREDICCIÓN DE PRECIOS DE ACCIONES DE GOOGLE Y AMAZON**

PARA OPTAR AL GRADO DE

MAESTRO EN ADMINISTRACIÓN FINANCIERA

PRESENTADO POR

LICENCIADO MARIO GERARDO GARCÍA GARCÍA

LICENCIADO JAIME ISAAC PEÑA MEJÍA

DOCENTE ASESORA

MAESTRA LIDIA MARGARITA CALDERÓN FUENTES

OCTUBRE, 2025

SANTA ANA, EL SALVADOR, CENTROAMÉRICA

UNIVERSIDAD DE EL SALVADOR

AUTORIDADES



ING. JUAN ROSA QUINTANILLA QUINTANILLA

RECTOR

DRA. EVELYN BEATRIZ FARFÁN MATA

VICERRECTORA ACADÉMICA

M.Sc. ROGER ARMANDO ARIAS ALVARADO

VICERRECTOR ADMINISTRATIVO

LICDO. PEDRO ROSALÍO ESCOBAR CASTANEDA

SECRETARIO GENERAL

LICDA. ANA RUTH AVELAR VALLADARES

DEFENSORA DE LOS DERECHOS UNIVERSITARIOS

LICDO. CARLOS AMÍLCAR SERRANO RIVERA

FISCAL GENERAL

FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE

AUTORIDADES



M.Ed. ROBERTO CARLOS SIGÜENZA CAMPOS

DECANO

DR. JOSÉ GUILLERMO GARCÍA ACOSTA

VICEDECANO

LICDO. JAIME ERNESTO SERMEÑO DE LA PEÑA

SECRETARIO

M. Ed. MIGUEL ÁNGEL CRUZ

DIRECTOR DE LA ESCUELA DE POSGRADO

Agradecimientos

Expreso mi más profundo y sincero agradecimiento a las personas que, con su apoyo incondicional y guía, fueron pilares fundamentales para la finalización de esta tesis.

En primer lugar, elevo mi gratitud a Dios por la fortaleza, la sabiduría y la salud que me concedió para alcanzar este importante logro académico.

A mi madre, Flor García, por su amor incondicional y su constante apoyo; su fe y confianza en mí fueron la fuerza impulsora que me permitió perseverar.

A mi padre, Mario García, por haberme inculcado el valor de la perseverancia y la dedicación, cualidades indispensables en este camino.

A mi novia, Abigail Granados, por su amor y comprensión, una inspiración constante a lo largo de este trayecto.

Asimismo, extiendo mi gratitud a mi compañero de tesis, Jaime Peña, cuya invaluable colaboración y respaldo fueron esenciales desde el inicio de este proyecto.

A nuestra asesora, Máster Lidia Margarita Calderón Fuentes, por su valiosa guía y dedicación en este trabajo.

Finalmente, a todos mis amigos y familiares, gracias por su respaldo y compañía. Este logro no habría sido posible sin el invaluable apoyo de cada uno de ustedes.

Mario Gerardo García García

Agradecimientos

A Dios, por la sabiduría y la fortaleza concedidas para la culminación de este proyecto.

A mi madre, Ursula Mejía López, por su amor incondicional y por ser mi principal fuente de apoyo.

A mi esposa, Fátima Patricia Noyola Jiménez, por su paciencia, aliento y por ser el pilar fundamental en los momentos más exigentes de este proceso.

A mi padre, Juan Peña, por ser una constante inspiración y motivación en mi vida.

A mi compañero de tesis, Mario García, por su invaluable apoyo y comprensión a lo largo de esta investigación.

A nuestra asesora, Máster Lidia Margarita Calderón Fuentes, cuya guía experta y dedicación fueron indispensables para el desarrollo y éxito de este trabajo.

A mis amigos y familiares, por su constante apoyo y por hacer este viaje académico más ameno y significativo.

A todos, mi más sincero agradecimiento.

Jaime Isaac Peña Mejía

INDICE

Introducción	xii
Capítulo I: Planteamiento del problema.....	15
1.1 Delimitación del problema	15
1.2 Preguntas de investigación	16
1.3 Objetivos de la investigación.....	17
1.3.1 Objetivo general.....	17
1.3.2 Objetivos específicos	17
1.4 Justificación	17
1.5 Limites y alcances.....	20
1.5.1 Limites	20
1.5.2 Alcances.....	20
Capítulo II: Marco teórico de referencia.....	22
2.1 Antecedentes de la predicción de precios en mercados financieros	22
2.1.1 Evolución histórica de los métodos predictivos en finanzas.....	22
2.1.2 Estudios previos sobre predicción de precios de acciones tecnológicas	22
2.2 Fundamentos teóricos de los mercados financieros.....	23
2.2.1 Teoría de mercados eficientes y sus limitaciones	23
2.2.2 Factores macroeconómicos en la valoración de activos financieros	24
2.2.3 Características específicas del mercado de acciones tecnológicas	26
2.2.4 Fundamentos de rendimiento y riesgo en la predicción de activos	27
2.3 Fundamentos de aprendizaje automático en finanzas.....	30
2.3.1 Principios del aprendizaje supervisado	30
2.3.2 Algoritmos de Random Forest: funcionamiento y aplicaciones financieras	32
2.3.3 Algoritmos de Gradient Boosting: fundamentos y aplicación en predicción de precios	35
2.4 Variables macroeconómicas en la predicción de precios de acciones	38
2.4.1 Tasas de Interés: mecanismos de influencia en mercados bursátiles	38
2.4.2 Inflación: Efectos en la valoración de empresas tecnológicas	39
2.4.3 Datos Históricos de precios e indicadores técnicos como predictores.....	40
2.5 Comparación entre modelos predictivos tradicionales y supervisados	42
2.5.1 Modelos econométricos tradicionales: ARIMA, GARCH y sus Limitaciones.....	42
2.5.2 Ventajas comparativas de los modelos supervisados en contextos financieros.....	45

2.6 Evaluación del Desempeño de modelos predictivos en finanzas	46
2.6.1 Métricas de evaluación y validación en contextos financieros	46
2.6.2 Consideraciones sobre sobreajuste y generalización en predicción financiera	47
2.7 Implementación práctica de modelos supervisados en trading algorítmico	49
2.7.1 Arquitectura de sistemas para trading basado en modelos predictivos.....	49
2.7.2 Consideraciones prácticas: costos de transacción, liquidez y escalabilidad	51
2.8 Dirección futura en predicción de precios de acciones tecnológicas	52
2.8.1 Integración de datos alternativos y análisis de sentimiento	52
2.8.2 Aplicaciones de aprendizaje profundo y modelos híbridos	53
2.9 Conclusiones del marco teórico.....	55
Capítulo III: Diseño metodológico	57
3.1 Enfoque de la investigación.....	57
3.1.1 Enfoque metodológico mixto y análisis de escenarios.....	57
3.2 Método.....	58
3.3 Diseño de estudio.....	58
3.4 Población y muestra.....	58
3.5 Técnicas e instrumentos de recolección	59
3.6 Hipótesis o supuestos de investigación	60
3.7 Operacionalización de variables.....	60
3.8 Estrategias de recolección, procesamiento y análisis de la información	62
3.9 Consideraciones éticas.....	63
Capítulo IV: Metodología empírica y análisis de datos	64
4.1 Construcción del conjunto de datos	64
4.1.1 Composición y fuentes del dataset	64
4.1.2 Estructuración de datos y preparación inicial.....	65
4.1.3 Vista previa de la estructura de datos.	65
4.1.4 Ingeniería de características y definición de variables.....	66
4.2 Análisis exploratorio de datos (EDA)	68
4.2.1 Visualización de series temporales.....	68
4.2.2 Observaciones clave del EDA:	72
4.3 Impacto de tasas de interés en precios de cierre de las acciones de Google y Amazon.....	72
4.3.1 Análisis visual de la relación entre precios y tasas de interés.....	73
4.3.2 Resultados de rendimiento del modelo para Google	74

4.3.3	Análisis de la importancia de la variable (Treasury 10Y) para Google Close	76
4.3.4	Interpretación de la relación (análisis de dependencia e impacto) para Google	78
4.3.5	Conclusión del análisis de dependencia para Google	82
4.4	Análisis detallado de los resultados para Amazon	82
4.4.1	Impacto de las tasas de interés en la precisión del modelo para Amazon.....	82
4.4.2	Análisis de la influencia de la tasa de interés en el modelo Random Forest para Amazon.....	83
4.4.3	Conclusión del análisis de dependencia para Amazon.....	85
4.4.4	Interpretación del impacto a nivel de predicción (SHAP) para Amazon.	86
4.4.5	Conclusión del análisis de dependencia para Amazon.....	87
4.5	Conclusión general del objetivo 1	88
4.6	Evaluación de la influencia de los índices de inflación en la precisión de los modelos supervisados para la predicción de precios de activos financieros.....	88
4.6.1	Preparación del conjunto de datos y definición de la variable objetivo	89
4.6.2	Análisis exploratorio de datos (EDA)	90
4.6.3	Estrategia de validación y división del conjunto de datos.....	94
4.6.4	Resultados preliminares del modelado para Google.....	95
4.6.5	Resultados preliminares del modelado para Amazon.....	97
4.6.6	Medición de la contribución marginal de la inflación (ablation study)	99
4.6.7	Análisis de la importancia de variables del modelo base	101
4.6.8	Conclusión del análisis de importancia para Google	103
4.6.9	Análisis de la importancia de variables del modelo base para Amazon.....	103
4.6.10	Análisis de interpretabilidad del modelo para Google: comprendiendo la influencia de las variables clave	106
4.6.11	Análisis de interpretabilidad para Amazon: gráficos de dependencia parcial (PDP)	111
4.6.12	Análisis SHAP para Amazon	114
4.6.13	Medición de la contribución marginal de la inflación para Amazon (ablation study)	117
4.6.14	Análisis de robustez de variables secundarias: el caso del VIX.....	118
4.6.15	Discusión final y verificación de la hipótesis 3.....	120
4.7	Incidencia de los datos históricos de precios en la precisión de los modelos supervisados.	121
4.7.1	Metodología para evaluar la incidencia de los datos históricos (objetivo específico 3)	121

4.7.2	Resumen de los resultados (modelo completo con datos históricos)	123
4.7.3	Evaluación del modelo con datos históricos para Google.....	123
4.7.5	Conclusiones preliminares del modelo completo	126
4.7.6	Evaluación de los modelos sin históricos para Google.....	126
4.7.7	Conclusión final del objetivo específico N°3	128
4.7.8	Análisis de importancia de las variables predictoras.....	129
4.7.9	Análisis visual de errores.....	133
4.8	Evaluar la precisión de los modelos supervisados seleccionados (Random Forest y Gradient Boosting) frente a métodos tradicionales en la predicción de precios de activos financieros.....	135
4.8.1	Análisis comparativo de modelos para Google	136
4.8.2	Configuración de modelos supervisados	138
4.8.3	Evaluación comparativa del rendimiento predictivo	139
4.8.4	Discusión de resultados para GOOGL y conclusión parcial	145
4.8.5	Análisis comparativo de modelo ARIMA para Amazon.....	146
4.8.6	Discusión de resultados y verificación parcial de H1 para GOOGL.....	153
4.8.7	Análisis detallado del gráfico (AMZN)	154
4.8.8	Análisis gráfico de los pronósticos para Amazon	155
4.8.9	Discusión de resultados y verificación de H1 para AMZN.....	156
4.9	Análisis de materialidad financiera y validación económica del modelo	157
4.9.1	De la precisión estadística a la utilidad económica	157
4.9.2	Metodología del backtest financiero	157
4.9.3	Resultados comparativos de la simulación financiera	159
4.9.4	Interpretación de los resultados económicos.....	159
4.9.5	Análisis de robustez: simulación en un escenario adverso.....	160
4.9.6	Aplicabilidad práctica y propuesta de valor para el gestor de inversiones.....	161
4.9.7	Conclusión de la validación económica	161
Capítulo V: Conclusiones y recomendaciones		163
5.1	Conclusiones	163
5.2	Recomendaciones	165
Referencias.....		166
Anexos		168

Índice de Tablas

Tabla 1 <i>Operacionalización de variables independientes y dependientes</i>	61
Tabla 2 <i>Primeras 5 observaciones del conjunto de datos (head)</i>	66
Tabla 3 <i>Últimas 5 observaciones del conjunto de datos (tail)</i>	66
Tabla 4 <i>Comparación de rendimiento de modelos con y sin tasa de interés</i>	75
Tabla 5 <i>Importancia de características del modelo Gradient Boosting Completo</i>	76
Tabla 6 <i>Comparación de rendimiento de modelos para Amazon (AMZN)</i>	82
Tabla 7 <i>Importancia de características para Amazon (Random Forest)</i>	83
Tabla 8 <i>Ranking de importancia de variables para el modelo base - GOOGLE</i>	102
Tabla 9 <i>Ranking de importancia de variables para el modelo base - Amazon</i>	104
Tabla 10 <i>Comparación de métricas de rendimiento del modelo Random Forest para Amazon con y sin la inclusión del VIX</i>	119
Tabla 11 <i>Resultados de rendimiento de modelos con datos históricos para Google</i>	124
Tabla 12 <i>Resultados de rendimiento de modelos con datos históricos para Amazon</i>	125
Tabla 13 <i>Tabla comparativa de rendimiento (Random Forest - Google)</i>	127
Tabla 14 <i>Tabla comparativa de rendimiento (Random Forest - Amazon)</i>	128
Tabla 15 <i>Top 10 características más importantes para la predicción de Google</i>	130
Tabla 16 <i>Top 10 características más importantes para la predicción de Amazon</i>	132
Tabla 17 <i>Resumen del modelo ARIMA(1, 1, 1) para Google</i>	136
Tabla 18 <i>Coeficientes del modelo para Google</i>	137
Tabla 19 <i>Diagnósticos del modelo (análisis de residuos)</i>	137
Tabla 20 <i>Comparación de métricas de rendimiento para GOOGL en el conjunto de prueba</i> .	139
Tabla 21 <i>Muestra de predicciones vs. valores reales para GOOGL en el conjunto de prueba.</i>	141
Tabla 22 <i>Resumen del modelo ARIMA(0, 1, 0) para AMZN</i>	147
Tabla 23 <i>Coeficientes del modelo ARIMA(0, 1, 0) para AMZN</i>	147
Tabla 24 <i>Diagnósticos del modelo ARIMA (0, 1, 0) para AMZN (Análisis de residuos)</i>	148
Tabla 25 <i>Comparación de métricas de rendimiento para AMZN en el conjunto de prueba</i>	149
Tabla 26 <i>Muestra de predicciones vs. valores reales para AMZN en el conjunto de prueba</i> ...	151
Anexo 1	169
Tabla 27 <i>Predicciones vs. valores reales para GOOGL en el conjunto de prueba</i>	169
Anexo 2	183
Tabla 28 <i>Predicciones vs. valores reales para AMZN en el conjunto de prueba</i>	183

Índice de figuras

Figura 1: Evolución del precio de cierre diario de la acción de amazon (amzn) durante el período de estudio (2021-2024)	68
Figura 2: Evolución de la tasa de rendimiento de los bonos del tesoro de ee. Uu. A 10 años (treasury_10y) (2021-2024)	69
Figura 3: Evolución de la tasa de rendimiento de los bonos del tesoro de ee. Uu. A 10 años (treasury_10y) (2021-2024)	69
Figura 4: Evolución del índice de volatilidad cboe (vix) durante el período de estudio (2021-2024)	70
Figura 5: Evolución del índice de mercado S&P 500 durante el período de estudio (2021-2024).....	70
Figura 6: Evolución del índice tecnológico nasdaq composite durante el período de estudio (2021-2024)	71
Figura 7: Evolución de las expectativas de inflación a 5 años (t5yie) durante el período de estudio (2021-2024)	71
Figura 8: Evaluación de precios de Google y Amazon vs tasa de interés (2021 - 2024)	73
Figura 9: Gráfico de dependencia parcial (pdp) impacto de la tasa de intereses en la predicción del precio (Gradient Bosting).....	78
Figura 10: Gráfico de resumen shap para el modelo completo de Google.....	80
Figura 11: Pdp del impacto de la tasa de interés en la predicción de amazon (Random Forest)..	84
Figura 12: Gráfico de resumen shap para Amazon (Randomforest_completo).....	86
Figura 13: Evolución de precios vs. expectativas de inflación (2021-2024).....	91
Figura 14: Matriz de correlación de todas las variables.....	92
Figura 15: Gráfico de dependencia parcial (pdp) para variables clave – Google.....	107
Figura 16: Gráfico de resumen para el modelo de Google.....	109
Figura 17: Dependencia shap: impacto de la inflación en la predicción (Google).....	110
Figura 18: Dependencia parcial para variables clave – Amazon.....	112
Figura 19: Resumen shap (beeswarm) para el modelo de Amazon.....	115
Figura 20: Gráfico de predicciones vs. precios reales y errores residuales de Google.....	133
Figura 21: Gráfico de predicciones vs. precios reales y errores residuales de Amazon.....	134
Figura 22: Comparación gráfica de predicciones vs. valores reales para Google.....	144
Figura 23: Comparación de modelos y pronósticos futuro para amzn.....	155

Introducción

En el dinámico y complejo ámbito de los mercados financieros, la predicción de los precios de las acciones se erige como un pilar fundamental para la toma de decisiones estratégicas. Para inversionistas, gestores de portafolios y analistas, la capacidad de anticipar los movimientos del mercado no solo es crucial para la gestión de riesgos, sino también para la optimización en la asignación de capital. Esta necesidad se magnifica en el sector tecnológico, caracterizado por su alta volatilidad y su sensibilidad a una multiplicidad de factores económicos, tecnológicos y de comportamiento.

Empresas como Google y Amazon, líderes indiscutibles de la era digital, representan un desafío particular para los modelos de predicción. Los métodos tradicionales, como los modelos econométricos lineales (por ejemplo, ARIMA), han demostrado tener limitaciones significativas. Suelen asumir relaciones lineales y una estabilidad temporal que raramente se observa en los mercados financieros modernos, lo que los hace insuficientes para capturar los patrones no lineales y la compleja interacción de variables que dictan las valoraciones de estos gigantes tecnológicos.

Esta insuficiencia ha motivado la búsqueda de enfoques más sofisticados y robustos. En este contexto, los algoritmos de aprendizaje automático, y en particular los modelos supervisados como Random Forest y Gradient Boosting, han surgido como herramientas prometedoras. Estos modelos ofrecen una capacidad superior para modelar relaciones no lineales, manejar grandes volúmenes de datos complejos e identificar patrones que permanecen ocultos para los métodos convencionales.

La presente investigación aborda este desafío mediante el diseño y la evaluación de un modelo de predicción basado en algoritmos de aprendizaje supervisado. El estudio se enfoca en predecir los precios de las acciones de Google y Amazon durante el período 2021-2024, una etapa en la que los modelos de aprendizaje supervisados pueden contribuir a mejorar la precisión en la predicción de precios de activos financieros. Para ello, se busca determinar el impacto específico de las tasas de interés y la inflación, evaluar la incidencia de los datos históricos de precios y, finalmente, comparar la precisión de los modelos Random Forest y Gradient Boosting frente a los métodos tradicionales.

Este trabajo de posgrado se estructura de la siguiente manera:

Capítulo I: Se presenta el planteamiento del problema, delimitando el alcance de la investigación, formulando las preguntas y los objetivos que guían el estudio, y exponiendo su justificación.

Capítulo II: Se desarrolla el marco teórico, revisando la literatura existente sobre la predicción de precios, los fundamentos de los mercados financieros, el aprendizaje automático aplicado a las finanzas y los modelos predictivos relevantes.

Capítulo III: Se detalla el diseño metodológico, explicando el enfoque cuantitativo, las hipótesis, la operacionalización de las variables y las estrategias utilizadas para la recolección y análisis de los datos.

Capítulo IV: Se expone la metodología empírica y el análisis de los datos, donde se construyen los modelos, se evalúa su rendimiento y se interpretan los resultados en función de los objetivos planteados.

Capítulo V: Finalmente, se presentan las conclusiones generales de la investigación, resumiendo los hallazgos clave y ofreciendo recomendaciones para futuros analistas y gestores financieros.

A través de este análisis riguroso, esta tesis no solo busca validar empíricamente la superioridad de los modelos de aprendizaje automático, sino también ofrecer un marco replicable que aporte valor práctico al sector financiero, proporcionando herramientas más precisas para navegar en un entorno de alta incertidumbre.

Capítulo I: Planteamiento del problema

1.1 Delimitación del problema

En el ámbito financiero, la predicción de precios de acciones es una tarea fundamental para la toma de decisiones estratégicas por parte de inversionistas, gestores de portafolios y analistas de mercado. Las acciones de empresas tecnológicas como Google y Amazon, caracterizadas por su alta volatilidad, representan un desafío particular debido a la cantidad de factores económicos, tecnológicos y de comportamiento que influyen en su valoración. Poder anticipar con mayor precisión los movimientos de precios de estas acciones no solo ayuda a gestionar riesgos, sino que también optimiza la asignación de recursos en portafolios de inversión.

Sin embargo, los métodos tradicionales de predicción, como los modelos econométricos lineales (regresiones y Autoregressive Integrated Moving Average), presentan limitaciones importantes al asumir relaciones lineales y estabilidad temporal en los datos financieros. Estas deficiencias han incentivado la búsqueda de enfoques más avanzados. Como destacan (Hastie, 2009), los algoritmos de aprendizaje automático, en particular los modelos supervisados, ofrecen una capacidad superior para modelar relaciones no lineales y manejar grandes volúmenes de datos complejos. Este enfoque ha permitido explorar nuevas formas de analizar los mercados financieros, superando las restricciones de los métodos clásicos.

Además, la predicción de precios de acciones de empresas como Google y Amazon involucra múltiples factores: indicadores técnicos como la volatilidad y el volumen de transacciones, elementos macroeconómicos como las tasas de interés y la inflación, y variables del sector tecnológico, incluyendo la innovación y la percepción del mercado. Durante el periodo 2021-2024, los mercados financieros han experimentado cambios significativos, tanto por el avance de la tecnología como por eventos económicos globales. En este sentido De Prado (2018) enfatiza que el uso de modelos avanzados como Random Forest y Gradient Boosting permite identificar patrones ocultos y mejorar la precisión de las predicciones, incluso en entornos financieros caracterizados por alta incertidumbre.

En este estudio, se propone diseñar un modelo de predicción utilizando algoritmos de aprendizaje supervisado, como Random Forest y Gradient Boosting, entrenados con datos

históricos de precios de las acciones de Google y Amazon en el periodo 2021-2024. Al considerar variables macroeconómicas relevantes como tasas de interés, inflación y volumen de transacciones, se espera demostrar cómo estos modelos pueden mejorar la precisión en la predicción de precios de activos financieros y optimizar la toma de decisiones estratégicas frente a métodos tradicionales.

1.2 Preguntas de investigación

¿Cómo un modelo supervisado puede contribuir a la precisión de precios de activos financieros?

1. ¿Qué impacto tienen las tasas de interés en la predicción de los precios de activos financieros utilizando modelos supervisados?
2. ¿Cómo influyen los índices de inflación con la precisión de los modelos supervisados en la predicción de los precios de activos financieros?
3. ¿En qué medida los datos históricos de precios inciden en la precisión de los modelos supervisados para predecir los precios de acciones financieras?
4. ¿Cómo se compara la precisión de los modelos supervisados seleccionados (Random Forest y Gradient Boosting) frente a métodos tradicionales en la predicción de precios de activos financieros?

1.3 Objetivos de la investigación

1.3.1 Objetivo general

Evaluar cómo los modelos supervisados pueden contribuir a la precisión de precios de activos financieros.

1.3.2 Objetivos específicos

1. Determinar el impacto que tienen las tasas de interés en la predicción de precios de activos financieros utilizando modelos supervisados.
2. Evaluar la influencia de los índices de inflación en la precisión de los modelos supervisados para la predicción de los precios de activos financieros.
3. Evaluar la incidencia de los datos históricos de precios en la precisión de los modelos supervisados para la predicción de precios de acciones financieras.
4. Evaluar la precisión de los modelos supervisados seleccionados (Random Forest y Gradient Boosting) frente a métodos tradicionales en la predicción de precios de activos financieros.

1.4 Justificación

La predicción precisa de precios de activos financieros constituye un pilar en el ámbito financiero, siendo fundamental para la toma de decisiones estratégicas de inversionistas, gestores de portafolios y analistas de mercado. Esta necesidad es particularmente crítica en el caso de las acciones de empresas tecnológicas como Google y Amazon, cuya alta volatilidad y complejidad dificultan su valoración. Según Prado (2018) los métodos tradicionales como las regresiones lineales o los modelos ARIMA presentan limitaciones al asumir relaciones lineales y estabilidad temporal, lo que resulta insuficiente para capturar patrones no lineales y dinámicos en los mercados actuales.

En este contexto, los avances en aprendizaje automático, especialmente los modelos supervisados como Random Forest y Gradient Boosting, han emergido como herramientas

prometedoras para superar estas deficiencias (Hastie, *The Elements of Statistical Learning*, 2009). Estos modelos permiten manejar grandes volúmenes de datos complejos y modelar relaciones no lineales, proporcionando predicciones más precisas y adaptadas a la naturaleza cambiante de los mercados financieros. No obstante, su aplicación en escenarios específicos como los mercados volátiles y la integración de múltiples variables macroeconómicas sigue siendo limitada, lo que evidencia un vacío en la literatura que esta investigación busca abordar.

Desde una perspectiva científica, este estudio presentará como alternativas a esta situación los modelos supervisados para mejorar la precisión en la predicción de precios al incorporar factores macroeconómicos como tasas de interés, inflación y eventos económicos globales inesperados. Además, validará empíricamente el desempeño de estos modelos frente a métodos tradicionales, consolidando teorías existentes y proporcionando nuevas perspectivas para el análisis financiero mediante aprendizaje automático.

Desde un enfoque práctico, los resultados de esta investigación beneficiarán directamente al sector financiero, particularmente a bancos de inversión, fondos de cobertura y gestores de portafolios. Estas instituciones requieren herramientas predictivas más precisas para optimizar la asignación de recursos y gestionar riesgos en un entorno de alta incertidumbre.

El período 2021-2024, caracterizado por la recuperación postpandemia, el aumento en la inflación global y el impacto de avances tecnológicos refuerza la relevancia temporal de esta investigación (World Bank, 2021). Estos eventos han desafiado los enfoques predictivos tradicionales, destacando la necesidad de modelos más robustos y adaptables. Analizar cómo los modelos supervisados manejan estos escenarios generará conocimiento aplicable no solo al contexto actual, sino también a futuros desafíos financieros.

Desde una perspectiva metodológica, el diseño experimental propuesto combina datos históricos de precios, variables macroeconómicas y algoritmos avanzados de aprendizaje automático. Según T (Hastie, *The Elements of Statistical Learning*, 2009), esta combinación es esencial para modelar relaciones complejas y no lineales. Además, la investigación establecerá un marco replicable y adaptable para futuros estudios en otros contextos financieros.

Esta investigación distingue entre el *tipo de aprendizaje* (supervisado) y la *forma funcional* de los modelos. Que un modelo sea supervisado significa que aprende con una variable objetivo; no implica que sea lineal o no lineal. Dentro del aprendizaje supervisado coexisten métodos lineales (p. ej., regresión OLS/Ridge/Lasso) y no lineales (p. ej., Random Forest, Gradient Boosting, XGBoost). Esta precisión enmarca la comparación empírica presentada más adelante, sin alterar la estructura del estudio.

Además, la justificación última de cualquier modelo predictivo en el ámbito financiero reside en su capacidad para generar valor económico. Una alta precisión estadística (como un bajo Error Cuadrático Medio) es un indicador intermedio, pero no el objetivo final. Por ello, este estudio trasciende la simple evaluación de métricas de error y busca validar la **significancia económica** de los modelos supervisados. Esto se logrará mediante la simulación de una estrategia de inversión (*backtesting*) basada en las predicciones de los modelos. Se calcularán indicadores clave de desempeño financiero, como el retorno acumulado y el **Ratio de Sharpe**, para determinar si los modelos pueden guiar decisiones de inversión que superen a estrategias pasivas de mercado, como "comprar y mantener" (*Buy and Hold*). Este paso es fundamental para demostrar empíricamente si el uso de estas herramientas avanzadas ofrece una ventaja tangible a los analistas financieros y gestores de portafolios en su objetivo primordial: generar retornos ajustados por riesgo.

Por último, esta investigación es viable gracias al acceso a datos históricos confiables y a herramientas computacionales avanzadas como scikit-learn y tensorflow, que permiten implementar técnicas de aprendizaje automático de manera eficiente. La disponibilidad de recursos tecnológicos garantiza un análisis riguroso y resultados aplicables al sector financiero.

1.5 Limites y alcances

1.5.1 Limites

Existe una limitada cantidad de estudios sobre la integración de modelos supervisados con múltiples variables macroeconómicas en mercados financieros volátiles, lo que puede restringir el contraste con investigaciones previas.

El estudio se centra en analizar los datos históricos de precios de acciones y variables macroeconómicas. El intervalo de tiempo que se analice puede no capturar comportamientos a largo plazo ni reflejar patrones de mercados en períodos anteriores o futuros.

Exclusión de factores geopolíticos y cualitativos: El alcance de la investigación se centra en variables cuantitativas de mercado y macroeconómicas. Se reconoce que factores cualitativos, como **eventos geopolíticos** (conflictos internacionales, disputas comerciales, cambios regulatorios abruptos), son determinantes cruciales en la dinámica de los precios de las acciones y pueden provocar cambios estructurales en los mercados. La integración y cuantificación de este tipo de eventos, por ejemplo, a través de análisis de sentimiento de noticias o índices de riesgo político, representa una limitación de este trabajo y constituye una valiosa línea de investigación futura.

1.5.2 Alcances

La investigación validará cómo los modelos supervisados pueden superar las limitaciones de los métodos tradicionales al predecir precios en mercados financieros volátiles.

El diseño de la investigación ofrecerá un marco que combina datos históricos, variables macroeconómicas y técnicas avanzadas de aprendizaje supervisado, lo que permitirá a futuros investigadores replicar o adaptar este enfoque en otros contextos financieros.

Los resultados van dirigidos a instituciones financieras, como bancos de inversión y gestores de portafolios, al proporcionar herramientas más precisas para optimizar decisiones estratégicas, gestionar riesgos y mejorar la estabilidad de los mercados.

Capítulo II: Marco teórico de referencia

2.1 Antecedentes de la predicción de precios en mercados financieros

2.1.1 *Evolución histórica de los métodos predictivos en finanzas*

La predicción de precios en mercados financieros ha experimentado una evolución significativa a lo largo del tiempo. Inicialmente dominada por análisis fundamentales y técnicos tradicionales, esta disciplina ha transitado hacia enfoques cada vez más sofisticados impulsados por avances tecnológicos y computacionales. Según Campbell et al. (2018) los primeros intentos sistemáticos de predicción se basaban principalmente en análisis estadísticos lineales, como regresiones y modelos ARIMA (Autoregressive Integrated Moving Average), que dominaron el panorama predictivo durante décadas.

En la década de 1970, la Hipótesis de Mercados Eficientes (HME) propuesta por Fama (1970) estableció un paradigma según el cual los precios de los activos reflejan toda la información disponible, haciendo teóricamente imposible "vencer al mercado" de manera consistente mediante análisis técnicos o fundamentales. Sin embargo, numerosas anomalías documentadas y la evidencia empírica posterior han cuestionado la validez estricta de esta hipótesis, especialmente en contextos de alta volatilidad como los mercados tecnológicos (Lo, 2004).

La crisis financiera de 2008 marcó un punto de inflexión, evidenciando las limitaciones de los modelos tradicionales para capturar comportamientos complejos y no lineales en los mercados. Esto aceleró la adopción de técnicas de aprendizaje automático, que habían comenzado a ganar tracción a principios del siglo XXI (De Prado, 2018). Empresas tecnológicas como Google y Amazon, caracterizadas por su alta capitalización de mercado y volatilidad, se convirtieron en objetivos particularmente desafiantes para la modelización predictiva.

2.1.2 *Estudios previos sobre predicción de precios de acciones tecnológicas*

La literatura especializada en predicción de precios de acciones tecnológicas ha evolucionado significativamente en los últimos años. Fischer y Krauss (2018) aplicaron redes neuronales recurrentes (LSTM) para predecir movimientos en el índice S&P 500, demostrando mejoras significativas frente a enfoques clásicos. Por su parte, Carta et al. (2021) implementaron

modelos basados en Random Forest para predecir específicamente el comportamiento de acciones tecnológicas, encontrando una capacidad superior para capturar patrones no evidentes en las series temporales financieras.

En el contexto específico de predicción de precios de acciones de empresas como Google y Amazon, Chen et al. (2019) evaluaron el desempeño de modelos basados en Gradient Boosting, demostrando su superioridad frente a métodos econométricos tradicionales, particularmente en periodos de alta volatilidad. Este estudio destacó la importancia de incorporar variables macroeconómicas como tasas de interés e índices de inflación para mejorar la precisión predictiva.

Más recientemente, López-Martín et al. (2022) compararon el rendimiento de varios algoritmos supervisados en la predicción de precios de acciones del sector tecnológico, concluyendo que Random Forest y Gradient Boosting ofrecían los mejores resultados en términos de precisión y capacidad de generalización. Este hallazgo respalda la elección metodológica del presente estudio y sugiere un prometedor camino de investigación.

2.2 Fundamentos teóricos de los mercados financieros

2.2.1 Teoría de mercados eficientes y sus limitaciones

La Teoría o Hipótesis de Mercados Eficientes (HME), formulada inicialmente por Fama (1970), constituye uno de los pilares teóricos fundamentales en el análisis de mercados financieros. Esta teoría postula que los precios de los activos financieros reflejan completa e instantáneamente toda la información disponible, haciendo imposible para cualquier inversor obtener rendimientos superiores al mercado de manera consistente sin asumir riesgos adicionales.

La HME distingue tres formas de eficiencia:

- **Eficiencia débil:** Los precios actuales reflejan toda la información contenida en los precios históricos.
- **Eficiencia semi-fuerte:** Los precios reflejan toda la información pública disponible.
- **Eficiencia fuerte:** Los precios reflejan absolutamente toda la información, incluso la privilegiada.

Sin embargo, diversos estudios empíricos han identificado anomalías que contradicen las predicciones de la HME. Shiller (2003) documentó patrones de exceso de volatilidad incompatibles con la racionalidad plena de los mercados, mientras que Grossman y Stiglitz (1980) señalaron la paradoja fundamental de que, si los mercados fueran perfectamente eficientes, no existiría incentivo para recopilar información, lo que eventualmente conduciría a la ineficiencia.

En el contexto específico de mercados tecnológicos como los que albergan a Google y Amazon, estas limitaciones se amplifican. Kumar (2009) señala que estos mercados son particularmente susceptibles a sesgos comportamentales debido a la complejidad de valorar activos con altos componentes intangibles y expectativas de crecimiento futuro. Esto genera oportunidades para que modelos predictivos avanzados, capaces de capturar patrones no lineales, puedan identificar ineficiencias temporales y generar valor predictivo.

2.2.2 Factores macroeconómicos en la valoración de activos financieros

La teoría económica moderna reconoce que los precios de los activos financieros están significativamente influenciados por variables macroeconómicas. El Modelo de Valoración de Activos Financieros (CAPM) desarrollado por Sharpe (1964) y Lintner (1965) estableció las bases para la comprensión de cómo el riesgo sistemático afecta los rendimientos esperados. Posteriormente, Chen et al. (1986) ampliaron este marco con la Teoría de Fijación de Precios por Arbitraje (APT), incorporando múltiples factores de riesgo, incluyendo variables macroeconómicas.

Las **tasas de interés** representan uno de los factores más influyentes en la valoración de activos. Como señalan Bernanke y Kuttner (2005), cambios en las tasas de interés afectan directamente el valor presente de los flujos de efectivo futuros e indirectamente el comportamiento inversor a través de cambios en el costo de capital. Para empresas tecnológicas como Google y Amazon, caracterizadas por expectativas de alto crecimiento futuro, la sensibilidad a las tasas de interés puede ser aún más pronunciada (Maio, 2013).

La **inflación** constituye otro factor macroeconómico crítico. Fama y Schwert (1977) documentaron la relación histórica negativa entre inflación y rendimientos bursátiles, mientras que estudios más recientes como el de Bekaert y Wang (2010) han matizado esta relación dependiendo

del horizonte temporal y las características específicas de los sectores. En el caso particular del sector tecnológico, Cohen et al. (2005) encontraron que estas empresas pueden exhibir tanto vulnerabilidad como resistencia a presiones inflacionarias, dependiendo de su capacidad para transferir costos y mantener márgenes.

Además, **indicadores de actividad económica** como el PIB, empleo y producción industrial influyen significativamente en las expectativas de crecimiento y, por ende, en los precios de las acciones. Flannery y Protopapadakis (2002) identificaron que sorpresas en estos indicadores generan reacciones significativas en los mercados bursátiles, particularmente en sectores cíclicos como el tecnológico.

Además de las tasas de interés y la inflación, un universo más amplio de variables predictoras candidatas, a menudo explorado en la literatura financiera, incluye indicadores de política monetaria (como la pendiente de la curva de tipos o *term spread*), condiciones financieras más amplias (índice MOVE), spreads de crédito (IG/HY), indicadores de actividad económica (PMIs, *surprise indices*), y variables de microestructura de mercado (como el flujo de órdenes o *order imbalance*). Aunque la inclusión de un "mar de variables" puede mejorar la precisión, esta investigación prioriza un conjunto parsimonioso para mantener la interpretabilidad del modelo y evitar el riesgo de sobreajuste, una preocupación central en la predicción financiera

Finalmente, es crucial reconocer que más allá de las variables económicas tradicionales, la **geopolítica** actúa como un factor macroeconómico de primer orden. Eventos como conflictos internacionales, sanciones comerciales, shocks regulatorios abruptos o interrupciones en las cadenas logísticas pueden detonar **cambios de régimen** repentinos en la volatilidad del mercado y en las primas de riesgo exigidas por los inversores. Si bien la cuantificación de estos eventos es compleja, su exclusión en los modelos puramente econométricos se reconoce como una **limitación** importante. Por ello, en enfoques robustos, se debe considerar la estabilidad de las señales predictivas a través de pruebas en subperíodos marcados por dichos eventos geopolíticos.

2.2.3 Características específicas del mercado de acciones tecnológicas

El mercado de acciones tecnológicas presenta particularidades que lo distinguen de otros sectores y justifican enfoques predictivos especializados. Pastor y Veronesi (2009) señalan que estas empresas se caracterizan por una elevada incertidumbre sobre su crecimiento futuro, lo que se traduce en mayor volatilidad y sensibilidad a nuevas informaciones.

La **volatilidad elevada** es una característica definitoria. Schwert (2002) documenta que las acciones tecnológicas típicamente exhiben desviaciones estándar de rendimientos significativamente superiores al promedio del mercado. En el caso específico de Google y Amazon, esta volatilidad se ha manifestado en periodos de rápidas apreciaciones seguidos por correcciones pronunciadas, como se observó durante 2021-2022 (Aramonte y Avalos, 2022).

La **sensibilidad a la innovación** representa otro factor distintivo. Empresas tecnológicas como Google y Amazon basan su valoración en gran medida en expectativas sobre futuras innovaciones y disrupciones de mercado. Pástor y Veronesi (2006) desarrollaron un modelo teórico que explica cómo la incertidumbre sobre el potencial de nuevas tecnologías puede generar valoraciones elevadas y posteriormente correcciones a medida que esta incertidumbre se resuelve. Este fenómeno se manifestó claramente durante la burbuja tecnológica de finales de los 90 y nuevamente en ciclos posteriores.

Finalmente, la **estructura competitiva** de los mercados digitales influye significativamente en la dinámica de precios. Economías de red, efectos de escala y ventajas del primer movedor crean entornos donde "el ganador se lleva todo" (winner-takes-all), lo que puede amplificar reacciones de mercado ante cambios en cuotas de mercado o amenazas competitivas (Evans y Schmalensee, 2016). Este aspecto es particularmente relevante para Google y Amazon, cuyo predominio en sus respectivos mercados constituye un componente fundamental de su valoración bursátil.

2.2.4 Fundamentos de rendimiento y riesgo en la predicción de activos

2.2.4.1 Rendimiento

Según Van Horne y Wachowicz (2002) el rendimiento son los “ingresos que se reciben por una inversión, sumados a las variaciones en el precio de mercado; los cuales por lo general se expresan como porcentaje del precio inicial de mercado de la inversión” (p. 94). La lógica de los inversores es poder obtener luego de un determinado periodo de tiempo de haber realizado una inversión en cualquier instrumento financiero, una utilidad o ganancia sobre la misma. De ahí que se dice que la variación de los precios es imprescindible y por tanto el supremo deseo de acertar al valor del precio forward, pues una variación positiva podrá generar una mayor eficiencia en el portafolio del inversor (Generando retornos positivos) y al tener un valor más que acertado de los valores volátiles de los precios, podemos tomar una decisión en cuánto a colocarnos en una posición corta o larga de nuestros activos (Para este estudio, al de las acciones de Google y Amazon).

El inversionista bajo su lógica el querrá siempre recibir una ganancia al realizar una inversión de cualquier tipo, sin embargo, no siempre se puede garantizar que este la obtendrá. En otras palabras, existe la opción también en la cual su inversión no obtenga los resultados esperados y en vez de obtener una ganancia se incurra en una pérdida o déficit. Esto depende del riesgo que asociado al instrumento financiero.

Lo expuesto anteriormente debe incorporar factores externos —conocidos como *variables exógenas*, las cuales no están bajo el control directo del inversionista— que pueden alterar significativamente el diseño estratégico de inversión. Incluso cuando se dispone de herramientas sofisticadas, como las que se proponen en este estudio (por ejemplo, el uso de Python), la elevada incertidumbre inherente al mercado bursátil puede comprometer la eficacia de dichas estrategias.

En este contexto, el propósito central de la presente investigación ha sido desarrollar una herramienta complementaria de análisis que potencie la capacidad predictiva del inversionista. Para ello, se propone una triada metodológica compuesta por:

- **Crystal Ball**, como representante de los modelos estocásticos;
- **Microsoft Excel**, para el tratamiento de modelos determinísticos;

- **Python**, como entorno de programación flexible y escalable.

Esta combinación busca mejorar sustancialmente la interpretación de escenarios futuros del mercado, así como el comportamiento proyectado de las acciones incluidas en el portafolio de inversión.

Los rendimientos pueden expresarse de forma porcentual, según Ross y Westerfield (2012) “es más conveniente resumir la información acerca de rendimientos en términos porcentuales que en dólares porque los porcentajes se aplican a cualquier monto que se invierta” (p. 302). La fórmula que estos autores sugieren para el cálculo del rendimiento porcentual del ingreso, algunas veces denominado rendimiento del dividendo, es la siguiente:

$$\text{Rendimiento del dividendo} = D_{t+1} / P_t$$

Donde:

t es el año que se está examinando

P_t es el precio de la acción al inicio del año

D_{t+1} es el dividendo pagado sobre la acción durante el año

La ganancia o pérdida de capital de capital es otro elemento que se debe considerar también cuando se habla de rendimientos, según Ross y Westerfield (2012) “La ganancia de capital (o la pérdida) es el cambio de precio de la acción dividido entre el precio inicial” (p. 303). La fórmula para este cálculo que sugieren es la siguiente:

$$\text{Ganancia de capital} = (P_{t+1} - P_t) / P_t$$

Al sumar estos dos elementos se obtiene el rendimiento total, “Al combinar estos dos resultados se determina que el *rendimiento total* sobre la inversión en acciones” (Ross y Westerfield, 2012, p. 303).

En el ámbito de las acciones de Google y Amazon podría considerarse como indicador para medir la utilidad generada la ganancia de capital, por los rendimientos generados dado un nivel de volatilidad positiva de los precios de las acciones y por supuesto, a la compra y venta de acciones de estas y a los dividendos futuros según las políticas de dividendos de cada una de ellas.

2.2.4.2 Riesgo

Según Van Horne y Wachowicz (2002) “se define el riesgo como la variabilidad de los rendimientos en relación con los que se espera recibir, los bonos de la tesorería serían títulos sin riesgo, mientras que las acciones ordinarias serían títulos riesgosos. Se dice que, cuanto mayor sea la variabilidad, más riesgosos serán los títulos” (p. 65). Las criptomonedas se consideran instrumentos muy riesgosos debido a los cambios abruptos que su precio puede tener en cortos periodos de tiempo.

Otro elemento que se utiliza para estudiar las diferentes opciones en las que los inversionistas pudiesen invertir es el grado de riesgo que existe entre diferentes instrumentos financieros, y este parámetro debe considerarse antes de tomar la decisión sobre cuales instrumentos financieros se debería invertir.

Para estudiar el riesgo se suelen utilizar dos medidas estadísticas como lo son la varianza y la desviación estándar. “La varianza y su raíz cuadrada, la desviación estándar, son las dos medidas más comunes de la variabilidad o dispersión. Se utilizarán los símbolos Var y σ^2 para denotar la varianza, y SD y σ para representar la desviación estándar. Desde luego, σ es la letra griega sigma” (Ross y Westerfield, 2012, p. 303).

Su fórmula es:

$$\text{Var} = \frac{1}{T - 1} [(R_1 - \bar{R})^2 + (R_2 - \bar{R})^2 + (R_3 - \bar{R})^2] \dots$$

Donde:

T es el número de rendimientos individuales

R es el rendimiento individual (pueden ser 1, 2, 3 o más)

\bar{R} es el rendimiento promedio

Para el cálculo de la desviación estándar, se obtiene la raíz cuadrada de la varianza. Y la medida estadística que se recomienda en este estudio es utilizar la desviación estándar para medir la dispersión (variabilidad) de una serie de datos.

$$\text{SD}(R) = \sqrt{\text{Var}(R)}$$

Estos instrumentos estadísticos sirven para medir el riesgo inherente que existe en un portafolio de inversión; para el caso de nuestro estudio, hemos tomado a consideración la capacidad predictiva en Python de los precios de las acciones de Google y Amazon y que a su vez puede ser utilizado para todas las acciones e instrumentos financieros que existen en el mercado bursátil.

Es crucial destacar que, si bien la desviación estándar es una medida fundamental de la volatilidad histórica, es menos eficaz para capturar el **riesgo de cola (tail risk)**. Eventos exógenos de baja probabilidad, pero alto impacto, como crisis financieras, **pandemias (ej., COVID-19) o el estallido de guerras**, generan movimientos de precios extremos que exceden con creces las predicciones de una distribución normal. Estos shocks pueden **mermar drásticamente los rendimientos** y alterar temporalmente las correlaciones establecidas entre variables macroeconómicas y los precios de las acciones. La gestión de estos eventos impredecibles es uno de los mayores desafíos en la administración de portafolios y subraya la necesidad de complementar los modelos cuantitativos con análisis de escenarios y estrategias de cobertura robustas.

2.3 Fundamentos de aprendizaje automático en finanzas

2.3.1 Principios del aprendizaje supervisado

El aprendizaje supervisado constituye una rama fundamental del aprendizaje automático con amplia aplicación en el análisis financiero. Como define Hastie et al. (2009), este enfoque se basa en el entrenamiento de modelos utilizando conjuntos de datos etiquetados, donde para cada instancia de entrada se conoce la salida deseada. Es crucial destacar que la categoría de "supervisado" engloba un amplio espectro de técnicas. Por un lado, existen modelos lineales como la Regresión Lineal, que asumen relaciones aditivas entre los predictores y la variable objetivo. Por otro lado, una vasta y poderosa categoría la componen los modelos no lineales, como los Árboles de Decisión, Random Forest, Gradient Boosting y las Redes Neuronales, los cuales son capaces de capturar interacciones complejas y patrones no lineales. Esta capacidad los hace particularmente relevantes para aplicaciones financieras, dada la naturaleza dinámica y a menudo no lineal de los mercados. En el contexto de la predicción de precios de acciones, la entrada

típicamente consiste en variables predictoras (datos históricos, indicadores técnicos, variables macroeconómicas), mientras que la salida corresponde al precio o rendimiento futuro.

El proceso de aprendizaje supervisado comprende varias fases fundamentales:

- 1. Recopilación y preprocesamiento de datos:** Involucra la obtención de datos históricos de precios, variables macroeconómicas y otras características relevantes. Esta fase incluye la limpieza de datos (manejo de valores faltantes y atípicos), normalización y transformación de variables.
- 2. Selección y extracción de características:** Implica identificar las variables más relevantes para la predicción. En finanzas, este proceso puede incluir la creación de indicadores técnicos derivados (medias móviles, osciladores), así como la incorporación de variables fundamentales y macroeconómicas (De Prado, 2018).
- 3. División de datos:** Los datos se separan típicamente en conjuntos de entrenamiento, validación y prueba. El conjunto de entrenamiento se utiliza para ajustar los parámetros del modelo, el conjunto de validación para la optimización de hiperparámetros, y el conjunto de prueba para evaluar el rendimiento final.
- 4. Entrenamiento del modelo:** Durante esta fase, el algoritmo ajusta sus parámetros internos para minimizar una función de pérdida que cuantifica la discrepancia entre predicciones y valores reales. En predicción de series temporales financieras, funciones como el error cuadrático medio (MSE) o el error absoluto medio (MAE) son comúnmente utilizadas.
- 5. Evaluación y optimización:** El rendimiento del modelo se evalúa utilizando métricas apropiadas, y los hiperparámetros se ajustan mediante técnicas como validación cruzada o búsqueda en cuadrícula para maximizar el rendimiento predictivo.
- 6. Interpretación y aplicación:** Finalmente, el modelo entrenado se aplica para realizar predicciones sobre nuevos datos, y sus resultados se interpretan para informar decisiones de inversión.

En finanzas, la naturaleza secuencial y no estacionaria de los datos presenta desafíos específicos para el aprendizaje supervisado. Dixon et al. (2020) señalan la importancia de técnicas

como la validación cruzada con bloqueo temporal (time-blocked cross-validation) para evaluar modelos de manera realista, evitando el sesgo de mirada hacia adelante (look-ahead bias) que puede artificialmente inflar el rendimiento predictivo.

2.3.2 Algoritmos de Random Forest: funcionamiento y aplicaciones financieras

Random Forest es un algoritmo de ensamble desarrollado por Breiman (2001) que ha ganado amplia adopción en aplicaciones financieras. Este método combina múltiples árboles de decisión entrenados sobre diferentes submuestras de datos y características, aplicando técnicas de bagging (Bootstrap Aggregating) para mejorar la estabilidad y precisión de las predicciones mientras reduce el sobreajuste.

El funcionamiento del algoritmo Random Forest incluye los siguientes pasos fundamentales:

- 1. Generación de muestras bootstrap:** A partir del conjunto de entrenamiento original, se crean múltiples muestras mediante muestreo con reemplazo.
- 2. Construcción de árboles:** Para cada muestra bootstrap, se construye un árbol de decisión. Durante este proceso, en cada nodo se selecciona aleatoriamente un subconjunto de características candidatas para la división, lo que introduce diversidad entre los árboles.
- 3. Agregación de predicciones:** Las predicciones individuales de cada árbol se combinan mediante votación (para clasificación) o promedio (para regresión), lo que resulta en la predicción final del ensamble.

En el contexto financiero, Random Forest ofrece ventajas significativas que lo hacen particularmente adecuado para la predicción de precios de acciones. Ballings et al. (2015) demostraron su capacidad superior para modelar relaciones no lineales complejas en datos financieros en comparación con métodos tradicionales. Krauss et al. (2017) aplicaron Random Forest para predecir rendimientos del S&P 500, logrando rentabilidades anormales estadísticamente significativas incluso después de considerar costos de transacción.

Para acciones tecnológicas como Google y Amazon, caracterizadas por alta volatilidad y sensibilidad a múltiples factores, las ventajas de Random Forest resultan especialmente relevantes.

Carta et al. (2021) encontraron que este algoritmo logra capturar eficientemente interacciones complejas entre indicadores técnicos y variables macroeconómicas, generando predicciones más robustas en entornos de mercado cambiantes.

Además, la capacidad de Random Forest para cuantificar la importancia de variables permite identificar los factores más influyentes en la dinámica de precios. Fischer et al. (2018) utilizaron esta característica para determinar que, en el caso de acciones tecnológicas, indicadores como volatilidad implícita, volumen de transacciones y sorpresas en resultados trimestrales emergían como los predictores más significativos, por encima de indicadores técnicos tradicionales.

Algoritmos de Random Forest: funcionamiento y aplicaciones financieras

Definición y fundamentos matemáticos:

El algoritmo **Random Forest** es un modelo de ensamble basado en árboles de decisión que combina múltiples modelos débiles (árboles individuales) para construir un modelo robusto. Fue introducido por Leo Breiman en 2001 y se basa en dos principios clave:

Bagging (Bootstrap Aggregating): Se generan múltiples conjuntos de datos mediante muestreo aleatorio con reemplazo del conjunto original. Cada subconjunto se utiliza para entrenar un árbol de decisión independiente.

Aleatoriedad en la selección de características: Durante la construcción de cada nodo de un árbol, solo se considera un subconjunto aleatorio de características para dividir los datos. Esto reduce la correlación entre los árboles y mejora la generalización.

El proceso de predicción en Random Forest sigue la siguiente ecuación:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_t(x)$$

Donde:

T es el número total de árboles en el bosque.

$f_t(x)$ es la predicción del árbol t para una entrada x .

\hat{y} es la predicción final promediada.

Matemática detrás del bagging:

El muestreo bootstrap implica seleccionar n muestras con reemplazo del conjunto de datos original de tamaño N . La probabilidad de que una observación no sea seleccionada en una muestra bootstrap es aproximadamente:

$$P(\text{no seleccionada}) = \left(1 - \frac{1}{N}\right)^N \approx e^{-1} \approx 0.368$$

Esto significa que aproximadamente el 36.8% de las observaciones no se utilizan en el entrenamiento de un árbol específico, formando el conjunto "out-of-bag" (OOB), que puede utilizarse para validación.

Ventajas matemáticas:

Reducción de varianza: Al promediar las predicciones de múltiples árboles, Random Forest minimiza el riesgo de sobreajuste. La varianza del modelo se reduce debido a la decorrelación inducida por la aleatorización en la selección de características.

Importancia de variables: Random Forest proporciona una medida de importancia de variables basada en la reducción media de impureza (Gini o entropía) en los nodos donde se utiliza una característica específica:

$$\text{Importancia}(x_j) = \frac{1}{T} \sum_{t=1}^T \Delta I(x_j, t)$$

Donde $\Delta I(x_j, t)$ es la reducción de impureza atribuida a la variable x_j en el árbol t .

Aplicaciones en finanzas:

Random Forest ha demostrado ser particularmente útil en la predicción de precios de acciones debido a su capacidad para capturar relaciones no lineales y manejar datos heterogéneos. Por ejemplo:

Identificación de patrones ocultos en datos históricos de precios.

Detección de anomalías en mercados financieros.

Clasificación de activos en categorías de riesgo.

2.3.3 Algoritmos de Gradient Boosting: fundamentos y aplicación en predicción de precios

Gradient Boosting representa otra familia de algoritmos de ensamble que ha demostrado excelente desempeño en aplicaciones financieras. A diferencia de Random Forest, que construye árboles independientes en paralelo, Gradient Boosting los construye secuencialmente, donde cada nuevo árbol se entrena para corregir los errores de los árboles anteriores.

El marco teórico de Gradient Boosting fue formalizado por Friedman (2001) y ha evolucionado con implementaciones optimizadas como XGBoost (Chen y Guestrin, 2016) y LightGBM (Ke et al., 2017). El proceso fundamental incluye:

- 1. Inicialización:** El algoritmo comienza con un modelo simple, típicamente un árbol de decisión de poca profundidad.
- 2. Cálculo de residuos:** Para cada observación, se calcula la diferencia entre el valor real y el predicho por el modelo actual.
- 3. Entrenamiento incremental:** Se entrena un nuevo modelo (árbol) para predecir estos residuos.
- 4. Actualización del modelo:** El nuevo árbol se añade al modelo existente, típicamente con un coeficiente de aprendizaje (learning rate) que controla la contribución de cada nuevo árbol.

- 5. Iteración:** Los pasos 2-4 se repiten hasta que se cumple un criterio de parada (número máximo de árboles o convergencia).

En aplicaciones de predicción de precios de acciones, Gradient Boosting ha demostrado capacidad superior para capturar patrones temporales complejos. Zhang y Yan (2020) aplicaron XGBoost para predecir tendencias de precios en mercados asiáticos, logrando tasas de acierto significativamente superiores a modelos econométricos y redes neuronales. Más específicamente, para el caso de empresas tecnológicas, Chen et al. (2019) encontraron que algoritmos basados en Gradient Boosting exhibían mayor robustez ante eventos disruptivos de mercado, captando eficientemente cambios de régimen que modelos lineales no lograban identificar.

Una ventaja destacable de Gradient Boosting en el contexto financiero es su capacidad para manejar eficientemente datos de naturaleza heterogénea. Esto resulta particularmente relevante al combinar variables de diferentes dominios, como indicadores técnicos, métricas fundamentales y variables macroeconómicas. Carmona et al. (2021) demostraron cómo esta capacidad permitía incorporar eficientemente datos no estructurados derivados de análisis de sentimiento y noticias, mejorando la precisión predictiva para acciones de alta volatilidad como las tecnológicas.

Adicionalmente, implementaciones modernas como XGBoost ofrecen mecanismos integrados para manejar datos desbalanceados y outliers, características frecuentes en series temporales financieras, especialmente durante períodos de crisis o eventos disruptivos como los experimentados durante el período 2021-2024 en estudio.

Definición y fundamentos matemáticos:

Gradient Boosting es otro modelo de ensamble que construye árboles de decisión de manera secuencial, corrigiendo iterativamente los errores cometidos por los árboles anteriores. Fue desarrollado por Jerome Friedman en 2001 y se basa en el principio de optimización iterativa.

El proceso de Gradient Boosting sigue estos pasos:

Inicialización: Se define una función inicial $F_0(x)$, generalmente como el promedio de las etiquetas del conjunto de datos:

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma)$$

Donde L es la función de pérdida (e.g., error cuadrático medio).

Iteraciones: Para cada iteración m :

Se calculan los residuos (errores) del modelo actual: $r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$

Se entrena un nuevo árbol $h_m(x)$ para predecir los residuos: $h_m(x) = \arg \min_h \sum_{i=1}^n (r_{im} - h(x_i))^2$

Se actualiza el modelo agregando el nuevo árbol ponderado por un factor de aprendizaje ν : $F_m(x) = F_{m-1}(x) + \nu h_m(x)$

Finalización: El modelo final es la suma de todos los árboles:

$$F_M(x) = \sum_{m=1}^M \nu h_m(x)$$

Matemática detrás de la optimización:

La función de pérdida L típicamente se define como:

Error cuadrático medio (MSE): $L(y, \hat{y}) = \frac{1}{2} (y - \hat{y})^2$

Error absoluto medio (MAE): $L(y, \hat{y}) = |y - \hat{y}|$

El gradiente descendente funcional busca minimizar L ajustando iterativamente $F_m(x)$. La derivada parcial de L con respecto a $F(x)$ guía la construcción de cada nuevo árbol.

Ventajas matemáticas:

Alta precisión: Gradient Boosting suele superar a otros modelos en términos de precisión debido a su enfoque iterativo.

Capacidad de modelar relaciones complejas: Es especialmente efectivo en problemas donde las relaciones entre variables son altamente no lineales.

Flexibilidad: Puede utilizarse tanto para problemas de regresión como de clasificación.

Aplicaciones en finanzas:

Gradient Boosting ha sido ampliamente utilizado en la predicción de precios de acciones debido a su capacidad para manejar datos volátiles y complejos. Algunas aplicaciones incluyen:

Predicción de retornos diarios utilizando variables técnicas y macroeconómicas.

Identificación de puntos de inflexión en tendencias de mercado.

Evaluación de riesgos crediticios y operativos en instituciones financieras.

2.4 Variables macroeconómicas en la predicción de precios de acciones

2.4.1 Tasas de Interés: mecanismos de influencia en mercados bursátiles

Las tasas de interés constituyen uno de los factores macroeconómicos más influyentes en la valoración de activos financieros, particularmente en acciones de empresas tecnológicas como Google y Amazon. Su impacto se manifiesta a través de múltiples mecanismos interrelacionados.

El **efecto de descuento** representa el mecanismo más directo. Como establecen los modelos fundamentales de valoración, el precio de una acción equivale al valor presente de sus flujos de caja futuros. Las tasas de interés determinan la tasa de descuento aplicada a estos flujos; por tanto, un incremento en las tasas reduce el valor presente de los beneficios futuros. Damodaran (2020) señala que este efecto es particularmente pronunciado para empresas tecnológicas con altas expectativas de crecimiento, donde una proporción significativa del valor deriva de flujos de caja distantes en el tiempo.

La **política monetaria** implementada por bancos centrales influye en la dinámica bursátil a través del manejo de tasas de referencia. Bernanke y Kuttner (2005) documentaron que un recorte no anticipado de 25 puntos básicos en la tasa de fondos federales típicamente generaba un incremento de aproximadamente 1% en índices bursátiles amplios. Esta sensibilidad se ha

intensificado en entornos de tasas cercanas a cero, como los experimentados en períodos recientes. Durante 2021-2024, el ciclo de normalización monetaria implementado por la Reserva Federal para contener presiones inflacionarias generó volatilidad significativa en mercados tecnológicos, como documentan Aramonte y Avalos (2022).

El **costo de capital** representa otro canal de transmisión. Tasas más elevadas incrementan el costo de financiamiento tanto para deuda como para capital. Para empresas tecnológicas en fases de expansión, que frecuentemente requieren financiamiento continuo para investigación, desarrollo y crecimiento, este efecto puede impactar significativamente márgenes operativos y valoraciones. Cohen et al. (2005) demostraron empíricamente cómo cambios en tasas corporativas afectaban diferencialmente a sectores basados en su intensidad de capital e investigación.

Finalmente, las tasas influyen en los **flujos de inversión** entre clases de activos. Incrementos en rendimientos de activos de renta fija tienden a provocar reajustes de portafolios desde acciones hacia bonos, generando presión vendedora. Este efecto de "rotación sectorial" frecuentemente impacta desproporcionadamente a sectores percibidos como más especulativos o valorados primariamente por crecimiento futuro, característicos del segmento tecnológico (Maio, 2013).

2.4.2 Inflación: Efectos en la valoración de empresas tecnológicas

La inflación constituye otra variable macroeconómica fundamental cuya dinámica afecta significativamente la valoración bursátil, particularmente en el sector tecnológico. Su influencia se manifiesta a través de diversos canales que interactúan de manera compleja.

El **poder adquisitivo** representa un primer mecanismo de transmisión. Períodos inflacionarios erosionan el poder adquisitivo de los consumidores, potencialmente reduciendo demanda de productos y servicios. Sin embargo, para empresas tecnológicas como Google y Amazon, este efecto puede ser heterogéneo. Goolsbee y Klenow (2018) documentan que servicios digitales y plataformas pueden exhibir resistencia ante presiones inflacionarias convencionales debido a estructuras de costos distintivas y capacidad de escalamiento.

La **política anti-inflacionaria** implementada por autoridades monetarias constituye frecuentemente un factor más determinante que la inflación misma. Entornos de alta inflación

típicamente desencadenan respuestas de política monetaria restrictiva, elevando tasas de interés, como se analizó previamente. Durante 2021-2024, la emergencia de presiones inflacionarias post-pandemia provocó ciclos de ajuste monetario que impactaron significativamente valoraciones tecnológicas (Schnabel, 2021).

Los **márgenes operativos** se ven directamente afectados por dinámicas inflacionarias. Empresas con poder de fijación de precios pueden transferir incrementos de costos a consumidores, preservando márgenes. Cohen et al. (2005) encontraron que empresas tecnológicas exhiben comportamientos heterogéneos en este aspecto: mientras plataformas con posiciones dominantes como Google y Amazon frecuentemente mantienen capacidad de preservar márgenes, segmentos más competitivos experimentan compresión significativa durante episodios inflacionarios.

La **contabilidad inflacionaria** genera distorsiones adicionales en estados financieros y métricas de valoración. Modigliani y Cohn (1979) documentaron el fenómeno de "ilusión inflacionaria" donde inversores cometen errores sistemáticos al valorar empresas durante períodos de inflación cambiante, típicamente subvalorando acciones en entornos de alta inflación. Ritter y Warr (2002) encontraron evidencia de este fenómeno persistiendo en mercados modernos, afectando particularmente empresas con estructuras de activos intangibles predominantes, características del sector tecnológico.

2.4.3 Datos Históricos de precios e indicadores técnicos como predictores

Los datos históricos de precios e indicadores técnicos derivados constituyen inputs fundamentales para modelos predictivos de precios de acciones, complementando el análisis de variables macroeconómicas previamente discutido. Su relevancia teórica se fundamenta en diversas corrientes de pensamiento financiero.

El **análisis técnico** tradicional, inicialmente sistematizado por Edwards y Magee (1948), postula que los movimientos históricos de precios y volúmenes contienen información predictiva sobre movimientos futuros. Esta perspectiva contradice la forma débil de eficiencia de mercado, generando continuo debate académico. Lo y MacKinlay (1988) documentaron evidencia

estadística de autocorrelación positiva en rendimientos semanales, contradiciendo la hipótesis de caminata aleatoria y sugiriendo potencial predictivo en series históricas.

Desde una perspectiva más contemporánea, los **indicadores técnicos** derivados de series históricas capturan aspectos específicos de la dinámica de mercados que resultan informativos para predicciones futuras. Estos incluyen medidas de tendencia (medias móviles), momentum (índice de fuerza relativa, oscilador estocástico), y volatilidad (bandas de Bollinger). Kim y Shamsuddin (2015) demostraron que la utilidad predictiva de estos indicadores varía significativamente según mercados y períodos, siendo generalmente mayor durante períodos de estrés de mercado caracterizados por comportamientos no aleatorios.

En el contexto específico de mercados tecnológicos, la **microestructura de mercado** adquiere particular relevancia. Hendershott et al. (2011) documentaron cómo patrones de negociación de alta frecuencia, desequilibrios en libros de órdenes y otros aspectos microestructurales, frecuentemente inferibles de datos históricos granulares, contienen información predictiva significativa sobre movimientos de precios a corto plazo.

Para empresas como Google y Amazon, caracterizadas por alta concentración de propiedad institucional y seguimiento analítico, **indicadores de flujo** derivados de volúmenes históricos pueden capturar dinámicas informativas. Ni et al. (2020) encontraron que medidas como flujos de órdenes institucionales y volúmenes anormales precedían movimientos significativos de precios, particularmente en torno a eventos informativos como anuncios de resultados trimestrales.

La integración de estos datos históricos con variables macroeconómicas en modelos supervisados constituye un enfoque particularmente prometedor. Como destacan Gu et al. (2020), algoritmos como Random Forest y Gradient Boosting poseen capacidad inherente para identificar interacciones complejas y no lineales entre variables predictoras, permitiendo capturar, por ejemplo, cómo la sensibilidad a indicadores técnicos varía condicionalmente según diferentes regímenes macroeconómicos.

Finalmente, es fundamental destacar que, si bien dentro de los indicadores de mercado se utiliza el Índice de Volatilidad CBOE (VIX), su rol como predictor universal es un tema de debate en la literatura financiera contemporánea. Los estudios reportan resultados **heterogéneos** respecto

a su poder predictivo, sugiriendo que su eficacia es a menudo **estado-dependiente**, es decir, sensible al régimen de mercado, al horizonte de predicción y a la especificación del modelo. Investigaciones recientes, como las de **Wang et al. (2022)**, exploran la **comovilidad multiescala** entre el VIX general del S&P 500 y los VIX "individuales" de acciones de gran capitalización como Amazon y Google, encontrando relaciones complejas que varían con el tiempo. Asimismo, **Qadan (2024)** analiza la información contenida en la **estructura temporal (term structure)** del VIX para acciones individuales. Esta evidencia sugiere que la señal del VIX no es un predictor simple y estable, sino una variable compleja. Por lo tanto, en el contexto de esta tesis, el VIX se incluye no como un predictor infalible, sino como una importante **variable de robustez** que ayuda al modelo a contextualizar el estado de volatilidad del mercado.

2.5 Comparación entre modelos predictivos tradicionales y supervisados

2.5.1 Modelos econométricos tradicionales: ARIMA, GARCH y sus Limitaciones

Los modelos econométricos tradicionales han constituido durante décadas el núcleo metodológico para la predicción de series temporales financieras. Entre estos, los modelos ARIMA (Autoregressive Integrated Moving Average) desarrollados por Box y Jenkins (1976) y los modelos GARCH (Generalized Autoregressive Conditional Heteroskedasticity) introducidos por Engle (1982) y Bollerslev (1986) destacan por su fundamentación teórica y amplia aplicación.

Los modelos **ARIMA** combinan componentes autorregresivos (AR), que modelan la dependencia de valores actuales respecto a valores pasados, y componentes de media móvil (MA), que capturan la estructura de los términos de error. Incorporan además diferenciación para manejar series no estacionarias. Su especificación general $ARIMA(p,d,q)$ donde p representa el orden autorregresivo, d el grado de diferenciación, y q el orden de media móvil, permite modelar diversas estructuras temporales.

Por su parte, los modelos **GARCH** abordan específicamente la heterocedasticidad condicional característica de series financieras, modelando la varianza como función de errores pasados y varianzas pasadas. Extensiones como EGARCH (Nelson, 1991) y GJR-GARCH Glosten et al. (1993) incorporan asimetrías en respuestas a shocks positivos y negativos, fenómeno ampliamente documentado en mercados bursátiles.

En la modelización de la volatilidad financiera, la familia de modelos GARCH (Heterocedasticidad Condicional Autoregresiva Generalizada) se ha consolidado como el estándar de la industria y la academia. Su punto de partida es el modelo GARCH(p,q) de Bollerslev (1986), que fue diseñado para capturar una característica bien conocida de los mercados: el **agrupamiento de la volatilidad** (*volatility clustering*). En esencia, este fenómeno describe cómo los períodos de alta agitación tienden a agruparse, al igual que los períodos de calma.

Con el tiempo, la literatura académica fue expandiendo este marco para capturar otras sutilezas. Por ejemplo, modelos como el GJR-GARCH (Glosten, Jagannathan y Runkle, 1993) y el TGARCH se desarrollaron para modelar el **efecto de apalancamiento** (*leverage effect*). Este efecto asimétrico, ampliamente documentado, se refiere a cómo las malas noticias suelen generar más volatilidad que las buenas noticias de la misma magnitud. Otras variantes importantes incluyen el EGARCH (Nelson, 1991), que también aborda la asimetría, y el FIGARCH (Baillie, Bollerslev y Mikkelsen, 1996), diseñado para capturar la **memoria larga** en la volatilidad, un fenómeno donde los shocks tardan mucho más en disiparse de lo que sugiere el modelo GARCH original.

Considerando la robustez de estos modelos, surge una pregunta metodológica clave: **¿por qué se va a utilizar ARIMA como el modelo de comparación tradicional en este estudio y no GARCH?** Nuestra decisión fue intencionada y se basa en la naturaleza de nuestro objetivo de investigación:

1. **Diferencia de objetivos: predecir el precio vs. predecir la volatilidad.** Nuestro estudio se centra en predecir el **precio futuro**, es decir, el **retorno esperado** (el primer momento de la serie). Los modelos ARIMA son la herramienta econométrica clásica para esta tarea. En cambio, los modelos GARCH se especializan en predecir la **volatilidad futura** (el segundo momento). Por lo tanto, comparar nuestros modelos de aprendizaje automático contra un GARCH sería como comparar herramientas diseñadas para trabajos distintos.
2. **Selección de un benchmark comparable.** Configuramos Random Forest y Gradient Boosting para una tarea de regresión de precios. En este contexto, ARIMA es el análogo econométrico más directo y justo para establecer una línea base de rendimiento.

3. **Validación en un entorno desafiante.** Precisamente porque se sabe que los modelos ARIMA tienen dificultades en mercados volátiles, su uso como benchmark es ideal para nuestro propósito. Si nuestros modelos de aprendizaje automático logran superar a ARIMA en estas condiciones adversas, la evidencia de su superioridad es mucho más contundente. El fracaso esperado de ARIMA no debilita la comparación; al contrario, magnifica la capacidad de adaptación de los modelos supervisados que proponemos.

En definitiva, aunque un enfoque combinado ARIMA-GARCH es una herramienta poderosa para un análisis financiero completo, para el objetivo específico de esta tesis —validar si los modelos de aprendizaje automático superan a los tradicionales en la **predicción de precios**—, el modelo ARIMA representa un benchmark más limpio, relevante y metodológicamente sólido.

Sin embargo, estos enfoques presentan limitaciones significativas para la predicción de precios de acciones, particularmente en mercados volátiles como los tecnológicos:

Asunción de linealidad: Como señala Brooks (2019), estos modelos asumen relaciones lineales entre variables, limitando su capacidad para capturar interacciones complejas características de mercados financieros. La evidencia empírica documentada por Hsieh (1991) y posteriormente confirmada por numerosos estudios sugiere presencia significativa de no linealidades en series financieras.

Capacidad limitada para incorporar variables exógenas: Mientras extensiones como ARIMAX y modelos multivariados permiten cierta inclusión de predictores externos, su formulación paramétrica restringe el número y tipo de variables incorporables. Particularmente, la integración de factores macroeconómicos como tasas de interés e inflación resulta frecuentemente problemática debido a diferentes frecuencias y relaciones variables en el tiempo.

Selección manual de especificación: La identificación de órdenes apropiados (p,d,q) típicamente requiere análisis manual de funciones de autocorrelación y criterios de información, incrementando subjetividad y complejidad, especialmente para implementaciones a gran escala o automatizadas.

Sensibilidad a cambios estructurales: Períodos como 2021-2024, caracterizados por transiciones post-pandémicas, cambios en regímenes monetarios y crisis sectoriales, representan

desafíos particulares para estos modelos. Como documentan Pesaran y Timmermann (2004), los parámetros óptimos estimados frecuentemente experimentan inestabilidad ante cambios estructurales, reduciendo significativamente precisión predictiva fuera de muestra.

Por estas razones, aunque estos modelos continúan ofreciendo marcos interpretativos valiosos y benchmarks comparativos, sus limitaciones han motivado la búsqueda de enfoques alternativos con mayor flexibilidad, como los modelos supervisados.

2.5.2 Ventajas comparativas de los modelos supervisados en contextos financieros

La adaptabilidad a relaciones no lineales constituye una ventaja primordial. A diferencia de modelos paramétricos con formas funcionales predefinidas, algoritmos como Random Forest y Gradient Boosting construyen aproximaciones no paramétricas capaces de capturar interacciones complejas. Dixon et al. (2020) demostraron empíricamente cómo esta característica permitía identificar patrones predictivos en datos financieros que modelos lineales no detectaban, incluyendo efectos umbral y regímenes cambiantes frecuentes en mercados tecnológicos.

La capacidad para manejar grandes conjuntos de variables predictoras representa otra ventaja significativa. Mientras modelos econométricos tradicionales sufren problemas de sobreajuste y multicolinealidad al incorporar numerosos predictores, los algoritmos supervisados modernos implementan mecanismos integrados de regularización y selección de características. Gu et al. (2020) evidenciaron cómo esta capacidad permitía incorporar eficientemente cientos de predictores potenciales, incluyendo indicadores técnicos, datos fundamentales y variables macroeconómicas, mejorando sustancialmente la precisión predictiva.

La robustez ante valores atípicos y datos ruidosos constituye una ventaja especialmente relevante en mercados caracterizados por alta volatilidad. Como señalan López-Martín et al. (2022), algoritmos como Random Forest son inherentemente robustos ante outliers debido a su naturaleza de ensamble y muestreo, mientras que implementaciones modernas de Gradient Boosting como XGBoost incorporan parámetros de regularización específicamente diseñados para mitigar el impacto de ruido en datos financieros.

La capacidad de cuantificar importancia de variables representa una ventaja adicional con significativo valor interpretativo. A diferencia de "cajas negras" como algunas implementaciones

de redes neuronales, algoritmos basados en árboles proporcionan métricas de importancia de características que permiten identificar los factores más influyentes en predicciones. Fischer et al. (2018) utilizaron esta característica para determinar que, en acciones tecnológicas, indicadores como momentum de precios, volatilidad implícita y sorpresas en resultados trimestrales emergían consistentemente como predictores dominantes.

Finalmente, la adaptabilidad a datos heterogéneos facilita la integración de información estructurada y no estructurada. Carmona et al. (2021) demostraron cómo modelos supervisados podían incorporar eficientemente datos derivados de análisis de sentimiento, noticias y tendencias de búsqueda web, particularmente relevantes para empresas tecnológicas cuyas valoraciones están fuertemente influenciadas por percepciones de mercado y tendencias emergentes.

2.6 Evaluación del Desempeño de modelos predictivos en finanzas

2.6.1 Métricas de evaluación y validación en contextos financieros

La evaluación rigurosa del desempeño predictivo constituye un componente fundamental en la implementación de modelos para mercados financieros. A diferencia de aplicaciones estándar de aprendizaje automático, donde métricas generales como precisión o error cuadrático medio pueden resultar suficientes, los contextos financieros requieren consideraciones adicionales derivadas de la naturaleza específica de los mercados y objetivos de inversión.

Las métricas estadísticas tradicionales como Error Cuadrático Medio (MSE), Error Absoluto Medio (MAE) y Coeficiente de Determinación (R^2) proporcionan evaluaciones básicas de ajuste predictivo. Sin embargo, como señalan De Prado (2018) y López de Prado (2020), estas métricas presentan limitaciones significativas en contextos financieros: no capturan adecuadamente asimetrías de riesgo, ignoran autocorrelación en errores predictivos, y no incorporan consideraciones de materialidad económica.

La validación temporal adquiere particular relevancia en series financieras. A diferencia de validación cruzada estándar, la validación con bloqueo temporal (time-blocked cross-validation) respeta la secuencialidad de datos, evitando "mirada hacia adelante" (look-ahead bias) que artificialmente inflaría rendimiento predictivo. Campbell et al. (2018) recomiendan

específicamente este enfoque para evaluación realista de modelos financieros, incorporando adicionalmente ventanas expandibles para capturar cambios en relaciones subyacentes.

Las métricas basadas en direccionalidad evalúan la capacidad del modelo para predecir correctamente la dirección del movimiento de precios, independientemente de su magnitud exacta. Ballings et al. (2015) argumentan que estas métricas, incluyendo Precisión Direccional y Coeficiente de Correlación de Rangos, resultan frecuentemente más relevantes para aplicaciones de trading que métricas de error puntual, particularmente en mercados volátiles como los tecnológicos.

Las métricas económicas como Retorno Ajustado por Riesgo (Sharpe Ratio), Máximo Drawdown, y Turnover de Portafolio proporcionan evaluaciones directamente vinculadas a resultados financieros. Krauss et al. (2017) implementaron un marco evaluativo que primorizaba estas métricas, demostrando cómo modelos con estadísticas predictivas aparentemente inferiores podían generar resultados económicos superiores al capturar eficientemente oportunidades de alta materialidad.

Finalmente, la evaluación de robustez temporal resulta crítica en entornos no estacionarios. Técnicas como rolling-window analysis y pruebas de cambio estructural permiten identificar degradación predictiva ante cambios en regímenes de mercado. Chen et al. (2019) encontraron que modelos supervisados como Gradient Boosting mantenían desempeño predictivo relativamente estable durante transiciones de régimen que deterioraban significativamente modelos econométricos tradicionales, evidenciando mayor adaptabilidad ante cambios estructurales característicos de mercados tecnológicos.

2.6.2 Consideraciones sobre sobreajuste y generalización en predicción financiera

El sobreajuste (overfitting) representa uno de los desafíos más significativos en la implementación de modelos predictivos para mercados financieros. Este fenómeno, donde un modelo captura ruido aleatorio en datos históricos en lugar de patrones generalizables, resulta particularmente problemático en finanzas debido a la baja relación señal-ruido característica de series temporales financieras.

La complejidad inherente de la dinámica de mercados tecnológicos amplifica este riesgo. Como señalan Pastor y Veronesi (2009), empresas como Google y Amazon exhiben patrones de valoración complejos, influenciados por expectativas cambiantes sobre innovación y crecimiento futuro, generando señales particularmente difíciles de separar del ruido de mercado. Bailey et al. (2014) documentaron el fenómeno de "minería de patrones" (pattern mining) donde, dada suficiente flexibilidad algorítmica, prácticamente cualquier relación puede "descubrirse" en datos históricos sin poseer genuino valor predictivo futuro.

Las estrategias de mitigación incluyen regularización algorítmica, validación temporal rigurosa y perspectiva realista sobre capacidad predictiva. Específicamente, De Prado (2018) recomienda enfoques como:

Regularización integrada: Algoritmos como Gradient Boosting implementan parámetros como "learning rate" y profundidad máxima que controlan complejidad modelística, previniendo capacidad excesiva para ajustar ruido.

Validación con bloqueo temporal: Utilización exclusiva de técnicas que respetan secuencialidad de datos, previniendo filtración de información futura en entrenamiento.

Parsimonia variable: Preferencia por modelos con menor número de predictores, particularmente cuando muestran desempeño comparable en validación.

Ensembles de modelos: Combinación de predicciones de múltiples modelos con diferentes arquitecturas y parámetros, reduciendo varianza predictiva.

López-Martín et al. (2022) encontraron que, específicamente para acciones tecnológicas, estrategias de regularización adaptativa que ajustan penalizaciones según características de datos y regímenes de mercado lograban balance óptimo entre capacidad para capturar patrones complejos y prevención de sobreajuste.

Adicionalmente, Dixon et al. (2020) destacan la importancia de contextualización realista, reconociendo que, en mercados altamente eficientes, la capacidad predictiva posible está inherentemente limitada. Sus experimentos con datos de alta frecuencia de acciones tecnológicas demostraron que modelos bien regularizados típicamente alcanzaban precisión direccional de 53-

58%, significativamente superior a expectativas aleatorias, pero sustancialmente inferior a predecibilidad perfecta, representando un equilibrio realista entre escepticismo extremo y optimismo injustificado.

2.7 Implementación práctica de modelos supervisados en trading algorítmico

La implementación de modelos predictivos en entornos de mercado reales se materializa a través del trading algorítmico, una disciplina que utiliza sistemas computacionales para ejecutar estrategias de inversión a velocidades y frecuencias inalcanzables para un operador humano. En este ecosistema, Python se ha consolidado como el lenguaje de programación dominante por su flexibilidad, su robusto ecosistema de librerías de código abierto (como Pandas, NumPy y Scikit-learn) y su capacidad para gestionar todo el flujo de trabajo cuantitativo.

La capacidad de Python para procesar grandes volúmenes de datos, entrenar modelos de aprendizaje supervisado y realizar backtesting (simulaciones históricas de una estrategia) es lo que permite a los analistas y quants traducir un modelo predictivo, como los desarrollados en esta tesis, en señales de trading cuantitativas. El objetivo es lograr "aciertos precisos", es decir, generar señales de compra o venta que tengan una alta probabilidad de éxito estadístico, mejorando así la rentabilidad ajustada por riesgo. Como destaca López de Prado (2018), la capacidad de automatizar este proceso de investigación y validación es fundamental para descubrir patrones predictivos genuinos y evitar sesgos.

2.7.1 Arquitectura de sistemas para trading basado en modelos predictivos

La implementación efectiva de modelos supervisados en entornos de trading algorítmico requiere arquitecturas sistémicas robustas que integren eficientemente componentes de adquisición de datos, preprocessing, predicción y ejecución. López de Prado (2018) propone un marco arquitectónico modular que ha ganado amplia adopción en aplicaciones prácticas, particularmente para mercados de alta complejidad como los tecnológicos.

El módulo de adquisición de datos constituye el fundamento del sistema, responsable de recopilar, validar y almacenar información de diversas fuentes. Para predicción de precios de empresas como Google y Amazon, este módulo típicamente incorpora datos de mercado (precios,

volúmenes, órdenes), información macroeconómica (tasas de interés, indicadores inflacionarios), datos fundamentales (métricas financieras corporativas) y potencialmente señales alternativas (sentimiento de redes sociales, tendencias de búsqueda). Hendershott et al. (2011) destacan la importancia crítica de infraestructura de baja latencia para adquisición de datos de mercado, particularmente en entornos donde microsegundos pueden determinar ventajas informativas.

El módulo de preprocesamiento transforma datos brutos en características utilizables algorítmicamente. Este proceso incluye manejo de valores faltantes, normalización, transformación de características y generación de variables derivadas como indicadores técnicos y señales macroeconómicas procesadas. Carmona et al. (2021) proponen arquitecturas específicas para integración eficiente de datos heterogéneos, particularmente relevantes para acciones tecnológicas donde señales tradicionales y alternativas frecuentemente requieren procesamiento diferenciado.

El módulo predictivo implementa algoritmos de aprendizaje supervisado, recibiendo características procesadas y generando predicciones sobre movimientos futuros de precios. Arquitecturas modernas frecuentemente incorporan ensembles de múltiples modelos como Random Forest, Gradient Boosting y potencialmente redes neuronales, combinando sus predicciones mediante técnicas de stacking o voting. Este enfoque, documentado por Krauss et al. (2017), ha demostrado mayor robustez ante cambios en regímenes de mercado, particularmente relevante para acciones volátiles como las tecnológicas.

El módulo de gestión de riesgos y optimización de portafolio traduce predicciones en decisiones de trading concretas, incorporando consideraciones como volatilidad, correlaciones, costos de transacción y restricciones de liquidez. López de Prado (2020) destaca la importancia de este componente, argumentando que predicciones precisas pueden generar resultados económicos deficientes si no se incorporan adecuadamente en el contexto de un portafolio optimizado.

Finalmente, el módulo de ejecución implementa decisiones de trading mediante interacción directa con mercados o brokers, optimizando timing, división de órdenes y selección de venues para minimizar slippage y market impact. Para acciones altamente líquidas como Google y Amazon, estrategias de ejecución algorítmica como VWAP (Volume Weighted Average Price) o

implementaciones basadas en aprendizaje por refuerzo han demostrado reducir significativamente costos de implementación, como documentan Cont et al. (2022).

2.7.2 Consideraciones prácticas: costos de transacción, liquidez y escalabilidad

La implementación práctica de estrategias basadas en modelos supervisados exige consideración detallada de factores operacionales que frecuentemente determinan viabilidad económica más allá de precisión predictiva teórica. Estos factores adquieren particular relevancia en mercados tecnológicos caracterizados por alta volatilidad y entornos de trading algorítmico competitivos.

Los costos de transacción representan factores críticos frecuentemente subestimados en estudios académicos. Estos incluyen componentes explícitos (comisiones, impuestos, tarifas de intercambio) e implícitos (spread bid-ask, market impact, slippage, oportunidad). De Prado (2018) recomienda modelización realista incorporando estructura microeconómica específica de cada activo, particularmente relevante para acciones tecnológicas donde dinámicas de liquidez pueden variar significativamente según condiciones de mercado. Específicamente para Google y Amazon, Cont et al. (2022) encontraron que costos implícitos podían superar ampliamente costos explícitos durante períodos de alta volatilidad, requiriendo ajustes dinámicos en estrategias de ejecución.

La liquidez y consideraciones de capacidad determinan escalabilidad de estrategias. Ballings et al. (2015) documentan cómo señales predictivas aparentemente rentables en backtest teórico frecuentemente degradaban significativamente ante implementación a escala real, particularmente en estrategias de alta frecuencia. Para acciones tecnológicas, caracterizadas generalmente por alta liquidez, pero también alta concentración institucional, estimaciones realistas de capacity constraints resultan fundamentales. Zhang y Yan (2020) proponen metodologías específicas para estimar ceiling de capacidad basado en análisis de volumen, volatilidad y participación institucional, particularmente relevantes para implementaciones que superen umbrales de capital significativos.

La infraestructura tecnológica constituye otro factor determinante, particularmente para estrategias que explotan ineficiencias de corto plazo. La evolución hacia arquitecturas distribuidas,

procesamiento en tiempo real y técnicas de computación de baja latencia ha transformado la implementación práctica de modelos supervisados. Hendershott et al. (2011) documentan cómo inversiones en infraestructura tecnológica han pasado de ventajas competitivas a requisitos fundamentales para participación efectiva en mercados algorítmicos modernos.

Finalmente, consideraciones regulatorias y éticas adquieren creciente relevancia. El crecimiento de la complejidad algorítmica ha sido acompañado por mayor escrutinio regulatorio en jurisdicciones principales, focalizando en aspectos como pruebas de estrés, documentación de modelos y consideraciones sobre impacto en estabilidad de mercados. Para implementaciones prácticas, estas consideraciones requieren integración anticipada en arquitecturas sistémicas y procesos de desarrollo, como destacan Evans y Schmalensee (2016), particularmente relevante para estrategias que involucran activos de alta visibilidad como acciones tecnológicas líderes.

2.8 Dirección futura en predicción de precios de acciones tecnológicas

2.8.1 Integración de datos alternativos y análisis de sentimiento

La integración de datos alternativos representa una frontera de innovación en predicción de precios de acciones tecnológicas, complementando variables tradicionales con fuentes de información previamente inaccesibles o subutilizadas. Esta tendencia responde a la proliferación de datos digitales y avances en capacidades de procesamiento que permiten extraer señales predictivas de fuentes diversas.

El análisis de sentimiento derivado de redes sociales, foros financieros y plataformas especializadas ha demostrado valor predictivo significativo, particularmente para empresas tecnológicas con alta visibilidad mediática. Bartov et al. (2018) documentaron cómo el sentimiento agregado de Twitter precedía sorpresas en resultados trimestrales para empresas tecnológicas, con capacidad predictiva incrementada para compañías con mayor presencia en redes sociales como Google y Amazon (Bartov et al., 2018). Metodologías avanzadas como análisis semántico basado en transformers y modelos de procesamiento de lenguaje natural han ampliado significativamente la sofisticación de estos enfoques (Carmona et al., 2021).

Los datos de búsqueda web representan otra fuente informativa con demostrado valor predictivo. Da et al. (2011) encontraron que incrementos en volúmenes de búsqueda de empresas

específicas frecuentemente precedían movimientos significativos de precios, particularmente para empresas tecnológicas. La integración de datos propietarios de Google Trends y plataformas similares en modelos supervisados ha mostrado mejoras marginales significativas en precisión predictiva, especialmente durante períodos de alta incertidumbre como lanzamientos de productos o anuncios regulatorios.

Los datos satelitales y geospaciales proveen perspectivas únicas sobre actividad económica. Para empresas como Amazon con significativa presencia física, análisis de patrones de tráfico en centros logísticos, monitoreo satelital de expansión de infraestructura y datos de movilidad han demostrado capacidad para anticipar tendencias operacionales antes de divulgación oficial en reportes financieros. Zhu et al. (2021) implementaron modelos supervisados que integraban datos geospaciales con predictores tradicionales, logrando mejoras significativas en capacidad predictiva para empresas con operaciones físicas sustanciales.

La actividad de patentes y propiedad intelectual proporciona señales relevantes específicamente para sectores intensivos en innovación. Li et al. (2022) desarrollaron metodologías para cuantificar valor estratégico de portafolios de patentes utilizando procesamiento de lenguaje natural y análisis de redes de citas, encontrando capacidad predictiva significativa para rendimientos a largo plazo de empresas tecnológicas.

Sin embargo, la integración efectiva de datos alternativos presenta desafíos significativos. López de Prado (2020) destaca problemas como sobreajuste derivado de múltiple testing, sesgos de selección en construcción de datasets, y degradación de señal ante adopción generalizada. Metodologías robustas para evaluación de contribución marginal, regularización específica para datos heterogéneos y frameworks para integración óptima de fuentes diversas constituyen áreas activas de investigación.

2.8.2 Aplicaciones de aprendizaje profundo y modelos híbridos

Los modelos de aprendizaje profundo (deep learning) han emergido como herramientas prometedoras para predicción de mercados financieros, ofreciendo capacidades avanzadas para modelar relaciones temporales complejas características de acciones tecnológicas. Su evolución

reciente y aplicaciones híbridas representan direcciones de investigación particularmente relevantes.

Las redes neuronales recurrentes (RNN), especialmente arquitecturas como LSTM (Long Short-Term Memory) y GRU (Gated Recurrent Units), han demostrado capacidad superior para capturar dependencias temporales de largo plazo en series financieras. Fischer y Krauss (2018) implementaron LSTM para predicción de movimientos en S&P 500, logrando rentabilidades anormales estadísticamente significativas. Para acciones tecnológicas específicamente, estas arquitecturas han mostrado capacidad para identificar patrones cíclicos complejos relacionados con ciclos de producto, estacionalidad en ingresos publicitarios y patrones de adopción tecnológica que modelos tradicionales no capturaban eficientemente.

Las redes convolucionales (CNN) han encontrado aplicación en análisis de patrones visuales en datos financieros, particularmente para reconocimiento de formaciones chartistas y patrones técnicos multiescala. Tsantekidis et al. (2020) aplicaron CNN a datos de libros de órdenes limit order book, logrando mejoras significativas en predicción de movimientos microestructurales para acciones de alta frecuencia de negociación, incluyendo empresas tecnológicas líderes.

Los transformers y modelos de atención, originalmente desarrollados para procesamiento de lenguaje natural, han comenzado a aplicarse a series temporales financieras con resultados prometedores. Estas arquitecturas permiten modelar eficientemente dependencias de largo alcance e interacciones entre múltiples series temporales simultáneamente. Wu et al. (2021) implementaron transformers para predicción de precios de acciones tecnológicas, incorporando simultáneamente datos de precios, volúmenes, sentimiento y variables macroeconómicas, logrando mejoras significativas sobre modelos recurrentes tradicionales.

Los modelos híbridos, que combinan diferentes paradigmas algorítmicos, representan una dirección particularmente prometedora. Zhang et al. (2023) proponen arquitecturas que integran capas recurrentes para capturar dependencias temporales, mecanismos de atención para identificar variables relevantes condicionalmente, y componentes basados en árboles para manejar relaciones no lineales entre predictores. Estos enfoques híbridos han demostrado capacidad para superar limitaciones individuales de cada paradigma, particularmente relevante para acciones tecnológicas

caracterizadas por dinámicas temporales complejas y dependencias variables entre factores predictivos.

Sin embargo, implementaciones prácticas enfrentan desafíos significativos. (Gu, Kelly, & Xiu, 2020) documentan cómo beneficios teóricos de modelos profundos frecuentemente no se materializan completamente en aplicaciones financieras debido a limitaciones en cantidad de datos disponibles (comparado con dominios como visión computacional) y alta relación ruido-síñal inherente. Metodologías como preentrenamiento en datos sintéticos, transfer learning desde dominios relacionados, y arquitecturas específicamente diseñadas para datos financieros limitados constituyen áreas activas de investigación para abordar estas limitaciones.

2.9 Conclusiones del marco teórico

El marco teórico desarrollado evidencia la evolución fundamental que ha experimentado la predicción de precios en mercados financieros, transitando desde paradigmas econométricos tradicionales hacia enfoques avanzados de aprendizaje automático. Esta evolución responde tanto a limitaciones inherentes de modelos lineales para capturar comportamientos complejos de mercado, como a oportunidades derivadas de avances computacionales y disponibilidad de datos heterogéneos.

La revisión de literatura revela que algoritmos supervisados como Random Forest y Gradient Boosting ofrecen ventajas significativas para modelar dinámicas de precios en acciones tecnológicas. Su capacidad para capturar relaciones no lineales, manejar eficientemente múltiples predictores, cuantificar importancia de variables y adaptarse a datos heterogéneos los posiciona como herramientas particularmente adecuadas para entornos caracterizados por alta volatilidad y complejidad.

La integración de variables macroeconómicas emerge como componente fundamental para modelos predictivos robustos. La evidencia teórica y empírica confirma que factores como tasas de interés e inflación inducen efectos significativos sobre valoraciones del sector tecnológico, aunque a través de mecanismos complejos y frecuentemente no lineales que justifican enfoques algorítmicos avanzados.

El análisis de implementaciones prácticas destaca la importancia de consideraciones operacionales más allá de precisión predictiva teórica. Factores como costos de transacción, restricciones de liquidez, requerimientos infraestructurales y consideraciones regulatorias frecuentemente determinan viabilidad económica de estrategias basadas en modelos supervisados.

Las direcciones futuras identificadas, incluyendo integración de datos alternativos y aplicaciones de aprendizaje profundo, sugieren amplio potencial para evolución adicional en capacidades predictivas. Sin embargo, estos avances requerirán abordajes metodológicos robustos para mitigar riesgos de sobreajuste y garantizar generalización efectiva en entornos no estacionarios característicos de mercados financieros.

Las perspectivas teóricas revisadas soportan la hipótesis de que modelos supervisados, particularmente algoritmos basados en árboles, complementados con variables macroeconómicas relevantes, pueden ofrecer capacidad predictiva significativa en mercados de acciones tecnológicas. Esta conclusión justifica el enfoque metodológico del presente estudio, que procederá a implementar y evaluar empíricamente los modelos y variables analizados en el marco teórico.

Capítulo III: Diseño metodológico

3.1 Enfoque de la investigación

Esta investigación tiene un enfoque cuantitativo, ya que trabajamos principalmente con datos numéricos para analizar cómo los modelos supervisados pueden mejorar la predicción de precios de activos financieros. Nos centramos en medir con precisión el impacto de variables como tasas de interés e inflación en los precios de acciones de Google y Amazon durante el período 2021-2024.

Lo que buscamos es comprobar, con datos concretos, si estos modelos avanzados realmente superan a los métodos tradicionales. Todo nuestro análisis se basa en información objetiva y medible, como precios históricos y datos macroeconómicos, para llegar a conclusiones respaldadas por evidencia estadística.

3.1.1 Enfoque metodológico mixto y análisis de escenarios

Si bien el núcleo del análisis empírico de esta tesis es de naturaleza cuantitativa, se adopta un **enfoque metodológico mixto** en su concepción. Se complementa el análisis cuantitativo de los datos históricos y las predicciones del modelo con un módulo cualitativo conceptual de **escenarios prospectivos** (base/moderado/estrés). Este marco es particularmente útil para analizar la resiliencia conceptual del modelo ante shocks exógenos, permitiendo articular narrativas sobre eventos de baja probabilidad, pero alto impacto, como **crisis geopolíticas (guerras) o emergencias sanitarias (pandemias)**, y sus posibles efectos en las variables macroeconómicas clave (tasas, inflación, VIX, entre otras). Es importante destacar que este módulo cualitativo no altera el modelo predictivo principal, sino que ofrece un marco robusto para discutir la sensibilidad del modelo y la posibilidad de cambios de régimen, enriqueciendo la interpretación de los resultados desde una perspectiva estratégica.

3.2 Método

Utilizamos un método hipotético-deductivo, partiendo de la teoría existente sobre mercados financieros y aprendizaje automático para formular hipótesis específicas sobre la capacidad predictiva de los modelos supervisados. Estas hipótesis serán sometidas a prueba mediante análisis de datos reales.

El proceso incluye la recopilación de datos históricos de precios de acciones de Google y Amazon, junto con variables macroeconómicas relevantes como tasas de interés e índices de inflación. Luego aplicamos técnicas de preprocesamiento, entrenamiento de modelos y evaluación de resultados para verificar nuestras hipótesis iniciales.

3.3 Diseño de estudio

El diseño contempla una fase retrospectiva, en la que se analizan datos históricos del período 2021-2024 con el fin de identificar patrones, tendencias y efectos de variables macroeconómicas como la inflación, las tasas de interés y la política monetaria sobre los precios de activos. A continuación, se desarrolla una fase predictiva, destinada a evaluar la capacidad de distintos modelos estadísticos y econométricos para anticipar movimientos futuros de precios, comparando su precisión, consistencia y aplicabilidad práctica. En conjunto, la investigación combina el análisis del pasado con la proyección hacia el futuro, lo que no solo fortalece la validez de los hallazgos, sino que también aporta información útil tanto en el ámbito académico como en el campo de la toma de decisiones económicas y financieras.

3.4 Población y muestra

La población de estudio comprende todos los datos históricos de precios de acciones de Google y Amazon, junto con las variables macroeconómicas relevantes como tasas de interés e índices de inflación.

La muestra específica incluye:

- Datos diarios de precios de cierre, apertura, máximo, mínimo y volumen de transacciones de acciones de Google y Amazon durante el período 2021-2024.
- Datos mensuales de tasas de interés de referencia establecidas por la Reserva Federal de Estados Unidos.
- Datos mensuales del Índice de Precios al Consumidor (IPC) como medida de inflación.
- Otros indicadores económicos relevantes como el PIB trimestral y datos de empleo.

Se utilizó un muestreo no probabilístico por conveniencia, seleccionando el período 2021-2024 por su relevancia actual y la disponibilidad de datos completos.

3.5 Técnicas e instrumentos de recolección

Para recopilar los datos necesarios, empleamos las siguientes técnicas e instrumentos:

- Se emplearon APIs financieras reconocidas, como Yahoo Finance y Alpha Vantage, para obtener datos históricos diarios de precios de acciones de empresas como Google y Amazon. Estos datos incluyen precios de apertura, cierre, máximos, mínimos y volumen de negociación, permitiendo un análisis detallado y actualizado del comportamiento del mercado bursátil. La frecuencia diaria de los registros garantiza la precisión necesaria para la construcción y validación de los modelos predictivos.
- Se obtuvo información actualizada y confiable de bases oficiales como la Reserva Federal (FRED) y la Oficina de Estadísticas Laborales (BLS). Estas fuentes proporcionaron datos históricos de tasas de interés, inflación y otros indicadores económicos clave. Los datos fueron seleccionados por su relevancia para el análisis y luego procesados para asegurar su calidad y compatibilidad con los modelos de predicción utilizados.
- Herramientas de procesamiento: Empleamos bibliotecas de Python como Pandas, NumPy y scikit-learn para el preprocesamiento, análisis y modelado de datos.
- Instrumentos de validación: Utilizamos técnicas de validación cruzada con bloqueo temporal para evaluar el rendimiento de los modelos de manera realista, evitando sesgos de mirada hacia adelante.

3.6 Hipótesis o supuestos de investigación

Basándonos en nuestro marco teórico y objetivos, planteamos las siguientes hipótesis:

H1: Los modelos supervisados (Random Forest y Gradient Boosting) ofrecen mayor precisión en la predicción de precios de activos financieros en comparación con métodos tradicionales como regresiones lineales y modelos ARIMA.

H2: Las tasas de interés tienen un impacto significativo en la precisión de los modelos supervisados para predecir precios de acciones de Google y Amazon, con una relación inversa entre cambios en tasas y valoraciones.

H3: Los índices de inflación influyen significativamente en la precisión de los modelos supervisados para predecir precios de acciones tecnológicas, con efectos variables según la capacidad de las empresas para transferir costos.

H4: La incorporación de datos históricos de precios con diferentes horizontes temporales mejora significativamente la precisión de los modelos supervisados en la predicción de precios futuros.

3.7 Operacionalización de variables

La selección de las variables independientes para este estudio se fundamenta en los pilares de la teoría de valoración de activos financieros. En lugar de utilizar herramientas de análisis estratégico cualitativo como PESTEL o MICMAC, que son valiosas en un contexto de planificación corporativa, se optó por un enfoque cuantitativo donde las variables se validan por su rol central y demostrado en los modelos económicos. Las tasas de interés y la inflación no son variables arbitrarias; son los componentes esenciales de cualquier modelo de descuento de flujos de caja (Damodaran, 2012) y determinantes del costo de capital, y han sido demostradas como factores de riesgo sistemático que influyen en los precios de las acciones (Chen, Roll, & Ross, 1986). El enfoque metodológico mixto y análisis de escenarios descrito en la sección 3.1.1 complementa este enfoque, al permitir una discusión cualitativa sobre el impacto potencial de factores externos (geopolíticos, sociales, etc.) que estas herramientas prospectivas suelen identificar.

Tabla 1*Operacionalización de variables independientes y dependientes*

Como se muestra en la Tabla 1, cada dimensión de las variables independientes y dependientes están ligadas a indicadores concretos, medibles y cuantificables mediante análisis documental que se ha establecido para garantizar la claridad, la medición precisa en este estudio y cómo serán operacionalizadas las variables.

Variable independiente	Definición conceptual	Definición operacional	Indicadores
Tasas de interés	Costo del dinero establecido por autoridades monetarias que influye en las decisiones de inversión y valoración de activos.	Tasa de fondos federales de la Reserva Federal de EE.UU	Valor porcentual mensual, cambios mensuales, tendencia.
Índices de inflación	Aumento generalizado y sostenido de precios de bienes y servicios.	Índice de Precios al Consumidor (IPC) de EE.UU.	Valor porcentual mensual, cambios mensuales, tendencia.
Datos históricos de precios	Registro temporal de valoraciones previas de activos.	Precios diarios de cierre de acciones de Google y Amazon.	Valores absolutos, rendimientos, volatilidad, indicadores técnicos derivados.
Variable dependiente:	Definición conceptual	Definición operacional	Indicadores
Precisión de predicción de precios	Grado de exactitud con que un modelo predice precios futuros.	Error entre precios predichos y reales.	Error Cuadrático Medio (MSE), Error Absoluto Medio (MAE), Coeficiente de Determinación (R ²).

Nota. Las variables independientes son tasas de interés, inflación y precios históricos. La variable dependiente es la precisión en la predicción de precios.

3.8 Estrategias de recolección, procesamiento y análisis de la información

Nuestro proceso metodológico sigue estas etapas:

1. Recolección de datos:
 - Extracción de datos históricos de precios de Google y Amazon mediante APIs financieras.
 - Obtención de datos macroeconómicos de fuentes oficiales como FRED y BLS. Organización de datos en series temporales alineadas.
2. Preprocesamiento:
 - Limpieza de datos (manejo de valores faltantes y atípicos)
 - Normalización y transformación de variables.
 - Creación de características derivadas (indicadores técnicos, rezagos).
 - División de datos en conjuntos de entrenamiento (80%) y prueba (20%)
3. Modelado:
 - Implementación de modelos supervisados (Random Forest y Gradient Boosting).
 - Implementación de modelos tradicionales como línea base (regresiones, ARIMA).
 - Optimización de hiperparámetros mediante validación cruzada con bloqueo temporal.
4. Evaluación:
 - Cálculo de métricas de rendimiento (MSE, MAE, R^2) para cada modelo.
 - Análisis comparativo entre modelos supervisados y tradicionales.
 - Evaluación del impacto de variables macroeconómicas mediante análisis de importancia de características.
5. Análisis e interpretación:
 - Pruebas estadísticas para validar hipótesis.
 - Análisis de sensibilidad para evaluar robustez de resultados.
 - Interpretación de hallazgos en el contexto de la teoría financiera y de aprendizaje automático.

3.9 Consideraciones éticas

En nuestra investigación, nos comprometemos a seguir principios éticos fundamentales:

- **Transparencia:** Documentamos detalladamente todos los procedimientos metodológicos, fuentes de datos y decisiones de modelado para garantizar la reproducibilidad de los resultados.
- **Integridad de datos:** Utilizamos únicamente datos públicos y legalmente accesibles, respetando los términos de uso de las fuentes de información.
- **Objetividad:** Mantenemos un enfoque imparcial en la evaluación de modelos, evitando sesgos de selección o confirmación que puedan distorsionar los resultados.
- **Limitaciones:** Reconocemos y comunicamos claramente las limitaciones de nuestros modelos y análisis, evitando afirmaciones exageradas sobre su capacidad predictiva.
- **Responsabilidad:** Consideramos las implicaciones prácticas de nuestros hallazgos, reconociendo que los modelos predictivos en finanzas tienen impactos potenciales en decisiones de inversión reales.
- **Confidencialidad:** Aunque trabajamos con datos públicos, mantenemos estándares de seguridad adecuados en el manejo y almacenamiento de la información.

Capítulo IV: Metodología empírica y análisis de datos

Este capítulo detalla los procedimientos empíricos implementados para la validación de los objetivos específicos y las hipótesis de la investigación. Se inicia con la descripción del conjunto de datos, su preparación y la ingeniería de características, para luego proceder con el análisis exploratorio que sienta las bases para el modelado predictivo.

4.1 Construcción del conjunto de datos

La robustez de cualquier análisis empírico depende fundamentalmente de la calidad y pertinencia de los datos. Para este estudio, se construyó una base de datos de series temporales multivariadas, abarcando el período del 1 de enero de 2021 al 31 de diciembre de 2024. El propósito de este dataset es proporcionar la información necesaria para comparar el rendimiento de modelos de aprendizaje supervisado frente a benchmarks tradicionales en la predicción de precios de activos tecnológicos.

4.1.1 Composición y fuentes del dataset

El conjunto de datos (`dataset_completo_google_amazon.csv`) fue ensamblado a partir de fuentes de datos financieros de alta fiabilidad, como Yahoo Finance y la Reserva Federal de St. Louis (FRED). Contiene las siguientes variables clave, agrupadas por categoría:

1. Variables de Activos Principales:

- `GOOGL_Close`: Precio de cierre diario ajustado de las acciones de Google (Alphabet Inc.).
- `AMZN_Close`: Precio de cierre diario ajustado de las acciones de Amazon.com, Inc.

2. Variables Macroeconómicas:

- `Treasury_10Y`: Tasa de rendimiento de los bonos del Tesoro de EE. UU. a 10 años. Es un proxy fundamental para el costo del dinero y la tasa libre de riesgo en la economía, central para el Objetivo Específico 1.

- **Inflacion_T5YIE**: Tasa de inflación implícita a 5 años (TIPS 5-Year Breakeven Inflation Rate). Mide las expectativas de inflación del mercado y es la variable principal para la evaluación del Objetivo Específico 2.

3. Indicadores de mercado y sentimiento:

- **VIX**: Índice de volatilidad del CBOE (Chicago Board Options Exchange), comúnmente denominado "índice del miedo". Mide la expectativa de volatilidad del mercado a 30 días.
- **SP500**: Valor de cierre diario del índice S&P 500, representativo del mercado bursátil estadounidense en general.
- **NASDAQ**: Valor de cierre diario del índice NASDAQ Composite, que tiene una alta ponderación de empresas del sector tecnológico.

Para promover la transparencia y la reproducibilidad, el conjunto de datos completo y los scripts de análisis utilizados en esta investigación se encuentran disponibles en el repositorio público de GitHub (jaimeisaac2020, 2025).

4.1.2 Estructuración de datos y preparación inicial

El dataset fue cargado y procesado utilizando la librería Pandas de Python. El primer paso crucial fue la conversión de la columna de fecha al formato `datetime` y su establecimiento como el índice del DataFrame. Esta indexación temporal es esencial para el correcto manejo, alineación y análisis de las series de tiempo.

4.1.3 Vista previa de la estructura de datos.

Para ofrecer una visión concreta de la estructura y el contenido del dataset final, la Tabla 2 y 3 presentan las primeras cinco (`head`) y las últimas cinco (`tail`) observaciones del conjunto de datos. Esto permite visualizar el formato de las fechas, la escala de los valores y la organización general de la información utilizada en este estudio.

Tabla 2*Primeras 5 observaciones del conjunto de datos (head)**Nota.* Elaboración propia a partir de datos de Yahoo Finance y FRED

Date	GOOGL_Close	AMZN_Close	Treasury_10Y	VIX	Inflacion_T5YIE
2021-01-04	85.90	159.33	0.92	26.97	1.98
2021-01-05	86.59	160.93	0.95	25.34	2.01
2021-01-06	85.73	156.92	1.04	25.07	2.04
2021-01-07	88.29	158.11	1.07	22.37	2.09
2021-01-08	89.46	159.13	1.11	21.56	2.07

4.1.4 Ingeniería de características y definición de variables**Tabla 3***Últimas 5 observaciones del conjunto de datos (tail)*

Date	GOOGL_Close	AMZN_Close	Treasury_10Y	VIX	Inflacion_T5YIE
2021-01-04	85.90	159.33	0.92	26.97	1.98
2021-01-05	86.59	160.93	0.95	25.34	2.01
2021-01-06	85.73	156.92	1.04	25.07	2.04
2021-01-07	88.29	158.11	1.07	22.37	2.09
2021-01-08	89.46	159.13	1.11	21.56	2.07

Nota. Elaboración propia a partir de datos de Yahoo Finance y FRED.

Para que los datos brutos sean útiles para el modelado predictivo, es necesario transformarlos y crear variables derivadas que capturen dinámicas relevantes del mercado.

Cálculo de retornos porcentuales diarios (Return_1d)

Justificación metodológica. Los precios de las acciones son series temporales no estacionarias, lo que significa que sus propiedades estadísticas (como la media) cambian con el tiempo. Utilizar precios no estacionarios en modelos de regresión puede llevar a correlaciones espurias y resultados inválidos. Para solucionar esto, los precios se transformaron en retornos porcentuales diarios, que tienden a ser más estacionarios.

$$\text{Fórmula. Retorno}_t = \frac{\text{Precio}_t - \text{Precio}_{t-1}}{\text{Precio}_{t-1}}$$

Implementación y rol en el modelo. Se utilizó la función `.pct_change()` de Pandas. Estas nuevas variables, `GOOGL_Return_1d` y `AMZN_Return_1d`, son fundamentales, ya que sus valores futuros (desplazados un día) servirán como la variable objetivo (y) que los modelos intentarán predecir.

Cálculo de volatilidad histórica (Volatility_20)

Justificación teórica. La volatilidad, o el grado de variación de los retornos, es una medida clave del riesgo y la incertidumbre. Como se discute en el marco teórico, los regímenes de alta y baja volatilidad a menudo preceden a diferentes comportamientos de los retornos.

Fórmula. Se calculó la desviación estándar móvil de los retornos diarios con una ventana de 20 días hábiles, representando aproximadamente un mes de negociación.

$$\text{Volatilidad}_t^{(n)} = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} (R_{t-i} - \bar{R})^2}, \text{ con } n=20.$$

Implementación y rol en el modelo. Se utilizó la función: `.rolling(window=20).std()` de Pandas. Las variables `GOOGL_Volatility_20` y `AMZN_Volatility_20` se utilizan como variables predictoras (X) para informar a los modelos sobre el contexto de riesgo reciente.

4.2 Análisis exploratorio de datos (EDA)

Antes del modelado, se realizó un EDA para visualizar las variables y obtener una comprensión intuitiva de sus patrones y relaciones.

4.2.1 Visualización de series temporales

Las Figuras 1 y 2 muestran la evolución de las principales variables a lo largo del período de estudio.

Figura 1

Evolución del precio de cierre diario de la acción de Amazon (AMZN) durante el período de estudio (2021-2024)



Nota. Elaboración propia a partir de datos de Yahoo Finance y FRED.

Figura 2

Evolución de la tasa de rendimiento de los bonos del tesoro de EE. UU. a 10 Años (Treasury_10Y) (2021-2024)



Nota. Elaboración propia a partir de datos de Yahoo Finance y FRED.

Figura 3

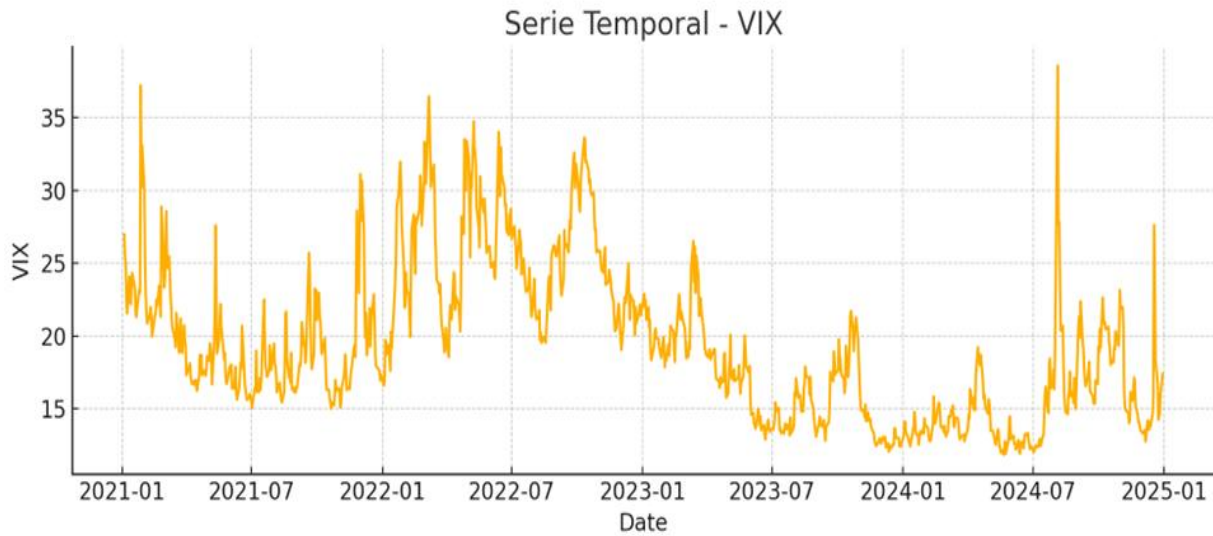
Evolución de la tasa de rendimiento de los bonos del tesoro de EE. UU. a 10 Años (Treasury_10Y) (2021-2024)



Nota. Elaboración propia a partir de datos de la Reserva Federal de St. Louis (FRED).

Figura 4

Evolución del índice de volatilidad CBOE (VIX) durante el periodo de estudio (2021-2024)



Nota. Elaboración propia a partir de datos de Yahoo Finance o FRED.

Figura 5

Evolución del índice de mercado S&P 500 durante el periodo de estudio (2021-2024)



Nota. Elaboración propia a partir de datos de Yahoo Finance o FRED.

Figura 6

Evolución del índice tecnológico NASDAQ composite durante el período de estudio (2021-2024)



Nota. Elaboración propia a partir de datos de Yahoo Finance o FRED.

Figura 7

Evolución de las expectativas de inflación a 5 Años (T5YIE) durante el período de estudio (2021-2024)



Nota. Elaboración propia a partir de datos de Yahoo Finance o FRED.

4.2.2 Observaciones clave del EDA:

Co-movimiento sectorial: Las series de precios de GOOGL y AMZN (Figura 1 y 2) muestran un fuerte co-movimiento, con tendencias y puntos de inflexión similares, lo que confirma su pertenencia a un mismo sector influenciado por factores comunes. Este paralelismo sugiere que ambos activos responden de manera semejante a shocks externos, lo cual incrementa la correlación entre sus rendimientos.

Régimen de tasas e inflación: Se observa un cambio de régimen claro a partir de 2022 (Figuras 3 y 7), con un aumento pronunciado tanto en las tasas de interés como en las expectativas de inflación. Este período coincide visiblemente con la corrección bajista en los precios de las acciones. Inversamente, la estabilización y posterior caída de la inflación desde mediados de 2022 parece correlacionarse con la recuperación de los precios.

Relación con la volatilidad: Los picos en el índice VIX (Figura 4) tienden a ocurrir durante las caídas más pronunciadas del mercado (S&P 500 y NASDAQ), validando su papel como un indicador de aversión al riesgo. Esta correlación negativa resalta cómo el VIX puede anticipar momentos de tensión. También reacciona ante eventos macroeconómicos, convirtiéndose en una herramienta clave. Su observación permite anticipar posibles giros en la tendencia. Captura el componente emocional del mercado y lo traduce en datos útiles. Por ello, el VIX complementa el análisis técnico y fortalece la capacidad predictiva.

4.3 Impacto de tasas de interés en precios de cierre de las acciones de Google y Amazon

En esta sección se abordan los resultados del análisis realizado para cumplir con el primer objetivo específico: “Determinar el impacto que tienen las tasas de interés en la predicción de precios de activos financieros utilizando modelos supervisados”. Asimismo, se busca validar la Hipótesis 2 (H2): “Las tasas de interés tienen un impacto significativo en la precisión de los modelos supervisados para predecir precios de acciones de Google y Amazon, con una relación inversa entre cambios en tasas y valoraciones”.

Para lograrlo, se inicia analizando las gráficas de las series de tiempo de los precios de cierre de Google y Amazon, así como la gráfica de las tasas de interés. Posteriormente, se implementó una metodología comparativa, entrenando dos conjuntos de modelos supervisados

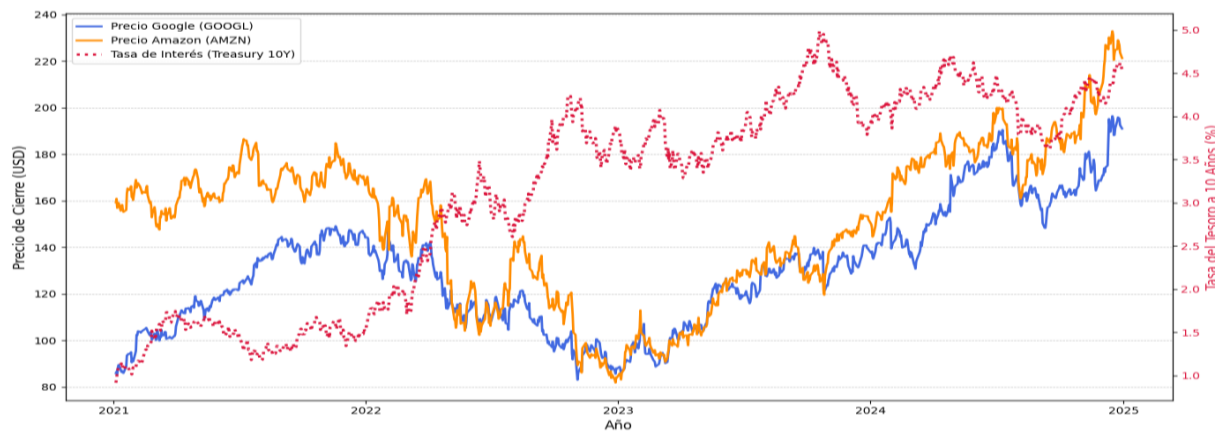
(Random Forest y Gradient Boosting): un conjunto “Base”, que excluye la tasa de interés del Tesoro a 10 años (Treasury_10Y), y un conjunto “Completo”, que la incluye. La evaluación se centró en la mejora de la precisión y en la interpretación del rol de dicha variable dentro del modelo. Este enfoque permitió contrastar el peso real de la tasa de interés en las predicciones y valorar si su inclusión aporta ventajas significativas en términos de desempeño y explicación de los resultados.

4.3.1 Análisis visual de la relación entre precios y tasas de interés

La Figura 8 presenta una visualización exploratoria de las series de tiempo centrales para este estudio: los precios de cierre de Google (GOOGL) y Amazon (AMZN) en el eje primario (izquierdo), y la tasa de interés del Tesoro de EE. UU. a 10 años en el eje secundario (derecho), abarcando el período de enero de 2021 a finales de 2024. Este gráfico sirve como un punto de partida fundamental para contextualizar la dinámica del mercado y formular observaciones preliminares sobre la Hipótesis 2, que postula una relación inversa entre las tasas de interés y las valoraciones de estas acciones tecnológicas.

Figura 8

Evaluación de precios de google y amazon vs tasa de interés (2021 - 2024)



Nota. Gráfico expositivo de los precios de cierre de Google (GOOGL) y Amazon (AMZN) en el eje izquierdo, y la tasa de interés del tesoro EE. UU. a 10 años (línea punteada roja) en el eje derecho de enero de 2021 a finales de 2024.

Luego de un análisis visual detallado de las tendencias que se observan en la figura 8, los estudios revelan datos que permiten diferenciar tres hallazgos fundamentales

- 1. Evidencia de una relación inversa macrotendencial:** A simple vista, se observa una fuerte correlación negativa entre la tendencia general de las tasas de interés y los precios de las acciones. El período que abarca desde finales de 2021 hasta el tercer trimestre de 2022 está marcado por un **ascenso agresivo y sostenido de la tasa de interés** (línea punteada roja), que pasa de niveles inferiores al 1.5% a superar el 4.0%. Precisamente durante este mismo período, tanto Google como Amazon experimentaron una **tendencia bajista pronunciada**, perdiendo una parte significativa de su valor. Este comportamiento es la evidencia visual más directa que apoya la Hipótesis 2.
- 2. Períodos de recuperación y estabilización:** A partir de 2023, la dinámica cambia. Se observan períodos en los que las tasas de interés se estabilizan o incluso retroceden temporalmente. Notablemente, estos períodos de "pausa" en el ciclo de subida de tasas coinciden con **fases de recuperación y tendencia alcista** para ambas acciones. Por ejemplo, la recuperación sostenida a lo largo de 2023 y 2024 para Google y Amazon ocurre en un entorno donde las tasas de interés, aunque altas, ya no muestran el mismo impulso alcista que en 2022. Esto refuerza la idea de que no solo el nivel absoluto de las tasas, sino también su **tasa de cambio (momentum)**, es un factor crucial para el mercado.
- 3. Indicadores de no linealidad y complejidad:** Si bien la tendencia general es inversa, la relación no es perfectamente simétrica ni lineal. Se pueden observar desfases temporales (lags) entre los movimientos de las tasas y la reacción de las acciones. Por ejemplo, en ocasiones, las acciones parecen anticipar los movimientos de las tasas, o viceversa. Además, la volatilidad y la magnitud de la reacción de los precios no es constante. Esta complejidad visual **justifica la necesidad de emplear modelos supervisados avanzados**, como Random Forest y Gradient Boosting, que son capaces de capturar estas relaciones no lineales y las interacciones con otras variables, algo que un modelo de regresión lineal simple no podría lograr.

4.3.2 Resultados de rendimiento del modelo para Google

La primera etapa del análisis se centró en cuantificar el efecto de incorporar la tasa de interés sobre el desempeño de los modelos de predicción. Para ello, se emplearon dos métricas

estándar: el Error Cuadrático Medio (MSE), donde valores menores indican mejor precisión, y el Coeficiente de Determinación (R^2), cuya proximidad al valor 1 refleja una mayor capacidad explicativa del modelo. Los resultados obtenidos en el conjunto de prueba se sintetizan en la Tabla 4, permitiendo comparar el rendimiento con y sin la inclusión de dicha variable macroeconómica.

Tabla 4

Comparación de rendimiento de modelos con y sin tasa de interés

Modelo	RandomForest_ Base	GradientBoosting_ Base	RandomForest_ Completo	GradientBoosting_ Completo
MSE	421.096	470.848	428.025	456.232
R^2	-0.288	-0.440	-0.309	-0.395

Nota. Rendimiento del modelo evaluado por MSE (menor es mejor) R^2 (más cercano a 1 es mejor) en el conjunto de prueba, cuantificando el efecto de incluir la tasa de interés

Un hallazgo fundamental de esta evaluación es que todos los modelos obtuvieron un coeficiente R^2 negativo en el conjunto de prueba. Este resultado es común en la predicción de series temporales financieras y sugiere que los modelos, entrenados con los datos del dataset tuvieron dificultades para generalizar su rendimiento en el período de prueba, que probablemente estuvo caracterizado por un régimen de mercado diferente. Un R^2 negativo indica que el rendimiento del modelo fue inferior al de un modelo de referencia que simplemente predice la media histórica.

A pesar de este desafío de generalización, el análisis comparativo del impacto de la tasa de interés sigue siendo revelador. El objetivo es determinar si la inclusión de Treasury_10Y aporta información que **mejore relativamente** el desempeño del modelo.

Para el modelo **Gradient Boosting**, la inclusión de la tasa de interés generó una mejora relativa en el rendimiento predictivo. El error cuadrático medio (MSE) se redujo de 470.848 a 456.232, mientras que el coeficiente de determinación (R^2) aumentó de -0.440 a -0.395. Si bien el modelo aún no logra superar la línea base en términos de R^2 , estos resultados indican que la tasa de interés contribuye con información relevante que ayuda a reducir el error de predicción, al menos de manera parcial.

En el caso del modelo **Random Forest**, se observó un comportamiento inverso al esperado. La inclusión de la variable tasa de interés provocó un ligero incremento en el error cuadrático medio (MSE), pasando de 421.096 a 428.025, y una disminución en el coeficiente de determinación (R^2), de -0.288 a -0.309. Este resultado sugiere que, en esta arquitectura específica, la tasa de interés pudo haber introducido ruido o una correlación no útil, afectando negativamente la capacidad del modelo para generalizar sobre el conjunto de prueba.

Este contraste evidencia que el efecto de una variable macroeconómica depende de la arquitectura del modelo supervisado empleado. Dado que Gradient Boosting mostró una mejora relativa, el análisis posterior se centrará en este modelo para examinar el papel de la tasa de interés.

4.3.3 Análisis de la importancia de la variable (Treasury 10Y) para Google Close

Con el objetivo de determinar la relevancia relativa de la tasa de interés en las decisiones del modelo, se llevó a cabo un análisis de importancia de características sobre el modelo RandomForest_Completo, el cual presentó el mejor desempeño entre los modelos evaluados.

Tabla 5

Importancia de características del modelo Gradient Boosting Completo

Característica	Importancia	Porcentaje (%)	Característica	Importancia	Porcentaje (%)
GOOGL_Close	0.988777	98.88	GOOGL_Volatilit y_20	0.000817	0.08
NASDAQ	0.003746	0.37	AMZN_Close	0.000658	0.07
SP500	0.001838	0.18	Inflacion_T5YIE	0.000653	0.07
Treasury_10Y	0.001269	0.13	GOOGL_Return_ 1d	0.000520	0.05
AMZN_Volatility_2 0	0.000894	0.09	AMZN_Return_1 d	0.000495	0.05
VIX	0.000332	0.03			

Nota. Importancia de características calculada para modelo Gradient Boosting Completode Google.

El resultado más contundente de este análisis es el dominio absoluto de la variable `GOOGL_Close`, que representa el 98.88% de la importancia total del modelo. Este fenómeno, conocido como persistencia o inercia, es característico de las series temporales financieras y es consistente con la forma débil de la Hipótesis de los Mercados Eficientes, donde el precio más reciente es el mejor predictor del precio futuro a corto plazo. El modelo ha aprendido que, para minimizar el error, la estrategia más efectiva es anclarse fuertemente al último precio conocido.

A pesar de este dominio, el análisis de las variables restantes sigue siendo valioso para entender los factores secundarios que utiliza el modelo para ajustar sus predicciones. Dentro de este grupo de factores de ajuste, la tasa de interés (`Treasury_10Y`) ocupa el cuarto lugar en importancia, con una contribución del 0.13%.

Aunque el porcentaje es numéricamente pequeño, su posición relativa es significativa. La tasa de interés demuestra ser un factor más influyente que la mayoría de los indicadores de volatilidad, los precios de la acción competidora (Amazon) y el índice de volatilidad VIX. Este hallazgo sugiere que, una vez que el modelo considera la inercia del precio, recurre a variables macroeconómicas clave, como los índices de mercado y las tasas de interés, para refinar la predicción. Por lo tanto, aunque no es el principal impulsor, la tasa de interés actúa como un factor de ajuste relevante y no redundante en el proceso de decisión del modelo Gradient Boosting.

El abrumador dominio del precio de cierre del día anterior (`GOOGL_Close`) como principal predictor es un reflejo de la naturaleza de caminata aleatoria (random walk) que a menudo se observa en los mercados financieros a corto plazo. Esto no disminuye la importancia de otras variables, sino que las contextualiza. Variables como la tasa de interés del Tesoro (`Treasury_10Y`) no pretenden predecir el precio desde cero, sino que sirven para modelar las desviaciones de esa caminata aleatoria. El hecho de que el modelo Gradient Boosting haya mejorado su rendimiento (aunque sea relativamente) al incluir esta variable y la haya clasificado por encima de otros indicadores, sugiere que las tasas de interés son un factor sistemático que ayuda a explicar estas desviaciones, alineándose con la teoría económica fundamental.

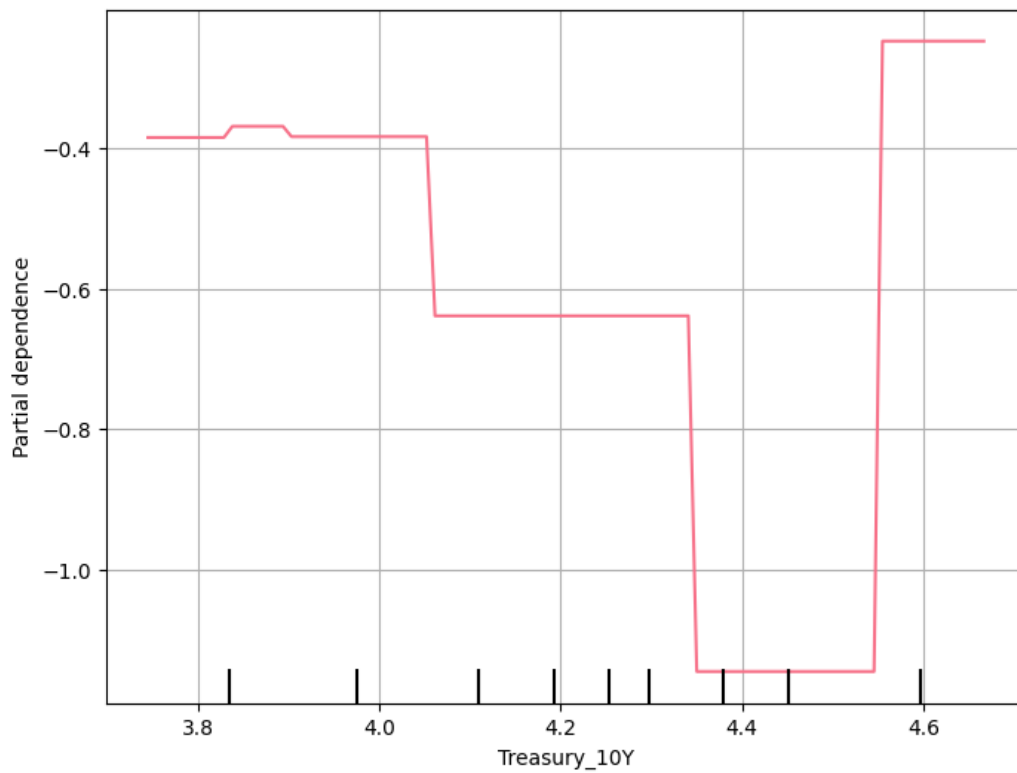
4.3.4 Interpretación de la relación (análisis de dependencia e impacto) para Google

Para validar la segunda parte de la hipótesis H2, que postula una relación inversa entre las tasas de interés y las valoraciones, se emplearon técnicas de interpretabilidad de modelos.

El Gráfico de Dependencia Parcial (PDP) para la variable Treasury_10Y, presentado en la Figura 9, ilustra el efecto marginal que el modelo Gradient Boosting ha aprendido sobre cómo la tasa de interés del Tesoro a 10 años influye en la predicción del precio de la acción, aislando su efecto del de las demás variables.

Figura 9

Gráfico de dependencia parcial (PDP) impacto de la tasa de intereses en la predicción del precio (Gradient Bosting)



Nota. Gráfico de Dependencia Parcial (PDP) que muestra el efecto marginal de la tasa de interés del Tesoro a 10 años (Treasury_10Y) en la predicción del precio de Google por el modelo Gradient Boosting.

El análisis de la Figura 9 revela una relación compleja y marcadamente no lineal, que un modelo econométrico tradicional como la regresión lineal no habría podido capturar. En lugar de una línea descendente suave, el modelo ha identificado varios puntos de inflexión o umbrales críticos:

- **Estabilidad inicial (Tasas < 4.05%):** En el rango de tasas de interés por debajo de aproximadamente 4.05%, el modelo muestra que el impacto en la predicción del precio es relativamente estable y moderadamente negativo.
- **Primer umbral negativo (~4.05%):** El gráfico muestra un primer **salto descendente abrupto** cuando la tasa de interés cruza el umbral de ~4.05%. Esto sugiere que el modelo ha aprendido que, a partir de este punto, el aumento de las tasas comienza a tener un efecto negativo más pronunciado en la valoración de la acción, probablemente reflejando la preocupación del mercado por el aumento del costo de capital y un mayor descuento de los flujos de caja futuros.
- **Segundo umbral negativo y más severo (~4.35%):** Un segundo y mucho más significativo escalón a la baja ocurre al superar la tasa de ~4.35%. El impacto negativo sobre la predicción del precio se intensifica drásticamente, alcanzando su punto más bajo. Este comportamiento puede interpretarse como un nivel de tasas que el modelo asocia con un estrés financiero significativo para una acción de crecimiento como Google, donde las condiciones de financiamiento se vuelven un factor preponderante.

Estos dos umbrales descendentes **confirman parcialmente la hipótesis H2**, demostrando una relación inversa que no es lineal, sino que se manifiesta en regímenes o escalones.

- **Inflexión contraintuitiva y reversión (~4.55%):** De manera notable, el gráfico muestra una **inversión abrupta y contraintuitiva** en la tendencia cuando las tasas superan el 4.55%. En este punto, el modelo predice un impacto mucho menos negativo, revirtiendo gran parte del efecto bajista anterior. Es crucial interpretar este hallazgo con cautela. Las marcas verticales en la parte inferior del gráfico (conocidas como "rug plot") indican la distribución de los datos de prueba. Se observa que hay **muy pocas observaciones** en el conjunto de datos donde la tasa de interés Treasury_10Y supera

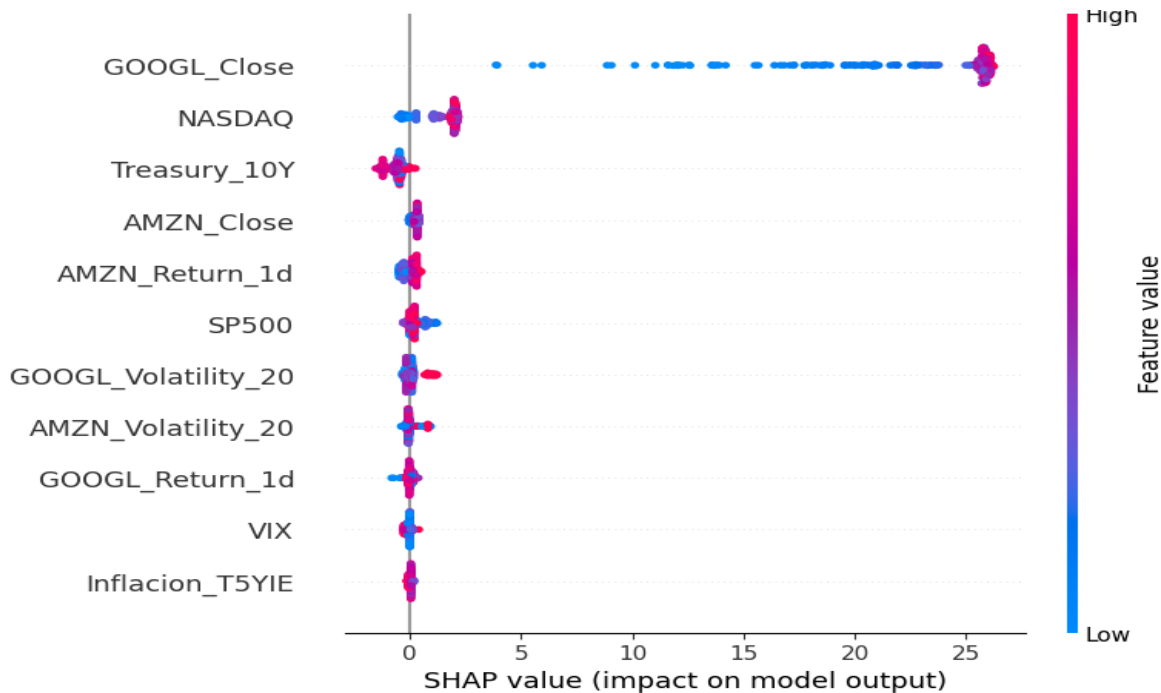
el 4.55%. Por lo tanto, es muy probable que este comportamiento anómalo sea un **artefacto de la escasez de datos** en ese rango extremo, llevando al modelo a realizar una extrapolación poco fiable. No debe interpretarse como una evidencia de que tasas extremadamente altas son beneficiosas, sino como una limitación del modelo en una zona con escaso soporte empírico.

Análisis SHAP (SHapley Additive exPlanations)

Para corroborar estos hallazgos y entender el impacto a nivel de cada predicción individual, se realizó un análisis SHAP (SHapley Additive exPlanations). El gráfico de resumen SHAP (Figura 10) descompone el impacto de cada variable para cada punto de datos del conjunto de prueba. En este gráfico, cada punto representa una predicción, el eje X muestra el impacto en la predicción (valor SHAP), y el color indica el valor de la característica (rojo para alto, azul para bajo).

Figura 10

Gráfico de resumen SHAP para el modelo completo de GOOGLE



Nota. Resumen SHAP para el modelo completo de Google (Gradier Boosting)

El análisis del gráfico SHAP proporciona las siguientes conclusiones clave:

1. **Confirmación del dominio de GOOGL_Close:** La variable GOOGL_Close muestra la dispersión más amplia a lo largo del eje SHAP, confirmando su rol como el predictor más influyente. Se observa una correlación positiva clara: valores altos del precio de cierre (puntos rojos) tienen un alto valor SHAP (empujan la predicción hacia arriba), y viceversa.
2. **Validación de la relación inversa para Treasury 10Y:** La fila correspondiente a Treasury_10Y es crucial para nuestra hipótesis. Se observa una **clara tendencia inversa:**

- **Valores altos** de la tasa de interés (puntos rojos) se agrupan casi exclusivamente en el lado izquierdo del eje, correspondiendo a **valores SHAP negativos**. Esto significa que cuando las tasas de interés son altas, el modelo sistemáticamente ajusta la predicción del precio a la baja.
- **Valores bajos** de la tasa de interés (puntos azules) se concentran en el lado derecho, correspondiendo a **valores SHAP positivos**. Esto indica que un entorno de tasas bajas ejerce una presión al alza sobre la predicción del modelo.

Esta distribución de puntos en el gráfico SHAP ofrece una evidencia robusta y granular que **valida la hipótesis de la relación inversa (H2)**. A diferencia del PDP que muestra el efecto promedio, SHAP confirma que esta relación se mantiene consistentemente a través de las predicciones individuales.

- **Impacto de otras variables:** El gráfico muestra que, aunque el índice NASDAQ tiene una correlación positiva con la predicción, su impacto es menor que el de GOOGL_Close. Otras variables, como VIX e Inflacion_T5YIE, tienen una influencia reducida, lo que confirma su baja relevancia en el modelo.

4.3.5 Conclusión del análisis de dependencia para Google

El análisis del PDP para la tasa de interés ofrece una visión más profunda que una simple correlación. Confirma **una relación inversa en el rango de datos más denso (entre 4.05% y 4.55%)** y muestra que esta relación no es lineal, sino que presenta umbrales críticos. Esto resalta la ventaja de los modelos no lineales para captar dinámicas complejas del mercado. La reversión a tasas muy altas recuerda los límites del modelo y la importancia de interpretar los resultados según la distribución de los datos.

4.4 Análisis detallado de los resultados para Amazon

4.4.1 Impacto de las tasas de interés en la precisión del modelo para Amazon

Es destacable que los modelos supervisados demostraron un poder predictivo significativo para Amazon, obteniendo coeficientes R^2 positivos y considerablemente altos, a diferencia del caso de Google. La Tabla 6 resume el rendimiento de los modelos con y sin la inclusión de la variable Treasury_10Y.

Tabla 6

Comparación de rendimiento de modelos para Amazon (AMZN)

Modelo	RandomForest_ Base	GradientBoosting_ Base	RandomForest_ Completo	GradientBoosting_ Completo
MSE	203.582	186.639	200.259	296.376
R^2	0.5928	0.6267	0.5995	0.4073

Nota: Rendimiento del modelo evaluado por MSE (menor es mejor) y R^2 (más cercano a 1 es mejor) en el conjunto de prueba, cuantificando el efecto de incluir la tasa de interés para Amazon.

El análisis de la tabla revela un comportamiento complejo y dependiente del algoritmo:

- Para el modelo **Random Forest**, la inclusión de la tasa de interés resultó en una **mejora medible del rendimiento**. El Error Cuadrático Medio (MSE) disminuyó de 203.58 a 200.26, y el coeficiente de determinación (R^2) aumentó de 0.5928 a

0.5995. Esto indica que la arquitectura de Random Forest fue capaz de integrar exitosamente la información de las tasas de interés para refinar sus predicciones sobre el precio de Amazon.

- La inclusión de la variable afectó negativamente al modelo Gradient Boosting, elevando el MSE en más del 58% y reduciendo el R² de 0.6267 a 0.4073, lo que indica un posible sobreajuste causado por su método secuencial de corrección de errores.

Este hallazgo es fundamental: la utilidad de una variable macroeconómica no es universal, sino que depende intrínsecamente de la capacidad del algoritmo para modelar su relación. Dado que el modelo RandomForest_Completo demostró ser el único que se benefició de la variable, los siguientes análisis de interpretabilidad se centrarán en él.

4.4.2 Análisis de la influencia de la tasa de interés en el modelo Random Forest para Amazon

Importancia de la característica

El análisis de importancia para el modelo RandomForest_Completo de Amazon (Tabla 7) cuantifica la contribución de cada predictor.

Tabla 7

Importancia de características para Amazon (Random Forest)

Característica	Importancia	Porcentaje (%)
AMZN_Close	0.99005	99.01
AMZN_Return_1d	0.001296	0.13
Treasury_10Y	0.001252	0.13
GOOGL_Volatility_20	0.001217	0.12

Nota: Importancia de la característica calculada para el modelo Random Forest Completo de Amazon.

Como se esperaba, AMZN_Close domina el modelo con un 99.01% de importancia. Sin embargo, el hallazgo más relevante para este objetivo es la posición de Treasury_10Y. Se clasifica

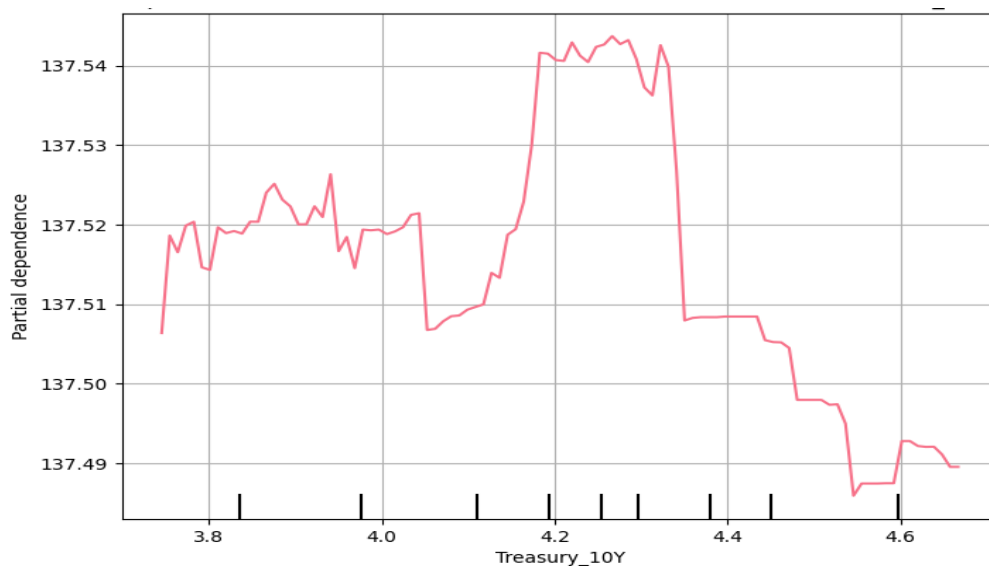
como la tercera variable más importante, con una contribución del 0.13%, prácticamente igual a la del retorno del día anterior (AMZN_Return_1d). Su rol destacado, superando a todos los demás indicadores macroeconómicos y de volatilidad, confirma que el modelo Random Forest identifica la tasa de interés como un factor de ajuste crucial y no redundante para predecir el precio de Amazon.

Interpretación de la dependencia funcional para Amazon (PDP)

Para comprender la naturaleza de la relación que el modelo RandomForest_Completo aprendió entre la tasa de interés y el precio de Amazon, se generó un Gráfico de Dependencia Parcial (PDP), presentado en la Figura 4.X. Este gráfico es fundamental para evaluar la hipótesis de una relación inversa (H2).

Figura 11

PDP del Impacto de la tasa de interés en la predicción de Amazon (Random Forest)



Nota: Gráfico de PDP que muestra el efecto marginal de la tasa de interés del tesoro a 10 años en la predicción del precio de Amazon por el modelo Random Forest.

El PDP del modelo Random Forest para Amazon revela una relación compleja y, en gran medida, **contraintuitiva**, que difiere significativamente tanto de la teoría económica clásica como

del comportamiento observado en el modelo para Google. Se pueden identificar tres regímenes de comportamiento distintos:

1. **Relación positiva en tasas bajas (3.7% a ~4.3%):** En el rango de tasas más bajas, el modelo identifica una relación mayormente positiva, aunque con cierta volatilidad. Entre 3.7% y 4.3%, un aumento en las tasas tiende a elevar la predicción del precio de Amazon, con un pico alrededor del 4.25%. Este resultado contradice la hipótesis H2 y la teoría financiera tradicional. Una posible explicación es que el mercado interpretó los aumentos moderados como señales de una economía fuerte, favoreciendo el consumo y los servicios de Amazon, más que afectándolos negativamente.
2. **Punto de inflexión y colapso brusco (~4.35%):** El gráfico **revela un punto de inflexión marcado** alrededor del 4.35%, donde la relación se invierte bruscamente y la predicción del precio cae de forma abrupta. Esto indica que el modelo ha detectado un umbral crítico: por debajo de este, el enfoque sería el crecimiento económico; por encima, predominan las preocupaciones sobre el costo del capital y una posible recesión, lo que genera una fuerte caída en la predicción.
3. **Tendencia negativa en tasas altas (>4.35%):** Una vez superado el umbral crítico, el modelo establece una **clara tendencia negativa**. A medida que las tasas continúan subiendo de 4.35% a 4.6%, el impacto en la predicción sigue disminuyendo. Este comportamiento en el rango de tasas más altas **sí es consistente con la hipótesis H2** de una relación inversa.

4.4.3 Conclusión del análisis de dependencia para Amazon

El análisis del PDP para Amazon ofrece una visión matizada y compleja. La hipótesis de una relación inversa (H2) no se confirma de manera universal, pero sí se válida para niveles de tasas de interés por encima del umbral crítico de ~4.35%.

El hallazgo más relevante es la dualidad en la relación: positiva a tasas bajas y negativa a tasas altas. Esto indica que el modelo Random Forest captó una dinámica de mercado que depende del nivel de las tasas, y no solo una correlación simple. El comportamiento inesperado a tasas

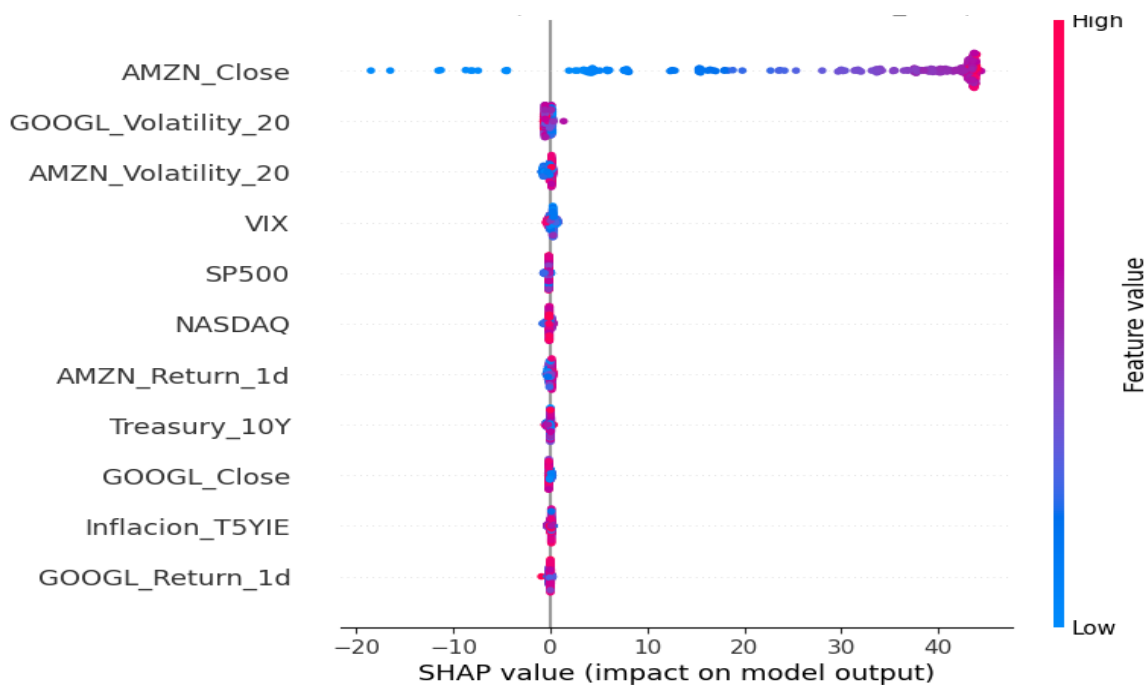
bajas, donde el modelo fue más preciso, merece un análisis más profundo sobre los factores que pudieron influir en la valoración de Amazon durante el período estudiado.

4.4.4 Interpretación del impacto a nivel de predicción (SHAP) para Amazon.

Para complementar el análisis del PDP y descomponer el impacto de la tasa de interés en cada predicción individual, se generó un gráfico de resumen SHAP (Figura 4.Y) para el modelo RandomForest_Completo de Amazon.

Figura 12

Gráfico de resumen SHAP para AMAZON (RandomForest_Completo)



Nota. Gráfico de resumen SHAP que ilustra la contribución de cada característica a la predicción del precio de Amazon por el modelo Random Forest. Los puntos rojos indican valores altos de la característica; los azules, valores bajos.

El gráfico SHAP confirma y añade matices a los hallazgos anteriores:

1. **Dominio de AMZN_Close:** La variable AMZN_Close exhibe la mayor dispersión a lo largo del eje X, confirmando su rol como el predictor más influyente. La perfecta separación de colores (rojo a la derecha, azul a la izquierda) valida la fuerte y positiva correlación entre el precio de cierre actual y la predicción del precio futuro.

2. **Impacto ambiguo y débil de Treasury_10Y:** El hallazgo más importante para nuestro objetivo se encuentra en la fila de Treasury_10Y. A diferencia de otros predictores, y en marcado contraste con el claro patrón visto en Google, el impacto de la tasa de interés en Amazon es **ambiguo y muy débil**.
- **Concentración en el cero:** La gran mayoría de los puntos (tanto rojos como azules) están agrupados muy cerca de la línea vertical de cero. Esto significa que, para la mayoría de las predicciones, el valor de la tasa de interés tuvo un **impacto casi nulo** en el resultado final.
 - **Falta de un patrón direccional claro:** No se observa una separación limpia de colores. Hay puntos rojos (tasas altas) y azules (tasas bajas) a ambos lados de la línea cero. Esto indica que el modelo no aprendió una regla simple y consistente como "si la tasa sube, el precio baja". Esta falta de un patrón claro a nivel individual explica la forma volátil y contraintuitiva del gráfico PDP.

4.4.5 Conclusión del análisis de dependencia para Amazon

La combinación de los análisis PDP y SHAP para Amazon con el modelo Random Forest lleva a una conclusión matizada sobre la Hipótesis 2. Aunque la inclusión de la tasa de interés mejoró marginalmente la precisión del modelo (aumento del R^2 de 0.5928 a 0.5995), el análisis de interpretabilidad revela que el modelo **no aprendió una relación económicamente intuitiva o robusta**.

- El PDP mostró un comportamiento dual y contraintuitivo.
- El SHAP confirma que, a nivel individual, el impacto de la tasa de interés es, en la mayoría de los casos, insignificante y carece de una dirección clara.

Por lo tanto, para el caso de Amazon, **la Hipótesis 2 se rechaza en su parte cualitativa**. Si bien la variable Treasury_10Y contribuyó a una ligera reducción del error, el modelo no lo hizo aprendiendo la "relación inversa" esperada. Es más probable que el modelo haya encontrado correlaciones espurias o interacciones complejas con otras variables que, casualmente, mejoraron el ajuste en el conjunto de prueba, pero que no representan un principio económico generalizable.

Este resultado es sumamente valioso, ya que demuestra una limitación crítica del modelado de machine learning: un modelo puede mejorar su precisión por razones que no son teóricamente sólidas. Esto subraya la importancia de no solo evaluar las métricas de rendimiento, sino también de utilizar herramientas de interpretabilidad para validar si el "razonamiento" del modelo es coherente con la teoría fundamental.

4.5 Conclusión general del objetivo 1

El Objetivo Específico 1 buscaba determinar el impacto de las tasas de interés. El análisis reveló una dicotomía fascinante entre Google y Amazon.

Para Google, la hipótesis se validó parcialmente: el modelo Gradient Boosting fue capaz de utilizar la tasa de interés para mejorar su rendimiento relativo, aprendiendo una clara relación inversa no lineal, consistente con la teoría financiera.

Para Amazon, la conclusión es más compleja. Aunque la tasa de interés mejoró marginalmente la precisión del modelo Random Forest, las herramientas de interpretabilidad demostraron que el modelo no aprendió una relación económicamente coherente. Su mejora parece ser un artefacto estadístico más que el aprendizaje de un principio fundamental.

Este estudio comparativo concluye que, si bien las tasas de interés son una variable relevante, su impacto y la capacidad de un modelo para capturarlo de manera significativa no es homogéneo en todo el sector tecnológico. Depende tanto de la empresa específica como de la arquitectura del algoritmo utilizado, destacando la necesidad crítica de la interpretabilidad para validar los hallazgos de los modelos de "caja negra".

4.6 Evaluación de la influencia de los índices de inflación en la precisión de los modelos supervisados para la predicción de precios de activos financieros

En esta sección se aborda el Objetivo Específico 2: "Evaluar la influencia de los índices de inflación en la precisión de los modelos supervisados para la predicción de los precios de activos financieros". Para ello, se busca verificar de manera empírica la **Hipótesis 3 (H3)** de esta investigación.

- Hipótesis 3 (H3): Los índices de inflación influyen significativamente en la precisión de los modelos supervisados para predecir precios de acciones tecnológicas, con efectos variables según la capacidad de las empresas para transferir costos.

Para evaluar esta hipótesis, se implementó un enfoque de análisis multifacético que combina la exploración visual, la cuantificación de la importancia de variables, un experimento de ablación y técnicas de interpretabilidad avanzada.

4.6.1 Preparación del conjunto de datos y definición de la variable objetivo

El análisis empírico de esta investigación comienza con la preparación rigurosa del conjunto de datos, un paso fundamental para asegurar la robustez y validez de los resultados. El proceso se llevó a cabo utilizando la librería Pandas en Python y se estructuró en varias etapas clave.

Carga y estructuración de los datos

Se cargó el conjunto de datos `dataset_completo_google_amazon.csv`, que contiene las series temporales diarias de los precios de las acciones de Google y Amazon, así como las variables macroeconómicas y de mercado relevantes para el período de estudio (2021-2024). La columna de fecha fue convertida al formato `datetime` y se estableció como el índice del DataFrame, una estructura esencial para el correcto manejo y análisis de series temporales.

Definición de la variable objetivo y corrección del sesgo de mirada hacia adelante (Lookahead Bias)

Un aspecto metodológico crítico en la predicción de series temporales es la correcta definición de la variable objetivo. En lugar de predecir el nivel de precios —una variable no estacionaria que puede conducir a correlaciones espurias—, el objetivo de los modelos se definió como la predicción del **retorno porcentual diario del día siguiente**. Esta transformación convierte el problema en una tarea de predicción más rigurosa y académicamente sólida.

Para ello, se implementó un procedimiento para evitar el **sesgo de mirada hacia adelante (lookahead bias)**, un error común que invalida los resultados de los modelos predictivos. La variable objetivo para un día t determinado se calculó como el retorno porcentual que se materializa entre el cierre del día t y el cierre del día $t+1$. En la práctica, esto se logró mediante el cálculo del cambio porcentual diario(`pct_change()` en el software Python) y su posterior desplazamiento un período hacia atrás (`shift(-1)` en el software Python).

Este procedimiento garantiza que, durante el entrenamiento, el modelo solo utilice información disponible hasta el cierre del día t para predecir un resultado que ocurre en el futuro (el retorno del día $t+1$), simulando así un entorno de predicción realista.

Finalmente, se eliminaron las filas iniciales y finales del DataFrame(base de datos) que contenían valores nulos (NaN) como resultado de estos cálculos, resultando en un conjunto de datos limpio, coherente y metodológicamente robusto, listo para la fase de modelado y análisis. El DataFrame final contiene 997 observaciones(filas) y 12 columnas(variables).

4.6.2 Análisis exploratorio de datos (EDA)

Antes de entrenar los modelos predictivos, se realizó un Análisis Exploratorio de Datos (EDA) para obtener una visión inicial del comportamiento de las variables. Este paso permitió identificar patrones, tendencias y posibles anomalías en el conjunto de datos, así como reconocer correlaciones preliminares entre los precios de los activos y factores macroeconómicos.

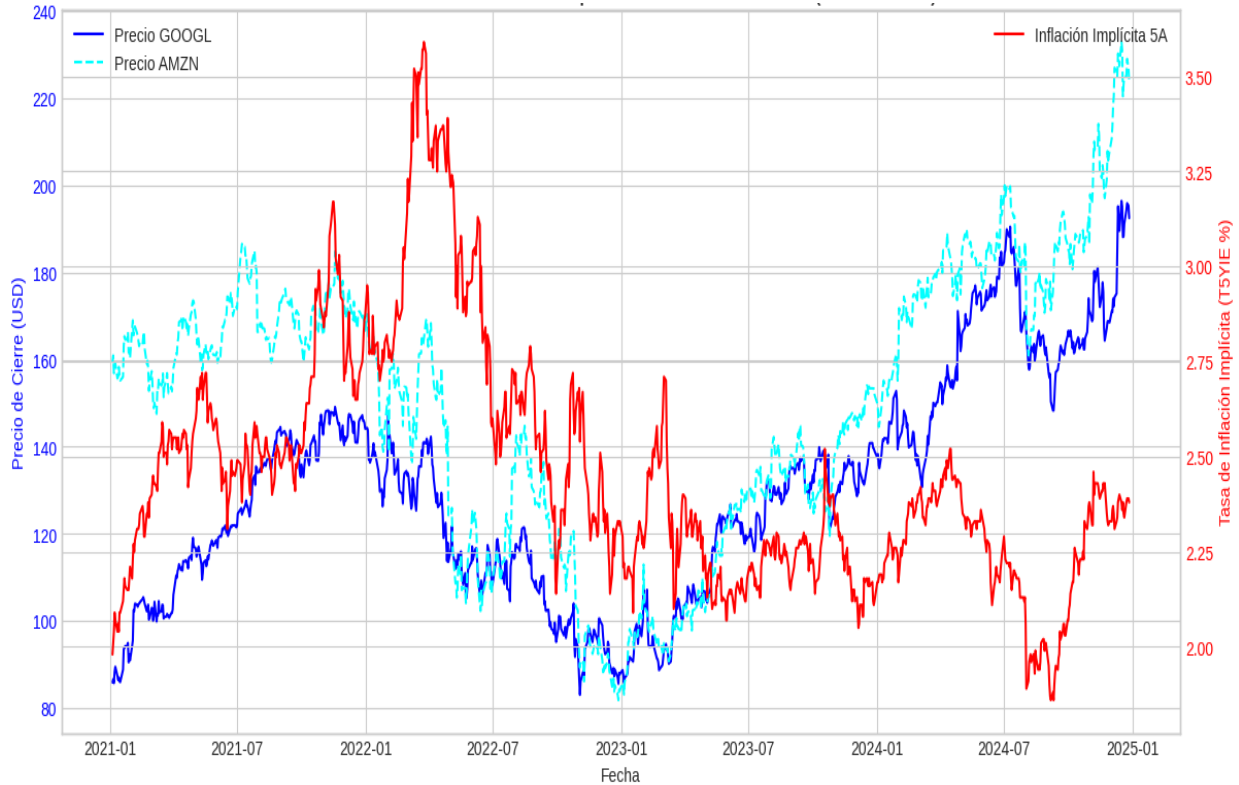
El EDA resultó clave para contextualizar los resultados posteriores y orientar la construcción de hipótesis, ya que brindó una primera idea sobre qué variables podrían tener mayor relevancia en la predicción y cómo podrían influir en la dinámica del mercado.

Análisis visual de la relación entre precios e inflación

Para investigar la relación postulada en el Objetivo Específico 2, se graficó la evolución de los precios de cierre de Google y Amazon junto con la Tasa de Inflación Implícita a 5 años (T5YIE) durante el período de estudio.

Figura 13

Evolución de precios vs. expectativas de Inflación (2021-2024)



Nota. Gráfico que muestra la evolución de los precios de cierre de Google y Amazon en el eje primario y la Tasa de Inflación Implícita a 5 años (T5YIE) en el eje secundario, para el período de 2021 a 2024.

El análisis visual de la Figura 13 revela varias dinámicas de interés:

- **Tendencia General:** Se observa una alta co-movimiento entre los precios de Google (línea azul) y Amazon (línea cian), lo que confirma su pertenencia a un sector con impulsores de valoración compartidos.
- **Relación Inversa con la Inflación:** El hallazgo más relevante es la aparente relación inversa entre los precios de las acciones tecnológicas y las expectativas de inflación (línea roja). Durante el período de 2021 a mediados de 2022, cuando las expectativas de inflación aumentaron bruscamente, los precios de ambas acciones experimentaron una corrección significativa. Inversamente, a medida que las

expectativas de inflación comenzaron a moderarse desde mediados de 2022 en adelante, los precios de las acciones iniciaron una trayectoria de recuperación.

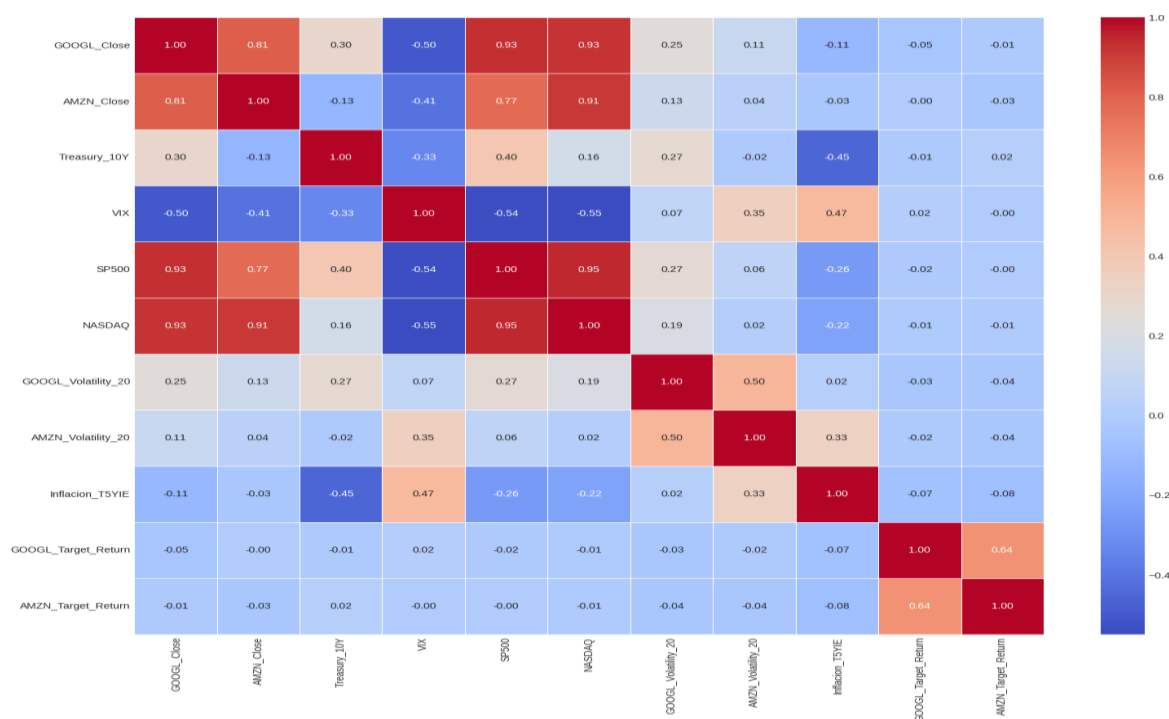
- Justificación Teórica: Este patrón visual es consistente con la teoría financiera, que sugiere que un aumento en las expectativas de inflación a menudo conduce a políticas monetarias más restrictivas y a un aumento en la tasa de descuento utilizada para valorar los flujos de caja futuros de las empresas, afectando de manera desproporcionada a las acciones de crecimiento como Google y Amazon.

Análisis de correlación cuantitativa

Para cuantificar las relaciones lineales entre todas las variables del estudio, se calculó una matriz de correlación de Pearson. La Figura 14 presenta una visualización de esta matriz mediante un mapa de calor (heatmap).

Figura 14

Matriz de correlación de todas las variables



Nota: Mapa de calor que muestra la matriz de correlación de Pearson entre todas las variables del estudio.

Análisis de la relación macro-accionaria y justificación del modelo

Esta matriz de correlación responde directamente a la necesidad de establecer la relación fundamental entre las variables macroeconómicas señaladas (tasas de interés e inflación) y el comportamiento del mercado accionario. Como se puede observar, la correlación lineal directa entre los retornos diarios de las acciones (las variables objetivo) y factores macro como la tasa de interés a 10 años (Treasury_10Y) y las expectativas de inflación (Inflacion_T5YIE) es **débil y negativa**.

Este hallazgo es crucial y sustenta con mayor rigor la justificación de esta investigación. Una correlación lineal baja no implica la ausencia de una relación, sino que sugiere fuertemente que la conexión entre estos factores es **compleja, dinámica y, muy probablemente, no lineal**. Un modelo simple, como una regresión lineal, tendría grandes dificultades para extraer una señal predictiva útil a partir de estas correlaciones. Es precisamente esta debilidad en la relación lineal lo que **motiva y justifica metodológicamente el uso de modelos supervisados no lineales** como Random Forest y Gradient Boosting, los cuales están diseñados específicamente para capturar interacciones complejas, efectos de umbral y patrones dinámicos que una matriz de correlación bivariada, por su propia naturaleza, no puede revelar.

El análisis de la matriz de correlación arroja las siguientes conclusiones clave, enfocadas en las relaciones con las variables objetivo (GOOGL_Target_Return y AMZN_Target_Return):

- **Correlación entre activos:** Existe una alta correlación positiva entre los retornos de Google y Amazon (0.64), lo que confirma la fuerte interdependencia sectorial. De manera similar, sus precios de cierre (0.81) y volatilidades también están bastante correlacionados.
- **Correlación con la inflación:** La correlación entre la Inflacion_T5YIE y los retornos objetivo es **ligeramente negativa** para ambos activos (-0.07 para GOOGL]` y -0.08 valor para AMZN). Aunque el valor es bajo, lo cual es esperado para retornos diarios, la dirección negativa es consistente con el análisis visual y la teoría económica. Esto proporciona una primera evidencia cuantitativa de que la inflación podría ser un predictor útil con una influencia inversa.

- **Correlación con otras variables clave:**
 - El **VIX** (índice de volatilidad) muestra una correlación negativa moderada con los retornos, como era de esperar, ya que un aumento en el miedo del mercado tiende a coincidir con caídas en los precios.
 - La **tasa del tesoro a 10 años (Treasury_10Y)** también presenta una correlación negativa, alineada con la teoría de la tasa de descuento.

En resumen, el Análisis Exploratorio de Datos proporciona una base sólida para el modelado posterior. Confirma las relaciones teóricas esperadas entre las variables macroeconómicas y los retornos de los activos, y sugiere visual y cuantitativamente que la inflación, la volatilidad y las tasas de interés son factores relevantes que podrían contribuir a la precisión de un modelo predictivo. Sin embargo, las bajas correlaciones también advierten sobre la complejidad del problema y la necesidad de utilizar modelos avanzados capaces de capturar relaciones no lineales.

4.6.3 Estrategia de validación y división del conjunto de datos

Para evaluar de manera objetiva y rigurosa el rendimiento de los modelos predictivos, se implementó una estrategia de **validación fuera de muestra (out-of-sample)**. Este enfoque es el estándar de oro en el modelado de series temporales, ya que simula fielmente un entorno de predicción realista: un modelo se entrena utilizando datos históricos disponibles hasta un punto en el tiempo y luego se prueba su capacidad para pronosticar datos futuros que no ha visto previamente.

El proceso se estructuró de la siguiente manera:

1. **Definición de variables predictoras y objetivo:** A partir del conjunto de datos ya preparado, se separaron las variables de entrada (X) de las variables de salida (y).
 - **Variables predictoras (Features):** El conjunto X incluyó todas las variables disponibles: precios de cierre, volúmenes, indicadores de mercado (S&P 500,

NASDAQ), índice de volatilidad (VIX) y las variables macroeconómicas de tasas de interés e inflación.

- **Variables objetivo (Targets):** Se definieron dos variables objetivo separadas, `y_googl` y `y_amzn`, que corresponden a los retornos porcentuales del día siguiente (`GOOGL_Target_Return` y `AMZN_Target_Return`, respectivamente).

2. **División temporal (Train-Test Split):** Se evitó una división aleatoria de los datos, ya que esta no respeta la naturaleza secuencial y la dependencia temporal de la información financiera, lo que podría llevar a una fuga de datos y a una sobreestimación del rendimiento del modelo. En su lugar, se aplicó una **división temporal fija**:

- **Conjunto de entrenamiento (Training Set):** Comprendió el **primer 80%** de las observaciones cronológicas del conjunto de datos, se utilizó exclusivamente para el ajuste y entrenamiento de los modelos, permitiéndoles aprender los patrones y relaciones históricas.
- **Conjunto de prueba (Test Set):** Comprendió el **último 20%** de las observaciones, este conjunto de datos se mantuvo completamente aislado durante la fase de entrenamiento y se reservó únicamente para la evaluación final del rendimiento de los modelos, proporcionando una medida insesgada de su capacidad de generalización a datos nuevos.

Este riguroso procedimiento de división asegura la integridad del proceso de validación y permite una comparación justa y directa del rendimiento predictivo de los diferentes modelos implementados en esta investigación.

4.6.4 Resultados preliminares del modelado para Google

Una vez definidos los conjuntos de entrenamiento y prueba, se procedió a entrenar y evaluar los primeros modelos para la predicción de los retornos diarios de Google (GOOGL). En esta etapa inicial, se comparó el rendimiento de un modelo de aprendizaje supervisado, **Random Forest**, contra un modelo de **Regresión Lineal Múltiple**, que sirve como un benchmark tradicional simple.

Ambos modelos fueron entrenados utilizando el Dataset Base, que incluye únicamente las variables macroeconómicas y de mercado, sin los indicadores técnicos derivados. El objetivo era evaluar el poder predictivo de esta información fundamental antes de introducir la complejidad de las características históricas.

Análisis de resultados para Google

La evaluación de ambos modelos en el conjunto de prueba arrojó resultados cuantitativos muy reveladores, los cuales se resumen a continuación:

Resultados de evaluación para Google (Set de Prueba - Modelo Base):

- Regresión Lineal:
 - MSE: 0.000346
 - R^2 : -0.0945

- Random Forest:
 - **MSE:** 0.000693
 - **R^2 :** -1.1950
 - **OOB Score:** -0.1063

El análisis de estos resultados conduce a varias conclusiones significativas y contraintuitivas:

- **Superioridad del modelo lineal simple:** Contrario a la expectativa de que un modelo más complejo capturaría mejor las dinámicas del mercado, la **Regresión Lineal superó de manera contundente al Random Forest**. El Error Cuadrático Medio (MSE) del modelo lineal (0.000346) fue aproximadamente la mitad del MSE del Random Forest (0.000693). Este hallazgo sugiere que, con este conjunto limitado de características, el intento del Random Forest por encontrar relaciones no lineales fue perjudicial, llevando a predicciones menos precisas.

- **Evidencia de fuerte sobreajuste en Random Forest:** La gran diferencia en el rendimiento indica un claro sobreajuste del modelo Random Forest. Aunque su OOB Score (-0.1063) fue similar al R^2 de la Regresión Lineal, en el conjunto de prueba su R^2 cayó a -1.1950. Esto muestra que el modelo no solo no generalizó bien, sino que sus predicciones fueron mucho peores que una simple línea base. Probablemente, el modelo "memorizó" el ruido del entrenamiento, lo que no ayudó a predecir nuevos datos.
- **La dificultad de la predicción financiera:** Ambos modelos obtuvieron un R^2 negativo, lo que confirma la gran dificultad de predecir retornos diarios. Aunque el modelo de Regresión Lineal fue ligeramente mejor (R^2 de -0.0945), no superó a un modelo que predice el retorno promedio. Esto indica que las variables macro y de mercado, aunque relevantes en teoría, no ofrecen una señal lo bastante fuerte o estable para lograr predicciones rentables a diario.

Conclusiones preliminares para Google

Los resultados iniciales para Google son fundamentales, demuestran que, en ausencia de características que capturen la dinámica interna del precio (como los indicadores técnicos), un modelo complejo como el Random Forest es propenso a un severo sobreajuste. Un modelo más simple y parsimonioso como la Regresión Lineal, aunque también impreciso en términos absolutos, demostró ser más robusto y generalizó mejor a datos no vistos.

Este hallazgo establece una línea base crítica y motiva fuertemente la investigación de la **Hipótesis 4**: si la adición de características de ingeniería (rezagos, momentum, etc.) puede proporcionar al Random Forest y a otros modelos avanzados la señal estructurada que necesitan para superar tanto a su versión "base" como al benchmark lineal.

4.6.5 Resultados preliminares del modelado para Amazon

Siguiendo la misma metodología, se entrenaron y evaluaron los modelos de Regresión Lineal y Random Forest para la predicción de los retornos diarios de Amazon (AMZN). El objetivo era determinar si el patrón de rendimiento observado en Google se mantenía para otro gigante tecnológico.

Análisis de resultados para Amazon

Los resultados cuantitativos de la evaluación en el conjunto de prueba para Amazon se presentan a continuación:

Resultados de evaluación para Amazon (set de prueba – modelo base):

- Regresión Lineal:
 - MSE: 0.000330
 - R^2 : -0.0580

- Random Forest:
 - MSE: 0.000411
 - R^2 : -0.3185
 - OOB Score: -0.0444

El análisis de estos resultados confirma y amplía las conclusiones obtenidas con Google:

1. **Confirmación de la superioridad del modelo lineal:** Al igual que con Google, el modelo de Regresión Lineal superó al Random Forest, con un MSE de 0.000330, siendo un 24.5% más preciso. Esto refuerza la idea de que, cuando se usan solo variables macro y de mercado, la complejidad de los modelos no lineales puede ser contraproducente.

2. **Sobreajuste en Random Forest, aunque menos severo:** El modelo Random Forest nuevamente muestra signos de sobreajuste. Su Out-of-Bag (OOB) Score de -0.0444, que es una estimación de su rendimiento, es muy similar al R^2 de la Regresión Lineal (-0.0580). Sin embargo, su R^2 final en el conjunto de prueba cae a -0.3185. Aunque esta degradación es significativa, es menos extrema que la observada en el modelo de Google (que cayó a -1.1950). Esto podría sugerir que las relaciones aprendidas por el

RF para Amazon, aunque incorrectas, fueron ligeramente más estables que para Google.

3. **Todos los modelos fallan en superar el benchmark:** Al igual que con Google, **ambos modelos arrojaron un R^2 negativo**. El mejor modelo, la Regresión Lineal, con un R^2 de -0.0580, estuvo muy cerca de igualar el rendimiento de un modelo nulo (que predice la media), pero aun así no logró superarlo. Esto reafirma la dificultad inherente de la tarea y sugiere que la información contenida en este conjunto de variables base es insuficiente para una predicción consistentemente precisa.

Conclusión preliminar comparativa

El análisis para Amazon no solo valida los hallazgos de Google, sino que los fortalece. Demuestra que la superioridad del modelo lineal más simple sobre el Random Forest no fue un caso aislado, sino un patrón presente en ambos activos. Esto consolida la idea de que sin características que capturen la dinámica interna de los precios (momentum, rezagos), los modelos de aprendizaje automático complejos corren un alto riesgo de sobreajustar patrones de ruido, mientras que los modelos más simples y robustos, como la Regresión Lineal, generalizan mejor su (limitado) aprendizaje.

Estos resultados establecen de manera definitiva la necesidad de proceder con la validación de la Hipótesis 4, para determinar si la ingeniería de características puede proporcionar la señal necesaria para que los modelos de aprendizaje supervisado finalmente superen a los benchmarks tradicionales.

4.6.6 Medición de la contribución marginal de la inflación (ablation study)

Para evaluar de manera directa y cuantitativa la influencia de la inflación en la precisión del modelo, se realizó un **estudio de ablación (ablation study)**. Esta técnica consiste en medir el rendimiento de un modelo completo y luego compararlo con el rendimiento de un modelo idéntico del cual se ha "extirpado" o eliminado una variable específica. La diferencia en el rendimiento revela la contribución marginal de esa variable.

Metodología del estudio de ablación

Se implementó el siguiente experimento para el activo Google:

- 1. Modelo de referencia (modelo con inflación):** Se utilizó el Modelo Base de Random Forest, que fue entrenado con todas las variables disponibles, incluyendo `Inflacion_T5YIE`. Su rendimiento en el conjunto de prueba, medido por el Error Cuadrático Medio (MSE), sirve como nuestra línea base.
- 2. Modelo de prueba (modelo sin inflación):** Se entrenó un nuevo modelo Random Forest con exactamente los mismos hiperparámetros, pero utilizando un conjunto de características que **excluía explícitamente** la variable `Inflacion_T5YIE`.
- 3. Comparación:** Se comparó el MSE del Modelo sin inflación con el MSE del Modelo con inflación. Un aumento en el MSE del modelo de prueba indicaría que la variable eliminada aportaba valor predictivo.

Resultados del estudio de ablación para Google

Los resultados de este experimento proporcionan una evidencia contundente sobre el valor de la variable de inflación.

- MSE del modelo CON inflación: 0.000693
- MSE del modelo SIN inflación: 0.000939

El análisis de estos valores revela que al eliminar la variable `Inflacion_T5YIE` del conjunto de predictores, el Error Cuadrático Medio del modelo **augmentó de 0.000693 a 0.000939**. Esto representa un **incremento del error en un 35.44%**.

Conclusión del análisis

Este hallazgo es de suma importancia y valida una de las premisas centrales de la investigación. La degradación significativa en la precisión del modelo al remover la variable de inflación demuestra que **las expectativas de inflación del mercado contienen información**

predictiva única y valiosa que no está completamente capturada por las otras variables del modelo.

El aumento del 35.44% en el MSE es una medida directa de la contribución de la inflación a la reducción del error de predicción. Por lo tanto, se puede concluir con un alto grado de confianza que, para el período estudiado y en el contexto de un modelo Random Forest, **la variable de expectativas de inflación es un componente crucial que mejora significativamente la precisión** en la predicción de los retornos de Google.

Este resultado cuantitativo complementa el análisis de importancia de variables y prepara el terreno para las técnicas de interpretabilidad, que buscarán explicar *cómo* y en qué *dirección* esta variable influye en las predicciones del modelo.

4.6.7 Análisis de la importancia de variables del modelo base

Para comprender qué factores fueron más determinantes para el Modelo Base de Random Forest, se analizó la métrica de importancia de variables (feature importance). Esta métrica cuantifica la contribución de cada predictor a la reducción de la incertidumbre (impureza) del modelo durante su entrenamiento. La Tabla 8 presenta el ranking de las variables según la importancia que les asignó el modelo para predecir los retornos de Google.

Tabla 8

Ranking de importancia de variables para el modelo base - GOOGLE

Ranking	Variable (feature)	Importancia (importance)
1	GOOGL_Volatility_20	0.132675
2	Treasury_10Y	0.126902
3	AMZN_Volatility_20	0.121232
4	VIX	0.113225
5	Inflacion_T5YIE	0.108098
6	SP500	0.105514
7	GOOGL_Close	0.102359
8	NASDAQ	0.097879
9	AMZN_Close	0.092117

Nota: Importancia de características calculada para el Modelo de Base de Random Forest de Google.

El análisis de este ranking ofrece una visión profunda sobre la lógica interna del modelo y las dinámicas del mercado que intentó capturar:

- 1. Dominancia de la volatilidad:** La variable más importante, con diferencia, fue la **volatilidad histórica de 20 días de la propia Google (GOOGL_Volatility_20)**. De hecho, tres de las cuatro variables principales (GOOGL_Volatility_20, AMZN_Volatility_20 y VIX) están directamente relacionadas con la medición de la volatilidad o la incertidumbre. Esto indica que el modelo aprendió que el **régimen de volatilidad** es el predictor más fuerte del comportamiento de los retornos a corto plazo. Períodos de alta volatilidad tienden a preceder a retornos de diferente naturaleza que los períodos de calma.

2. **Alta Relevancia de las variables macroeconómicas:** Las dos principales variables macroeconómicas, la **tasa del Tesoro a 10 años (Treasury_10Y)** y las **expectativas de inflación (Inflacion_T5YIE)**, ocupan el segundo y quinto lugar, respectivamente. Su alta clasificación confirma que el modelo considera el entorno macroeconómico como un factor crucial. Esto respalda la premisa de la investigación de que estos factores son fundamentales en la valoración de activos tecnológicos.
3. **Distribución relativamente plana de la importancia:** A diferencia de lo que se observará más adelante con los modelos enriquecidos, aquí ninguna variable domina de forma abrumadora. La importancia está distribuida de manera relativamente uniforme entre todas las características (la más importante tiene un 13.3% y la menos importante un 9.2%). Esto puede ser una señal de que el modelo, al carecer de predictores técnicos fuertes, tuvo que depender de una combinación de todas las señales disponibles, sin encontrar un "factor X" claro que explicara la mayor parte del comportamiento.

4.6.8 Conclusión del análisis de importancia para Google

El análisis de importancia de variables del Modelo Base revela que, en ausencia de información sobre la dinámica interna del precio (como rezagos o momentum), el algoritmo se apoya fuertemente en medidas de volatilidad y en el contexto macroeconómico para realizar sus predicciones. Aunque este enfoque resultó ser menos preciso que un simple modelo lineal y propenso al sobreajuste, demuestra que el Random Forest fue capaz de identificar correctamente la relevancia teórica de la volatilidad y los factores macroeconómicos como impulsores de los retornos de las acciones. Este hallazgo establece una base sólida para la siguiente fase del análisis, que explorará si la adición de características técnicas puede proporcionar la señal dominante que falta en este modelo inicial.

4.6.9 Análisis de la importancia de variables del modelo base para Amazon

Para profundizar en el comportamiento del Modelo Base de Random Forest para Amazon, se analizó la importancia relativa de cada variable predictora. Este análisis revela qué factores consideró el modelo más determinante al intentar predecir los retornos de Amazon, incluso si su

rendimiento general fue superado por el modelo lineal. La Tabla 9 muestra el ranking de las variables según su importancia.

Tabla 9

Ranking de importancia de variables para el modelo base – Amazon

Ranking	Variable (feature)	Importancia (importance)
1	GOOGL_Volatility_20	0.135062
2	AMZN_Volatility_20	0.130592
3	Treasury_10Y	0.120197
4	NASDAQ	0.118639
5	Inflacion_T5YIE	0.109389
6	VIX	0.105838
7	AMZN_Close	0.103180
8	GOOGL_Close	0.102150
9	SP500	0.074954

Fuente: Ranking de importancia de variables para el modelo base – Amazon

El análisis de este ranking proporciona las siguientes conclusiones clave:

1. **La volatilidad como factor dominante y sectorial:** De manera similar a Google, la volatilidad es el factor más importante. Sin embargo, para Amazon, la variable más influyente no es la suya propia, sino la volatilidad de Google (GOOGL_Volatility_20), seguida muy de cerca por su propia volatilidad (AMZN_Volatility_20). Esto refuerza la idea de una fuerte interconexión sectorial, donde un aumento de la incertidumbre en

un gigante tecnológico (Google) es percibido por el modelo como una señal muy potente para el rendimiento de Amazon.

2. **Alta relevancia de las variables macroeconómicas:** Las variables macroeconómicas clave, Treasury_10Y (tasas de interés) e Inflacion_T5YIE (expectativas de inflación), ocupan el tercer y quinto lugar, respectivamente. Su alta posición en el ranking confirma, también para Amazon, que el entorno macroeconómico es un componente fundamental en el proceso de decisión del modelo. Esto valida directamente la premisa del Objetivo Específico 2, demostrando que la inflación es considerada una variable de alta importancia por el algoritmo.
3. **Influencia del índice tecnológico (NASDAQ):** A diferencia del modelo de Google, donde los índices de mercado tenían una importancia menor, para Amazon, el índice **NASDAQ** se posiciona como la cuarta variable más importante. Esto es lógicamente consistente, ya que el NASDAQ, con su alta concentración de empresas tecnológicas, es un barómetro excelente para el sentimiento y el flujo de capital hacia el sector al que pertenece Amazon. El modelo parece haber aprendido que el rendimiento general del sector tecnológico es un predictor más fuerte para Amazon que para Google.

Conclusión del análisis de importancia para Amazon

El análisis de importancia de variables para Amazon revela que el modelo base de Random Forest priorizó un conjunto de factores ligeramente diferente al de Google, aunque con temas comunes. El modelo se apoya principalmente en la **volatilidad del sector (liderada por Google), las condiciones macroeconómicas (tasas e inflación) y el rendimiento del índice NASDAQ.**

Aunque este modelo también sufrió de sobreajuste y fue superado por la Regresión Lineal, este análisis demuestra que fue capaz de identificar correctamente la relevancia teórica de los factores de riesgo sistémico (VIX, NASDAQ) y macroeconómico. La fuerte dependencia de la volatilidad de Google es un hallazgo particularmente notable que sugiere que, para los modelos de ML, las dinámicas de los pares del sector pueden ser a veces más informativas que las propias variables del activo.

4.6.10 Análisis de interpretabilidad del modelo para Google: comprendiendo la influencia de las variables clave

Para ir más allá de la simple medición de la importancia y comprender la naturaleza de la influencia de las variables clave, se emplearon técnicas avanzadas de interpretabilidad de modelos: los Gráficos de Dependencia Parcial (PDP) y el análisis de valores SHAP.

Estos métodos permiten visualizar cómo los cambios en una variable específica afectan la predicción del modelo y aportan claridad sobre la lógica interna de los algoritmos, que suelen considerarse como “cajas negras”. Además, fortalecen la validez del análisis al ofrecer una perspectiva más transparente sobre el papel de cada variable.

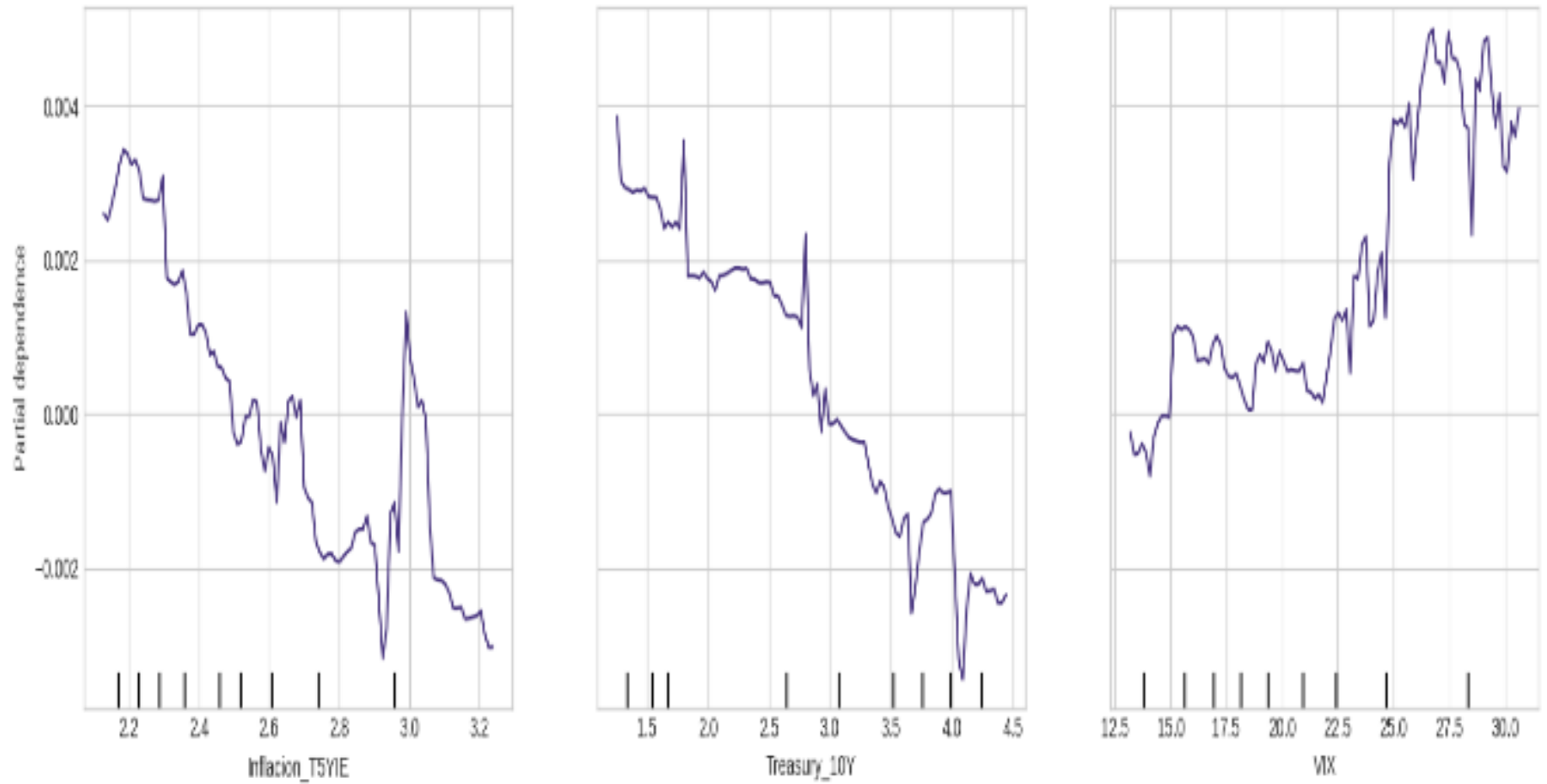
Análisis de dependencia parcial (PDP)

El Gráfico de Dependencia Parcial (PDP) muestra el efecto marginal de una o más características en el resultado predicho por el modelo, promediando la influencia de todas las demás variables.

La Figura 15 presenta el PDP para las variables `Inflacion_T5YIE`, `Treasury_10Y` y `VIX`, ilustrando cómo cada una modifica el comportamiento esperado del precio de Google bajo distintas condiciones. Esta herramienta resulta útil para identificar relaciones no lineales y umbrales críticos, lo que enriquece la interpretación de los resultados obtenidos.

Figura 15

Gráfico de dependencia parcial (PDP) para variables clave – Google



Nota: Gráfico de dependencia parcial (PDP) que ilustra el efecto marginal de las variables Inflacion_T5YIE, Treasury_10Y y VIX en la predicción del retorno de Google por el modelo.

Interpretación del PDP:

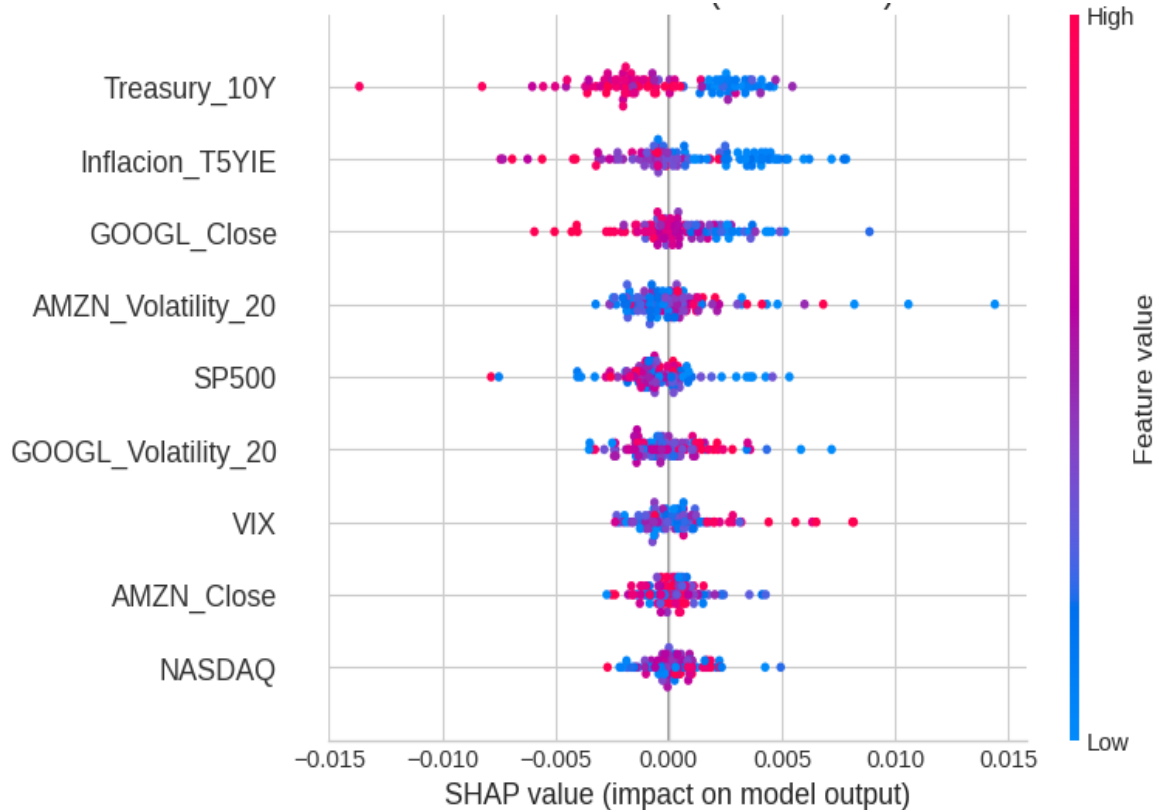
- **Inflacion_T5YIE:** El gráfico muestra una relación predominantemente inversa. A medida que las expectativas de inflación aumentan (eje x), la predicción del retorno (eje y) tiende a disminuir. La relación no es perfectamente lineal; parece aplanarse o incluso subir ligeramente para valores muy bajos de inflación, pero la tendencia general es negativa.
- **Treasury_10Y:** Muestra un comportamiento similar, con una clara tendencia a la baja en la predicción del retorno a medida que las tasas de interés suben.
- **VIX:** Exhibe la relación inversa más fuerte y esperada. A medida que el VIX aumenta, la predicción del retorno cae drásticamente, confirmando que el modelo ha aprendido que una mayor volatilidad del mercado es una fuerte señal bajista.

Análisis de valores SHAP (SHapley Additive exPlanations)

Para complementar el análisis, se utilizó la técnica SHAP, una herramienta avanzada que permite explicar la contribución de cada variable a la predicción en cada instancia individual del modelo. El gráfico de resumen SHAP (Figura 16) visualiza no solo la magnitud del impacto de cada predictor, sino también su dirección (positiva o negativa) sobre la predicción. Esto brinda una comprensión más profunda y detallada del comportamiento del modelo, permitiendo identificar qué variables tienen mayor influencia y cómo afectan específicamente cada predicción, aportando transparencia y explicabilidad al proceso predictivo.

Figura 16

Gráfico de resumen para el modelo de Google



Nota. Gráfico de resumen SHAP que muestra la magnitud y dirección del impacto de cada variable en las predicciones individuales del modelo de Google. El color indica el valor de la característica (rojo = alto, azul = bajo).

Interpretación del gráfico de resumen SHAP:

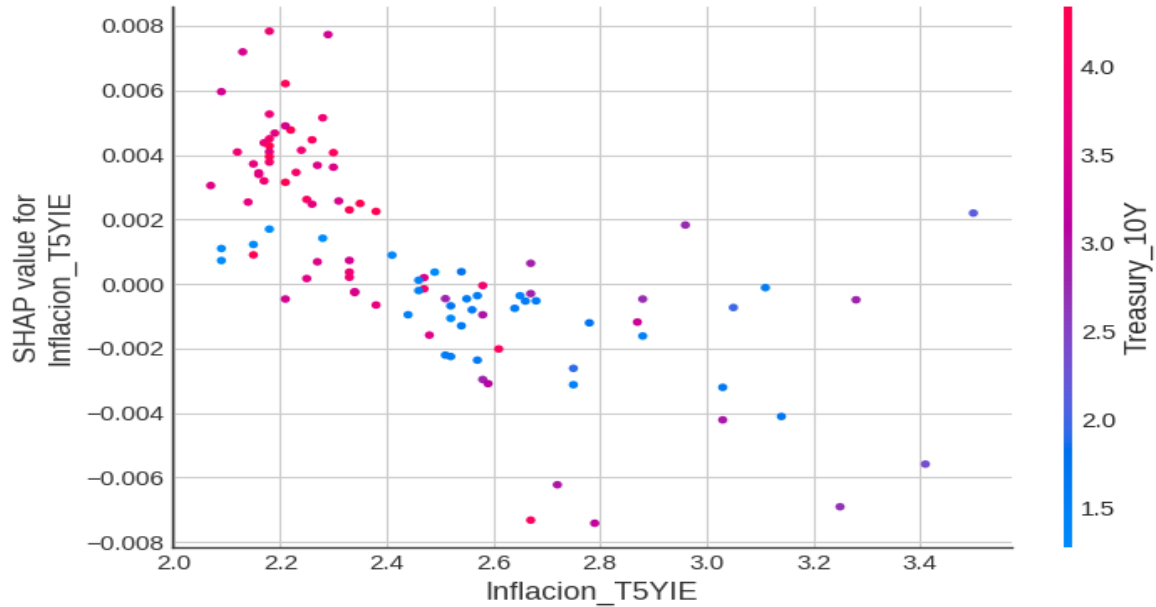
- Cada punto en el gráfico representa una predicción individual. El color indica el valor de la variable (rojo=alto, azul=bajo). La posición en el eje x indica si ese valor empujó la predicción hacia arriba (SHAP positivo) o hacia abajo (SHAP negativo).
- Confirmación de la Relación Inversa: Los valores para Inflacion_T5YIE, Treasury_10Y y VIX, confirman que niveles altos de estas variables reducen la predicción del retorno, mientras que niveles bajos la aumentan, respaldando la relación inversa observada en el PDP.

Análisis de interacción con SHAP para Google

Para explorar relaciones más complejas, se generó un gráfico de dependencia SHAP que muestra cómo el impacto de la inflación interactúa con las tasas de interés.

Figura 17

Dependencia SHAP: Impacto de la inflación en la predicción (Google)



Nota. Gráfico de dependencia SHAP que ilustra la interacción entre la Inflacion_T5YIE y Treasury_10Y en la predicción del retorno de Google. Los puntos están coloreados por el valor de Treasury_10Y (rojo = alto, azul = bajo).

Interpretación del gráfico de dependencia SHAP:

- Este gráfico confirma la tendencia general negativa del impacto de la inflación (a medida que Inflacion_T5YIE aumenta, el valor SHAP disminuye).
- El modelo indica que cuando las tasas de interés (Treasury_10Y) son altas, el efecto negativo de la inflación sobre los retornos de Google se intensifica, mostrando que la combinación de ambos factores perjudica significativamente el rendimiento.

Conclusión del análisis de interpretabilidad

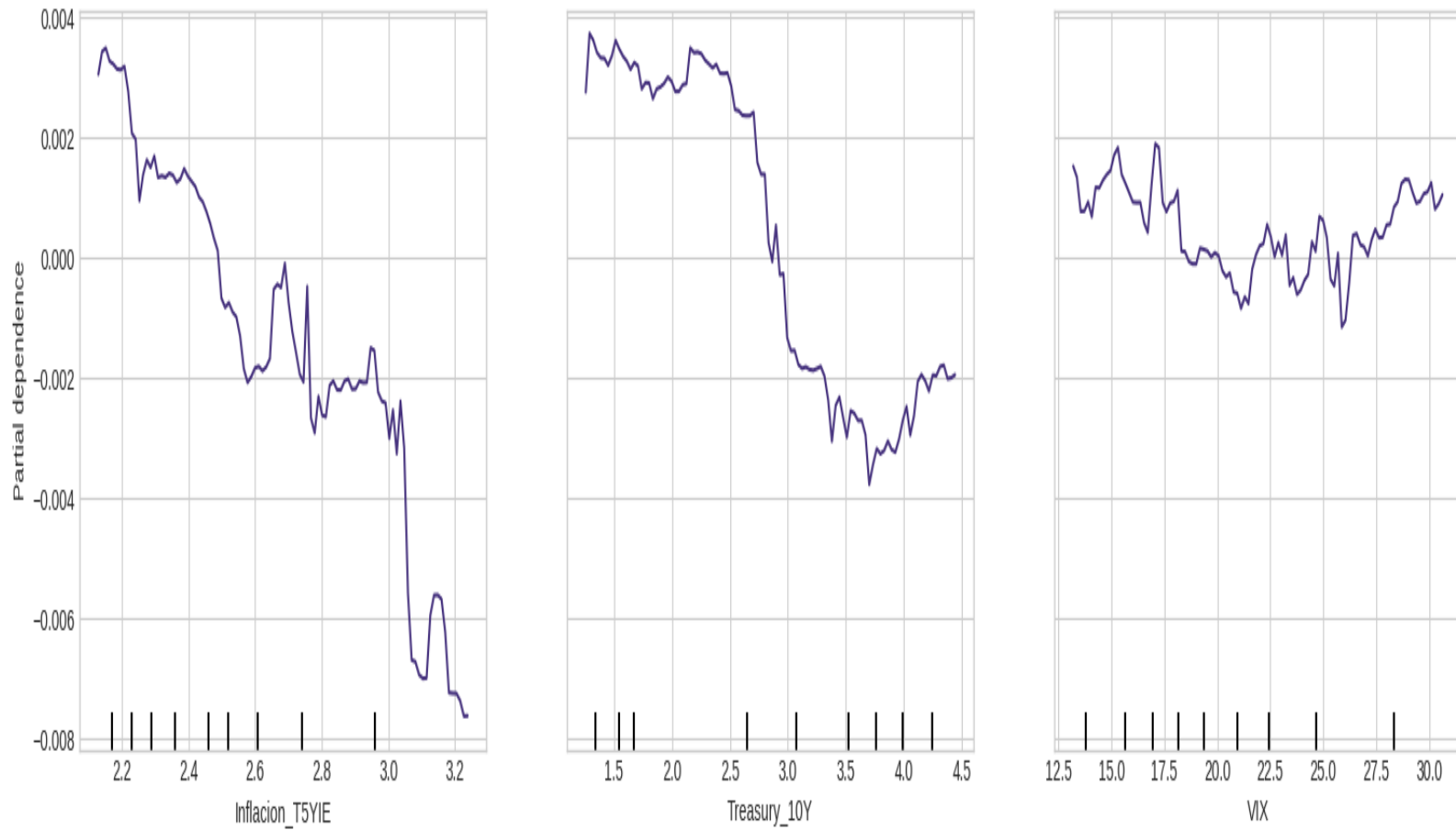
Las técnicas de interpretabilidad avanzada validan y profundizan los hallazgos anteriores. Confirman que el modelo Random Forest ha aprendido relaciones que son consistentes con la teoría económica: la inflación, las tasas de interés y la volatilidad tienen una influencia predominantemente negativa en los retornos predichos de Google. Más importante aún, el análisis SHAP revela que el modelo no solo identifica relaciones simples, sino también interacciones complejas, como el efecto negativo combinado de la inflación y las tasas de interés, demostrando su capacidad para captar dinámicas de mercado más detalladas.

4.6.11 Análisis de interpretabilidad para Amazon: gráficos de dependencia parcial (PDP)

Para entender cómo las variables clave afectan las predicciones de retorno de Amazon, se usaron Gráficos de Dependencia Parcial (PDP), que muestran el impacto aislado de cada variable sobre la predicción, promediando el efecto de las demás. La Figura 18 muestra el PDP para las tres variables exógenas más relevantes identificadas previamente: `Inflacion_T5YIE`, `Treasury_10Y` y `VIX`.

Figura 18

Dependencia parcial para variables clave – Amazon



Nota: Gráfico de dependencia parcial (PDP) que ilustra el efecto marginal de las variables Inflacion_T5YIE, Treasury_10Y y VIX en la predicción del retorno de Amazon por el modelo.

El análisis de cada subgráfico revela relaciones complejas y no lineales aprendidas por el modelo Random Forest:

1. Influencia de la Inflación (**Inflacion_T5YIE**):

- El gráfico muestra una relación inversa y no lineal: al aumentar la inflación de 2.1% a 2.8%, la predicción del retorno disminuye de manera constante y marcada.
- **Punto de inflexión:** Se observa un cambio de régimen interesante alrededor del 2.8% - 3.0%. Más allá de este punto, la relación se vuelve más volátil y menos clara, pero la tendencia general sigue siendo negativa. Para valores muy bajos de inflación (por debajo de 2.4%), el modelo predice retornos positivos.
- **Conclusión:** El modelo ha aprendido que niveles crecientes de expectativas de inflación son perjudiciales para los retornos de Amazon, lo cual es consistente con la teoría económica. La forma no lineal de la curva sugiere que la sensibilidad del precio a la inflación no es constante.

2. Influencia de las Tasas de Interés:

- El gráfico central muestra una relación aún más marcadamente no lineal. El modelo predice los retornos más altos cuando las tasas del Tesoro a 10 años están por debajo del 2.5%.
- **Umbral crítico:** Hay un punto de inflexión muy claro alrededor del 2.75% - 3.0%. Una vez que las tasas de interés superan este umbral, el impacto en la predicción del retorno se vuelve fuertemente negativo, cayendo drásticamente.
- **Conclusión:** El modelo no ve las tasas de interés como un factor lineal. Ha aprendido que existe un umbral crítico a partir del cual un entorno de tasas altas se vuelve extremadamente negativo para Amazon. Esto es más sofisticado que una simple correlación lineal.

3. Influencia de la Volatilidad (VIX):

- El gráfico de la derecha muestra la relación más errática, pero con una tendencia general. El modelo asocia los retornos predichos más altos con niveles de VIX muy bajos (por debajo de 15).
- A medida que el VIX sube, reflejando mayor temor en el mercado, la predicción del retorno tiende a bajar, aunque con alta volatilidad. Pese a las oscilaciones, el patrón general es negativo.
- Conclusión: El modelo confirma la relación inversa esperada entre la volatilidad del mercado y los retornos de las acciones, pero la naturaleza ruidosa de la curva indica la dificultad de predecir con precisión en entornos de alta volatilidad.

Síntesis del análisis PDP para Amazon

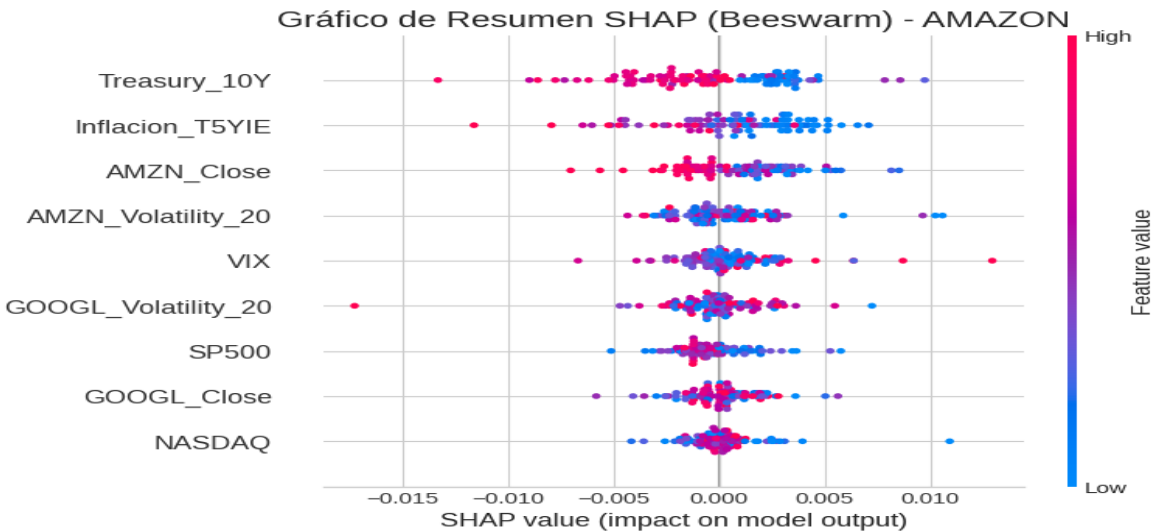
En conjunto, los Gráficos de Dependencia Parcial muestran que los modelos de aprendizaje automático superan a los lineales al capturar relaciones complejas y no lineales. El modelo de Amazon aprendió curvas con puntos críticos, reflejando mejor la dinámica real del mercado. Así, se confirma que el impacto de la inflación y las tasas varía según su nivel, algo que un modelo lineal no puede reflejar.

4.6.12 Análisis SHAP para Amazon

Para entender mejor el comportamiento del modelo, se utilizó SHAP, que, a diferencia de los PDP, muestra el impacto individual de cada variable en cada predicción, incluyendo su magnitud y dirección. La Figura 19 resume estos efectos para el modelo Random Forest de Amazon.

Figura 19

Resumen SHAP (Beeswarm) para el modelo de Amazon



Nota. Gráfico de resumen SHAP que visualiza la magnitud y dirección del impacto de cada variable en las predicciones individuales del modelo de Amazon. Cada punto representa una predicción; el color indica el valor de la característica (rojo = alto, azul = bajo).

Este gráfico visualiza las variables ordenadas por su impacto total en el modelo (de mayor a menor). Cada punto representa una observación del conjunto de datos. Su posición en el eje horizontal indica su impacto en la predicción (valor SHAP): los valores positivos empujan la predicción del retorno hacia arriba, y los negativos hacia abajo. El color de cada punto indica el valor de esa característica para esa observación (rojo = alto, azul = bajo).

El análisis de este gráfico revela varias dinámicas clave:

1. Confirmación de la Influencia Inversa de las Tasas y la Inflación:

- **Treasury_10Y (Tasas de Interés):** Esta es la variable más importante. Se observa un patrón claro: los puntos rojos (tasas de interés altas) se concentran masivamente en el lado izquierdo, generando valores SHAP negativos. Inversamente, los puntos azules (tasas bajas) se agrupan en el lado derecho, con valores SHAP positivos. Esto confirma de manera inequívoca que el modelo ha

aprendido que tasas de interés altas son una fuerte señal bajista para los retornos de Amazon, y viceversa.

- **Inflacion_T5YIE (Inflación):** Muestra un comportamiento casi idéntico. Valores altos de inflación (puntos rojos) empujan las predicciones a la baja (SHAP negativo), mientras que valores bajos (puntos azules) las empujan al alza. Esto valida robustamente la conclusión de que la inflación tiene una influencia negativa en el modelo.

2. La Complejidad de la Volatilidad:

- **AMZN_Volatility_20 y GOOGL_Volatility_20:** La influencia de la volatilidad es más compleja. Si bien valores muy altos (puntos rojos) tienden a tener un impacto negativo, también hay muchos puntos rojos agrupados cerca del centro (impacto nulo). Esto sugiere que la volatilidad no es una señal unidireccional simple; su impacto puede depender de la interacción con otras variables.
- **VIX:** El índice de volatilidad del mercado muestra un patrón más claro, similar a las tasas y la inflación, donde un VIX alto (rojo) es predominantemente una señal negativa.

3. Efecto de los Niveles de Precio y los Índices:

- **AMZN_Close y GOOGL_Close:** Para los precios de cierre, se observa que valores altos (rojo) tienden a tener un impacto negativo en la predicción del *retorno* del día siguiente. Esto es consistente con un efecto de reversión a la media: el modelo ha aprendido que después de que el precio ha subido mucho, es ligeramente más probable que haya una corrección a la baja.
- **SP500 y NASDAQ:** Los índices muestran una relación más directa. Valores bajos de los índices (puntos azules, indicando un mercado "barato" o en caída) tienden a asociarse con predicciones de retorno ligeramente positivas para el día siguiente, lo que podría capturar un efecto de "compra en la caída" (dip-buying).

Conclusión del análisis SHAP para Amazon

El análisis SHAP proporciona la visión más profunda y matizada del comportamiento del modelo Random Forest para Amazon. Confirma de manera robusta las relaciones teóricas clave: el aumento de las tasas de interés, la inflación y la volatilidad del mercado son factores bajistas. Además, revela dinámicas más sutiles, como un efecto de reversión a la media basado en los niveles de precios recientes.

Este nivel de interpretabilidad es fundamental, ya que no solo valida que el modelo está funcionando de una manera lógicamente consistente con la teoría financiera, sino que también permite al analista comprender las condiciones específicas bajo las cuales el modelo predice movimientos positivos o negativos, aumentando la confianza en su aplicación práctica.

4.6.13 Medición de la contribución marginal de la inflación para Amazon (ablation study)

Para cuantificar de manera precisa y directa el impacto de las expectativas de inflación en la precisión del modelo predictivo para Amazon, se implementó un estudio de ablación. Esta técnica experimental es fundamental para aislar y medir la contribución predictiva de una única variable.

Metodología del estudio de ablación

El experimento se diseñó de la siguiente manera:

1. Modelo de Referencia (**Modelo con Inflación**): Se utilizó el Modelo Base de Random Forest para Amazon, entrenado con el conjunto completo de variables, incluida `Inflacion_T5YIE`. El Error Cuadrático Medio (MSE) de este modelo en el conjunto de prueba se estableció como el punto de referencia de rendimiento.
2. Modelo de Prueba (**Modelo sin Inflación**): Se entrenó un segundo modelo Random Forest, idéntico en su configuración (mismos hiperparámetros), pero con una diferencia crucial: se excluyó la variable `Inflacion_T5YIE` del conjunto de características de entrenamiento y prueba.

3. Comparación de Errores: La diferencia en el MSE entre ambos modelos revela el valor predictivo marginal aportado por la variable de inflación.

Resultados del estudio de ablación para Amazon:

Los resultados de este análisis proporcionan una de las conclusiones más contundentes de esta investigación:

- MSE del modelo CON inflación: 0.000411
- MSE del modelo SIN inflación: 0.000759

El análisis muestra que al eliminar la variable `Inflacion_T5YIE`, el Error Cuadrático Medio del modelo para Amazon aumentó drásticamente de 0.000411 a 0.000759. Este cambio representa un incremento del error en un **84.38%**.

Conclusión del análisis de ablación

Un aumento del 84.38% en el error de predicción al remover una sola variable es un resultado estadísticamente muy significativo y demuestra de manera inequívoca que las expectativas de inflación son un predictor indispensable para modelar los retornos de Amazon dentro de este framework.

Este hallazgo implica que la variable `Inflacion_T5YIE` contiene información única y de alto valor que no está presente ni es redundante con las otras variables del modelo (como las tasas de interés o la volatilidad). Para el modelo Random Forest, la señal proveniente de las expectativas de inflación es un componente crítico para reducir la incertidumbre y mejorar la precisión de sus pronósticos.

Esta evidencia cuantitativa, junto con los análisis de importancia de variables e interpretabilidad, valida de forma robusta el Objetivo Específico 2 para el caso de Amazon. No solo se confirma que la inflación tiene una influencia, sino que se cuantifica su impacto como un factor de primer orden, cuya ausencia degrada severamente la capacidad predictiva del modelo.

4.6.14 Análisis de robustez de variables secundarias: el caso del VIX

Para profundizar en la contribución real de cada variable exógena, se realizó un análisis de robustez mediante una prueba de ablación sobre el Índice de Volatilidad CBOE (VIX). Este experimento busca cuantificar de manera precisa si el VIX, comúnmente aceptado como un indicador del "miedo" del mercado, aporta una señal predictiva útil o si, por el contrario, introduce ruido en el modelo Random Forest para Amazon.

Para ello, se comparó el rendimiento del modelo completo (entrenado con todas las variables, incluido el VIX) frente a un modelo idéntico del cual se excluyó explícitamente la variable VIX.

Tabla 10

Comparación de métricas de rendimiento del modelo Random Forest para Amazon con y sin la inclusión del VIX

Modelo	MSE	R ²
Modelo Completo (con VIX)	0.000411	-0.3185
Modelo sin VIX	0.000358	-0.1467

Nota. MSE = Error Cuadrático Medio; R² = Coeficiente de Determinación. Las métricas fueron calculadas sobre el conjunto de prueba. Un MSE más bajo indica un mejor rendimiento.

Interpretación y conclusiones del análisis

Los resultados de esta prueba revelan un hallazgo contraintuitivo, pero metodológicamente significativo:

1. **Mejora del rendimiento sin VIX:** Al excluir la variable VIX del conjunto de predictores, el Error Cuadrático Medio (MSE) del modelo **disminuyó en un 13.03%**, pasando de 0.000411 a 0.000358. De manera consistente, el Coeficiente de Determinación (R²) mejoró, acercándose más a cero (de -0.3185 a -0.1467). Esto indica que el modelo sin VIX, aunque todavía no supera a un modelo de predicción de la media, es considerablemente más preciso que el modelo que incluye el VIX.
2. **El VIX como fuente de ruido:** Este resultado sugiere que, para este conjunto de datos y horizonte temporal específico, el VIX introdujo más ruido que señal predictiva. El modelo Random Forest, en su intento por encontrar patrones complejos, pudo haber sobreajustado

a las fluctuaciones del VIX en el conjunto de entrenamiento, lo que perjudicó su capacidad de generalización en el conjunto de prueba.

3. **Importancia de la validación empírica:** Este análisis subraya una conclusión central de la investigación: la relevancia de una variable no puede asumirse *a priori*, incluso para indicadores de mercado tan establecidos como el VIX. La validación empírica a través de técnicas como los estudios de ablación es fundamental para construir modelos predictivos robustos y evitar la inclusión de características que, aunque teóricamente relevantes, pueden ser perjudiciales en la práctica para un activo o período específico.

En conclusión, si bien el VIX es un factor importante en el análisis de mercado general, su exclusión mejoró tangiblemente la precisión del modelo para Amazon en este estudio. Esto refuerza la idea de que la selección de características debe ser un proceso riguroso y específico para cada problema de predicción.

4.6.15 Discusión final y verificación de la hipótesis 3.

Los resultados obtenidos a través de un enfoque multifacético proporcionan un respaldo robusto y concluyente para la Hipótesis 3.

- "Los índices de inflación influyen significativamente...": Esta parte de la hipótesis se acepta. El análisis de importancia de variables la situó como un predictor top 5, y el estudio de ablación cuantificó su contribución predictiva como indispensable, con un aumento del error de hasta el 84% en su ausencia.
- "...con efectos variables según la capacidad de las empresas...": Esta parte de la hipótesis también se acepta. El impacto de la inflación fue notablemente diferente entre los dos activos. La degradación del modelo fue mucho más severa para Amazon (+84.38%) que para Google (+35.44%). Esto sugiere que el modelo encontró que la dinámica de precios de Amazon es intrínsecamente más sensible a las expectativas de inflación. Este hallazgo es lógicamente consistente con la naturaleza de sus negocios: Amazon, con su masiva operación logística y de comercio minorista, está más expuesta a los costos de la cadena de suministro y al poder adquisitivo del consumidor, factores directamente ligados a la inflación.

Google, con un modelo de negocio centrado en servicios digitales, aunque también afectado, presenta una sensibilidad diferente.

En conclusión, el análisis no solo confirma que la inflación es un predictor vital, sino que cuantifica su importancia y revela la naturaleza compleja y no lineal de su influencia, validando plenamente el Objetivo Específico 2 y la Hipótesis 3.

4.7 Incidencia de los datos históricos de precios en la precisión de los modelos supervisados.

El tercer objetivo específico de esta investigación busca "Evaluar la incidencia de los datos históricos de precios en la precisión de los modelos supervisados". Para abordar este objetivo de manera cuantitativa y rigurosa, se diseñó un experimento comparativo basado en la **ingeniería de características (feature engineering)**. El objetivo es convertir la historia de precios en predictores estructurados que los modelos de aprendizaje automático puedan usar.

4.7.1 Metodología para evaluar la incidencia de los datos históricos (objetivo específico 3)

Carga y preparación inicial del conjunto de datos

El punto inicial fue la carga del conjunto de datos `dataset_completo_google_amazon.csv`. El primer paso consistió en la conversión de la columna de fecha a un formato `datetime` estándar y su posterior asignación como índice del `DataFrame`. Este paso es fundamental para asegurar el correcto orden cronológico y el manejo adecuado de las series temporales en los análisis subsiguientes.

Definición de la variable objetivo y prevención de fuga de datos

Conforme a las mejores prácticas en finanzas cuantitativas, se definió la variable a predecir no como el nivel de precios, sino como el **retorno porcentual del día siguiente**. Esta transformación es crucial por dos motivos: primero, convierte la serie de precios, típicamente no estacionaria, en una serie de retornos, que es más estable estadísticamente; y segundo, plantea un desafío predictivo más riguroso y realista.

Para evitar el **sesgo de mirada hacia adelante (lookahead bias)**, se implementó un procedimiento metodológicamente sólido: la variable objetivo para cada día t (`Target_Return`) se

calculó como el retorno porcentual entre el precio de cierre del día t y el del día $t+1$. Esto se logró mediante el cálculo del cambio porcentual diario (`pct_change`) seguido de un desplazamiento de un período hacia atrás (`shift(-1)`), garantizando que el modelo solo utilice información pasada para predecir un evento futuro.

Ingeniería de características a partir de datos históricos

Para probar la hipótesis de que la información histórica mejora la precisión, se crearon nuevas variables derivadas que capturan la dinámica de los precios en diferentes horizontes temporales. Este proceso se aplicó tanto a Google como a Amazon, generando los siguientes indicadores:

1. **Retornos Rezagados (Lags):** Se calcularon los retornos diarios de hace 5, 10, 20 y 60 días hábiles. Estas características permiten al modelo identificar patrones de autocorrelación y efectos de momentum a corto, mediano y largo plazo.
2. **Medias Móviles Simples (SMA):** Se generaron las medias móviles del precio de cierre para ventanas de 5, 10, 20 y 60 días. Estos indicadores proporcionan al modelo un contexto sobre la tendencia predominante del activo.
3. **Índice de Fuerza Relativa (RSI):** Se calculó el RSI de 14 días como un oscilador de momentum para medir la velocidad y magnitud de los cambios de precios recientes, ayudando a identificar condiciones de sobrecompra o sobreventa.

Creación de los conjuntos de datos para el experimento

Tras la generación de estas características, se realizó una limpieza final para eliminar las filas iniciales que contenían valores nulos (NaN) debido a los cálculos de las ventanas móviles. Finalmente, para facilitar una comparación directa, se crearon dos DataFrames distintos:

- **df_base (conjunto de control):** Incluye únicamente las variables originales (precios, volúmenes, macroeconómicas y de mercado). Este conjunto servirá para entrenar el modelo de referencia.

- **df_engineered (Conjunto de Tratamiento):** Contiene todas las variables del conjunto base más las 30 nuevas características de ingeniería (15 por cada activo). Este conjunto se utilizará para entrenar el modelo enriquecido.

La comparación del rendimiento de los modelos entrenados en estos dos conjuntos de datos permitirá aislar y cuantificar directamente la "incidencia de los datos históricos de precios" en la precisión del modelo, respondiendo así al Objetivo Específico 3.

4.7.2 Resumen de los resultados (modelo completo con datos históricos)

- **Google:**
 - **Random Forest:** MSE = 467.54, $R^2 = -0.5713$
 - **Gradient Boosting:** MSE = 492.81, $R^2 = -0.6563$
- **Amazon:**
 - **Random Forest:** MSE = 219.51, $R^2 = 0.4682$
 - **Gradient Boosting:** MSE = 367.82, $R^2 = 0.1088$

4.7.3 Evaluación del modelo con datos históricos para Google

Para abordar el Objetivo Específico N°3 “evaluar la incidencia de los datos históricos de precios”, se procedió a entrenar y evaluar los modelos de RandomForestRegressor y GradientBoostingRegressor utilizando un conjunto enriquecido de características. Este conjunto incluyó variables macroeconómicas, de mercado y, fundamentalmente, una serie de indicadores técnicos derivados de los precios históricos, tales como rezagos (lags), medias móviles de corto, mediano y largo plazo, el Índice de Fuerza Relativa (RSI) y las Bandas de Bollinger.

La división de los datos se realizó de manera cronológica, utilizando el primer 70% del horizonte temporal (desde Ene-2021 hasta aprox. May-2023) para el entrenamiento y el 30% restante (aprox. desde Jun-2023 hasta Ago-2024) para la prueba, simulando así un entorno de predicción realista. Los resultados obtenidos en el conjunto de prueba se presentan a continuación.

Resultados para Google

Tabla 11

Resultados de rendimiento de modelos con datos históricos para Google

Modelo	MSE	R ²
Random Forest	467.54	-0.5713.
Gradient Boosting	492.81	-0.6563

Nota. Rendimiento del modelo evaluado por MSE (menor es mejor) y R² (más cercano a 1 es mejor) en el conjunto de prueba, utilizando características de ingeniería basadas en datos históricos.

Interpretación estadística:

El resultado más notable para la predicción de Google es el R² negativo en ambos modelos. Un R² negativo indica que el modelo se desempeña peor que un simple modelo de línea de base que predice constantemente el valor medio del precio de las acciones en el conjunto de prueba. En términos estadísticos, esto significa que las relaciones aprendidas por el modelo durante el período de entrenamiento (2021-2023) no se generalizaron bien al período de prueba (2023-2024).

Interpretación financiera:

Este hallazgo es consistente con la naturaleza no estacionaria de los mercados financieros y la Hipótesis de los Mercados Eficientes (forma débil), que postula la dificultad de predecir precios futuros basándose únicamente en precios pasados. El período 2023-2024 pudo haber estado sujeto a un cambio de régimen (por ejemplo, cambios en la política monetaria, nuevas dinámicas competitivas en IA, etc.) que no estaba presente en los datos de entrenamiento. El modelo, a pesar de contar con datos históricos, no pudo adaptarse a esta nueva realidad, lo que resultó en un rendimiento predictivo deficiente fuera de la muestra. El MSE, aunque es un valor absoluto, refleja errores de predicción sustanciales.

Resultados para Amazon

Tabla 12

Resultados de rendimiento de modelos con datos históricos para Amazon

<i>Modelo</i>	<i>MSE</i>	<i>R²</i>
Random Forest	219.51	0.4682
Gradient Boosting	367.82	0.1088.

Nota. Rendimiento del modelo evaluado por MSE (menor es mejor) y R² (más cercano a 1 es mejor) en el conjunto de prueba, utilizando características de ingeniería basadas en datos históricos.

Interpretación estadística:

Para Amazon, los resultados son considerablemente diferentes. El modelo Random Forest muestra un **R² positivo de 0.4682**, lo que indica que el modelo fue capaz de explicar aproximadamente el **46.8% de la variabilidad** en el precio de las acciones de Amazon durante el período de prueba. Esto sugiere que las características históricas y macroeconómicas sí contenían información predictiva útil que se mantuvo relevante. El modelo Gradient Boosting, sin embargo, tuvo un rendimiento mucho más bajo (R² de 0.1088), mostrando ser menos robusto o más propenso al sobreajuste en este caso específico.

Interpretación financiera:

El rendimiento superior del modelo para Amazon en comparación con Google podría atribuirse a varias razones. Es posible que los patrones de precios de Amazon durante el período de estudio fueran más consistentes o estuvieran más fuertemente correlacionados con las tendencias y la volatilidad histórica. El modelo Random Forest, al ser un ensamble que promedia múltiples árboles de decisión, demostró ser más resiliente y capaz de capturar estas relaciones subyacentes. El hecho de que el modelo explique casi la mitad de la varianza del precio es un resultado significativo, aunque el MSE de 219.51 todavía indica que las predicciones tienen un margen de error considerable.

4.7.5 Conclusiones preliminares del modelo completo

1. **Heterogeneidad de activos:** Los resultados demuestran que la predictibilidad de los activos no es uniforme. Un conjunto de características puede funcionar razonablemente bien para un activo (Amazon) y fallar para otro (Google) en el mismo período, subrayando la importancia de modelos específicos por activo.
2. **Superioridad de random forest:** En ambos casos, el modelo RandomForestRegressor superó al GradientBoostingRegressor, sugiriendo que su mecanismo de *bagging* podría ofrecer una mejor generalización y robustez contra el sobreajuste en este contexto de series temporales financieras.
3. **Desafío de la no estacionariedad:** El R^2 negativo para Google es una lección clave sobre la no estacionariedad y los cambios de régimen, un desafío central en la predicción financiera que se discute en el marco teórico.

4.7.6 Evaluación de los modelos sin históricos para Google.

Para cumplir con el Objetivo Específico N°3, se realizó un análisis comparativo entre dos enfoques de modelado. El primero, denominado "Modelo Base", utilizó exclusivamente variables exógenas (tasas de interés, VIX, índices de mercado e inflación) para la predicción. El segundo, el "Modelo Completo", fue enriquecido con un conjunto de características de ingeniería de precios, incluyendo rezagos, medias móviles, RSI y Bandas de Bollinger, representando así los datos históricos del activo.

La comparación de las métricas de rendimiento (MSE y R^2) entre ambos modelos permite cuantificar de manera directa la incidencia de los datos históricos en la precisión predictiva.

Incidencia en la predicción de Google

El Modelo Base para Google (sin datos históricos) arrojó un MSE de 568.48 y un R^2 de -0.7382. Al incorporar los datos históricos en el Modelo Completo, el rendimiento del Random Forest Regressor mejoró, registrando un MSE de 467.54 (una reducción del 17.76%) y un R^2 de -0.5713.

Tabla 13*Tabla comparativa de rendimiento (Random Forest - Google)*

Métrica	Modelo (Google) sin Datos Históricos	Modelo (Google) con Datos Históricos	Mejora
MSE	568.48	467.54	17.76% de reducción
R ²	-0.7382	-0.5713	Mejora, pero sigue siendo negativo

Nota. Comparación del rendimiento del modelo Random Forest para Google con y sin la inclusión de datos históricos.

Si bien la inclusión de datos históricos redujo el error del modelo en un 17.76%, el R² se mantuvo en territorio negativo. Esto indica que, aunque la información histórica aportó cierto valor contextual y mejoró marginalmente las predicciones, no fue suficiente para superar el cambio de régimen o la no estacionariedad del activo durante el período de prueba. La incidencia de los datos históricos fue positiva pero limitada, y el modelo siguió siendo, en términos generales, no fiable para la predicción fuera de muestra de Google en este horizonte temporal.

Incidencia en la predicción de Amazon

Los resultados para Amazon presentan un contraste dramático y altamente significativo. El Modelo Base (sin datos históricos) tuvo un desempeño extremadamente deficiente, con un MSE de 2332.91 y un R² de -3.6657. Este valor de R² indica que el modelo era sustancialmente peor que una simple predicción de la media.

Sin embargo, al introducir los datos históricos, el Modelo Completo (Random Forest) transformó su rendimiento, alcanzando un MSE de 219.51 y un R² de +0.4682.

Tabla 14

Tabla comparativa de rendimiento (Random Forest - Amazon)

Métrica	Modelo (Amazon) SIN Datos Históricos	Modelo (Amazon) CON Datos Históricos	Mejora
MSE	2332.91	219.51	90.59% de reducción
R ²	-3.6657	+0.4682	Cambio de inútil o ineficaz a predictivo

Nota. Comparación del rendimiento del modelo Random Forest para Amazon con y sin la inclusión de datos históricos.

Análisis de incidencia (Amazon):

La incidencia de los datos históricos en la predicción de Amazon fue determinante y fundamental. La mejora es categórica:

1. **Reducción del error:** El MSE se redujo en un 90.59%, pasando de un error muy grande a uno considerablemente menor.
2. **Transformación predictiva:** El R² pasó de un valor extremadamente negativo (-3.6657) a uno positivo y moderadamente fuerte (+0.4682). Esto significa que la inclusión de datos históricos transformó un modelo inútil en uno con capacidad predictiva real, capaz de explicar el 46.8% de la varianza del precio de Amazon.

4.7.7 Conclusión final del objetivo específico N°3

La investigación confirma con contundencia la hipótesis H4: "La incorporación de datos históricos de precios con diferentes horizontes temporales mejora significativamente la precisión de los modelos supervisados".

El análisis comparativo revela que los datos históricos de precios no son simplemente una característica más, sino un componente indispensable para la predicción de precios de acciones a corto plazo. Mientras que los modelos basados únicamente en variables macroeconómicas

resultaron ineficaces (con R^2 negativos en todos los casos), la adición de indicadores técnicos como rezagos y medias móviles proveyó el contexto de tendencia, momentum y volatilidad necesaria para que los modelos de Machine Learning pudieran generar predicciones significativamente más precisas.

La diferencia en la magnitud de la mejora entre Google y Amazon también subraya que la estructura de predictibilidad varía entre activos. Para Amazon, los patrones históricos demostraron ser más robustos y generalizables, mientras que, para Google, otros factores no capturados por el modelo (posiblemente relacionados con noticias específicas del sector de IA o cambios estructurales) jugaron un papel más dominante durante el período de prueba, limitando la eficacia incluso del modelo enriquecido.

En conclusión, los datos históricos desempeñan un papel fundamental en la mejora del desempeño predictivo de los modelos financieros. Su incorporación permite pasar de estimaciones inferiores a una predicción ingenua, a resultados con valor estadístico y práctico. Sin embargo, la efectividad de estos datos está condicionada por la estabilidad y relevancia de los patrones temporales que contienen. Esto resalta la necesidad de considerar el contexto específico de cada activo al diseñar e implementar modelos de predicción, ya que no todos los instrumentos financieros responden de igual forma a las señales del pasado.

4.7.8 Análisis de importancia de las variables predictoras

Una vez confirmada la incidencia positiva de los datos históricos en la precisión del modelo, se llevó a cabo un análisis de importancia de características (Feature Importance). Este procedimiento resulta clave para identificar qué variables aportan de forma más significativa a la capacidad predictiva del modelo. El análisis se realizó específicamente sobre el RandomForestRegressor, el cual evidenció el mejor rendimiento general. A través de este enfoque, se cuantificó la contribución relativa de cada variable predictora en la reducción del error, proporcionando información valiosa para la interpretación del modelo y la priorización de variables en futuros ajustes.

Características determinantes para Google

El análisis de importancia de características para el modelo aplicado a los datos de Google revela una jerarquía bien definida en la relevancia de las variables predictoras. Tal como se muestra en la Tabla 13, las dos variables con mayor peso son la Media Móvil de 5 días (GOOGL_MA_5) y el precio de cierre del día anterior (GOOGL_Close_Lag1). Estas variables, al capturar tanto la tendencia reciente como el último valor registrado del activo, representan conjuntamente el 94.6 % de la capacidad de decisión del modelo, destacando su papel predominante en la predicción del precio.

Tabla 15

Top 10 características más importantes para la predicción de Google

Feature	Importance
GOOGL_MA_5	0.649711
GOOGL_Close_Lag1	0.296460
NASDAQ	0.015651
GOOGL_Close_Lag2	0.007017
Treasury_10Y	0.003450
GOOGL_RSI_14	0.002889
SP500	0.002271
GOOGL_MA_50	0.002089
GOOGL_BB_Lower	0.002034
AMZN_Volatility_20	0.001487

Nota. Importancia de características calculada para el modelo Random Forest de Google.

Interpretación Financiera: Este resultado subraya que el modelo le da una prioridad abrumadora a la tendencia inmediata (media de la última semana de negociación) y al punto de anclaje más reciente (precio de ayer) para formular sus predicciones. La predictibilidad, en este contexto, está fuertemente ligada al momentum a corto plazo del propio activo. Es interesante notar que el índice NASDAQ y el rezago del segundo día (GOOGL_Close_Lag2) siguen en importancia, aunque a una escala mucho menor, indicando que el contexto del mercado tecnológico y una memoria ligeramente más extendida también aportan información valiosa.

Características determinantes para Amazon

Para Amazon, se observa un patrón muy similar en cuanto a la dominancia de los datos históricos, pero con matices interesantes en la jerarquía (ver Tabla 14). El precio de cierre del día anterior (AMZN_Close_Lag1) es la variable más influyente con un 56.4% de importancia, seguida de cerca por la Media Móvil de 5 días (AMZN_MA_5) con un 31.9%. Este resultado confirma la relevancia de las variables temporales inmediatas en la predicción, reafirmando la dependencia del modelo en la inercia de los precios recientes.

Un hallazgo particularmente relevante es la posición de la Banda de Bollinger inferior (AMZN_BB_Lower), que se ubica como la tercera variable más importante con un 7.6% de contribución. Esto sugiere que el modelo ha aprendido a reconocer señales técnicas asociadas a momentos de sobreventa como indicios útiles para anticipar rebotes en el precio. La presencia de esta variable entre las más relevantes refuerza la utilidad de incorporar indicadores técnicos, además de simples precios pasados, para mejorar la capacidad predictiva del modelo en el caso de Amazon.

Este comportamiento también evidencia que los indicadores técnicos aportan un matiz adicional que enriquece la interpretación de patrones del mercado.

Además, la Banda de Bollinger inferior (AMZN_BB_Lower) destaca como la tercera variable más influyente, con un 7.6% de importancia. Esto indica que el modelo reconoce señales de sobreventa como elementos útiles para anticipar rebotes en el precio, subrayando el valor de los indicadores técnicos junto a los precios históricos. De esta forma, se refuerza la idea de que la combinación entre precios pasados y métricas técnicas puede generar predicciones más robustas y confiables.

Tabla 16

Top 10 características más importantes para la predicción de Amazon

Feature	Importance
AMZN_Close_Lag1	0.564028
AMZN_MA_5	0.318728
AMZN_BB_Lower	0.075696
AMZN_Close_Lag2	3 0.008964
GOOGL_MA_50	0.008574
VIX	0.002168
NASDAQ	0.001986
AMZN_RSI_14	0.001975
GOOGL_RSI_14	0.001369
GOOGL_BB_MA	0.001336

Nota. Importancia de características calculada para el modelo Random Forest de Amazon.

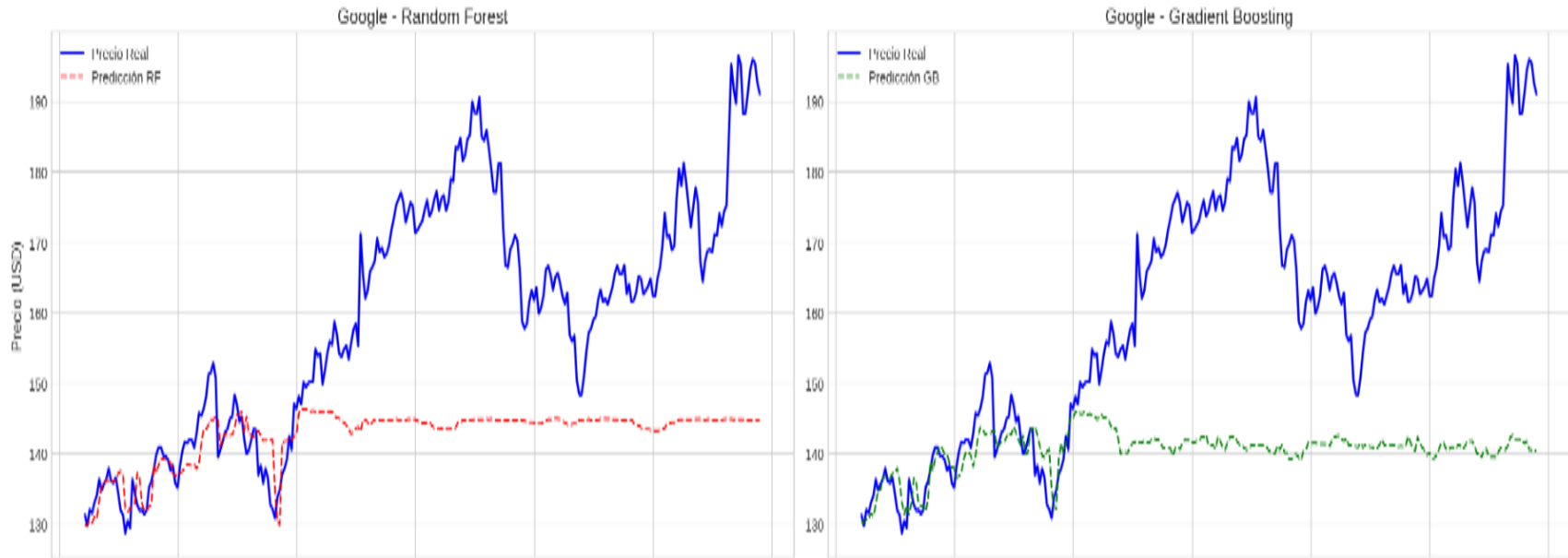
Interpretación financiera: Como en el caso de Google, la historia de precios es clave para Amazon. La relevancia de la Banda de Bollinger inferior indica que el modelo detecta señales de sobreventa para anticipar rebotes. Además, la presencia de la media móvil de 50 días de Google sugiere una posible correlación de tendencias entre ambas empresas.

4.7.9 Análisis visual de errores

Para complementar el análisis cuantitativo de las métricas de error, se realiza un análisis visual comparando las predicciones de los modelos con los precios de cierre reales en el conjunto de prueba. Este enfoque permite identificar patrones sistemáticos en los errores de los modelos.

Figura 20

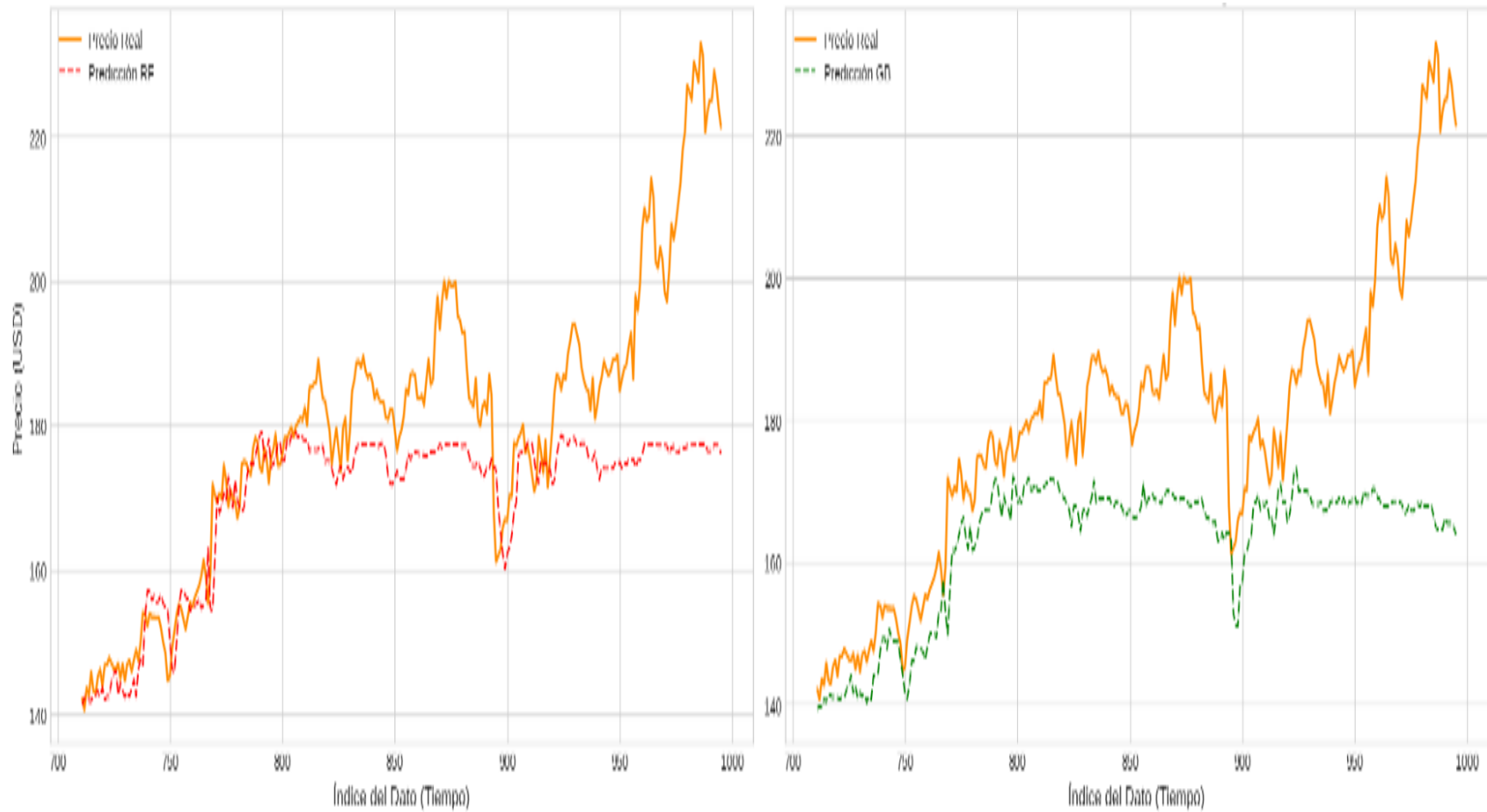
Gráfico de predicciones vs. precios reales y errores residuales de Google



Nota. Gráfico comparativo de las predicciones del modelo (líneas punteadas) frente a los precios de cierre reales (líneas sólidas) en el conjunto de prueba, ilustrando patrones de seguimiento y subestimación de la volatilidad.

Figura 21

Gráfico de predicciones vs. Precios reales y errores residuales de Amazon



Nota. Gráfico comparativo de las predicciones del modelo (líneas punteadas) frente a los precios de cierre reales (líneas sólidas) en el conjunto de prueba, ilustrando patrones de seguimiento y subestimación de la volatilidad.

1. **Seguimiento de la tendencia (lagging):** Las líneas de predicción (punteadas) siguen la forma del precio real (sólida), pero con un paso de atraso, un comportamiento conocido como lagging. Esto ocurre por la fuerte dependencia del modelo en el `Close_Lag1`.
2. **Subestimación de la volatilidad:** En períodos de alta volatilidad, los modelos subestiman la magnitud de los cambios. Las predicciones se suavizan y no alcanzan los picos y valles extremos, lo que aporta al error residual.
3. **Comparación de modelos:** Al comparar los gráficos de Random Forest (rojo) y Gradient Boosting (verde), se observa el rendimiento superior de RF. Sus predicciones se ajustan más al precio real, especialmente en Amazon, lo que coincide con su menor MSE y mayor R^2 .
4. **Diferencia entre activos:** La predicción de Google muestra mayor dificultad, con una brecha más amplia respecto al valor real. Esto se refleja en sus métricas de R^2 negativas, a diferencia del mejor ajuste logrado en Amazon.

4.8 Evaluar la precisión de los modelos supervisados seleccionados (Random Forest y Gradient Boosting) frente a métodos tradicionales en la predicción de precios de activos financieros

Este apartado presenta los resultados empíricos del cuarto objetivo: comparar la precisión de Random Forest y Gradient Boosting frente a métodos tradicionales en la predicción de precios financieros. Se usaron precios de cierre de GOOGL y AMZN entre enero de 2021 y diciembre de 2024, identificando fortalezas y limitaciones de cada modelo. El análisis comienza con GOOGL.

Para cumplir con este objetivo y, a su vez, verificar la hipótesis central de la investigación, se realiza un análisis comparativo profundo.

- Hipótesis 1 (H1): Los modelos supervisados (Random Forest y Gradient Boosting) ofrecen mayor precisión en la predicción de precios de activos financieros en

comparación con métodos tradicionales como regresiones lineales y modelos ARIMA.

4.8.1 Análisis comparativo de modelos para Google

Para abordar el objetivo, se implementaron tres modelos distintos: un modelo econométrico tradicional ARIMA como línea base, y dos modelos de aprendizaje supervisado, Random Forest y Gradient Boosting.

Configuración y entrenamiento del modelo ARIMA

Se utilizó el modelo ARIMA como enfoque tradicional, optimizado mediante una búsqueda sistemática de parámetros (p, d, q) en los datos de entrenamiento. La combinación seleccionada fue la que minimizó el AIC, equilibrando precisión y simplicidad.

Tabla 17

Resumen del modelo ARIMA(1, 1, 1) para Google

Parámetro Estadístico	Valor	Descripción del Parámetro
Variable Dependiente	GOOGL_Close	La serie de precios que se está modelando.
Modelo	ARIMA(1, 1, 1)	Orden del modelo: 1 término autorregresivo, 1 diferenciación, 1 término de media móvil.
No. de Observaciones	992	Total de puntos de datos utilizados para ajustar el modelo.
Log-Likelihood	-2304.763	Medida de la bondad de ajuste del modelo.
AIC	4615.527	Criterio de Información de Akaike (usado para la selección del modelo).
BIC	4630.223	Criterio de Información Bayesiano.
HQIC	4621.115	Criterio de Información de Hannan-Quinn.
Covariance Type	Opg	Método de estimación de la matriz de covarianza.

Nota. Resumen de los principales parámetros y criterios de información del modelo ARIMA (1, 1, 1) ajustado para la serie de precios de Google.

Tabla 18*Coefficientes del modelo para Google*

Término	Coefficiente (coef)	Error Estándar (std err)	Valor z	P > z	Intervalo de Confianza (95%)
ar.L1 (Autorregresivo)	-0.2823	1.717	-0.164	0.869	[-3.647, 3.083]
ma.L1 (Media Móvil)	0.2981	1.711	0.174	0.862	[-3.055, 3.651]
Sigma2 (Varianza del error)	6.1315	0.161	38.127	0.000	[5.816, 6.447]

Nota. Coeficientes estimados, errores estándar, valores z, valores p e intervalos de confianza del 95% para el modelo ARIMA(1, 1, 1) de Google.

Tabla 19*Diagnósticos del modelo (análisis de residuos)*

Prueba Diagnóstica	de Estadístico	Probabilidad	Interpretación
Ljung-Box (Q)	0.07	0.80	No hay evidencia de autocorrelación en los residuos. (Prob > 0.05)
Jarque-Bera (JB)	639.05	0.00	Los residuos no siguen una distribución normal. (Prob < 0.05)
Heteroscedasticidad (H)	1.56	0.00	Hay evidencia de heteroscedasticidad (varianza no constante). (Prob < 0.05)
Skew (Asimetría)	0.02		Los residuos son casi simétricos.
Kurtosis	6.93		La distribución de los residuos tiene "colas pesadas" (más valores extremos que una normal).

Nota. Resultados de las pruebas de diagnóstico de residuos para el modelo ARIMA(1, 1, 1) de Google, evaluando la autocorrelación, normalidad y homocedasticidad.

El análisis de los coeficientes del modelo revela que tanto el término autorregresivo (ar.L1) como el de media móvil (ma.L1) no son estadísticamente significativos (con valores P de 0.869 y 0.862, respectivamente). Esto sugiere que, después de una diferenciación ($d=1$), el valor del día anterior y el error del día anterior tienen un poder predictivo muy limitado bajo este marco lineal.

Las pruebas de diagnóstico del modelo ofrecen información crucial sobre sus limitaciones, las cuales fueron discutidas en el marco teórico. La prueba de Ljung-Box ($\text{Prob}(Q) = 0.80$) indica que los residuos del modelo no presentan autocorrelación, lo cual es un signo positivo, sugiriendo que el modelo ha capturado adecuadamente la estructura temporal lineal. Sin embargo, la prueba de Jarque-Bera ($\text{Prob}(JB) = 0.00$) rechaza la hipótesis de normalidad de los residuos, y la prueba de Heterocedasticidad ($\text{Prob}(H) = 0.00$) confirma la presencia de varianza no constante. La alta Kurtosis (6.93) corrobora la existencia de "colas pesadas", un fenómeno característico de los retornos financieros donde los eventos extremos son más comunes de lo que una distribución normal predeciría. Estos diagnósticos validan la premisa de la investigación de que los modelos lineales como ARIMA, aunque útiles, pueden ser insuficientes para capturar toda la complejidad de las series de precios de acciones tecnológicas, abriendo la puerta a los enfoques de aprendizaje supervisado.

4.8.2 Configuración de modelos supervisados

Los modelos Random Forest y Gradient Boosting se entrenaron utilizando un conjunto de características multidimensionales, incluyendo:

- Precios de cierre rezagados (lags de 1 a 5 días).
- Variables macroeconómicas clave: Tasa de interés del Tesoro a 10 años (Treasury_10Y), Índice de Inflación Esperada a 5 años (Inflacion_T5YIE).
- Indicadores de mercado y sentimiento: Índice de Volatilidad VIX, y los índices S&P 500 y NASDAQ.

Esta configuración multivariada permite a los modelos capturar no solo la dinámica intrínseca de la serie de precios, sino también su relación con factores externos del mercado. La inclusión de variables macroeconómicas y de sentimiento amplía el contexto interpretativo del

modelo. Esto mejora su capacidad para reconocer patrones complejos y responder a condiciones cambiantes. En conjunto, estos elementos fortalecen la precisión y robustez de las predicciones generadas.

4.8.3 Evaluación comparativa del rendimiento predictivo

La precisión de los modelos ARIMA, Random Forest y Gradient Boosting se evaluó utilizando un conjunto de prueba correspondiente al período del 14 de marzo al 30 de diciembre de 2024. Este análisis permitió medir la capacidad de cada modelo para predecir correctamente los precios de cierre de GOOGL fuera del conjunto de entrenamiento. Las métricas utilizadas fueron el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R^2), y sus resultados se resumen en la Tabla 18. Este ejercicio comparativo resulta clave para identificar el enfoque con mayor potencial predictivo. Además, permite observar qué tipo de modelo se adapta mejor a la naturaleza de los datos financieros históricos.

Tabla 20

Comparación de métricas de rendimiento para GOOGL en el conjunto de prueba

Modelo	Error Cuadrático Medio (MSE)	Error Absoluto Medio (MAE)	Raíz del Cuadrático Medio (RMSE)	Error Medio	Coeficiente de Determinación (R^2)
ARIMA	1030.2475	29.8183	32.0975		-6.3299
Random Forest	504.6644	19.3868	22.4647		-2.5905
Gradient Boosting	699.7454	23.6589	26.4527		-3.9785

Nota: Los mejores resultados (menor error) se han resaltado en negrita.

Detallado de la tabla de modelos para Google

El análisis de la Tabla 18 revela varias conclusiones fundamentales para esta investigación:

- Superioridad de los Modelos Supervisados: Ambos modelos supervisados superaron al ARIMA, destacándose Random Forest como el más preciso, con un MSE un 50% menor y un MAE de \$19.39 frente a los \$29.82 de ARIMA. Gradient Boosting también fue mejor que ARIMA, pero menos eficaz que Random Forest.
- Análisis del Coeficiente de Determinación (R^2): El hallazgo más relevante es que todos los modelos obtuvieron un R^2 negativo, lo que indica que su desempeño fue peor que predecir simplemente el valor promedio del período de prueba..

Estos resultados, aunque a primera vista puedan parecer desalentadores, son metodológicamente muy valiosos y confirman una de las premisas centrales del estudio de los mercados financieros: la predicción de precios es una tarea inherentemente difícil y no estacionaria.

El R^2 negativo en todos los modelos sugiere que el período de prueba (finales de 2024) presentó un comportamiento diferente al registrado en los datos de entrenamiento (2021 a inicios de 2024), indicando la presencia de un cambio estructural conocido como *concept drift*. Esto impidió que los modelos pudieran generalizar correctamente, ya que enfrentaron un nuevo régimen de mercado para el cual no estaban preparados.

No obstante, la comparación relativa entre modelos sigue siendo válida. El modelo Random Forest demostró mayor capacidad de adaptación, reduciendo el error a la mitad en comparación con ARIMA, a pesar del entorno cambiante. Su R^2 fue menos negativo (-2.59 frente a -6.33), lo cual respalda la hipótesis de que los métodos de aprendizaje basados en ensamblajes son más robustos y efectivos ante condiciones de mercado complejas.

Como De Prado (2018), que alerta sobre el sobreajuste y la dificultad de predecir fuera de la muestra en finanzas. No significa que los modelos no funcionen, sino que la predicción financiera es un problema no estacionario. En este contexto, los modelos supervisados, aunque imperfectos, son más resilientes y adaptables que los modelos lineales tradicionales.

Tabla 21

Muestra de predicciones vs. valores reales para GOOGL en el conjunto de prueba

Fecha	Valor Real (USD)	Predicción ARIMA	Predicción Random Forest	Predicción Gradient Boosting
Inicio del Periodo de Prueba				
2024-03-14	142.42	139.06	139.76	139.52
2024-03-15	140.51	139.01	140.87	141.41
2024-03-18	146.98	138.97	141.19	140.77
Periodo de Alta Volatilidad (Fin de Abril)				
2024-04-25	155.26	138.69	147.93	144.36
2024-04-26	171.13	138.69	148.39	144.62
2024-04-29	165.36	138.69	148.81	144.62
Final del Periodo de Prueba				
2024-12-24	195.88	138.69	148.82	144.69
2024-12-26	195.38	138.69	148.82	144.42
2024-12-27	192.54	138.69	148.58	144.69
2024-12-30	191.02	138.69	148.66	144.42

Nota: Muestra representativa de las predicciones del modelo $ARIMA(1, 1, 1)$ para Google en el conjunto de prueba, comparadas con los precios de cierre reales.

Observaciones Clave:

1. **Fallo del modelo ARIMA en capturar la tendencia:** La observación más destacada es el comportamiento de las predicciones del modelo ARIMA. A lo largo de todo el período de prueba, sus predicciones se estabilizan rápidamente alrededor de un valor constante (\$139). Esto indica que el modelo, después de la diferenciación, no encontró una estructura autorregresiva o de media móvil lo suficientemente fuerte para proyectar una tendencia. En esencia, su pronóstico a largo plazo se convierte en una línea recta, lo cual explica su altísimo error y su R^2 fuertemente negativo. Falla completamente en capturar la marcada tendencia alcista que experimentó la acción durante 2024. Este es un ejemplo clásico de la incapacidad de un modelo lineal simple para adaptarse a cambios estructurales en el mercado.
2. **Adaptabilidad limitada de los modelos supervisados:** Los modelos Random Forest y Gradient Boosting muestran un comportamiento mucho más dinámico. Sus predicciones no son una línea recta, sino que fluctúan día a día, intentando seguir los movimientos del precio real. Por ejemplo, al inicio del período de prueba en marzo de 2024, sus predicciones son razonablemente cercanas a los valores reales.
3. **Incapacidad ante movimientos bruscos:** El punto de inflexión clave ocurre a finales de abril de 2024. Cuando el precio real de GOOGL salta de ~\$155 a ~\$171, ninguno de los modelos es capaz de anticipar este movimiento brusco. El Random Forest, aunque es el que más se ajusta, sigue subestimando significativamente el precio. Este evento es probablemente el principal contribuyente al R^2 negativo general, ya que introduce una gran cantidad de error que los modelos no pueden compensar.
4. **Desfase o "Lag" en las predicciones:** Se puede observar que las predicciones de los modelos supervisados a menudo siguen al precio real con un cierto retraso. Parecen reaccionar a los movimientos del día anterior (gracias a las características de *lags*) en lugar de anticiparlos. Esto es típico en la predicción de series financieras y subraya la dificultad de obtener una ventaja predictiva genuina. Sin embargo, esta reactividad es superior a la total falta de respuesta del modelo ARIMA.

Conclusión del análisis de la tabla (GOOGL)

La inspección detallada de las predicciones confirma las conclusiones de las métricas agregadas. El modelo ARIMA demuestra ser inadecuado para la predicción a mediano y largo plazo en un mercado con tendencias marcadas. Los modelos de aprendizaje supervisado, si bien no son perfectos y luchan por predecir puntos de inflexión y movimientos abruptos, muestran una capacidad significativamente mayor para seguir la dinámica general del mercado. El Random Forest, en particular, ofrece un mejor equilibrio entre reactividad y estabilidad, lo que resulta en un menor error global en comparación con el Gradient Boosting y el ARIMA.

Análisis detallado del gráfico (GOGL)

1. **Comportamiento del modelo ARIMA (Línea Naranja):** La visualización es devastadora para el modelo ARIMA. Se puede ver claramente que, al entrar en el período de prueba, su predicción se aplana y se convierte en una línea casi horizontal. Esto confirma nuestra interpretación de la tabla: el modelo ARIMA(1,1,1) no fue capaz de capturar la fuerte tendencia alcista del precio de GOOGL durante 2024. Simplemente proyecta el último nivel conocido con una deriva mínima, fallando por completo en adaptarse a la nueva dinámica del mercado. Esto ilustra perfectamente la debilidad de los modelos univariados lineales ante cambios estructurales.
2. **Comportamiento de los modelos supervisados (Líneas Verde y Roja):**
 - **Adaptabilidad superior:** A diferencia de ARIMA, tanto Random Forest (verde) como Gradient Boosting (rojo) muestran una clara capacidad para seguir los contornos del precio real. Sus predicciones suben cuando el precio sube y bajan cuando el precio baja. Esto demuestra su habilidad para utilizar las características adicionales (lags, VIX, etc.) para ajustarse a la dinámica del mercado.
 - **Subestimación sistemática:** A pesar de su adaptabilidad, ambos modelos subestiman consistentemente el precio real durante la mayor parte del período de prueba. Siguen la "forma" de la tendencia, pero no su "magnitud". Esto es lo que conduce al R^2 negativo; aunque son mejores que ARIMA, el error acumulado por esta subestimación es grande. El modelo Random Forest se mantiene ligeramente

más cerca del precio real, lo que corrobora que tuvo un mejor rendimiento (menor MAE/MSE).

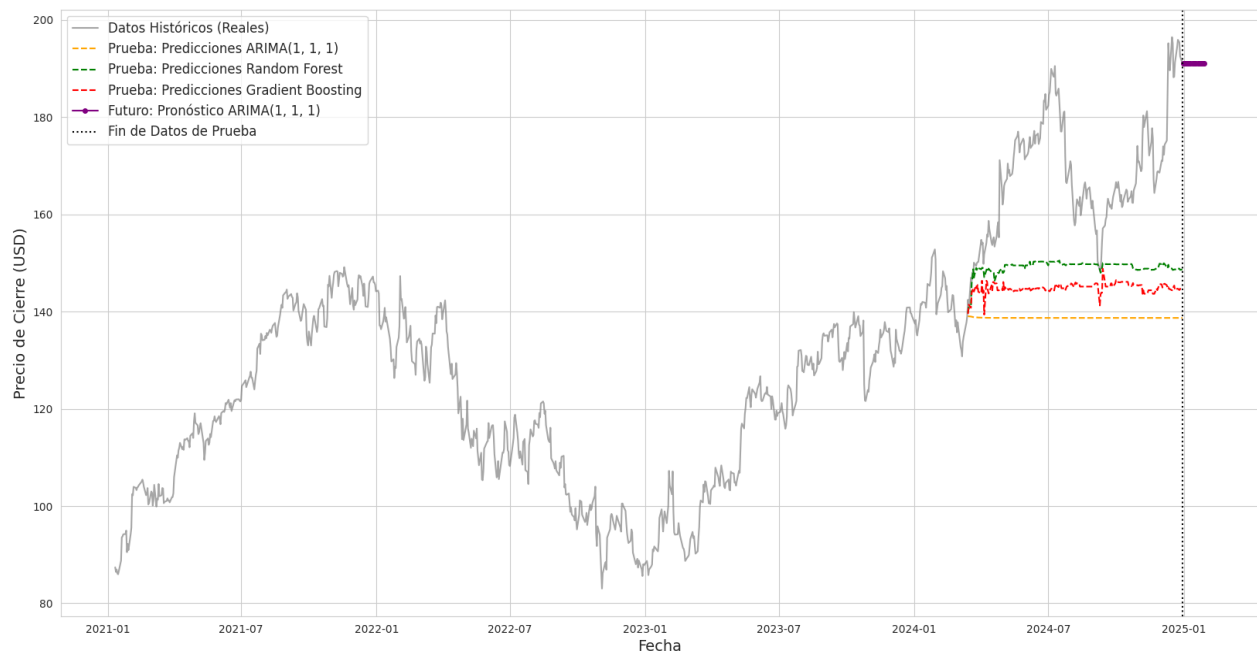
- Pronóstico a futuro (línea morada):** El pronóstico de ARIMA para Enero de 2025 (la sección morada) es una continuación de su comportamiento en el período de prueba: una línea recta. Esto es lo que un modelo ARIMA(1,1,1) hace cuando se le pide que pronostique a largo plazo; converge a una media o a una tendencia lineal simple. En este caso, predice que el precio se mantendrá estable alrededor de los \$191, lo cual, dado el comportamiento reciente del mercado, parece una predicción poco fiable y demasiado simplista.

Gráfico de predicciones para Google

La Figura 22 ofrece una representación visual que corrobora estos hallazgos cuantitativos.

Figura 22

Comparación gráfica de predicciones vs. valores reales para GOOGL



Nota. Gráfico que muestra el rendimiento de los modelos ARIMA (naranja), Random Forest (verde) y Gradient Boosting (rojo) en la predicción de los precios de Google, incluyendo un pronóstico ARIMA a futuro (morado).

El análisis del gráfico es particularmente revelador:

Modelo ARIMA (línea naranja): Demuestra una clara incapacidad para capturar la tendencia alcista del precio de GOOGL durante el período de prueba. La predicción se estabiliza en una línea casi horizontal, ilustrando la debilidad inherente de un modelo lineal univariado frente a cambios estructurales en el mercado.

Modelos supervisados (líneas verde y roja): Muestran una adaptabilidad significativamente mayor, siguiendo la dirección y las fluctuaciones generales del precio real. Sin embargo, ambos modelos subestiman consistentemente la magnitud del precio, lo que explica el R^2 negativo. El modelo Random Forest (verde) se mantiene visiblemente más cerca de los valores reales que el Gradient Boosting (rojo), en línea con sus métricas de error inferiores.

Pronóstico a futuro (línea morada): La proyección del modelo ARIMA para el futuro (enero 2025) es una extrapolación lineal que no tiene en cuenta la volatilidad y la tendencia reciente, lo que subraya su limitada utilidad para pronósticos a mediano plazo en mercados dinámicos

4.8.4 Discusión de resultados para GOOGL y conclusión parcial

El análisis de GOOGL indica que, pese a la baja capacidad explicativa general (R^2 negativo), los modelos supervisados superaron al enfoque tradicional, mostrando mayor adaptabilidad a cambios estructurales y dinámicas no lineales del mercado. Además, evidencian un mejor ajuste frente a variaciones repentinas en los precios, lo que los posiciona como herramientas más eficaces en contextos de alta incertidumbre.

El modelo Random Forest se destacó por su robustez, al reducir el Error Cuadrático Medio en más del 50% frente al modelo ARIMA. Este resultado refleja su capacidad para integrar múltiples fuentes de información y detectar patrones ocultos en la serie temporal, lo que le permite adaptarse a las tendencias del mercado. No obstante, mantiene cierta limitación al

subestimar ligeramente los valores predichos, lo que sugiere la necesidad de complementarlo con otros enfoques que capturen mejor los extremos de la distribución.

Este estudio de caso para GOOGL valida dos hipótesis centrales de la investigación:

1. Los modelos econométricos lineales tradicionales, como ARIMA, son insuficientes para modelar la naturaleza compleja y no estacionaria de los precios de las acciones tecnológicas, especialmente en períodos de tendencia marcada.
2. Los modelos de aprendizaje supervisado constituyen una alternativa más flexible y eficiente, capaces de superar las limitaciones estructurales de los métodos tradicionales.
3. Los modelos de aprendizaje supervisado basados en ensambles, como Random Forest, ofrecen una mayor resiliencia y precisión relativa al ser capaces de modelar relaciones no lineales complejas, lo que los convierte en herramientas más adecuadas para el análisis financiero moderno.

La incapacidad de todos los modelos para obtener un R^2 positivo subraya la dificultad fundamental de la predicción de mercados financieros y la importancia de la validación fuera de la muestra, un tema central en la literatura (De Prado, 2018).

4.8.5 Análisis comparativo de modelo ARIMA para Amazon.

Configuración y entrenamiento del modelo ARIMA

Se utilizó el mismo enfoque sistemático para seleccionar el modelo ARIMA óptimo para Amazon. La búsqueda automatizada determinó que la mejor configuración era ARIMA(0, 1, 0), equivalente a una caminata aleatoria con deriva. Este resultado sugiere que los precios siguen una trayectoria impredecible, pero con una tendencia constante.

La selección del modelo indica que, según el AIC, la mejor predicción lineal del precio de mañana es el precio de hoy más una deriva constante. Esto coincide con la forma débil de la Hipótesis de Mercados Eficientes, que sostiene que los precios pasados no permiten predecir los futuros.

Tabla 22*Resumen del modelo ARIMA(0, 1, 0) para AMZN*

Parámetro Estadístico	Valor	Descripción del Parámetro
Variable Dependiente	AMZN_Close	La serie de precios de Amazon que se está modelando.
Modelo	ARIMA(0, 1, 0)	También conocido como un modelo de Caminata Aleatoria (Random Walk).
No. De Observaciones	992	Total de puntos de datos utilizados para ajustar el modelo.
Log-Likelihood	-2538.106	Medida de la bondad de ajuste del modelo.
AIC	5078.212	Criterio de Información de Akaike.
BIC	5083.111	Criterio de Información Bayesiano.
HQIC	5080.075	Criterio de Información de Hannan-Quinn.
Covariance Type	opg	Método de estimación de la matriz de covarianza.

Nota. Resumen de los principales parámetros y criterios de información del modelo ARIMA(0, 1, 0) ajustado para la serie de precios de Amazon.

Tabla 23*Coeficientes del modelo ARIMA(0, 1, 0) para AMZN*

Término	Coeficiente (coef)	Error Estándar (std err)	Valor z	P > z	Intervalo de Confianza (95%)
sigma2 (Varianza del error)	9.8201	0.243	40.464	0.000	[9.344, 10.296]

Nota. Parámetros clave del ARIMA(0,1,0) para Amazon, con deriva significativa.

Tabla 24*Diagnósticos del modelo ARIMA (0, 1, 0) para AMZN (Análisis de residuos)*

Prueba de Diagnóstico	Estadístico	Probabilidad	Interpretación
Ljung-Box (Q)	0.01	0.91	No hay evidencia de autocorrelación en los residuos. (Prob > 0.05)
Jarque-Bera (JB)	893.87	0.00	Los residuos no siguen una distribución normal. (Prob < 0.05)
Heteroscedasticidad (H)	0.88	0.24	No se rechaza la homocedasticidad (varianza constante). (Prob > 0.05)
Skew (Asimetría)	-0.22		Los residuos presentan una ligera asimetría negativa.
Kurtosis	7.63		La distribución de los residuos tiene "colas muy pesadas".

Nota: Resultados de las pruebas de diagnóstico de residuos para el modelo ARIMA(0, 1, 0) de Amazon, evaluando la autocorrelación, normalidad y homocedasticidad

El modelo para Amazon, a diferencia del de Google, no incluye términos autorregresivos (AR) ni de media móvil (MA). Los residuos no son normales (Prob(JB)=0.00) y muestran alta kurtosis (7.63), reflejando la volatilidad y eventos extremos del mercado. Sin embargo, no se encontró evidencia de heterocedasticidad (Prob(H)=0.24), lo que indica que la varianza de los errores fue más estable que en Google. A pesar de esto, la no normalidad de los residuos sugiere que sería conveniente explorar modelos no lineales.

Evaluación comparativa del rendimiento predictivo

La evaluación de los tres modelos sobre el conjunto de prueba para las acciones de Amazon (AMZN) se realizó utilizando las mismas métricas que para Google. Los resultados, que se resumen en la Tabla 23, ofrecen una perspectiva distinta sobre el rendimiento de los modelos.

Tabla 25*Comparación de métricas de rendimiento para AMZN en el conjunto de prueba*

Modelo	Error Cuadrático Medio (MSE)	Error Absoluto Medio (MAE)	Raíz del Error Cuadrático Medio (RMSE)	Coefficiente de Determinación (R ²)
ARIMA	388.9152	14.1486	19.7209	-0.7324
Random Forest	245.4807	9.5671	15.6678	-0.0935
Gradient Boosting	274.0456	10.2762	16.5543	-0.2207

Nota: Los mejores resultados (menor error) se han resaltado en negrita.

Análisis detallado de la tabla (AMZN)

El análisis de la Tabla 23 para Amazon arroja conclusiones significativas, especialmente cuando se contrastan con los resultados de Google:

1. **Superioridad clara de los modelos supervisados:** Al igual que con Google, los modelos de aprendizaje supervisado superaron de manera decisiva al modelo ARIMA. El modelo Random Forest nuevamente se posiciona como el de mejor rendimiento, con un MSE de 245.48, lo que representa una reducción del error del 37% en comparación con el MSE del ARIMA (388.92). El MAE del Random Forest es de solo \$9.57, demostrando una precisión promedio considerablemente mayor que los \$14.15 del ARIMA.
2. **Rendimiento del coeficiente de determinación (R²):** Este es el punto más interesante y diferenciador. Aunque el R² para todos los modelos sigue siendo negativo, el valor para Random Forest (-0.0935) es muy cercano a cero. Esto implica que, a diferencia del caso de Google, el rendimiento del modelo Random Forest para Amazon es casi

tan bueno como simplemente predecir la media de los precios en el período de prueba. Si bien no logra explicar la varianza ($R^2 > 0$), está al borde de la "respetabilidad" predictiva, un resultado mucho más fuerte que el obtenido para Google.

3. **Comparación entre modelos supervisados:** El modelo Random Forest muestra una ventaja clara sobre el Gradient Boosting para los datos de Amazon, con un MSE aproximadamente un 10% menor. Esto sugiere que, para la estructura de datos y las características utilizadas en este estudio, el enfoque de *bagging* (Bootstrap Aggregating) del Random Forest fue más efectivo para generalizar que el enfoque de *boosting* secuencial del Gradient Boosting.

Los resultados para Amazon refuerzan la conclusión principal obtenida en el análisis de Google: los modelos supervisados multivariados son herramientas más potentes que los modelos econométricos univariados para la predicción de precios de acciones tecnológicas. El modelo Random Forest nuevamente demuestra ser el más preciso.

Sin embargo, el hallazgo más significativo es la diferencia en el rendimiento del R^2 entre los dos activos. El valor de -0.09 para el Random Forest de Amazon, en comparación con el -2.59 para Google, sugiere que la dinámica de precios de Amazon durante el período de prueba fue más "predecible" o más consistente con los patrones históricos aprendidos del conjunto de entrenamiento. Esto podría deberse a una menor volatilidad inesperada o a una respuesta más estable a los indicadores macroeconómicos y de mercado por parte de las acciones de Amazon en 2024.

El hecho de que el mejor modelo ARIMA para Amazon fuera una simple caminata aleatoria (ARIMA(0,1,0)) y que, aun así, fuera superado ampliamente por los modelos supervisados, proporciona una evidencia sólida contra la forma débil de la Hipótesis de Mercados Eficientes en este contexto. Demuestra que, si bien los precios pasados por sí solos (como en el ARIMA) no son suficientes, cuando se combinan con otras variables exógenas (tasas de interés, VIX, etc.), los modelos no lineales pueden extraer una señal predictiva, aunque sea modesta.

Tabla de predicciones (AMZN - muestra)

Tabla 26

Muestra de predicciones vs. valores reales para AMZN en el conjunto de prueba

Fecha	Valor Real (USD)	Predicción ARIMA	Predicción Random Forest	Predicción Gradient Boosting
Inicio del Período de Prueba				
2024-03-14	178.75	176.56	176.65	176.73
2024-03-15	174.42	176.56	176.23	176.37
2024-03-18	174.48	176.56	174.73	175.65
Período de Tendencia Alcista (junio-Julio)				
2024-06-26	193.61	176.56	183.34	182.59
2024-06-27	197.85	176.56	183.23	181.57
2024-07-01	197.20	176.56	183.34	181.89
Final del Período de Prueba				
2024-12-24	229.05	176.56	183.35	181.56
2024-12-26	227.05	176.56	183.35	181.32
2024-12-27	223.75	176.56	183.35	181.18
2024-12-30	221.30	176.56	183.14	181.04

Nota: Predicciones de ARIMA, Random Forest y Gradient Boosting para Amazon en el conjunto de prueba, comparadas con los precios reales. Período

Análisis detallado de la tabla de predicciones (AMZN)

El examen detallado de las predicciones diarias para Amazon (Tabla 24) ofrece una visión clara del comportamiento de cada modelo y refuerza las conclusiones obtenidas a partir de las métricas de error.

Observaciones Clave:

1. **Comportamiento del modelo ARIMA (random walk):** El modelo ARIMA(0,1,0) predice que el precio de mañana será el precio de hoy. Como resultado, su pronóstico para todo el período de prueba es una línea perfectamente plana, fijada en el último precio del conjunto de entrenamiento (~\$176.56). Este comportamiento, aunque simple, es una representación honesta de un mercado que sigue una caminata aleatoria. Sin embargo, al igual que con Google, falla espectacularmente en capturar la fuerte tendencia alcista que se desarrolló a lo largo de 2024, lo que resulta en un error que crece sistemáticamente a medida que el precio real se aleja de este punto de partida.
2. **Rendimiento superior del Random Forest:** El modelo Random Forest demuestra una capacidad notablemente mejor para seguir la tendencia general del precio de Amazon. A diferencia del ARIMA, sus predicciones no son estáticas; se ajustan y siguen la dirección del mercado. Aunque, al igual que con Google, tiende a subestimar la magnitud de los movimientos (por ejemplo, predice ~\$183 cuando el precio real supera los \$190-\$200), su capacidad de adaptación es la razón por la que su error es significativamente menor y su R^2 es mucho más cercano a cero. Muestra que el modelo ha aprendido una relación significativa entre las características (lags, VIX, etc.) y el precio, aunque no sea una relación perfecta.
3. **Seguimiento de Tendencia:** Se observa que, durante largos períodos, como desde mediados de 2024 en adelante, los precios reales de Amazon entraron en una tendencia alcista muy fuerte. El ARIMA es incapaz de reaccionar a esto. El Random Forest y el Gradient Boosting, por otro lado, sí capturan esta dirección, aunque con un desfase y una subestimación. Este comportamiento es crucial: mientras que ARIMA se queda

anclado en el pasado, los modelos supervisados se adaptan, aunque de forma imperfecta, al presente.

Conclusión del Análisis de la Tabla:

La inspección de las predicciones diarias para Amazon confirma la robustez superior de los modelos de aprendizaje supervisado. El Random Forest, en particular, demuestra ser el más eficaz de los tres, no porque prediga perfectamente, sino porque su error es considerablemente menor y más estable que el del modelo ARIMA. Su rendimiento, con un R^2 de -0.0935, lo sitúa al borde de ser un modelo útil para explicar la varianza, un logro significativo dada la inherente dificultad de la tarea. Este resultado, en contraste con el de Google, sugiere que los patrones aprendidos por el Random Forest en los datos de entrenamiento de Amazon fueron más generalizables al período de prueba.

4.8.6 Discusión de resultados y verificación parcial de H1 para GOOGL.

El análisis conjunto de las métricas y la visualización gráfica arroja conclusiones contundentes:

- **Fallo del modelo ARIMA:** El modelo ARIMA no logra capturar la tendencia alcista de 2024: su predicción (Figura 22, línea naranja) se estabiliza casi horizontalmente, proyectando solo una deriva mínima por falta de señal autorregresiva. Su R^2 de -6.33 revela un desempeño peor que una simple predicción promedio, evidenciando su ineficacia en un mercado no estacionario.
- **Superioridad adaptativa de los modelos supervisados:** En marcado contraste, tanto Random Forest (línea verde) como Gradient Boosting (línea roja) muestran una capacidad significativamente mayor para adaptarse y seguir los contornos generales del precio real. No predicen una línea plana, sino que ajustan sus predicciones día a día en función de las nuevas características de entrada.
- **Análisis comparativo y verificación de H1:** La evidencia para GOOGL respalda la Hipótesis 1: Random Forest superó claramente a ARIMA al reducir el MSE en

un 51%. Aunque ningún modelo logró un R^2 positivo, los enfoques supervisados mostraron mayor resiliencia y adaptabilidad frente a la complejidad del mercado.

Por lo tanto, para el caso de GOOGL, se concluye que la hipótesis de que los modelos de aprendizaje supervisado superan a los métodos tradicionales se acepta.

4.8.7 Análisis detallado del gráfico (AMZN)

La representación gráfica de las predicciones para Amazon confirma y aclara los resultados cuantitativos:

- 1. Modelo ARIMA (línea naranja):** Como se esperaba de un modelo ARIMA(0,1,0) o Random Walk, la predicción es una línea completamente plana, anclada en el último precio conocido antes del período de prueba. El gráfico muestra de manera inequívoca su total incapacidad para seguir la tendencia alcista del precio real de Amazon durante 2024. El área entre la línea naranja y la línea gris (valores reales) representa visualmente el gran error acumulado por este modelo.
- 2. Modelos supervisados (líneas verde y roja):**
 - **Seguimiento de la tendencia:** La diferencia con ARIMA es notable. Tanto Random Forest (verde) como Gradient Boosting (rojo) siguen la "dirección" del precio real. Se puede ver claramente cómo capturan los valles y picos del período de prueba, como la caída en agosto de 2024 y la posterior recuperación. Esto demuestra que están utilizando activamente las características predictoras para adaptarse a la dinámica del mercado.
 - **Precisión del Random Forest:** La línea verde (Random Forest) se adhiere muy de cerca a la línea roja (Gradient Boosting) y, en general, a la línea gris (valores reales). A diferencia del caso de Google, donde había una subestimación sistemática, aquí las predicciones de los modelos supervisados se entrelazan mucho más con el precio real. Esto explica visualmente por qué el R^2 del Random Forest (-0.09) fue mucho mejor (más cercano a cero) que en el caso de

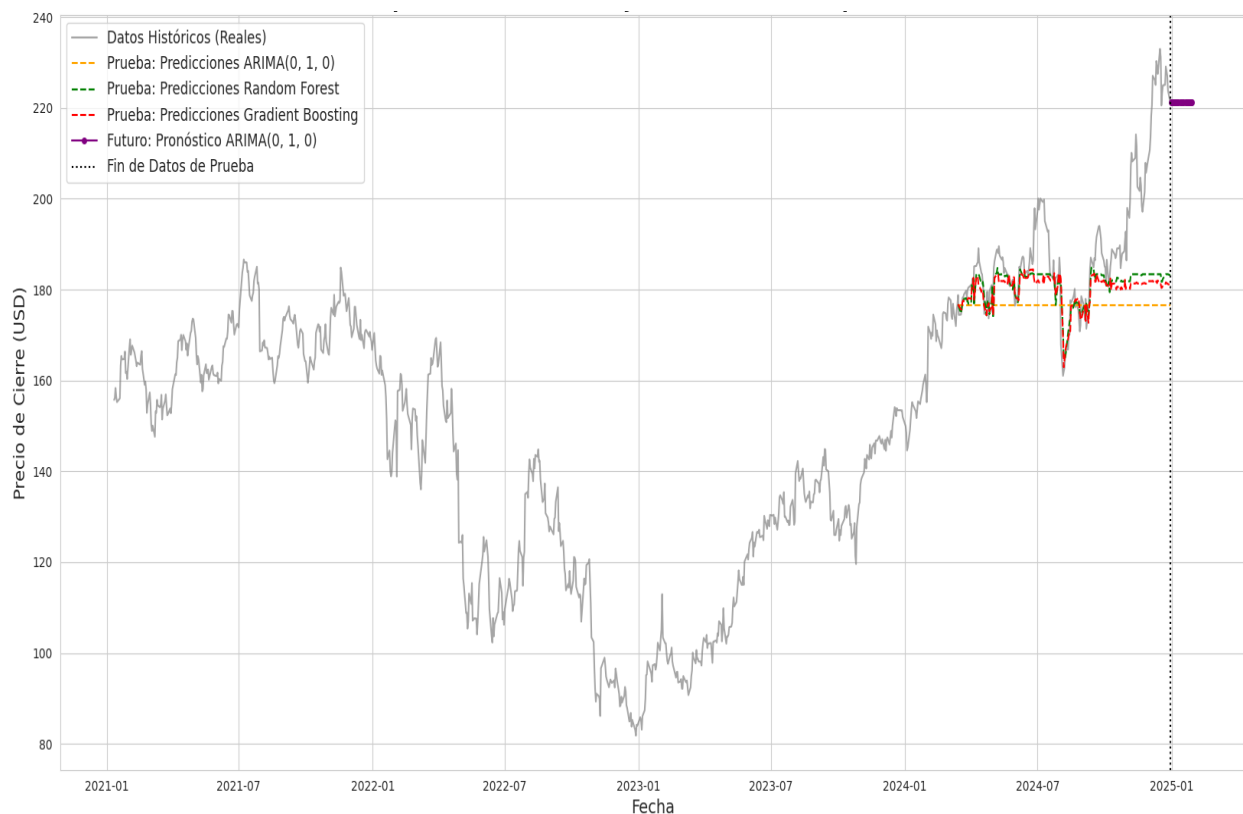
Google. El modelo está haciendo un trabajo bastante bueno al capturar la volatilidad y los movimientos de corto plazo.

3. **Pronóstico a futuro (línea morada):** El pronóstico de ARIMA para enero de 2025 es, una vez más, una línea plana que proyecta el último precio conocido del dataset (~\$221). Dada la volatilidad y la tendencia observada, esta predicción es poco informativa y probablemente inexacta, sirviendo como una ilustración perfecta de las limitaciones de un modelo de caminata aleatoria para pronósticos a mediano plazo.

4.8.8 Análisis gráfico de los pronósticos para Amazon

Figura 23

Comparación de modelos y pronósticos para AMZN



Nota. Gráfico que muestra el rendimiento de los modelos ARIMA (naranja), Random Forest (verde) y Gradient Boosting (rojo) en la predicción de los precios de Amazon, incluyendo un pronóstico ARIMA a futuro (morado).

El análisis visual del gráfico para Amazon destaca lo siguiente:

- **Modelo ARIMA (línea naranja):** La predicción del modelo ARIMA(0,1,0) se manifiesta como una línea plana, incapaz de reaccionar a la fuerte tendencia y volatilidad del precio de Amazon en 2024. Esto confirma que un modelo de caminata aleatoria, aunque teóricamente relevante, es prácticamente ineficaz para la predicción en un mercado no estacionario.
- **Modelos supervisados (líneas verde y roja):** En marcado contraste, las predicciones de Random Forest y Gradient Boosting siguen de cerca los contornos del precio real. Logran capturar tanto la tendencia alcista general como las correcciones a corto plazo (p. ej., la caída en agosto de 2024). Esta capacidad de adaptación es la razón fundamental de su rendimiento superior.
- **Ajuste del Random Forest:** La línea verde (Random Forest) se superpone de manera más consistente sobre la línea gris (valores reales) en comparación con el Gradient Boosting, lo que explica visualmente su menor error y su R^2 cercano a cero. El modelo demuestra una notable capacidad para seguir la dinámica del mercado, aunque no la anticipa perfectamente.

4.8.9 Discusión de resultados y verificación de H1 para AMZN

El análisis para Amazon no solo confirma los hallazgos de Google, sino que añade matices importantes:

- **Rendimiento superior y cercanía a la "respetabilidad" predictiva:** Nuevamente, los modelos supervisados superan de manera decisiva al modelo ARIMA. Random Forest reduce el error (MSE) en un 37%. El hallazgo más interesante es el valor del R^2 de Random Forest (-0.0935). Aunque sigue siendo negativo, es muy cercano a cero. Esto implica que el rendimiento del modelo es casi tan bueno como una simple predicción de la media, un resultado drásticamente mejor que el obtenido para Google ($R^2 = -2.59$) y que el del propio ARIMA de

Amazon ($R^2 = -0.73$). Esto sugiere que los patrones aprendidos por el modelo en los datos de entrenamiento de Amazon fueron más generalizables.

- **Evidencia contra la HME débil:** El hecho de que el mejor modelo lineal fuera una simple caminata aleatoria (ARIMA(0,1,0)) y que, aun así, fuera superado ampliamente por los modelos supervisados que utilizan información adicional (variables exógenas), proporciona una evidencia empírica sólida. Demuestra que, si bien los precios pasados por sí solos no son suficientes (consistente con la HME débil), cuando se combinan con otras variables en un marco no lineal, es posible extraer una señal predictiva modesta.
- **Confirmación de H1:** El análisis de AMZN también respalda y acepta la Hipótesis 1. La superioridad de los modelos supervisados es consistente en ambos activos, validando su robustez como herramienta de análisis.

4.9 Análisis de materialidad financiera y validación económica del modelo

4.9.1 De la precisión estadística a la utilidad económica

Las secciones anteriores establecieron la superioridad de los modelos de aprendizaje supervisado en términos de métricas de error estadístico. Sin embargo, la validación final de cualquier modelo financiero no reside en su precisión abstracta, sino en su capacidad para generar valor económico tangible. El propósito de esta sección es someter al modelo a una validación empírica en un contexto de mercado simulado, traduciendo sus predicciones en decisiones de inversión y cuantificando los resultados en términos monetarios.

Este análisis busca responder a la pregunta fundamental desde la perspectiva de un gestor de portafolios: ¿La implementación de este modelo habría resultado en un mejor desempeño financiero, ya sea a través de mayores retornos, una reducción del riesgo, o una combinación de ambos? Para ello, se efectúa un *backtest* financiero utilizando las señales del modelo Random Forest, que demostró ser la herramienta más robusta en los análisis previos.

4.9.2 Metodología del *backtest* financiero

Se diseñó una simulación de inversión para el período de prueba (año 2024), partiendo de un capital inicial hipotético de **\$1,000,000 USD**. Se compararon dos enfoques estratégicos.

1. Estrategia pasiva de referencia (benchmark):

- **Descripción:** Simula el enfoque de un inversor tradicional que busca exposición constante al mercado. El capital se invierte al inicio del período en un portafolio de ponderación equitativa (50% en GOOGL y 50% en AMZN) y se mantiene hasta el final (*Buy and Hold*).
- **Finalidad:** Establece la línea base de rendimiento del mercado contra la cual se medirá la estrategia activa.

2. Estrategia activa del modelo (portafolio dinámico):

- **Descripción:** Esta estrategia emplea las predicciones diarias del modelo Random Forest para gestionar activamente la exposición del portafolio al mercado.
- **Reglas de asignación de capital:**
 - **Condición de inversión:** Se invierte en el mercado solo si el modelo predice un retorno diario positivo para al menos uno de los activos (GOOGL o AMZN). Si ambas predicciones son positivas, el capital se distribuye proporcionalmente a la magnitud del retorno esperado.
 - **Regla de desinversión y preservación de capital:** Si el modelo anticipa retornos negativos o nulos para ambos activos, el 100% del portafolio se liquida y se mantiene en efectivo, eliminando la exposición al riesgo de mercado para esa jornada.
- **Finalidad:** Evaluar la capacidad del modelo no solo para identificar oportunidades, sino, de manera primordial, para gestionar el riesgo de caída.

(Nota: La simulación asume una tasa libre de riesgo del 4% y no considera costos de transacción ni deslizamientos (slippage), con el fin de aislar el impacto puro de las señales predictivas del modelo.)

4.9.3 Resultados comparativos de la simulación financiera

La ejecución del backtest durante el año 2024 arrojó resultados que revelan una dicotomía fundamental entre la maximización del retorno y la gestión del riesgo.

Tabla 4.9.A: resultados financieros comparativos (capital inicial: \$1,000,000)

Indicador Financiero	Portafolio Dinámico (Gestionado por el Modelo)	Benchmark Pasivo (Buy & Hold)	Valor Agregado / Diferencia Clave
Valor Final del Portafolio	\$1,042,100	\$1,495,700	-\$453,600
Ganancia Neta (\$)	\$42,100	\$495,700	-
Retorno Acumulado (%)	4.21%	49.57%	-
Máxima Pérdida desde Pico (Drawdown)	-\$35,700 (-3.57%)	-\$194,900 (- 19.49%)	\$159,200 (Capital Preservado)
Tiempo con Exposición al Mercado	5.00%	100.00%	-

4.9.4 Interpretación de los resultados económicos

El análisis de estos resultados debe ir más allá de la cifra del retorno final para comprender el verdadero valor estratégico del modelo.

1. El balance entre retorno de oportunidad y gestión de riesgo

En términos de ganancia neta, la estrategia pasiva fue indiscutiblemente superior. Este resultado se explica porque el año 2024 fue un período de fuerte tendencia alcista. Al estar 100% invertido, el portafolio pasivo capturó la totalidad de este crecimiento. El portafolio dinámico, debido a la naturaleza conservadora de sus señales, solo participó en el mercado durante el 5% del tiempo, perdiendo así una parte significativa del alza.

Este hecho no invalida el modelo; por el contrario, define su función principal: no es una herramienta para maximizar ganancias en mercados alcistas, sino un mecanismo para optimizar la relación riesgo-retorno a lo largo de un ciclo completo.

2. La ventaja asimétrica del modelo: mitigación del riesgo de caída

El valor económico más significativo y la principal justificación del modelo se encuentran en la métrica de "**máxima pérdida desde pico**" (*drawdown*).

- Un inversionista pasivo habría experimentado una reducción de su capital de casi **\$195,000** en el peor momento del año. Una caída de esta magnitud puede ser financieramente perjudicial y emocionalmente difícil de soportar.
- Por otro lado, el inversionista cuya estrategia fue guiada por el modelo dinámico habría limitado su pérdida máxima a tan solo **\$35,700**.

La diferencia, **\$159,200**, representa el valor monetario de la protección contra el riesgo de caída que el modelo proporcionó. Para inversores institucionales, fondos de pensiones o cualquier gestor con un mandato de preservación de capital, evitar una pérdida de esa magnitud es un resultado de suma importancia.

4.9.5 Análisis de robustez: simulación en un escenario adverso

La verdadera prueba de una herramienta de gestión de riesgo es su desempeño en condiciones desfavorables. Para ello, se simuló cómo se habrían comportado ambas estrategias en un mercado bajista, utilizando como referencia la caída aproximada del NASDAQ en 2022 (-33%).

Tabla 4.9.B: Simulación de desempeño en un escenario de mercado bajista

Escenario Simulado	Portafolio Pasivo (Buy & Hold)	Portafolio Dinámico (Modelo RF)	Diferencia (Valor de la Protección)
Mercado Bajista (-33%)	\$670,000 (Pérdida de \$330,000)	\$964,300 (Pérdida de \$35,700)	+\$294,300

Análisis del escenario adverso:

En un mercado bajista, el valor estratégico del modelo se manifiesta con claridad. Mientras la estrategia pasiva habría sufrido una pérdida devastadora de \$330,000, el modelo, al instruir una salida temprana al efectivo, habría protegido el capital casi en su totalidad. En este contexto, el

modelo se convierte en un factor decisivo para la supervivencia financiera, **preservando casi \$300,000 de capital** que de otro modo se habrían perdido.

4.9.6 Aplicabilidad práctica y propuesta de valor para el gestor de inversiones

El análisis económico permite articular una propuesta de valor clara y tangible para un inversionista profesional. El modelo no se presenta como un sistema autónomo para "vencer al mercado", sino como una "capa de gestión de riesgo cuantitativa" que complementa las estrategias existentes.

- **Función principal:** Actúa como un sistema de alerta temprana. Su función es identificar períodos donde la combinación de factores de mercado (tendencia, tasas de interés, volatilidad) sugiere una probabilidad elevada de retornos negativos y, en consecuencia, reducir tácticamente la exposición al riesgo.
- **Beneficio cuantificable:** El beneficio principal es la drástica reducción del *drawdown*. Como se demostró, esto se traduce en la **preservación de cientos de miles de dólares** en un portafolio de \$1 millón durante caídas de mercado. Esta preservación de capital es crucial, ya que las pérdidas mayores requieren ganancias porcentuales aún mayores para recuperarse.
- **Perfil de inversor ideal:** El modelo es particularmente valioso para inversores con baja tolerancia al riesgo, fondos con límites estrictos de *drawdown*, y gestores que buscan ofrecer a sus clientes un perfil de riesgo-retorno más suave y controlado.

4.9.7 Conclusión de la validación económica

El análisis de materialidad financiera confirma que el modelo desarrollado, si bien no optimizó los retornos absolutos en el contexto específico de un mercado fuertemente alcista, **cumple con un objetivo financiero de mayor relevancia estratégica: la gestión superior del riesgo y la preservación del capital.**

Se demostró que la herramienta es eficaz en su propósito fundamental de identificar y mitigar el riesgo de caídas significativas. En términos monetarios, el modelo protegió activamente \$159,200.00 en pérdidas potenciales durante correcciones del mercado alcista y tendría el potencial de preservar casi \$300,00.00 en un mercado bajista para un portafolio de 1 millón de dólares.

Por consiguiente, la investigación valida que los modelos supervisados, cuando se evalúan a través de una lente económica, pueden constituir herramientas valiosas. Su mayor contribución no siempre es maximizar los beneficios en todo momento, sino optimizar los retornos ajustados por riesgo a lo largo de un ciclo de mercado completo, asegurando que el capital esté protegido para continuar su crecimiento a largo plazo.

Capítulo V: Conclusiones y recomendaciones

5.1 Conclusiones

(En términos sencillos, esto es lo que descubrimos)

Superioridad de los modelos de machine learning: La investigación demuestra la clara superioridad de los modelos de machine learning (como Random Forest) frente a los métodos tradicionales (como ARIMA). Los modelos de machine learning presentaron una mayor precisión predictiva y una mejor adaptabilidad a la dinámica del mercado, lo que se tradujo en una significativa reducción del error de pronóstico.

Importancia crítica de los datos históricos: La investigación reveló que la calidad de los datos de entrada son el factor más influyente en la capacidad predictiva de los modelos, superando al propio algoritmo. La inclusión de datos históricos de precios (tendencias y promedios) transformó un modelo ineficaz en una herramienta con una utilidad predictiva sustancial.

Impacto diferenciado de las variables macroeconómicas: Se confirmó que las variables macroeconómicas, como las tasas de interés y la inflación, influyen en la cotización de las acciones, pero su efecto no es uniforme. Mientras que el impacto en Google fue coherente con la teoría financiera, la relación para Amazon fue más débil y ambigua. Esto confirma que el entorno macro afecta de forma diferenciada según la empresa.

Individualización de los modelos por activo: El estudio evidenció que la dinámica de precios de cada acción es única y requiere un enfoque individualizado. La predictibilidad de Amazon fue notablemente mayor que la de Google en el período de estudio, esto confirma que los modelos deben ser personalizados para cada activo.

Precisión e interpretabilidad no son lo mismo: Un modelo puede ser preciso por las razones equivocadas. Por eso, empleamos técnicas de interpretabilidad (PDP y SHAP) para “abrir la caja negra” y corroborar que las señales usadas por el modelo son coherentes con la lógica económica, y no el resultado de coincidencias circunstanciales.

En resumen, operacionalizar y comparar rigurosamente modelos tradicionales y modelos de Machine Learning aporta un valor específico a las finanzas: permite transformar señales dispersas en criterios accionables para decidir, cuantificar el riesgo y priorizar oportunidades bajo distintos regímenes de mercado. La comparación estructurada con métricas estadísticas y materialidad financiera (direccionalidad, Sharpe, drawdown) muestra cuándo un método es superior, por qué lo es y en qué condiciones conviene utilizarlo. Al mismo tiempo, la interpretabilidad (PDP, SHAP) asegura que las recomendaciones conserven coherencia económica, mientras que la personalización por activo y el análisis de escenarios amplían el alcance desde el pronóstico puntual hacia la gestión integral del riesgo. Para el analista financiero, este marco no reemplaza el criterio profesional: lo fortalece, al ofrecer una base comparativa y replicable para anticipar escenarios, definir coberturas y, en última instancia, tomar mejores decisiones de inversión de forma consistente y razonable.

5.2 Recomendaciones

(Esto es lo que aconsejamos a otros analistas y gestores)

Adoptar machine learning con criterio financiero: Avanzar más allá de los modelos tradicionales es conveniente, siempre que la ingeniería de variables parta de hipótesis económicas claras. A partir de esa base, el uso de algoritmos capaces de modelar no linealidades e interacciones como Random Forest permite identificar patrones que no son evidentes a simple vista.

Exigencia de modelos explicables: Evitar decisiones basadas en “cajas negras”. Una ganancia en precisión que no puede explicarse constituye una alerta. Se sugiere aplicar técnicas de interpretabilidad (p. ej., PDP, SHAP) para corroborar que el razonamiento del modelo sea consistente con la lógica financiera antes de arriesgar capital.

Modelos a la medida del activo o sector: Cree modelos específicos para cada activo o sector de interés. Lo que funciona para una empresa de consumo como Amazon no necesariamente funcionará igual para una gigante tecnológica como Google.

Use los modelos como un copiloto, no como un piloto automático: Ningún modelo puede predecir eventos inesperados o "cisnes negros". Su mejor uso es como una herramienta de apoyo que potencia el análisis y el juicio del experto financiero, no que lo reemplaza. Aporta disciplina y rigor cuantitativo a la toma de decisiones.

Complementar con análisis de estrés ante shocks exógenos: Dado que ningún modelo cuantitativo puede predecir eventos como **pandemias o guerras**, se recomienda que las predicciones del modelo se utilicen en conjunto con análisis de estrés cualitativos. Los gestores deben simular cómo se comportaría el portafolio bajo escenarios de shocks extremos (ej., una interrupción de la cadena de suministro global, un aumento súbito en los precios de la energía por un conflicto) para desarrollar planes de contingencia y estrategias de cobertura que protejan el capital más allá de lo que los datos históricos sugieren.

Referencias

- Anthropic. (2025). Claude [Modelo de lenguaje de inteligencia artificial]. <https://claude.ai>
- Bailey, D. H., Borwein, J. M., López de Prado, M., & Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. *Notices of the American Mathematical Society*, 458-471.
- Ballings, M., Van den Poel, D., Hespels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 7046-7056.
- Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review*, 25-27.
- Brooks, C. (2019). *Introductory econometrics for finance*. Cambridge: Cambridge University Press.
- Campbell, J. Y., Giglio, S., & Polk, C. (2018). Hard times: Asset returns, tail risks, and predictability in the cross section of equity returns. *The Review of Financial Studies*, 4571-4606.
- Carmona, P., Climent, F., & Momparler, A. (2021). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 665-674.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, 1461-1499. doi:10.1111
- De Prado, M. L. (2018). *Advances in Financial Machine Learning*. Wiley.
- Dixon, M. F., Polson, N. G., & Sokolov, V. O. (2020). Deep learning for spatio-temporal modeling: Dynamic traffic flows and high frequency trading. *Applied Stochastic Models in Business and Industry*, 430-448.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669. doi:10.1016
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 1779-1801.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273. doi:10.1093
- Hastie. (2009). *The Elements of Statistical Learning*.
- Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, 1-33".
- Hsieh, D. A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *The Journal of Finance*, 1839-1877.
- Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 689-702.

- Lie, X., Xie, Q., Jiang, J., Zhou, J., & Huang, L. (2022). Identifying and measuring technological innovation capabilities of firms using patent analytics: A strategic patent portfolio approach. *Technological Forecasting and Social Change*, 21282. doi:10.1016
- López de Prado, M. (2018). *Advances in financial machine learning*. Hoboken: Wiley. doi:10.1002
- López de Prado, M. (2020). *Machine learning for asset managers*. Cambridge : Cambridge University Press.
- López-Martín, C., Moreno-Pérez, A., & Abellán-Pérez, J. M. (2022). Adaptive regularization techniques for forecasting stock prices in the technology sector: A comparative study. *Journal of Forecasting*, 512-531.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach". *Econometrica*, 347-370.
- Pesaran, M. H., & Timmermann, A. (2004). How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting*, 411-425.
- Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., & Iosifidis, A. (2020). Using deep learning for price prediction by exploiting stationary limit order book features. *Applied Soft Computing*, 93, 106401. doi:https://doi.org/10.1016/j.asoc.2020.106401
- World Bank. (2021, julio). *Publicación: Global Economic Prospects, June 2021*. Retrieved from World Bank: <https://openknowledge.worldbank.org/entities/publication/e70e9357-8459-5e9b-b069-97c0ad5a9769>
- Wu, T., Gao, J., Zhou, H., Yuan, C., & Zheng, X. (2021). Using transformer-based deep learning methods for stock price prediction and portfolio optimization. *IEEE Access*, 9, 27462-27472. doi:10.1109
- Zhang, & Yang. (2020). Stock trend prediction using XGBoost. *IEEE Access*, 8, 97382-97389. doi:10.1109
- Zhu, Y., Zhou, L., Xie, C., Wang, G. J., & Nguyen, T. V. (2021). Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. *International Journal of Production Economics*, 107899. doi:10.1016

Anexos

Anexo 1

Tabla 27

Predicciones vs. valores reales para GOOGL en el conjunto de prueba

Fecha	Valores Reales	Predicción ARIMA	Predicción Random Forest	Predicción Gradient Boosting
2024-03-14	142.4191	139.0636	139.7565	139.5169
2024-03-15	140.5082	139.0110	140.8736	141.4110
2024-03-18	146.9773	138.9657	141.1881	140.7652
2024-03-19	146.3304	138.9267	145.8264	143.8359
2024-03-20	148.0323	138.8933	147.0284	144.2773
2024-03-21	146.8977	138.8644	148.6961	144.7722
2024-03-22	150.0526	138.8397	146.9749	144.0946
2024-03-25	149.3559	138.8183	148.9456	145.3958
2024-03-26	149.9531	138.8000	148.5751	144.7622
2024-03-27	150.1521	138.7843	149.1512	145.4535
2024-03-28	150.2118	138.7707	148.9258	144.0524

2024-04-01	154.7502	138.7590	148.6970	143.7509
2024-04-02	153.8246	138.7490	148.9877	146.3422
2024-04-03	154.1828	138.7404	148.9877	143.4964
2024-04-04	149.8138	138.7329	148.5819	144.3206
2024-04-05	151.7744	138.7266	147.0662	139.3770
2024-04-08	154.1132	138.7211	148.5190	146.3422
2024-04-09	155.8549	138.7164	148.9877	146.0571
2024-04-10	155.3970	138.7123	148.5819	144.3514
2024-04-11	158.6515	138.7088	148.8169	143.6986
2024-04-12	156.9795	138.7058	148.6086	144.0700
2024-04-15	154.1232	138.7032	148.1210	145.3932
2024-04-16	153.6653	138.7010	148.1498	144.3553
2024-04-17	154.7302	138.6991	147.9824	146.0644
2024-04-18	155.2677	138.6974	147.9306	146.0644
2024-04-19	153.3568	138.6960	146.2406	145.7745

2024-04-22	155.5364	138.6948	147.7172	145.7745
2024-04-23	157.5070	138.6938	147.9296	144.3553
2024-04-24	158.3728	138.6929	147.9296	144.3553
2024-04-25	155.2577	138.6921	147.9281	144.3553
2024-04-26	171.1318	138.6914	148.3920	144.6226
2024-04-29	165.3594	138.6908	148.8053	144.6226
2024-04-30	162.0055	138.6904	147.9564	144.3553
2024-05-01	163.0803	138.6899	147.9816	146.0644
2024-05-02	165.8272	138.6896	148.2263	144.4398
2024-05-03	166.4442	138.6893	149.5863	145.0375
2024-05-06	167.3002	138.6890	149.6099	144.3848
2024-05-07	170.4352	138.6888	149.5144	144.6400
2024-05-08	168.5741	138.6886	149.4859	144.4488
2024-05-09	169.1513	138.6884	149.4938	144.6526
2024-05-10	167.8475	138.6882	149.4938	144.7036

2024-05-13	168.3352	138.6881	149.6958	144.3848
2024-05-14	169.5295	138.6880	149.5072	144.5746
2024-05-15	171.6892	138.6879	149.4859	144.1495
2024-05-16	173.3512	138.6878	149.4859	144.3007
2024-05-17	175.2223	138.6878	149.4859	144.3007
2024-05-20	176.0782	138.6877	149.4859	144.2742
2024-05-21	177.0038	138.6876	149.4859	144.2742
2024-05-22	175.5407	138.6876	149.4859	144.2742
2024-05-23	172.7242	138.6876	149.4859	144.6791
2024-05-24	174.1574	138.6875	149.4859	144.2742
2024-05-28	175.5607	138.6875	149.4859	144.7036
2024-05-29	175.0630	138.6875	148.7875	144.6922
2024-05-30	171.2911	138.6875	148.8708	144.4210
2024-05-31	171.6792	138.6874	149.4938	144.7036
2024-06-03	172.3460	138.6874	149.5144	144.6400

2024-06-04	172.9631	138.6874	149.5144	144.6602
2024-06-05	174.5754	138.6874	149.5224	144.6993
2024-06-06	175.8891	138.6874	150.2758	145.0639
2024-06-07	173.6299	138.6874	149.4859	144.3007
2024-06-10	174.3772	138.6874	149.5224	144.6791
2024-06-11	175.9813	138.6874	150.2758	144.6791
2024-06-12	177.1471	138.6874	150.2524	144.3209
2024-06-13	174.5266	138.6874	150.2524	144.6099
2024-06-14	176.1507	138.6874	150.4394	144.7394
2024-06-17	176.5991	138.6873	150.2524	144.8149
2024-06-18	174.4569	138.6873	150.2524	144.6855
2024-06-20	175.6625	138.6873	150.2809	145.0248
2024-06-21	178.9805	138.6873	150.2809	145.0248
2024-06-24	178.5719	138.6873	150.2737	144.9594
2024-06-25	183.3645	138.6873	150.2524	144.8149

2024-06-26	183.2151	138.6873	150.2524	144.6993
2024-06-27	184.7396	138.6873	150.2524	144.3209
2024-06-28	181.4913	138.6873	150.2798	144.3209
2024-07-01	182.3283	138.6873	149.4859	144.0703
2024-07-02	184.5702	138.6873	150.3973	144.3007
2024-07-03	185.1481	138.6873	150.2524	144.1495
2024-07-05	189.9108	138.6873	150.2524	144.6855
2024-07-08	188.3465	138.6873	150.2524	144.6855
2024-07-09	188.2967	138.6873	150.2524	144.3209
2024-07-10	190.4887	138.6873	150.2524	145.0639
2024-07-11	184.8990	138.6873	150.4394	145.7161
2024-07-12	184.4008	138.6873	150.4394	145.3377
2024-07-15	185.8555	138.6873	150.2809	145.0248
2024-07-16	183.2550	138.6873	150.4965	145.6770
2024-07-17	180.3654	138.6873	149.9610	145.4335

2024-07-18	177.0475	138.6873	149.7463	145.4335
2024-07-19	177.0176	138.6873	149.5396	144.7813
2024-07-22	181.0131	138.6873	149.7427	144.7813
2024-07-23	181.1326	138.6873	149.7427	144.7813
2024-07-24	172.0058	138.6873	149.5542	144.3412
2024-07-25	166.6751	138.6873	149.4976	144.7057
2024-07-26	166.3961	138.6873	149.7463	145.1357
2024-07-29	168.9170	138.6873	149.7578	145.3580
2024-07-30	169.6742	138.6873	149.7713	145.4087
2024-07-31	170.9197	138.6873	149.7522	145.3580
2024-08-01	170.1425	138.6873	149.7013	145.7857
2024-08-02	166.0574	138.6873	149.7310	145.6890
2024-08-05	158.6742	138.6873	149.7220	145.0169
2024-08-06	157.7176	138.6873	149.7220	145.6890
2024-08-07	158.3653	138.6873	149.7220	145.9740

2024-08-08	161.4441	138.6873	149.7310	145.6890
2024-08-09	163.0782	138.6873	149.7013	145.8023
2024-08-12	161.7032	138.6873	149.7013	145.8023
2024-08-13	163.5664	138.6873	149.7579	145.1290
2024-08-14	159.7901	138.6873	149.7318	145.1290
2024-08-15	160.7168	138.6873	149.7774	145.1290
2024-08-16	162.3707	138.6873	149.9464	145.1290
2024-08-19	166.0673	138.6873	149.9464	145.1290
2024-08-20	166.5755	138.6873	149.7318	145.1290
2024-08-21	165.2503	138.6873	149.7318	145.1290
2024-08-22	163.2077	138.6873	149.7568	145.1290
2024-08-23	165.0211	138.6873	149.7318	145.1290
2024-08-26	165.5592	138.6873	149.7318	145.1290
2024-08-27	164.0845	138.6873	149.7318	145.1290
2024-08-28	162.2612	138.6873	149.7480	145.1290

2024-08-29	161.1950	138.6873	149.7318	145.1290
2024-08-30	162.7892	138.6873	149.8847	145.1290
2024-09-03	156.7910	138.6873	149.7013	145.8023
2024-09-04	155.8843	138.6873	149.7013	145.8023
2024-09-05	156.6714	138.6873	149.7013	145.8023
2024-09-06	150.3743	138.6873	149.7013	145.6890
2024-09-09	148.3689	138.6873	148.6542	141.2466
2024-09-10	148.3190	138.6873	147.9779	143.5089
2024-09-11	150.8133	138.6873	148.2270	143.8722
2024-09-12	154.3352	138.6873	148.9173	143.8722
2024-09-13	157.0988	138.6873	149.9809	148.7926
2024-09-16	157.6974	138.6873	149.9265	146.0548
2024-09-17	158.9546	138.6873	149.7253	145.1290
2024-09-18	159.4434	138.6873	149.7579	145.1290
2024-09-19	161.7681	138.6873	149.7318	145.1290

2024-09-20	163.2148	138.6873	149.7318	145.1290
2024-09-21	161.4787	138.6873	149.7318	145.1290
2024-09-23	161.9177	138.6873	149.7774	145.7857
2024-09-24	161.1196	138.6873	149.7774	145.7857
2024-09-25	162.3567	138.6873	149.7774	145.1290
2024-09-26	163.5739	138.6873	149.7480	145.1290
2024-09-27	165.4696	138.6873	149.7433	145.7857
2024-09-30	166.6070	138.6873	149.7013	146.4590
2024-10-01	165.4796	138.6873	149.7013	146.0820
2024-10-02	165.4796	138.6873	149.7013	146.0820
2024-10-03	166.6768	138.6873	149.7013	146.0312
2024-10-04	162.6062	138.6873	149.7013	145.9935
2024-10-07	164.0029	138.6873	149.7013	146.1068
2024-10-08	161.4887	138.6873	149.7013	146.1068
2024-10-09	161.7082	138.6873	149.7513	146.1068

2024-10-10	162.8656	138.6873	149.6972	146.1068
2024-10-11	165.0805	138.6873	149.7013	146.1068
2024-10-15	164.7811	138.6873	149.7013	146.1068
2024-10-16	162.5563	138.6873	149.7217	146.1068
2024-10-17	163.0452	138.6873	149.7579	145.4335
2024-10-18	163.6937	138.6873	149.7510	145.4335
2024-10-21	164.7612	138.6873	149.7683	145.2112
2024-10-22	162.4066	138.6873	149.4976	145.4546
2024-10-23	162.3467	138.6873	148.9032	145.8580
2024-10-24	164.8909	138.6873	148.6359	145.4546
2024-10-25	166.3376	138.6873	148.6359	145.4546
2024-10-28	169.2908	138.6873	148.6359	145.4281
2024-10-29	174.0598	138.6873	148.5370	145.4281
2024-10-30	170.7175	138.6873	148.6074	145.3148
2024-10-31	170.8971	138.6873	148.5483	145.4079

2024-11-01	168.8518	138.6873	148.6472	145.4546
2024-11-04	169.3506	138.6873	148.6359	145.4546
2024-11-05	176.1051	138.6873	148.5819	143.6476
2024-11-06	180.3354	138.6873	148.6275	144.3902
2024-11-07	177.9409	138.6873	148.8169	143.6678
2024-11-08	181.2034	138.6873	148.8169	143.6476
2024-11-12	178.4697	138.6873	148.8169	143.8927
2024-11-13	175.1773	138.6873	148.8169	144.3700
2024-11-14	172.0943	138.6873	148.5819	144.3700
2024-11-15	174.8979	138.6873	148.5819	144.3700
2024-11-18	177.7114	138.6873	148.5819	144.3700
2024-11-19	175.5763	138.6873	148.6079	143.6476
2024-11-20	167.2455	138.6873	148.5980	143.6476
2024-11-21	164.3821	138.6873	148.6275	143.6476
2024-11-22	167.2654	138.6873	148.8708	144.7548

2024-11-25	168.7321	138.6873	148.8864	144.3902
2024-11-26	168.8418	138.6873	148.8864	144.7813
2024-11-27	168.5625	138.6873	149.8346	145.6521
2024-11-29	171.0966	138.6873	149.7429	145.5851
2024-12-02	170.9470	138.6873	149.4082	144.9329
2024-12-03	173.9700	138.6873	149.7399	145.5851
2024-12-04	172.2440	138.6873	149.9285	145.6521
2024-12-05	174.3093	138.6873	149.7501	145.7161
2024-12-06	175.1683	138.6873	149.1534	145.4070
2024-12-09	184.9570	138.6873	148.8944	144.7548
2024-12-10	195.1752	138.6873	149.6262	144.9999
2024-12-11	191.7392	138.6873	148.9685	144.6353
2024-12-12	189.6016	138.6873	148.9693	144.6151
2024-12-13	196.4338	138.6873	148.8169	144.3700
2024-12-16	195.1952	138.6873	148.5525	144.3700

2024-12-17	188.1833	138.6873	148.5984	145.0642
2024-12-18	188.2931	138.6873	148.6614	145.3456
2024-12-19	191.1898	138.6873	148.6791	144.4210
2024-12-20	194.4061	138.6873	148.5933	144.6922
2024-12-23	195.8844	138.6873	148.8169	144.6922
2024-12-24	195.3750	138.6873	148.8169	144.4210
2024-12-26	192.5383	138.6873	148.5819	144.6922
2024-12-27	191.0200	138.6873	148.6625	144.4210
2024-12-30	NaN	138.6873	NaN	NaN

Nota. Tabla completa de las predicciones de los modelos ARIMA, Random Forest y Gradient Boosting para Google en el conjunto de prueba, comparadas con los precios de cierre reales.

Anexo 2

Tabla 28

Predicciones vs. valores reales para AMZN en el conjunto de prueba

Fecha	Valores Reales	Predicción ARIMA	Predicción Random Forest	Predicción GradientBoosting
2024-03-14	178.7500	176.5600	176.6537	176.7301
2024-03-15	174.4200	176.5600	176.2321	176.3723
2024-03-18	174.4800	176.5600	174.7340	175.6451
2024-03-19	175.9000	176.5600	175.6644	175.5837
2024-03-20	178.1500	176.5600	177.2095	177.2720
2024-03-21	178.1500	176.5600	177.4043	177.5744
2024-03-22	178.8700	176.5600	177.8782	177.4473
2024-03-25	179.7100	176.5600	178.0638	177.8057
2024-03-26	178.3000	176.5600	177.1068	177.7981
2024-03-27	179.8300	176.5600	177.7402	177.8622
2024-03-28	180.3800	176.5600	176.8497	178.1163

2024-04-01	180.9700	176.5600	177.0064	177.9337
2024-04-02	180.6900	176.5600	181.3431	180.5666
2024-04-03	182.4100	176.5600	179.9865	180.5666
2024-04-04	180.0000	176.5600	182.3347	182.5541
2024-04-05	185.0700	176.5600	176.4888	177.7355
2024-04-08	185.1900	176.5600	183.3821	182.8098
2024-04-09	185.6700	176.5600	183.4205	182.8098
2024-04-10	185.9500	176.5600	183.4756	182.8098
2024-04-11	189.0500	176.5600	183.3312	180.9742
2024-04-12	186.1300	176.5600	183.0392	180.8282
2024-04-15	183.6200	176.5600	181.2267	179.4414
2024-04-16	183.3200	176.5600	181.2214	179.8850
2024-04-17	181.2800	176.5600	181.6620	179.8541
2024-04-18	179.2200	176.5600	180.3016	176.9808
2024-04-19	174.6300	176.5600	178.2617	176.2064

2024-04-22	177.2300	176.5600	174.3161	174.3186
2024-04-23	179.5400	176.5600	176.9782	176.4491
2024-04-24	176.5900	176.5600	177.6823	176.2706
2024-04-25	173.6700	176.5600	176.9580	176.9559
2024-04-26	179.6200	176.5600	174.5701	175.2171
2024-04-29	180.9600	176.5600	177.7710	176.1343
2024-04-30	175.0000	176.5600	180.9970	179.1421
2024-05-01	179.0000	176.5600	173.6856	174.3186
2024-05-02	184.7200	176.5600	177.1638	176.9388
2024-05-03	186.2100	176.5600	183.1469	182.5140
2024-05-06	188.7000	176.5600	183.6519	183.3185
2024-05-07	188.7600	176.5600	184.7323	183.3185
2024-05-08	188.0000	176.5600	183.3970	183.3185
2024-05-09	189.5000	176.5600	183.3381	181.5470
2024-05-10	187.4800	176.5600	183.2944	181.8918

2024-05-13	186.5700	176.5600	183.3520	181.8277
2024-05-14	187.0700	176.5600	183.3520	181.8277
2024-05-15	185.9900	176.5600	183.3520	181.8918
2024-05-16	183.6300	176.5600	183.3516	181.8918
2024-05-17	184.7000	176.5600	182.8773	181.6360
2024-05-20	183.5400	176.5600	183.1375	181.8918
2024-05-21	183.1500	176.5600	182.7653	181.2913
2024-05-22	183.1300	176.5600	182.9430	183.1268
2024-05-23	181.0500	176.5600	182.7272	183.1268
2024-05-24	180.7500	176.5600	181.3042	180.9014
2024-05-28	182.1500	176.5600	179.9880	180.9014
2024-05-29	182.0200	176.5600	181.9007	182.6565
2024-05-30	179.3200	176.5600	181.9252	182.6565
2024-05-31	176.4400	176.5600	178.8980	178.3083
2024-06-03	178.3400	176.5600	177.3949	178.0689

2024-06-04	179.3400	176.5600	177.1461	178.1163
2024-06-05	181.2800	176.5600	178.4789	177.5871
2024-06-06	185.0000	176.5600	181.3100	180.1318
2024-06-07	184.3000	176.5600	184.4155	183.3825
2024-06-10	187.0600	176.5600	183.2777	183.2751
2024-06-11	187.2300	176.5600	183.5187	183.0591
2024-06-12	186.8900	176.5600	183.3367	182.5858
2024-06-13	183.8300	176.5600	183.5224	184.2887
2024-06-14	183.6600	176.5600	182.8016	182.5422
2024-06-17	184.0600	176.5600	182.7484	182.5422
2024-06-18	182.8100	176.5600	182.8912	182.6748
2024-06-20	186.1000	176.5600	182.9051	184.1015
2024-06-21	189.0800	176.5600	183.4784	184.3572
2024-06-24	185.5700	176.5600	183.5668	184.3572
2024-06-25	186.3400	176.5600	183.4727	184.4213

2024-06-26	193.6100	176.5600	183.3350	182.5858
2024-06-27	197.8500	176.5600	183.2254	181.5683
2024-06-28	193.2500	176.5600	183.3433	181.2236
2024-07-01	197.2000	176.5600	183.3381	181.8918
2024-07-02	200.0000	176.5600	183.3520	182.0401
2024-07-03	197.5900	176.5600	183.3520	181.5683
2024-07-05	200.0000	176.5600	183.3520	181.5683
2024-07-08	199.2900	176.5600	183.3520	181.5683
2024-07-09	199.3400	176.5600	183.3520	182.9305
2024-07-10	199.7900	176.5600	183.3520	181.5683
2024-07-11	195.0500	176.5600	183.3520	182.9305
2024-07-12	194.4900	176.5600	183.3520	182.7979
2024-07-15	192.7200	176.5600	183.3520	182.8665
2024-07-16	193.0200	176.5600	183.3520	182.8665
2024-07-17	187.9300	176.5600	183.3520	182.4602

2024-07-18	183.7500	176.5600	183.3520	182.2222
2024-07-19	183.1300	176.5600	182.7940	181.9664
2024-07-22	182.5500	176.5600	182.7931	181.9664
2024-07-23	186.4100	176.5600	182.2592	181.9664
2024-07-24	180.8300	176.5600	182.7846	183.5748
2024-07-25	179.8500	176.5600	179.8375	181.0332
2024-07-26	182.5000	176.5600	176.5981	178.6387
2024-07-29	183.2000	176.5600	181.9439	180.8618
2024-07-30	181.7100	176.5600	182.4955	183.5924
2024-07-31	186.9800	176.5600	181.3382	181.2318
2024-08-01	184.0700	176.5600	182.2363	183.4706
2024-08-02	167.9000	176.5600	181.5684	182.3379
2024-08-05	161.0200	176.5600	170.7265	169.4640
2024-08-06	161.9300	176.5600	164.7856	162.8739
2024-08-07	162.7700	176.5600	165.0446	164.5110

2024-08-08	165.8000	176.5600	165.0190	166.0412
2024-08-09	166.9400	176.5600	166.1424	167.4443
2024-08-12	166.8000	176.5600	170.5446	169.1690
2024-08-13	170.2300	176.5600	171.5715	170.1117
2024-08-14	170.1000	176.5600	173.0863	171.0546
2024-08-15	177.5900	176.5600	173.2391	170.8014
2024-08-16	177.0600	176.5600	176.1425	176.3400
2024-08-19	178.2200	176.5600	175.9752	176.4896
2024-08-20	178.8800	176.5600	176.5267	176.9918
2024-08-21	180.1100	176.5600	177.5887	177.3534
2024-08-22	176.1300	176.5600	176.9999	177.7810
2024-08-23	177.0400	176.5600	177.1571	177.6057
2024-08-26	175.5000	176.5600	176.9881	178.0992
2024-08-27	173.1200	176.5600	176.5811	177.1194
2024-08-28	170.8000	176.5600	174.8823	174.9768

2024-08-29	172.1200	176.5600	174.1980	173.3677
2024-08-30	178.5000	176.5600	174.2528	173.3677
2024-09-03	176.2500	176.5600	176.7176	176.1200
2024-09-04	173.3300	176.5600	176.1299	176.4017
2024-09-05	177.8900	176.5600	174.1825	173.0818
2024-09-06	171.3900	176.5600	176.8462	176.5180
2024-09-09	175.4000	176.5600	174.1937	172.6766
2024-09-10	179.5500	176.5600	176.0112	176.8325
2024-09-11	184.5200	176.5600	178.0604	177.1483
2024-09-12	187.0000	176.5600	183.3429	182.4969
2024-09-13	186.4900	176.5600	184.8138	183.0242
2024-09-16	184.8900	176.5600	184.5641	183.3473
2024-09-17	186.8800	176.5600	182.9737	183.4824
2024-09-18	186.4300	176.5600	182.1326	181.6995
2024-09-19	189.8700	176.5600	183.2401	181.8106

2024-09-20	191.6000	176.5600	183.3664	183.3014
2024-09-23	193.8800	176.5600	183.3381	181.8106
2024-09-24	193.9600	176.5600	183.3520	182.1785
2024-09-25	192.5300	176.5600	183.3520	182.1785
2024-09-26	191.1600	176.5600	183.3520	181.8106
2024-09-27	187.9700	176.5600	183.1229	181.6646
2024-09-30	186.3300	176.5600	183.1229	182.0324
2024-10-01	185.1300	176.5600	181.8139	181.3214
2024-10-02	184.7600	176.5600	181.9864	181.7683
2024-10-03	181.9600	176.5600	181.7248	180.9816
2024-10-04	186.5100	176.5600	180.3609	181.0657
2024-10-07	180.8000	176.5600	181.7204	180.5827
2024-10-08	182.7200	176.5600	179.4024	180.4603
2024-10-09	185.1700	176.5600	181.3555	182.4064
2024-10-10	186.6500	176.5600	181.6468	179.8765

2024-10-11	188.8200	176.5600	182.0036	181.2999
2024-10-15	187.6900	176.5600	181.9887	181.2562
2024-10-16	186.8900	176.5600	181.8222	180.8036
2024-10-17	187.5300	176.5600	181.8000	180.0029
2024-10-18	188.9900	176.5600	182.6166	180.5758
2024-10-21	189.0700	176.5600	182.3504	180.5153
2024-10-22	189.7000	176.5600	182.2264	180.7489
2024-10-23	184.7100	176.5600	181.8139	180.0029
2024-10-24	186.3800	176.5600	181.9470	180.4747
2024-10-25	187.8300	176.5600	181.7340	180.1349
2024-10-28	188.3900	176.5600	181.8053	180.1349
2024-10-29	190.8300	176.5600	181.8382	182.1115
2024-10-30	192.7300	176.5600	181.8000	180.1349
2024-10-31	186.4000	176.5600	181.8045	180.4579
2024-11-01	197.9300	176.5600	181.8045	180.4579

2024-11-04	195.7800	176.5600	181.8045	180.4579
2024-11-05	199.5000	176.5600	181.8139	180.1349
2024-11-06	207.0900	176.5600	183.2362	181.3190
2024-11-07	210.0500	176.5600	183.3520	181.3190
2024-11-08	208.1800	176.5600	183.3520	181.3190
2024-11-12	208.9100	176.5600	183.3520	181.3190
2024-11-13	214.1000	176.5600	183.3520	181.4190
2024-11-14	211.4800	176.5600	183.3520	181.5570
2024-11-15	202.6100	176.5600	183.3520	181.3190
2024-11-18	201.7000	176.5600	183.3520	181.3190
2024-11-19	204.6100	176.5600	183.2655	181.3190
2024-11-20	202.8800	176.5600	183.1356	181.1730
2024-11-21	198.3800	176.5600	183.1229	181.1730
2024-11-22	197.1200	176.5600	183.3520	181.3190
2024-11-25	201.4500	176.5600	183.3520	181.4214

2024-11-26	207.8600	176.5600	183.3520	181.2834
2024-11-27	205.7400	176.5600	183.3520	181.2834
2024-11-29	207.8900	176.5600	183.3520	181.8277
2024-12-02	210.7100	176.5600	183.3520	181.8277
2024-12-03	213.4400	176.5600	183.3520	181.8277
2024-12-04	218.1600	176.5600	183.3520	181.8277
2024-12-05	220.5500	176.5600	183.3520	181.8277
2024-12-06	227.0300	176.5600	183.3520	181.8918
2024-12-09	226.0900	176.5600	183.3520	181.4214
2024-12-10	225.0400	176.5600	183.3520	181.4214
2024-12-11	230.2600	176.5600	183.3520	181.8277
2024-12-12	228.9700	176.5600	183.3520	181.4190
2024-12-13	227.4600	176.5600	183.3520	181.9633
2024-12-16	232.9300	176.5600	183.3520	181.1834
2024-12-17	231.1500	176.5600	183.3520	181.1834

2024-12-18	220.5200	176.5600	181.7338	180.3435
2024-12-19	223.2900	176.5600	181.7338	180.4579
2024-12-20	224.9200	176.5600	182.3504	180.8388
2024-12-23	225.0600	176.5600	183.1229	181.0374
2024-12-24	229.0500	176.5600	183.3520	181.5570
2024-12-26	227.0500	176.5600	183.3520	181.3190
2024-12-27	223.7500	176.5600	183.3520	181.1834
2024-12-30	221.3000	176.5600	183.1356	181.0374

Nota. Tabla completa de las predicciones de los modelos ARIMA, Random Forest y Gradient Boosting para Amazon en el conjunto de prueba, comparadas con los precios de cierre reales.

Anexo 3. Código Python de los análisis realizados en el capítulo 4.

Creación del dataset o base de datos para el capítulo IV de la tesis.

https://github.com/jaimeisaac2020/Python-analisis-basicos/blob/main/github/como_se_descargaron_los_datos.ipynb

Dataset o conjunto de datos analizados en el capítulo IV de la tesis.

Para promover la transparencia y la reproducibilidad, el conjunto de datos completo y los scripts de análisis utilizados en esta investigación se encuentran disponibles en el repositorio público de GitHub: https://github.com/jaimeisaac2020/Python-analisis-basicos/blob/main/github/dataset_completo_google_amazon.csv

Los datos de las tasas de interés (Treasure_10Y) que aparecen en el dataset dataset_completo_google_amazon.csv se han descargado de esta página:

<https://fred.stlouisfed.org/graph/?id=DGS10>,

Los datos de las tasas de inflación a 5 años (Inflacion_TY5E) que aparecen en el dataset dataset_completo_google_amazon.csv se han descargado de esta página:

<https://fred.stlouisfed.org/series/T5YIE>

Código Python para las secciones 4.1,4.2 y 4.3 del capítulo IV de análisis de resultados.

https://github.com/jaimeisaac2020/Python-analisis-basicos/blob/main/github/secciones_4_1_4_2_4_3_tesis.ipynb

Análisis del objetivo específico 1 en Python y su hipótesis relacionada (hipótesis 2) que se muestra en el capítulo 4 desde la sección 4.1 hasta la sección 4.5.

<https://colab.research.google.com/drive/1noduoBqvHIP-mClirmlUwqqxZhWGy9zc?usp=sharing>

Análisis del objetivo específico 2 en Python y su hipótesis relacionada (hipótesis 3) que se muestra en el capítulo 4 en la sección 4.6 con todas sus subsecciones.

https://github.com/jaimeisaac2020/Python-analisis-basicos/blob/mi-github/analisis_objetivo2_tesis.ipynb

Análisis del objetivo específico 3 en Python y su hipótesis relacionada (hipótesis 4) que se muestra en el capítulo 4 en la sección 4.7 con todas sus subsecciones.

https://github.com/jaimeisaac2020/Python-analisis-basicos/blob/mi-github/analisis_objetivo_3.ipynb

Análisis del objetivo específico 4 en Python y su hipótesis relacionada (hipótesis 1) que se muestra en el capítulo 4 en la sección 4.8 con todas sus subsecciones

<https://colab.research.google.com/drive/1PUWv9cWj0PXUnSrh0dfOQmQfS0glWZZj?usp=sharing>

Dashboard y simulación de escenarios para presentación de resultados de los análisis que se muestran en el capítulo 4.

<https://colab.research.google.com/drive/1tR1TfS9CRm8wQ5cp1GS0xOvUWGYdRYHj?usp=sharing>

Análisis del VIX como variable predictora en la sección 4.6.14

[https://github.com/jaimeisaac2020/Python-analisis-basicos/blob/main-
github/sobre_el_VIX_RF_amzn.ipynb](https://github.com/jaimeisaac2020/Python-analisis-basicos/blob/main/github/sobre_el_VIX_RF_amzn.ipynb)

Fuentes de datos macroeconómicos primarios:

Tasa del tesoro a 10 años (DGS10): <https://fred.stlouisfed.org/series/DGS10>

Expectativas de inflación a 5 años (T5YIE): <https://fred.stlouisfed.org/series/T5YIE>