

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

**Comparación entre Modelos Tradicionales y Nuevas Tecnologías de IA en
la Creación de Portafolios Diversificados de Renta Variable en el
Mercado de Estados Unidos entre el 2014 y 2024**

Tesis entorno a hipótesis o problema de investigación y su contrastación

Aarón Guillermo Castro Rodríguez

**Juan Carlos Chanabá Guerrero Ph.D.
Director de Trabajo de Titulación**

Trabajo de titulación de posgrado presentado como requisito
para la obtención del título de Máster en Finanzas

Quito, 28 de junio de 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

**Comparación entre Modelos Tradicionales y Nuevas Tecnologías de IA en
la Creación de Portafolios Diversificados de Renta Variable en el
Mercado de Estados Unidos entre el 2014 y 2024**

Aarón Guillermo Castro Rodríguez

Nombre del Director del Programa:	Juan José Espinosa de los Monteros
Título académico:	MBA
Director del programa de:	Escuela de Empresas

Nombre del Decano del colegio Académico:	Ana María Novillo
Título académico:	Ph. D.
Decano del Colegio:	Colegio de Administración de Empresas

Nombre del Decano del Colegio de Posgrados:	Dario Niebieskikwiat
Título académico:	Ph. D.

Quito, junio 2025

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombre del estudiante: Aarón Guillermo Castro Rodríguez

Código de estudiante: 00344257

C.I.: 1718899956

Lugar y fecha: Quito, 28 de junio de 2025.

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>.

UNPUBLISHED DOCUMENT

Note: The following graduation project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>.

DEDICATORIA

Esta tesis esta dedicada a mis padres, Guillermo y Judith, por ser el pilar fundamental en cada etapa de mi vida. Por su amor incondicional, su ejemplo de perseverancia y los valores que me han guiado hasta aquí. Esta tesis es un reflejo del camino que he recorrido gracias a su apoyo constante y su confianza en mí.

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a todas las personas que han hecho posible la realización de esta tesis.

En primer lugar, a mi familia, que han sido mi mayor respaldo. A mi padre Guillermo Castro, por su ejemplo de esfuerzo y compromiso. A mi madre Judith Rodríguez, por su amor incondicional y apoyo permanente. Y a mi hermano Mateo Castro, por su apoyo incondicional.

A a mi novia, Fernanda Betancourt, gracias por tu paciencia, comprensión y por estar a mi lado en los momentos de mayor presión. Tu apoyo fue fundamental para seguir adelante.

Agradezco también a mi tutor de tesis, Juan Carlos Chanabá, por su orientación, disposición y valiosos aportes a lo largo de este trabajo.

A mis amigos que me deja la maestría, Erick, Martín y Enrique, gracias por su amistad, apoyo académico y risas durante todo este recorrido. Compartir esta experiencia con ustedes la hizo más valiosa y llevadera.

Finalmente, agradezco a Dios por darme la fortaleza, la salud y la claridad para llegar hasta aquí.

Este logro no habría sido posible sin cada uno de ustedes.

RESUMEN

Esta investigación compara la efectividad de los modelos tradicionales de optimización de portafolios con enfoques basados en inteligencia artificial, específicamente mediante el uso del algoritmo Random Forest Regressor. El análisis se centra en el mercado de renta variable de Estados Unidos durante el periodo 2014–2024, utilizando datos históricos de precios y volúmenes de acciones del índice S&P 500. El objetivo es determinar si las nuevas tecnologías pueden mejorar el rendimiento y la diversificación de las inversiones frente a métodos clásicos como la Teoría Moderna de Portafolios.

Se construyeron dos tipos de portafolios: uno reducido, conformado por empresas seleccionadas con base en criterios de sostenibilidad (ESG) y desempeño histórico; y otro más amplio, que abarcó una muestra extensa del mercado. Ambos enfoques fueron evaluados bajo criterios de eficiencia ajustada al riesgo. Los resultados muestran que, si bien el enfoque tradicional mantiene su solidez, el modelo basado en machine learning ofreció una mayor capacidad de adaptación y eficiencia en escenarios más amplios y complejos.

Aunque el uso de inteligencia artificial presenta desafíos de interpretabilidad, sus capacidades predictivas abren nuevas posibilidades para la gestión de carteras en mercados volátiles. Este estudio concluye que los modelos de machine learning pueden complementar, e incluso superar en ciertos contextos, a los métodos tradicionales, constituyéndose en herramientas útiles para la toma de decisiones financieras basadas en datos.

ABSTRACT

This research compares the effectiveness of traditional portfolio optimization models with approaches based on artificial intelligence, specifically using the Random Forest Regressor algorithm. The analysis focuses on the U.S. equity market during the period from 2014 to 2024, using historical data on stock prices and trading volumes from the S&P 500 index. The main objective is to assess whether emerging technologies can enhance investment performance and diversification when compared to classical methods such as Modern Portfolio Theory.

Two types of portfolios were developed: a reduced portfolio, composed of companies selected based on sustainability (ESG) criteria and historical performance, and a broader one covering a more extensive market sample. Both approaches were evaluated using risk-adjusted efficiency metrics. The results indicate that while the traditional model remains solid, the machine learning approach provided greater adaptability and efficiency in broader and more complex scenarios.

Although artificial intelligence presents challenges in terms of interpretability, its predictive capabilities offer new possibilities for portfolio management in volatile market environments. The study concludes that machine learning models can complement and, in certain contexts, outperform traditional methods, offering valuable tools for data-driven financial decision-making.

Tabla de contenido

<i>CAPÍTULO 1</i>	11
1.1 Introducción.....	11
1.2 Planteamiento del problema	13
1.3 Propósito	15
1.4 Naturaleza del estudio	16
1.5 Introducción al marco teórico.....	17
1.6 Pregunta de investigación	18
1.7 Significado del estudio	19
1.8 Conclusión	20
<i>CAPÍTULO 2: REVISIÓN DE LITERATURA</i>	21
2.1. Fundamentos del mercado de renta variable en Estados Unidos.....	21
2.1.1. Introducción a la Renta Variable.....	21
2.1.3. Regulación y organismos supervisores.....	23
2.1.5 Factores Macroeconómicos que Impactan el Mercado.....	25
2.2. Modelos tradicionales en la construcción de portafolios	26
2.2.1. Teoría Moderna de Portafolios	27
2.2.2. Modelo de Valoración de Activos de Capital (CAPM).....	28
2.2.3 Otros Enfoques Tradicionales	29
2.2.4 Limitaciones de los Modelos Tradicionales	30
2.3. Avances en inteligencia artificial y machine learning en finanzas.....	32

2.4. Comparaciones de modelos tradicionales y de IA: ventajas y limitaciones.....	33
2.5. Uso de datos alternativos y capacidades computacionales en la optimización.....	35
<i>CAPÍTULO 3:</i>	36
3.1 Enfoque de la investigación	36
3.2 Tipo y diseño de la investigación	38
3.3 Población y muestra	39
3.4 Técnicas e instrumentos de recolección de datos	40
3.6 Técnicas de análisis de datos.....	42
3.7 Presentación de datos	43
3.8 Análisis descriptivo	47
3.8.1 Resultados pesos del portafolio	48
3.8.2 Simulación masiva con Random Forest	50
3.8.3 Comparación visual de portafolios optimizados	53
3.8.4 Validación con datos reales del mercado	55
<i>CAPÍTULO 4:</i>	57
4.1 Conclusiones	57
4.2 Limitaciones del estudio.....	59
4.3 Recomendaciones.....	60
5. Referencias Bibliográficas	62
6. Anexos.....	65

CAPÍTULO 1

1.1 Introducción

El mercado de renta variable en Estados Unidos es un referente global por su liquidez, solidez regulatoria e innovación tecnológica. Índices como el S&P 500 y el NASDAQ reflejan su dinamismo, convirtiéndolo en un espacio clave para la diversificación y el rendimiento de inversiones (SIFMA, 2023).

La construcción de portafolios de inversión diversificados es esencial para optimizar el rendimiento ajustado al riesgo en los mercados financieros. Tradicionalmente, la Teoría Moderna de Portafolios (Markowitz) y el Modelo de Valoración de Activos de Capital (CAPM) han sido las herramientas más utilizadas para lograr la diversificación adecuada. Estos enfoques se basan en supuestos simplificados, como la normalidad de los retornos y la estabilidad de las correlaciones entre activos. Sin embargo, en mercados financieros complejos como el de Estados Unidos, donde las condiciones económicas y las fluctuaciones del mercado son altamente dinámicas, estos modelos tradicionales pueden no ser suficientes para adaptarse a los desafíos actuales (CFA Institute, 2022).

A medida que las tecnologías emergentes han avanzado, la IA ha mostrado un gran potencial para transformar la gestión de portafolios. La IA, en especial el machine learning, permite analizar grandes volúmenes de datos, detectar patrones complejos y adaptar estrategias en tiempo real. La IA tiene la capacidad de hacer ajustes rápidos y acertados. Esto plantea una oportunidad única para mejorar la toma de decisiones y optimizar el rendimiento ajustado al riesgo, pero también plantea interrogantes sobre su efectividad comparada con las metodologías clásicas (CFA Institute, 2020).

El Machine Learning es una tecnología clave en la toma de decisiones financieras. Esta herramienta permite a las máquinas aprender y adaptarse mediante el análisis de grandes volúmenes de datos sin la necesidad de programación explícita. Los algoritmos de Machine Learning identifican patrones ocultos y tendencias históricas en el mercado financiero, lo que ayuda a seleccionar activos con alto potencial de rentabilidad. Como resultado, esta tecnología facilita decisiones bien fundamentadas y acertadas, optimizando tanto la asignación de recursos como la gestión del riesgo en los portafolios (RankiaPro, 2023).

Según Amat Rodrigo (2020), un modelo Random Forest combina múltiples árboles de decisión, cada uno entrenado con muestras de datos generadas mediante bootstrapping. Este enfoque permite capturar interacciones complejas y no lineales entre variables, superando las limitaciones de modelos globales que aplican una única ecuación a todo el espacio muestral. Los algoritmos basados en árboles segmentan el espacio de predictores en regiones simples, donde las relaciones son más manejables, ofreciendo mayor flexibilidad y precisión en casos con múltiples predictores interrelacionados. Esta capacidad de segmentación es clave para su efectividad en problemas de predicción.

En el contexto financiero actual, la IA está revolucionando el análisis de datos y la toma de decisiones. Avances en machine learning y análisis de datos están mejorando la precisión de las predicciones y la eficiencia en la gestión de carteras. Estas tecnologías permiten a los inversores procesar información al momento, facilitando tanto la detección de oportunidades como una gestión más eficiente del riesgo. Esto demuestra que la IA no solo optimiza el rendimiento, sino que también mejora la capacidad para adaptarse a un mercado financiero complejo (Bloomberg, 2023).

El objetivo de esta investigación es analizar cómo la inteligencia artificial puede complementar o incluso superar a los modelos tradicionales de optimización de portafolios,

específicamente en términos de rendimiento ajustado al riesgo y diversificación. Para ello, se implementarán herramientas como Python y Power BI, que permitirán realizar simulaciones de diferentes estrategias de inversión y evaluar su desempeño en escenarios reales. A través de estos métodos, se podrá comparar el rendimiento de los portafolios generados por IA con los contruidos bajo los enfoques convencionales, utilizando indicadores como el Ratio Sharpe. Esta comparación ayudará a entender si la IA ofrece una ventaja significativa en la gestión de carteras dentro de un entorno financiero volátil y competitivo.

1.2 Planteamiento del problema

El planteamiento de un portafolio de inversión eficiente es clave para maximizar el rendimiento ajustado al riesgo, especialmente en mercados dinámicos como el de renta variable. Aunque modelos tradicionales como la Teoría Moderna de Portafolios y el CAPM han sido ampliamente utilizados, dependen de supuestos simplificados que no siempre se cumplen en entornos financieros volátiles (Fabozzi et al., 2002). Ante estas limitaciones, la inteligencia artificial surge como una alternativa más flexible y eficiente para optimizar portafolios y enfrentar los desafíos actuales del mercado.

Uno de los principales problemas con los métodos tradicionales de optimización de portafolios es su incapacidad para procesar grandes volúmenes de datos y adaptarse a cambios abruptos del mercado. Los modelos tradicionales, como el CAPM, asumen que el mercado es eficiente y que los riesgos pueden ser sistemáticamente controlados, pero no logran capturar dinámicas complejas como los cambios abruptos en los precios o los eventos inesperados. En contraste, la inteligencia artificial tiene la capacidad de analizar datos en

tiempo real, identificar patrones no evidentes y ajustar las estrategias de inversión de manera más flexible. Sin embargo, a pesar de sus ventajas, aún existen dudas sobre si la implementación de IA puede realmente superar a los enfoques tradicionales en cuanto a rendimiento ajustado al riesgo, en mercados altamente volátiles (CFA Institute, 2022).

La inteligencia artificial optimiza la gestión de portafolios al analizar una gran cantidad de datos en tiempo real, identificar patrones complejos y ajustar estrategias de inversión de manera flexible. Métodos como las redes neuronales profundas y el aprendizaje reforzado han mejorado la predicción de tendencias y la asignación de activos en mercados volátiles (Gu, Kelly, & Xiu, 2020). Gracias a esta capacidad adaptativa, los inversionistas pueden maximizar la rentabilidad ajustada al riesgo y fortalecer la resiliencia de sus portafolios. Así, la IA no solo complementa los modelos tradicionales, sino que también ofrece una ventaja competitiva en la inversión moderna.

Un criterio importante para evaluar la efectividad de estos modelos tradicionales es el Ratio Sharpe, que mide el rendimiento ajustado al riesgo de un portafolio. Este índice permite comparar la eficiencia de los portafolios creados con modelos convencionales, como el CAPM, frente a aquellos que incorporan tecnologías más avanzadas, como la IA. La inteligencia artificial ha demostrado ser una herramienta potente, capaz de procesar un gran volumen de datos en tiempo real y detectar patrones complejos, cambios abruptos y nuevas variables que los modelos tradicionales no pueden abordar. Esto plantea la posibilidad de que los portafolios construidos mediante IA puedan ofrecer un mejor rendimiento ajustado al riesgo en comparación con los modelos clásicos (Bloomberg, 2023).

Para comparar modelos tradicionales con aquellos basados en inteligencia artificial, se analiza su desempeño en términos de riesgo y rendimiento usando herramientas como el Ratio Sharpe. Este índice, según Ross, Westerfield y Jordan (2016), mide la relación entre el

rendimiento de un portafolio y el riesgo asumido. Aunque la IA promete optimizar la gestión de portafolios al superar las limitaciones de los modelos tradicionales como el CAPM, es necesario evaluar cuidadosamente si realmente logra mejores resultados antes de adoptarla como alternativa preferida.

El objetivo principal de este estudio es utilizar el Ratio Sharpe como un criterio para determinar si los portafolios contruidos con inteligencia artificial superan a los tradicionales en términos de rendimiento y riesgo. Se busca evaluar la eficacia de los enfoques modernos en la optimización de portafolios, especialmente en un entorno financiero altamente volátil como el de Estados Unidos. Este análisis permitirá evaluar cómo la IA podría mejorar las estrategias de diversificación y gestión de activos en comparación con los métodos convencionales. Con esto, se pretende arrojar luz sobre la utilidad de las tecnologías emergentes en la mejora de los portafolios de inversión.

1.3 Propósito

Este estudio se enfoca en analizar la efectividad de los modelos tradicionales de optimización de portafolios en comparación con las estrategias basadas en IA para la creación de portafolios diversificados de renta variable en el mercado financiero de Estados Unidos durante el periodo 2014-2024. Dada la naturaleza dinámica y volátil de los mercados actuales, se pretende evaluar si las metodologías clásicas, como la Teoría Moderna de Portafolios y el CAPM, son suficientes para maximizar el rendimiento ajustado al riesgo o si la IA, especialmente el machine learning, ofrece ventajas en términos de eficiencia, diversificación y adaptación a cambios rápidos del mercado.

Un enfoque común en los modelos financieros tradicionales es el uso de datos basados en métodos estadísticos sencillos que asumen una relación lineal entre riesgo y retorno. Por

medio de datos generados en Python y Power BI que incorporan múltiples variables y técnicas complejas, lo que puede llevar a una mejor adaptación y predicción en un entorno de mercado cambiante, teniendo en cuenta los riesgos de agrupar muchas variables (Amat R, 2020).

Este estudio tiene como fin aportar evidencia empírica sobre las ventajas y limitaciones de los métodos tradicionales frente a los enfoques basados en inteligencia artificial, ofreciendo así nuevas perspectivas sobre la optimización de portafolios en un entorno financiero cada vez más complejo y competitivo.

1.4 Naturaleza del estudio

Para abordar las preguntas de investigación, este estudio adoptará un enfoque cuantitativo, utilizando herramientas tecnológicas avanzadas para evaluar y comparar los modelos tradicionales de optimización de portafolios con los basados en IA. Se seleccionó este enfoque debido a su capacidad para procesar grandes volúmenes de datos y realizar simulaciones precisas para comparar rendimiento ajustado al riesgo y diversificación.

Las variables dependientes clave en este estudio incluyen el rendimiento ajustado al riesgo, medido a través del Ratio Sharpe, y la diversificación, evaluada mediante indicadores como la correlación y la varianza del portafolio. Las variables independientes del estudio incluyen el tipo de modelo (tradicional o IA), los datos utilizados, las herramientas tecnológicas empleadas y el periodo de análisis 2014-2024, todos ellos determinantes para evaluar la efectividad en la creación de portafolios diversificados.

El estudio empleará plataformas tecnológicas como Python, Excel y Power BI para implementar y simular ambos enfoques. Estas herramientas permitirán ejecutar modelos

tradicionales, así como algoritmos de machine learning para la construcción de portafolios basados en IA.

El análisis se llevará a cabo utilizando técnicas estadísticas y simulaciones para comparar los resultados de los portafolios generados por ambos métodos, evaluando su desempeño en el periodo comprendido entre 2014 y 2024. Este diseño metodológico permitirá proporcionar una visión objetiva de la efectividad de la IA frente a los enfoques tradicionales en el contexto del mercado financiero de Estados Unidos.

1.5 Introducción al marco teórico

La creación y gestión de portafolios diversificados ha sido un elemento esencial en las finanzas, especialmente en el mercado de renta variable. Desde que Harry Markowitz (1952) desarrolló la Teoría Moderna de Portafolios, los modelos tradicionales han sido clave para optimizar las inversiones, buscando maximizar el rendimiento esperado a partir de un nivel específico de riesgo. Estos modelos, que se basan en supuestos como la distribución normal de los rendimientos y la racionalidad de los inversores, han sido ampliamente utilizados en el sector (Fabozzi et al., 2002).

El mercado de renta variable de los Estados Unidos es uno de los más robustos, líquidos y competitivos a nivel global. Su estructura ha sido diseñada para promover la eficiencia y la confianza de los inversionistas mediante reglas claras y una constante innovación tecnológica. Este mercado desempeña un papel fundamental en la formación de capital, permitiendo a las empresas públicas acceder a recursos y a los inversores participar en los beneficios del crecimiento económico (SIFMA, 2023).

El reporte del CFA Institute (2020) examina el impacto de la inteligencia artificial en la industria de la gestión de activos, especialmente en lo relacionado con el análisis financiero

y la toma de decisiones. En ese marco, se destaca cómo la IA se está aplicando en la construcción de portafolios, aprovechando su habilidad para analizar grandes cantidades de datos, detectar patrones no evidentes y optimizar las estrategias de inversión. Este enfoque proporciona una visión valiosa para comprender el papel de la IA en la optimización y diversificación de portafolios de renta variable, especialmente al compararlo con los métodos tradicionales de gestión de activos.

Para responder las preguntas planteadas, este estudio empleará un enfoque cuantitativo respaldado por herramientas tecnológicas avanzadas. Se utilizarán Power BI, Python y Yahoo Finance para implementar y evaluar tanto los modelos tradicionales de optimización de portafolios como las estrategias basadas en inteligencia artificial. Estas herramientas permitirán el acceso a datos históricos y la realización de simulaciones en tiempo real, facilitando así una comparación sólida del rendimiento y la diversificación de ambas metodologías, ajustadas al riesgo.

1.6 Pregunta de investigación

La cuestión clave es si las tecnologías de IA pueden crear portafolios de inversión más eficientes y con mayor diversificación de riesgo que los modelos tradicionales, específicamente en el contexto del mercado de Estados Unidos, uno de los más grandes y volátiles a nivel global. Se busca determinar si la IA supera las limitaciones de los modelos clásicos al integrar más variables y adaptarse mejor a las condiciones cambiantes del mercado, sin aumentar significativamente los costos o la complejidad de la gestión del portafolio.

Por lo tanto este documento se plantea la siguiente pregunta: ¿cuáles son las diferencias en el rendimiento y la diversificación del riesgo entre los portafolios generados por IA y los construidos con modelos tradicionales durante el periodo 2014-2024?

Este análisis busca comparar las fortalezas y debilidades de ambos enfoques, determinando en qué contexto es más efectivo cada uno para la diversificación del riesgo y la maximización de retornos en el diseño de carteras de inversión.

1.7 Significado del estudio

El propósito de este estudio es evaluar la efectividad de los métodos tradicionales y las estrategias basadas en inteligencia artificial para optimizar portafolios de renta variable en el contexto del mercado financiero estadounidense, mediante un enfoque que integra tanto análisis teóricos como empíricos.

El estudio proporcionará a inversionistas y estudiantes de finanzas una comprensión sobre como las nuevas tecnologías pueden mejorar e incluso superar a los enfoques tradicionales, proporcionando una comparación entre ambos métodos.

Además, se espera que este estudio sirva como una referencia para estudiantes, profesionales y cualquier persona interesada en el mercado de renta variable de EEUU, con el objetivo de mejorar las estrategias de inversión y maximizar el rendimiento ajustado al riesgo en el mercado financiero.

Este estudio se alinea con el Objetivo de Desarrollo Sostenible número 9, enfocado en fortalecer infraestructuras resilientes, impulsar una industrialización sostenible y fomentar la innovación. Asimismo, responde a la meta 9b, que promueve el avance tecnológico, la investigación y la innovación en países en desarrollo, facilitando un entorno normativo que

favorezca la diversificación industrial y el aumento del valor agregado en los productos básicos (Naciones Unidas, 2015).

1.8 Conclusión

Los modelos tradicionales han sido herramientas fundamentales en la gestión de inversiones al proporcionar un enfoque estructurado para optimizar la relación entre riesgo y rendimiento. Basados en supuestos clave, como la normalidad de los retornos, la eficiencia de los mercados y la estabilidad de las correlaciones entre activos, estos modelos han marcado un precedente en la construcción de portafolios. Sin embargo, la complejidad creciente de los mercados financieros, su alta volatilidad, eventos inesperados y la gran cantidad de datos, ha puesto en evidencia las limitaciones de estos enfoques.

La integración de la inteligencia artificial en la gestión de activos representa una evolución significativa en las metodologías tradicionales de análisis y construcción de portafolios. A través del uso de técnicas avanzadas de procesamiento de datos y algoritmos de aprendizaje automático, la IA se ofrece como una herramienta poderosa para la mejora de precisión y eficiencia en la toma de decisiones de inversión.

CAPÍTULO 2: REVISIÓN DE LITERATURA

2.1. Fundamentos del mercado de renta variable en Estados Unidos

El mercado de renta variable en Estados Unidos es uno de los más desarrollados y líquidos del mundo, desempeñando un papel crucial en la economía global al facilitar la financiación de empresas y ofrecer oportunidades de inversión a individuos e instituciones. Según datos de la Federación Mundial de Bolsas, la capitalización bursátil de las empresas estadounidenses representó aproximadamente el 40% de la capitalización bursátil mundial en 2023. Esta preeminencia se atribuye a factores como la estabilidad económica, un marco regulatorio sólido y la presencia de mercados financieros eficientes que fomentan la confianza de los inversores. Además, la diversidad sectorial y la innovación constante en industrias clave, como la tecnología y la salud, han consolidado la posición dominante de Estados Unidos en el ámbito bursátil. Comprender los fundamentos del mercado de renta variable estadounidense es esencial para analizar y comparar diferentes estrategias de construcción de portafolios, ya sean tradicionales o basadas en inteligencia artificial.

2.1.1. Introducción a la Renta Variable

El mercado de renta variable es un pilar fundamental del sistema financiero, ya que permite a los inversionistas adquirir participación en empresas y beneficiarse de su crecimiento. La renta variable se refiere a instrumentos financieros como las acciones, que representan la propiedad parcial de una empresa y otorgan derechos sobre sus ganancias y activos (Bodie, Kane y Marcus, 2018). A diferencia de la renta fija, donde los intereses pagados son previsible, la renta variable no garantiza rendimientos fijos, ya que sus retornos

están influenciados tanto por los resultados financieros de la empresa como por las dinámicas del entorno económico

Si bien la renta variable conlleva un mayor nivel de riesgo debido a la volatilidad del mercado, también ofrece un mayor potencial de rentabilidad en el largo plazo. Factores como el crecimiento económico, la política monetaria y la percepción de los inversionistas influyen en el precio de las acciones. Por esta razón, la renta variable sigue siendo una pieza clave en la construcción de portafolios de inversión, y su análisis ha evolucionado con el uso de modelos tradicionales y tecnologías avanzadas como la inteligencia artificial.

2.1.2. Estructura del Mercado de Acciones

El mercado accionario estadounidense se distingue por la coexistencia de dos de los mercados bursátiles más importantes: la New York Stock Exchange (NYSE) y el NASDAQ. La NYSE opera bajo un modelo híbrido que combina la negociación electrónica con la participación de Designated Market Makers (DMMs), quienes tienen la responsabilidad de mantener mercados justos y ordenados para los valores asignados, facilitando la liquidez y estabilidad durante las aperturas, cierres y períodos de volatilidad .

Por otro lado, el NASDAQ es una plataforma completamente electrónica que se especializa en empresas tecnológicas y de crecimiento, operando múltiples mercados en América del Norte que ofrecen diversos espacios para que los inversores encuentren la liquidez que complementa su estrategia de negociación . La estructura diferenciada de estas bolsas brinda a los inversores diversas oportunidades de inversión, permitiendo la diversificación de carteras según sectores y perfiles de riesgo. Comprender las particularidades operativas y estructurales de la NYSE y el NASDAQ es esencial para desarrollar estrategias de inversión efectivas en el mercado estadounidense.

2.1.3. Regulación y organismos supervisores

Uno de los pilares fundamentales del mercado accionario estadounidense es la existencia de un marco regulatorio sólido, el cual es liderado por la Comisión de Bolsa y Valores de los Estados Unidos (Securities and Exchange Commission – SEC). Esta entidad fue creada en 1934 como una respuesta a la crisis bursátil de 1929, con el fin de restaurar la confianza del público en los mercados financieros. Desde entonces, la SEC ha evolucionado hasta convertirse en la autoridad central encargada de regular y supervisar todos los actores relevantes del mercado de valores, incluyendo empresas emisoras, corredores de bolsa, fondos de inversión, asesores financieros, agencias calificadoras y las propias bolsas de valores. La principal responsabilidad de la SEC es resguardar los intereses de los inversionistas y garantizar que los mercados financieros funcionen de manera justa, transparente y accesible en cuanto a la información. Para ello, implementa normativas como la Securities Act de 1933 y la Securities Exchange Act de 1934, las cuales obligan a las empresas que cotizan en bolsa a divulgar regularmente estados financieros auditados. Esto fomenta la responsabilidad corporativa y brinda a los actores del mercado herramientas confiables para tomar decisiones fundamentadas.

Además de establecer lineamientos legales, la SEC ejerce una función fiscalizadora activa mediante investigaciones, sanciones y mecanismos de cumplimiento, con especial atención a delitos financieros como el uso de información privilegiada (insider trading), fraudes contables o manipulación de precios. En años recientes, la comisión ha intensificado su enfoque en el monitoreo algorítmico de transacciones y en el fortalecimiento de la ciberseguridad, dados los avances tecnológicos que han transformado la forma en que operan los mercados. Asimismo, colabora con otros entes como la Financial Industry Regulatory

Authority (FINRA) y la Commodity Futures Trading Commission (CFTC) para abarcar áreas complementarias del sistema financiero. Esta coordinación interinstitucional es crucial para enfrentar desafíos modernos como la volatilidad inducida por algoritmos, la negociación de activos digitales y los riesgos sistémicos globales. En resumen, la existencia de reguladores especializados y actualizados no solo garantiza el cumplimiento normativo, sino que también ofrece un entorno de mayor seguridad y confianza para los inversores nacionales e internacionales, lo que a su vez promueve la estabilidad y el crecimiento del mercado de capitales estadounidense.

2.1.4 Índices Bursátiles como Referencia de Portafolios

Los índices bursátiles desempeñan un papel fundamental en la evaluación del rendimiento de los portafolios de inversión, actuando como barómetros del comportamiento del mercado y proporcionando puntos de referencia para los inversores. Entre estos, el S&P 500 destaca como uno de los más representativos, ya que mide el desempeño de 500 empresas líderes en Estados Unidos, abarcando aproximadamente el 80% de la capitalización bursátil disponible. Este índice es ampliamente reconocido por su capacidad para reflejar la salud y dirección del mercado de valores estadounidense, y su metodología de ponderación por capitalización bursátil ajustada por flotación asegura que las empresas más grandes tengan una mayor influencia en el índice.

Utilizar el S&P 500 como referencia permite a los inversores comparar el rendimiento de sus carteras con el del mercado en general, facilitando la identificación de desviaciones y la evaluación de la eficacia de sus estrategias de inversión. Sin embargo, es importante considerar que la concentración de grandes empresas tecnológicas en el S&P 500 puede

influir en su comportamiento y, por ende, en las comparaciones realizadas. Por lo tanto, comprender la composición y metodología de los índices bursátiles es esencial para interpretar adecuadamente las comparaciones de rendimiento y tomar decisiones de inversión informadas.

2.1.5 Factores Macroeconómicos que Impactan el Mercado

El desempeño del mercado de acciones estadounidense está intrínsecamente ligado a diversos factores macroeconómicos que reflejan la salud y dirección de la economía en general. Dentro de estos indicadores, el Producto Interno Bruto (PIB) representa una medida fundamental que refleja el valor agregado de todos los bienes y servicios generados en el territorio nacional. Un crecimiento robusto del PIB suele asociarse con un aumento en las ganancias corporativas, lo que puede impulsar al alza los precios de las acciones. Por el contrario, una desaceleración o contracción del PIB puede generar preocupaciones sobre la rentabilidad futura de las empresas, ejerciendo presión a la baja en el mercado accionario. Además, la tasa de desempleo es otro indicador crucial; niveles bajos de desempleo sugieren una economía fuerte con mayor poder adquisitivo, mientras que tasas elevadas pueden indicar debilidad económica, afectando negativamente la confianza de los inversores y, por ende, el rendimiento del mercado de valores.

Las tasas de interés, influenciadas principalmente por las decisiones de política monetaria de la Reserva Federal, ejercen una influencia relevante sobre el comportamiento del mercado bursátil. Incrementos en las tasas de interés encarecen el costo del crédito para las empresas, lo que puede limitar sus inversiones y reducir las expectativas de crecimiento, afectando negativamente los precios de las acciones. Por el contrario, un descenso en las

tasas de interés facilita el endeudamiento, potencialmente estimulando la expansión empresarial y elevando el valor de las acciones. La inflación es otro factor determinante; niveles moderados de inflación pueden ser indicativos de una economía en crecimiento, pero una inflación elevada puede erosionar el poder adquisitivo y aumentar los costos operativos de las empresas, reduciendo sus márgenes de ganancia y afectando negativamente al mercado accionario. Comprender la interacción de estos factores macroeconómicos es esencial para los inversores, ya que influyen directamente las valoraciones de las acciones y las decisiones estratégicas de inversión.

2.2. Modelos tradicionales en la construcción de portafolios

La Teoría Moderna de Portafolios (MPT), propuesta por Markowitz en 1952, se consolidó como un pilar fundamental en el campo de las inversiones al introducir un método cuantitativo para seleccionar activos de manera eficiente. Su aporte más destacado es el concepto de la frontera eficiente, que identifica aquellos portafolios capaces de ofrecer el mayor rendimiento posible para un nivel de riesgo específico. Markowitz evidenció que es posible disminuir el riesgo total de una cartera mediante la diversificación, siempre que los activos incluidos no estén perfectamente correlacionados. Esta idea ha influido significativamente en la gestión de inversiones y ha sido ampliamente adoptada en la industria financiera (Fabozzi, Gupta & Markowitz, 2002). Sin embargo, la MPT parte de varios supuestos teóricos, como la aversión al riesgo de los inversores y la distribución normal de los rendimientos, lo que limita su aplicabilidad en escenarios de mercado más complejos y dinámicos. A pesar de ello, sigue siendo una referencia clave en la optimización de portafolios y ha servido como base para modelos más avanzados de inversión.

Si bien la MPT ha sido fundamental en la teoría financiera, sus limitaciones han sido objeto de debate, especialmente en relación con la volatilidad extrema y los eventos de cola en los mercados financieros. Fabozzi, Gupta y Markowitz (2002) señalan que el supuesto de correlación lineal entre activos subestima el impacto de interacciones no lineales, lo que puede llevar a decisiones subóptimas en entornos de alta incertidumbre. Además, la estimación de rendimientos esperados y matrices de covarianza es altamente sensible a los datos históricos utilizados, lo que puede generar sesgos en la construcción de portafolios óptimos. Estas limitaciones han motivado el desarrollo de enfoques alternativos, como los modelos de optimización robusta y el uso de inteligencia artificial en la gestión de carteras. Integrar técnicas modernas de análisis de datos con los principios de la MPT podría ofrecer una solución más adaptable a las condiciones actuales del mercado, permitiendo la construcción de portafolios más eficientes en términos de riesgo y rendimiento.

2.2.1. Teoría Moderna de Portafolios

La Teoría Moderna de Portafolios (Modern Portfolio Theory, MPT), formulada por Harry Markowitz en 1952, se ha convertido en uno de los enfoques más influyentes para el diseño eficiente de portafolios de inversión. Esta teoría parte del supuesto de que los inversionistas prefieren evitar el riesgo y, por tanto, buscan optimizar la rentabilidad esperada de su cartera para un nivel determinado de riesgo, o en su defecto, reducir al mínimo el riesgo asumido para alcanzar una rentabilidad deseada. Para lograrlo, la MPT plantea que la diversificación eficiente entre activos con distintas correlaciones permite reducir la volatilidad total del portafolio, sin necesidad de sacrificar retornos.

Desde la perspectiva cuantitativa desarrollada por Markowitz, el análisis del riesgo de un portafolio no debe limitarse al comportamiento individual de cada activo, sino

abordarse como un fenómeno integrado, determinado por las varianzas y covarianzas de los retornos de los activos que lo conforman. Esta visión dio paso al desarrollo del concepto de “frontera eficiente”, que representa el conjunto de carteras que maximizan el rendimiento esperado para cada nivel de riesgo asumido. Con el tiempo, esta teoría fue perfeccionada y ampliada tanto en el entorno académico como en su aplicación práctica en los mercados. A pesar de los años transcurridos desde su formulación, continúa siendo un referente esencial en el estudio y la implementación de estrategias de inversión, sirviendo de base para diversas metodologías actuales de asignación de activos.

2.2.2. Modelo de Valoración de Activos de Capital (CAPM)

El Modelo de Valoración de Activos de Capital (CAPM, por sus siglas en inglés) nace como una evolución de la Teoría Moderna de Portafolios y se ha vuelto una herramienta clave para calcular el rendimiento esperado de un activo según el nivel de riesgo que representa frente al mercado. Fue desarrollado en los años 60 por autores como William Sharpe, John Lintner y Jan Mossin. El modelo parte de la idea de que los inversionistas construyen carteras bien diversificadas y, por eso, solo deberían ser compensados por el riesgo que no puede eliminarse mediante diversificación, es decir, el riesgo sistemático. Este tipo de riesgo se representa mediante el beta (β), que muestra qué tanto varía el rendimiento de un activo en relación con los movimientos del mercado.

La fórmula principal del CAPM indica que el rendimiento que se espera obtener de un activo se calcula sumando la tasa libre de riesgo con la prima de riesgo del mercado, ajustada según el beta del activo, que mide cuánto riesgo tiene ese activo en comparación con el mercado. Esta relación permite a los inversores comparar de manera sistemática

activos con distintos grados de sensibilidad al riesgo del mercado, y proporciona un marco de referencia para evaluar si el precio de un activo es justo considerando el nivel de riesgo que implica. Como explican Bodie, Kane y Marcus (2018), el CAPM fue fundamental para establecer una conexión clara entre teoría financiera y prácticas de gestión de portafolios, dando origen al uso de benchmarks como el índice de mercado para evaluar el desempeño relativo de gestores de inversión.

No obstante, si bien el CAPM ha sido ampliamente adoptado tanto en la teoría como en la práctica, también ha sido objeto de diversas críticas y revisiones. Investigaciones empíricas posteriores han evidenciado que el modelo no siempre predice con precisión los rendimientos observados en el mercado, y que existen otros factores, como el tamaño de la empresa o el ratio de valor en libros, que también influyen en los retornos. A pesar de ello, su simplicidad y capacidad explicativa han convertido al CAPM en un referente clave para la valoración de activos, la evaluación de portafolios y la determinación del costo de capital en finanzas corporativas.

2.2.3 Otros Enfoques Tradicionales

Además de la Teoría Moderna de Portafolios y el modelo CAPM, existen otros enfoques tradicionales que han contribuido al desarrollo de la teoría de la inversión y a la construcción de carteras diversificadas. Uno de los más destacados es el Modelo de Valoración por Arbitraje (APT), propuesto por Stephen Ross, que plantea una alternativa al CAPM al permitir que múltiples factores macroeconómicos influyan en el rendimiento de los activos. A diferencia del CAPM, que se basa en un único factor de mercado, el APT considera una combinación de factores como tasas de interés, inflación, crecimiento económico y otros

indicadores relevantes. Esta flexibilidad metodológica lo convierte en un modelo más adaptable a distintas condiciones de mercado, aunque requiere una mayor complejidad estadística y estimación empírica para su implementación práctica (Ross et al., 2016).

Otro enfoque tradicional relevante es el modelo de asignación estratégica de activos con restricciones, que parte del modelo de media-varianza pero incorpora restricciones prácticas como límites de inversión por activo, exposición máxima por sector o requerimientos regulatorios. Estos modelos permiten adaptar la teoría a contextos institucionales reales, como fondos de pensiones o aseguradoras, donde no es posible construir portafolios estrictamente eficientes por las restricciones operativas o normativas existentes (Bodie et al., 2018). Asimismo, enfoques como el análisis fundamental, desarrollado en gran medida por Benjamin Graham y continuado por Warren Buffett, han promovido estrategias de inversión basadas en la selección de acciones subvaloradas (value investing), lo que representa una alternativa más cualitativa pero tradicionalmente influyente en la gestión de carteras. Todos estos modelos, si bien presentan diferencias metodológicas, comparten la intención de maximizar la rentabilidad ajustada por riesgo, y amplían el marco conceptual de los inversores que buscan superar las limitaciones de los modelos puramente cuantitativos.

2.2.4 Limitaciones de los Modelos Tradicionales

A pesar de su enorme influencia en la teoría financiera, los modelos tradicionales como la Teoría Moderna de Portafolios y el CAPM han sido objeto de numerosas críticas que cuestionan su aplicabilidad en entornos reales. Una de las principales limitaciones radica en los supuestos que sustentan estos modelos, tales como la racionalidad absoluta del

inversor, la eficiencia total del mercado, la normalidad en la distribución de los rendimientos y la disponibilidad ilimitada de información. En la práctica, estos supuestos rara vez se cumplen, ya que los mercados financieros están influenciados por sesgos conductuales, eventos imprevistos y restricciones estructurales. Por ejemplo, el modelo de media-varianza de Markowitz asume que la varianza es una medida adecuada del riesgo, sin embargo, numerosos estudios empíricos han demostrado que los retornos financieros presentan colas más pesadas (kurtosis) de lo que este modelo puede capturar adecuadamente (Fabozzi et al., 2002).

Además, el CAPM, aunque sencillo y elegante en su formulación, ha sido criticado por su bajo poder explicativo al momento de predecir rendimientos reales. Diversas investigaciones empíricas han demostrado que factores como el tamaño de la empresa (small vs. large caps) o su ratio valor/libro influyen significativamente en los retornos, pero no son considerados en el modelo clásico. Esto llevó al desarrollo de modelos multifactoriales, como el de Fama y French, que buscan subsanar esas omisiones. Por otro lado, modelos como el APT presentan dificultades prácticas en la selección de factores adecuados y en la estimación de sus sensibilidades. También es importante señalar que la implementación operativa de estos modelos puede volverse altamente dependiente de datos históricos, lo cual los vuelve vulnerables a cambios estructurales en el mercado. En conjunto, estas limitaciones han impulsado la búsqueda de enfoques más dinámicos, flexibles y adaptables, como los modelos basados en machine learning y técnicas de inteligencia artificial, los cuales prometen capturar patrones más complejos y no lineales presentes en los datos financieros actuales.

2.3. Avances en inteligencia artificial y machine learning en finanzas

La inteligencia artificial (IA) y el machine learning han transformado el sector financiero al mejorar la eficiencia en la gestión de inversiones. Empresas como BlackRock han implementado IA para analizar grandes volúmenes de datos, optimizar carteras y gestionar riesgos de manera más rápida y precisa que los métodos tradicionales (Financial Times, 2018). Estas tecnologías permiten automatizar procesos y detectar patrones en los mercados con mayor exactitud, facilitando decisiones estratégicas en tiempo real. Sin embargo, su éxito depende de la disponibilidad de datos de calidad y de una correcta integración en las estrategias de inversión.

El uso de técnicas de inteligencia artificial, como los algoritmos de *Random Forest*, ha abierto nuevas posibilidades para la construcción dinámica de portafolios financieros. En el estudio de Pinelis y Ruppert (2022), se propuso una estrategia que integra predicciones simultáneas de retorno y riesgo utilizando variables macroeconómicas y financieras, lo que permite ajustar mensualmente la exposición al mercado en función de las condiciones previstas. A diferencia de los enfoques tradicionales, esta metodología aplica criterios de maximización de utilidad y emplea pronósticos generados por modelos de aprendizaje automático, logrando una mayor capacidad de adaptación a entornos volátiles. Los resultados empíricos muestran que esta estrategia no solo mejora la eficiencia del portafolio, sino que también permite mitigar el impacto de periodos de alta incertidumbre, consolidando así su valor práctico en contextos reales de inversión.

Debido a los resultados prometedores de este tipo de enfoques, el presente estudio adopta el modelo Random Forest Regressor como herramienta principal para estimar

rendimientos esperados en la construcción de portafolios. Esta elección se fundamenta en su capacidad para modelar relaciones no lineales, adaptarse a distintos entornos de mercado y manejar un alto volumen de variables predictoras, características que resultan especialmente útiles en el contexto del mercado bursátil estadounidense. Su implementación permitirá contrastar de forma más robusta el desempeño de los portafolios generados frente a los contruidos con modelos tradicionales, alineándose con el objetivo comparativo central de esta investigación.

A pesar de sus ventajas, el uso de IA en finanzas enfrenta desafíos como la falta de transparencia en los algoritmos y el riesgo de amplificar la volatilidad si muchos actores dependen de modelos similares. Además, los sesgos en los datos pueden afectar la precisión de las predicciones y generar resultados inesperados (Deloitte, 2019). Para aprovechar su potencial sin comprometer la estabilidad del mercado, es clave combinar estas herramientas con supervisión humana y establecer regulaciones que mitiguen sus riesgos. A medida que la IA evoluciona, su impacto en la inversión seguirá creciendo, ofreciendo nuevas oportunidades en la construcción de portafolios. Lo anterior resalta la necesidad de evaluar cómo se comparan estos enfoques frente a los modelos tradicionales de inversión.

2.4. Comparaciones de modelos tradicionales y de IA: ventajas y limitaciones

Los modelos tradicionales de construcción de portafolios, como la Teoría Moderna de Portafolios (MPT), han sido ampliamente utilizados para optimizar la relación riesgo-rendimiento mediante principios estadísticos. Sin embargo, el auge de la inteligencia artificial (IA) y el machine learning ha introducido métodos alternativos que analizan grandes volúmenes de datos y detectan patrones que los modelos tradicionales no pueden captar

(Heaton et al., 2017). Mientras que los enfoques tradicionales ofrecen un marco teórico bien establecido y fácil de interpretar, los modelos de IA tienen la ventaja de adaptarse a cambios del mercado en tiempo real, mejorando la precisión en la toma de decisiones.

Una de las principales ventajas de los modelos basados en aprendizaje automático frente a los enfoques tradicionales es su capacidad para capturar relaciones no lineales entre múltiples variables financieras. A diferencia de modelos lineales como el CAPM, algoritmos como Random Forest pueden adaptarse de forma más eficaz a entornos no estacionarios, característicos de los mercados financieros actuales. Según Pinelis y Ruppert (2022), estos modelos no solo mejoran la precisión predictiva, sino que también permiten ajustar dinámicamente la exposición al riesgo ante escenarios adversos, lo que reduce la vulnerabilidad del portafolio durante periodos de crisis. Esta evidencia sugiere que la inteligencia artificial puede complementar, e incluso superar en ciertos contextos, a los enfoques tradicionales en la optimización de portafolios.

A pesar de sus beneficios, ambos enfoques presentan desafíos. Los modelos tradicionales pueden ser rígidos ante mercados volátiles y suposiciones simplificadas pueden afectar su desempeño. Por otro lado, la IA depende de datos de calidad y algoritmos complejos, lo que puede generar problemas de transparencia y dificultades en su interpretación (Gu et al., 2020). Para aprovechar lo mejor de ambos mundos, la combinación de técnicas tradicionales con herramientas de IA podría mejorar la diversificación y el rendimiento de los portafolios, logrando una gestión más eficiente y flexible ante condiciones de mercado cambiantes.

2.5. Uso de datos alternativos y capacidades computacionales en la optimización

El uso de datos alternativos ha revolucionado la gestión de inversiones al permitir análisis más precisos y personalizados. Fuentes como redes sociales, patrones de consumo y registros de movilidad ofrecen información valiosa que complementa los datos financieros tradicionales, mejorando las predicciones de mercado y la asignación de activos (Kolanovic & Krishnamachari, 2017). Además, los avances en computación en la nube y machine learning han facilitado el procesamiento de estos datos en tiempo real, permitiendo estrategias de inversión más dinámicas y adaptadas a cambios del mercado.

A pesar de sus beneficios, el uso de datos alternativos conlleva desafíos como la calidad de la información, sesgos y regulaciones que pueden afectar su fiabilidad (Jensen et al., 2021). Sin embargo, a medida que las capacidades computacionales continúan evolucionando, la integración de estos datos en la optimización de portafolios se vuelve cada vez más eficiente. La combinación de estas herramientas con enfoques tradicionales o basados en inteligencia artificial ofrece oportunidades para mejorar la diversificación y el rendimiento de las inversiones, adaptándose mejor a las condiciones del mercado.

CAPÍTULO 3:

3.1 Enfoque de la investigación

Este estudio se desarrolla bajo un enfoque cuantitativo, ya que se basa en la recolección y análisis de datos numéricos con el objetivo de evaluar la efectividad de distintos modelos de construcción de portafolios. La investigación se centra en comparar el rendimiento y la diversificación obtenidos a partir de modelos tradicionales frente a modelos basados en inteligencia artificial. Para esto, se utilizan datos históricos de precios de acciones, indicadores de riesgo y rendimiento, y herramientas tecnológicas que permiten realizar simulaciones y medir resultados de forma objetiva.

El enfoque cuantitativo fue elegido porque permite trabajar con series temporales y aplicar técnicas estadísticas para medir variables como el retorno ajustado al riesgo (medido por el ratio de Sharpe) o la varianza de los portafolios. Además, este tipo de análisis facilita la comparación directa entre ambos enfoques, utilizando métricas que pueden replicarse y validarse. A diferencia de un enfoque cualitativo o mixto, que se centra más en percepciones o casos particulares, este estudio requiere resultados concretos que se puedan evaluar numéricamente y generalizar en un contexto de mercado real.

Este enfoque también está relacionado directamente con el problema de investigación y los objetivos planteados. Al buscar comprobar si los modelos de IA ofrecen un mejor rendimiento y diversificación que los tradicionales, se necesita medir el comportamiento de portafolios reales durante un periodo determinado, en este caso del 2014 al 2024. El uso de modelos como Random Forest Regressor permite estimar los rendimientos esperados con base en múltiples variables, y compararlos con las carteras generadas a partir de métodos

clásicos como Markowitz o el CAPM. Esto requiere una estructura de análisis clara, sistemática y replicable, lo que refuerza la elección del enfoque cuantitativo.

En un entorno financiero cada vez más dinámico y volátil, la aplicación de modelos cuantitativos permite tomar decisiones basadas en datos objetivos y evidencia empírica. Los mercados actuales se ven influenciados por una gran cantidad de factores simultáneos, desde noticias económicas hasta indicadores técnicos y cambios en la política monetaria, lo cual hace que las decisiones de inversión basadas solo en intuición o experiencia histórica sean cada vez menos efectivas. Por eso, este estudio considera necesario apoyarse en metodologías que permitan procesar de forma rigurosa la información disponible, y que puedan adaptarse a los patrones de comportamiento del mercado mediante análisis sistemáticos, como lo permite el enfoque cuantitativo.

Además, el uso de herramientas como Python, Excel y Power BI facilita la implementación de ambos tipos de modelos, así como la visualización de los resultados. Estas plataformas permiten automatizar cálculos, manejar grandes volúmenes de información y aplicar modelos de machine learning de forma eficiente. Gracias a ellas, es posible construir portafolios bajo distintas metodologías y evaluar su desempeño en distintos escenarios del mercado, lo cual enriquece el análisis y fortalece la validez de las conclusiones obtenidas.

3.2 Tipo y diseño de la investigación

El presente estudio se enmarca dentro de un enfoque cuantitativo de tipo explicativo y correlacional, ya que busca identificar y analizar la relación entre el tipo de modelo utilizado para la construcción de portafolios (tradicional o basado en inteligencia artificial) y el desempeño de dichos portafolios en términos de riesgo, rendimiento y diversificación. El carácter explicativo se justifica porque no solo se describe el comportamiento de los portafolios bajo distintos métodos, sino que se busca determinar si el modelo aplicado influye significativamente en los resultados obtenidos. Asimismo, el componente correlacional radica en la evaluación de relaciones entre variables como los retornos, el ratio de Sharpe y la varianza, en función del modelo utilizado.

En cuanto al diseño, la investigación es no experimental y de corte transversal, ya que no se manipulan las variables independientes de forma directa, sino que se observan los efectos derivados de la aplicación de modelos ya existentes sobre datos históricos del mercado. El estudio se desarrolla sobre un conjunto definido de datos correspondientes al período comprendido entre 2014 y 2024, sin realizar intervención alguna sobre las condiciones del entorno. Además, el análisis es de carácter transversal porque se evalúan los datos en una sola etapa de análisis, sin un seguimiento en el tiempo más allá del horizonte de los datos históricos analizados.

Este tipo de diseño se adapta bien al objetivo del estudio, que es comparar dos formas de construir portafolios usando herramientas estadísticas y tecnológicas. Al trabajar con datos reales y simulaciones, se puede evaluar de forma objetiva la efectividad de cada modelo, sin necesidad de modificar el entorno del mercado o aplicar un experimento.

3.3 Población y muestra

La población del estudio está compuesta por todas las acciones que forman parte del mercado accionario estadounidense, específicamente aquellas que integraron el índice S&P 500 durante el periodo 2014–2024. Este índice fue elegido por representar una muestra amplia y diversificada del mercado de renta variable, abarcando empresas de distintos sectores y con alta capitalización bursátil.

La muestra se definió mediante un enfoque no probabilístico, priorizando la estabilidad y disponibilidad de datos de empresas del S&P 500 activas entre 2014 y 2024. En una primera etapa se analizarán alrededor de 300 acciones de empresas activas durante todo el periodo, y en una segunda se eligieron cinco empresas específicas: Apple Inc. (AAPL), Microsoft Corporation (MSFT), Linde plc (LIN), Starbucks Corporation (SBUX) y The Mosaic Company (MOS). La selección fue empírica, buscando representar distintos sectores económicos (tecnología, industria, consumo y materias primas) y considerando también criterios de diversificación y desempeño en sostenibilidad (ESG). Estas compañías tienen trayectoria sólida, buena liquidez y perfiles adecuados para comparar el comportamiento de portafolios generados con modelos tradicionales y de inteligencia artificial.

Según datos de agencias como MSCI y Sustainalytics, las cinco empresas presentan calificaciones ESG superiores al promedio de sus respectivas industrias: Apple y Microsoft destacan en tecnología con ratings “AAA” y “Líderes”, impulsadas por sus políticas de energía renovable y transparencia corporativa; Linde obtiene una calificación “AA” dentro del sector industrial, reflejando prácticas sólidas en seguridad y emisiones; Starbucks se posiciona como “Superior al promedio” en el sector de consumo discrecional por sus iniciativas laborales y cadena de suministro sostenible; y Mosaic, aunque en un sector

históricamente más expuesto a riesgos ambientales, presenta mejoras continuas en sus prácticas de minería responsable, logrando una calificación “Promedio”, por encima de muchos de sus competidores directos. Este enfoque busca no solo la diversificación sectorial, sino también una alineación con principios de inversión responsable (MSCI, 2024; Sustainalytics, 2024).

Este tipo de selección permite asegurar que las empresas analizadas cuentan con series temporales completas y comparables, evitando distorsiones por entradas o salidas del índice, adquisiciones o datos faltantes. Al tratarse de un análisis cuantitativo basado en series históricas, no se requiere una muestra probabilística, sino una selección intencional basada en criterios de calidad de los datos y estabilidad del activo en el mercado durante el horizonte temporal definido.

3.4 Técnicas e instrumentos de recolección de datos

La técnica principal de recolección de datos en este estudio fue el análisis documental y de series históricas, ya que se trabaja con datos financieros secundarios obtenidos de fuentes públicas. Los precios históricos diarios de acciones del índice S&P 500 entre los años 2014 y 2024 fueron extraídos principalmente desde Yahoo Finance, lo que permitió construir una base de datos confiable y adecuada para los objetivos del estudio. Además de los precios de cierre, se consideraron datos como el volumen negociado y se complementó con algunos indicadores de rendimiento ajustado al riesgo.

En cuanto a los instrumentos, se utilizaron principalmente Power BI, Excel y Python para procesar, limpiar y organizar los datos. Power BI fue clave en la etapa de depuración y estructuración de series temporales, permitiendo identificar valores faltantes, inconsistencias y facilitar la preparación de los datos para su análisis posterior. Excel se empleó

principalmente para la implementación del modelo tradicional (como media-varianza), así como para la visualización de resultados, elaboración de gráficos y cálculos estadísticos como retornos simples, varianza y correlaciones. Finalmente, en Python se aplicaron técnicas de machine learning mediante librerías como pandas, numpy y sklearn, enfocadas en entrenar y ejecutar el algoritmo Random Forest Regressor. Esta combinación de herramientas permitió automatizar procesos, mantener el control sobre la calidad de los datos y comparar con claridad los resultados obtenidos entre ambos enfoques metodológicos.

Para asegurar la calidad y validez de los datos, se aplicaron criterios de depuración como la eliminación de series con datos faltantes, valores atípicos evidentes o inconsistencias temporales. También se verificó que las fechas estuvieran alineadas y que todas las acciones utilizadas tuvieran información completa a lo largo de todo el periodo analizado. El uso de herramientas computacionales facilitó la detección de errores y la estandarización de las series de tiempo, lo cual fue clave para la confiabilidad del análisis.

Aunque no se aplicaron instrumentos como cuestionarios o entrevistas, el manejo riguroso de datos financieros y la aplicación de modelos en entornos controlados permitieron obtener información precisa y replicable. En conjunto, la metodología empleada para la recolección, limpieza y estructuración de datos aporta una base sólida para evaluar el comportamiento de los portafolios contruidos bajo diferentes enfoques.

3.5 Procedimiento

El desarrollo del estudio se realizó en tres etapas principales. En la primera, se definió la población objetivo y se construyó la muestra, seleccionando empresas del índice S&P 500 que contaran con información completa para el período 2014–2024. Esta etapa incluyó la validación de datos disponibles y la identificación de compañías representativas para un análisis más detallado.

En la segunda etapa se organizó la información recolectada, preparando las series históricas para su análisis. Aquí se trabajó en la limpieza de los datos, estructuración de variables clave y validación de consistencia temporal. Esta fase fue clave para asegurar que todos los portafolios se construyeran con datos comparables y libres de errores.

Finalmente, en la tercera etapa se aplicaron los modelos. Primero, se construyeron los portafolios bajo métodos tradicionales, y luego se implementó el modelo de machine learning (Random Forest Regressor) para estimar rendimientos esperados y construir portafolios alternativos. A partir de ahí, se realizó la comparación de resultados, con base en métricas como el rendimiento ajustado al riesgo y la diversificación.

Este procedimiento permitió abordar de forma ordenada cada fase del estudio, desde la selección de datos hasta la evaluación de los resultados, manteniendo una lógica coherente con los objetivos de la investigación.

3.6 Técnicas de análisis de datos

El análisis de datos de este estudio fue de tipo cuantitativo y se realizó en dos niveles. Primero, se aplicaron técnicas estadísticas descriptivas para analizar los datos históricos de precios de las acciones seleccionadas. Esto incluyó el cálculo de retornos diarios, medias, desviación estándar, varianza y correlaciones entre activos, con el objetivo de evaluar la

volatilidad, la dispersión de los rendimientos y el grado de relación entre ellos. Estos cálculos fueron realizados principalmente en Excel, y sirvieron como base para construir los portafolios bajo el enfoque tradicional.

En una segunda fase, se implementó un modelo de aprendizaje automático mediante regresión: el Random Forest Regressor. Este modelo fue utilizado para estimar el rendimiento esperado de cada acción a partir de variables históricas. Para ello, se entrenó el modelo utilizando la librería scikit-learn en Python, y se evaluó su rendimiento fuera de muestra para asegurar su capacidad predictiva. Una vez obtenidas las predicciones, se utilizó esa información para construir portafolios optimizados y compararlos con los generados por el método tradicional, aplicando criterios como el ratio de Sharpe y la diversificación basada en correlaciones cruzadas.

Esta combinación de métodos permitió realizar un análisis robusto, comparando el rendimiento y la eficiencia de los portafolios desde dos enfoques distintos: uno clásico, basado en estadísticas financieras, y otro moderno, apoyado en técnicas de inteligencia artificial. El uso de herramientas como Python y Excel facilitó el procesamiento de los datos y la visualización de los resultados, permitiendo llegar a conclusiones claras y comparables.

3.7 Presentación de datos

Con el objetivo de sustentar el análisis comparativo entre los modelos tradicionales y los basados en inteligencia artificial, esta sección presenta de forma organizada los datos utilizados, así como los resultados intermedios que fundamentan la construcción de los portafolios. La presentación se apoya en tablas, gráficos y matrices estadísticas que permiten visualizar las características esenciales de los activos seleccionados y los comportamientos observados a lo largo del periodo de estudio (2014–2024).

La estructura de esta sección responde a una lógica analítica que parte de la exploración inicial de datos, continúa con el desarrollo de métricas financieras clave (como retornos, varianzas y correlaciones), y culmina con la exposición de los insumos generados por los modelos predictivos. Esta disposición facilita la comparación entre enfoques, ofreciendo una base sólida para interpretar los resultados obtenidos en el capítulo siguiente.

La matriz inicial contiene precios de cierre y volúmenes mensuales de aproximadamente 300 acciones del S&P 500 entre 2014 y 2024, descargados desde Yahoo Finance. Se depuraron las empresas con datos incompletos, conservando únicamente aquellas con series consistentes de precio y volumen, lo que permitió construir una base uniforme para el análisis de ambos modelos.

La Tabla 1 presenta un fragmento representativo de esta matriz depurada, con precios de cierre y volúmenes mensuales de dos acciones seleccionadas del índice S&P 500: Apple Inc. (AAPL) y Microsoft Corp. (MSFT). La tabla corresponde a los primeros meses de 2014 y ejemplifica la estructura de la base consolidada, en la cual cada fila representa un mes calendario y cada columna corresponde a una variable específica (precio o volumen) de un activo.

Tabla 1. Fragmento de matriz de Precios y Volúmenes de Empresas S&P 500

Date	AAPL_Close	AAPL_Volume	MSFT_Close	MSFT_Volume
1/1/14	\$ 15,58	8765954400	\$ 31,56	930226200
1/2/14	\$ 16,38	5880366800	\$ 31,96	705304500
1/3/14	\$ 16,81	5001698800	\$ 34,45	778425700
1/4/14	\$ 18,48	6435060800	\$ 33,95	746112500
1/5/14	\$ 19,82	5735668400	\$ 34,41	574362900

Fuente: Código 1. Descarga base S&P 500 de YH Finance
Elaboración: Propia

Por otro lado, la Tabla 2 muestra un extracto de la matriz de precios y rendimientos mensuales de las cinco acciones seleccionadas para este estudio: Apple Inc. (AAPL), Microsoft Corp. (MSFT), The Mosaic Company (MOS), Linde plc (LIN) y Starbucks Corp. (SBUX). Estas compañías fueron elegidas con base en un criterio empírico y considerando su desempeño en sostenibilidad (ESG), así como su representatividad sectorial. A partir de los precios de cierre, se calcularon las variaciones porcentuales mensuales, que sirvieron como base para las métricas utilizadas en el análisis tradicional y para el entrenamiento del modelo de machine learning.

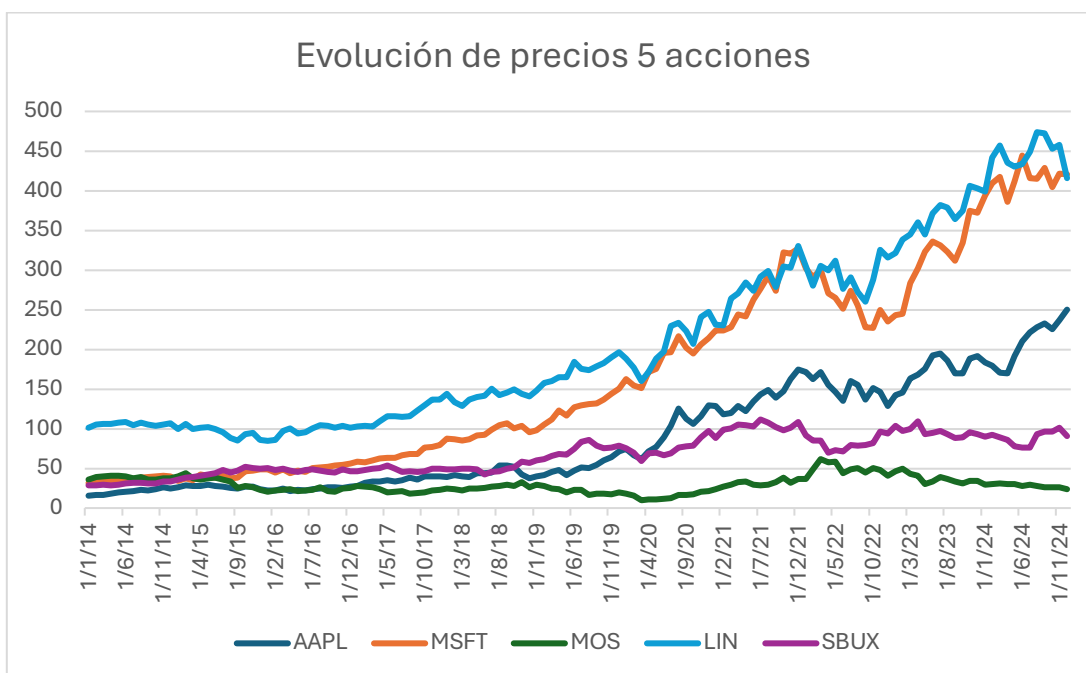
Tabla 2. Fragmento de matriz de Precios y Rendimientos de Empresas S&P 500

Fecha	AAPL	Variación AAPL	MSFT	Variación MSFT	MOS	Variación MOS	LIN	Variación LIN	SBUX	Variación SBUX
1/1/14	15,580		31,564		35,893		101,050		28,688	
1/2/14	16,378	5,12%	31,956	1,24%	39,268	9,40%	105,628	4,53%	28,624	-0,23%
1/3/14	16,805	2,61%	34,448	7,80%	40,184	2,33%	106,114	0,46%	29,712	3,80%
1/4/14	18,476	9,94%	33,952	-1,44%	40,423	0,59%	106,301	0,18%	28,594	-3,76%
1/5/14	19,819	7,27%	34,406	1,34%	40,382	-0,10%	107,677	1,29%	29,655	3,71%

Elaboración: Propia

El precio de cierre fue la variable clave tanto para calcular retornos como para alimentar los modelos predictivos. Su estabilidad mensual permite una medición adecuada del desempeño en horizontes de mediano plazo. Por su parte, el volumen negociado se incorporó como variable adicional especialmente en el modelo de machine learning, aportando información sobre la liquidez y el interés del mercado por cada acción. Ambas variables se integraron posteriormente en el proceso de predicción de rendimientos con el algoritmo Random Forest.

Figura 1. Evolución mensual del precio de las acciones (2014–2024)



Elaboración: Propia

La Figura 1 muestra la evolución de los precios de cierre mensuales para las cinco acciones seleccionadas en el estudio. Se puede observar una clara tendencia creciente en activos como LIN y MSFT, mientras que otros, como MOS, presentan un comportamiento más volátil y lateral. Este gráfico permite visualizar las diferencias en escalas y trayectorias individuales, lo que será relevante en la interpretación del desempeño de los portafolios contruidos con cada enfoque.

Con esta presentación de los datos, se establece la base necesaria para aplicar los modelos de análisis. A partir de esta información, se desarrollará el análisis descriptivo de las variables clave, lo que permitirá identificar patrones y características relevantes en el comportamiento de las acciones seleccionadas.

3.8 Análisis descriptivo

En la Tabla 3 se observan diferencias claras en el comportamiento de los retornos mensuales de las cinco acciones. AAPL y MSFT destacan por tener los promedios más altos, aunque también presentan una desviación estándar considerable, lo que indica un buen rendimiento pero con cierto nivel de riesgo.

En cambio, MOS muestra la mayor volatilidad del grupo, con un bajo retorno promedio y la desviación estándar más alta. Aunque su moda es positiva, esto parece deberse a unos pocos valores extremos que no reflejan un comportamiento estable.

Por su parte, LIN y SBUX presentan retornos más moderados y menor dispersión, lo que podría hacerlas útiles para equilibrar el riesgo dentro del portafolio.

Tabla 3. Estadísticas descriptivas de los retornos mensuales

	AAPL	MSFT	MOS	LIN	SBUX
Promedio	0,0246	0,0219	0,0045	0,0126	0,0114
Mediana	0,0289	0,0223	0,0139	0,0142	0,0134
Moda	-0,0067	-0,0661	0,1100	-0,0123	-0,0238
Desviación estándar	0,0768	0,0602	0,1125	0,0548	0,0656
Varianza	0,0059	0,0036	0,0127	0,003	0,0043
Mínimo	-0,184	-0,1302	-0,3647	-0,1144	-0,1795
Máximo	0,2144	0,1963	0,3534	0,1637	0,2132

Fuente: Excel Modelo tradicional

Elaboración: Propia

También se notan diferencias entre la media, mediana y moda, especialmente en AAPL y MSFT, donde la moda es negativa pero la media positiva. Esto sugiere que los valores más frecuentes tienden a ser bajos, aunque algunos meses con buenos resultados elevan el promedio general.

3.8.1 Resultados pesos del portafolio

En el modelo tradicional de Markowitz, los pesos óptimos se calcularon a partir de los retornos y la matriz de varianzas y covarianzas. Como se muestra en la Tabla 4, MSFT y AAPL concentraron la mayor parte del portafolio, mientras que MOS, LIN y SBUX tuvieron una participación marginal. Esta asignación refleja la relación histórica entre rendimiento y riesgo de cada activo.

Tabla 4. Pesos asignados por el modelo tradicional

AAPL	MSFT	MOS	LIN	SBUX	desvío objetivo	rendim	validación	Sharpe ratio	CML
32,7%	62,1%	1,0%	2,2%	2,0%	19,82%	29,92%	OK	140,84%	29,9%

Fuente: Excel Modelo tradicional
Elaboración: Propia

El criterio usado para optimizar el portafolio fue el ratio de Sharpe, que mide cuánto retorno adicional se obtiene por cada unidad de riesgo. En este modelo, se alcanzó un Sharpe del 140,84 %, lo que indica una buena eficiencia en la combinación seleccionada.

Por otro lado, el modelo basado en machine learning utilizó el algoritmo Random Forest Regressor para estimar rendimientos esperados en función de variables históricas. Estos valores se utilizaron luego como insumo para una optimización similar a la tradicional, pero basada en predicciones. Como se observa en la Tabla 5, MSFT vuelve a destacarse con una participación dominante del 69 %, mientras que AAPL se mantiene con un 23 %. En este caso, LIN y SBUX recibieron pesos mínimos, y MOS fue completamente excluida.

Tabla 5. Pesos asignados por el modelo de Machine Learning

	AAPL	LIN	MSFT	SBUX
Peso	23%	2%	69%	6%

Fuente: Código 4. Pesos
Portafolio con ML
Elaboración: Propia

Mientras que el modelo tradicional se basa en estadísticas históricas de retorno y varianza, el modelo de machine learning toma en cuenta patrones más complejos en los datos, priorizando acciones con mejores perspectivas según las variables entrenadas. Esto explica por qué MOS, con alta volatilidad, fue excluida por el modelo predictivo a pesar de tener una leve participación en el enfoque clásico.

Además, los rendimientos estimados por el modelo Random Forest mostraron un buen nivel de ajuste en su fase de validación. Como se resume en la Tabla 6, los valores de R^2 para las cinco acciones seleccionadas superan el 80 %, lo que indica una buena capacidad explicativa del modelo. Acciones como SBUX y MSFT alcanzaron los mejores niveles de ajuste, con R^2 de 0,86 y 0,85 respectivamente. Aunque el MAE fue ligeramente más alto en el caso de MOS, este activo fue finalmente descartado en el portafolio optimizado, lo cual coincide con su bajo rendimiento histórico. Estos resultados respaldan la solidez del modelo predictivo y justifican su uso como base para la construcción del portafolio alternativo.

Tabla 6. Métricas de validación del modelo Random Forest

Ticker	AAPL	MSFT	LIN	SBUX	MOS
MAE	0,0244	0,0173	0,0144	0,0149	0,0326
R²	0,8199	0,8543	0,8335	0,8606	0,8421

Fuente: Código 2. Modelo Rendimiento MAE R^2
Elaboración: Propia

Adicionalmente, se calculó el ratio de Sharpe del portafolio optimizado con los rendimientos proyectados por el modelo de machine learning. Este alcanzó un valor de 127,6 %, inferior al obtenido por el modelo tradicional (140,84 %), pero aún reflejando una eficiencia razonable considerando que la optimización se basó en predicciones y no en datos históricos realizados. Esta diferencia refuerza la importancia de evaluar no solo la capacidad

predictiva del modelo, sino también su desempeño al momento de construir portafolios reales.

3.8.2 Simulación masiva con Random Forest

A partir de la base de datos previamente descrita, que contiene precios y volúmenes mensuales de más de 300 acciones del índice S&P 500, se implementó una simulación predictiva utilizando el algoritmo Random Forest Regressor. El objetivo fue evaluar la capacidad del modelo de machine learning para estimar rendimientos esperados de forma masiva, replicando un escenario más amplio y realista dentro del análisis cuantitativo de portafolios.

El archivo resultante contiene más de 3.000 predicciones mensuales generadas a partir de datos históricos. En promedio, el modelo proyectó rendimientos mensuales moderados, con una media cercana al 1,5 % y una distribución centrada, sin sesgos extremos. Al agrupar las predicciones por año, se observa una relativa estabilidad, aunque con ligeras variaciones asociadas al comportamiento del mercado en ciertos periodos.

Los resultados de la Tabla 7 que las acciones con mayores rendimientos estimados pertenecen, en su mayoría, a sectores como tecnología, salud y consumo. Esta concentración sectorial sugiere que el modelo logró identificar patrones de comportamiento positivos en industrias con alto dinamismo durante el periodo analizado. En conjunto, estos hallazgos refuerzan la escalabilidad del enfoque de machine learning para evaluar grandes volúmenes de datos financieros y facilitar decisiones estratégicas de inversión a nivel masivo.

Tabla 7. Top 10 acciones con mayor retorno proyectado por el modelo

Ticker	Empresa	Sector aproximado	Retorno Proyectado
TSLA	Tesla	Tecnología / Automotriz	35,06 %
AMD	Advanced Micro Devices	Semiconductores	7,14 %
PODD	Insulet Corporation	Tecnología médica	6,67 %
PAYC	Paycom Software	Software empresarial	6,63 %
WBD	Warner Bros. Discovery	Medios / Entretenimiento	6,59 %
ANSS	ANSYS Inc.	Software de simulación	5,72 %
CPRT	Copart Inc.	Servicios automotrices	5,57 %
NEE	NextEra Energy	Energía renovable	5,53 %
LUV	Southwest Airlines	Aerolíneas	5,51 %
BIIB	Biogen Inc.	Biotecnología	5,40 %

Fuente Código 3. Rendimiento Portafolio con ML
Elaboración: Propia

Los hallazgos presentados a continuación complementan el análisis anterior con un enfoque más técnico sobre la calidad de las predicciones generadas por el modelo de machine learning y su aplicación en la construcción de portafolios eficientes. En la Tabla 8 se listan las 10 acciones seleccionadas por el modelo con los mayores niveles de ajuste, medidos mediante el coeficiente R^2 , junto con su retorno promedio proyectado. Estas acciones fueron identificadas a partir de un análisis sistemático de más de 300 empresas del S&P 500, priorizando tanto la calidad del modelo predictivo como la liquidez del activo. Los resultados destacan a empresas de sectores dinámicos como tecnología, comunicaciones y bienes raíces, evidenciando patrones consistentes de comportamiento positivo en sus precios históricos.

Tabla 8. Top acciones R^2 y retorno promedio

Ticker	Empresa	Industria	R^2	Retorno Promedio
FOXA	Fox Corporation	Communication Services	0,87526	3%
CRM	Salesforce Inc.	Technology	0,87222	3%
INVH	Invitation Homes Inc.	Real Estate	0,88499	2%
XOM	Exxon Mobil Corp.	Energy	0,88255	0%
PYPL	PayPal Holdings Inc.	Financials / Tech	0,87940	1%
VST	Vistra Corp.	Utilities	0,87829	-2%
DD	DuPont de Nemours Inc.	Materials	0,87809	1%
UNH	UnitedHealth Group Inc.	Health Care	0,86833	-5%
JNPR	Juniper Networks Inc.	Technology	0,86605	-2%
CCL	Carnival Corp	Consumer Discretionary	0,86426	3%

Fuente: Código 2. Modelo Rendimiento MAE R^2
Elaboración: Propia

A partir de esta selección, se procedió a construir un portafolio óptimo aplicando la función `max_sharpe()` de la librería `PyPortfolioOpt` (Martin, 2020), la cual permite encontrar la combinación de activos que maximiza el rendimiento ajustado por riesgo. El resultado, mostrado en la Tabla 9, fue un portafolio altamente concentrado en tres de las diez acciones evaluadas, reflejo de su rentabilidad esperada y su bajo nivel de correlación con las demás.

Tabla 9. Portafolio Óptimo con ML de todas las acciones

PORTAFOLIO OPTIMO ML		
	Nombre	Peso (%)
FOXA	Fox Corporation	59.03%
CRM	Salesforce	26.16%
INVH	Invitation Homes	14.81%

Código 7. Modelo ML estatico, todas las acciones
Elaboración: Propia

La Tabla 10 presenta un resumen comparativo de los tres modelos evaluados en este estudio, simulando una inversión inicial de \$100. Esta metodología busca facilitar la interpretación de los resultados a partir de una métrica intuitiva y estandarizada, permitiendo

observar la evolución del capital invertido bajo distintos enfoques. El modelo tradicional optimizado, basado en rendimientos históricos, logró convertir los \$100 en \$129,92, con una rentabilidad acumulada de 29,92% y un Sharpe Ratio de 1,40. Por su parte, el modelo ML aplicado a las mismas cinco acciones obtuvo un rendimiento menor (24,88%) y un Sharpe de 1,27, lo cual sugiere que las predicciones del modelo por sí solas no aportan una ventaja significativa si no están acompañadas de una correcta selección de activos. Finalmente, el modelo ML con libertad total de selección y optimización logró un rendimiento acumulado de 59,71% y un Sharpe Ratio de 2,45, lo que representa una mejora considerable tanto en rentabilidad como en eficiencia ajustada al riesgo.

Tabla 10. Comparación 3 modelos con 100\$

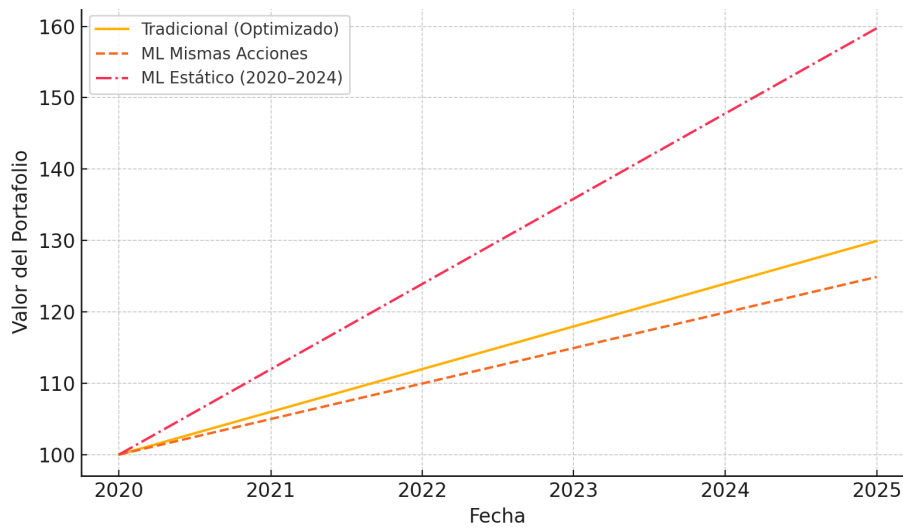
Modelo / Periodo	Inversión Inicial (\$)	Valor Final (\$)	Rentabilidad (%)	Sharpe (2%)
Tradicional (Optimizado)	100	129,92	29,92	1,4
ML Mismas acciones	100	124,88	24,88	1,27
ML Estático (2020– 2024)	100	159,71	59,71	2,45

Elaboración: Propia

3.8.3 Comparación visual de portafolios optimizados

La Figura 2 presenta una comparación gráfica simplificada de la evolución de una inversión inicial de \$100 en los tres modelos de portafolio analizados, sin considerar capitalización diaria. Esta simulación parte de los valores finales reportados en la tabla anterior, los cuales reflejan las rentabilidades acumuladas estimadas a lo largo del período 2020–2024: 29,92% para el modelo tradicional, 24,88% para el modelo de machine learning aplicado a las mismas cinco acciones, y 59,71% para el modelo ML libre (portafolio estático).

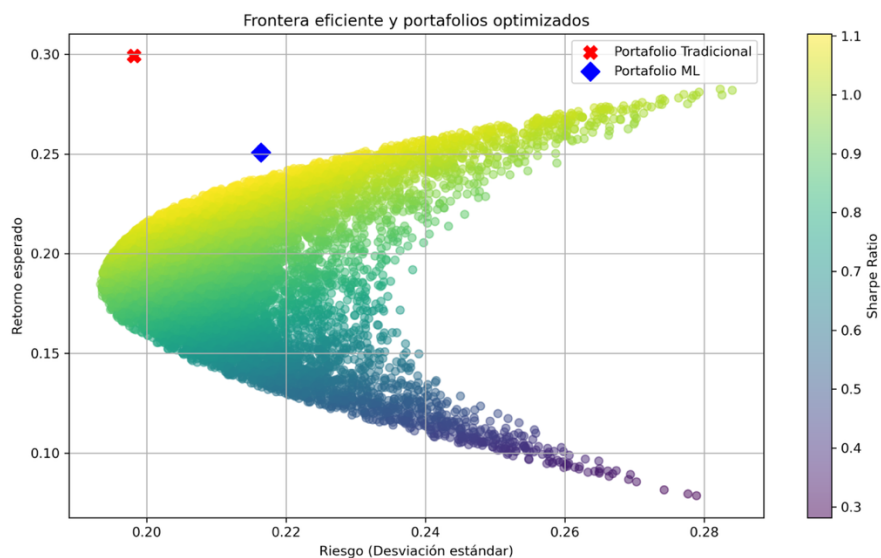
Figura 2. Evolución de una inversión de \$100 en los 3 Portafolio (2020-2024)



Fuente: Código 7. Modelo ML estatico, todas las acciones
Elaboración: Propia

Con el objetivo de contrastar de manera gráfica el rendimiento y riesgo de los portafolios generados por ambos enfoques, en la Figura 3 se presenta la frontera eficiente construida a partir de simulaciones con rendimientos estimados por el modelo Random Forest.

Figura 3. Frontera eficiente y portafolios optimizados



Fuente: Código 5. MonteCarlo portafolios tradicional y ML
Elaboración: Propia

Dentro de esta curva, se identificó el portafolio con el mayor ratio de Sharpe, señalado con un punto distintivo. Esta visualización refuerza la capacidad del modelo predictivo para explorar combinaciones eficientes, incluso en contextos complejidad de datos.

La Tabla 3.8 resume los valores clave de los portafolios generados por ambos métodos. Si bien el modelo tradicional de Markowitz alcanzó un rendimiento del 29,92 % con una desviación del 19,82 %, el portafolio basado en machine learning obtuvo un rendimiento del 24,88 % con un riesgo similar. Aunque la rentabilidad fue menor, el modelo ML mantuvo un Sharpe competitivo del 127,6 %, validando su eficiencia relativa bajo los supuestos de predicción.

Tabla 11. Comparación entre portafolios óptimos por método

Modelo	AAPL	MSFT	MOS	LIN	SBUX	Rendimiento	Riesgo	Sharpe Ratio
Tradicional	32,7 %	62,1 %	1,0 %	2,2 %	2,0 %	29,92%	19,82%	140,84%
Machine Learning	23,0 %	69,0 %	0,0 %	2,0 %	6,0 %	24,88%	19,49%	127,60%

Elaboración: Propia

3.8.4 Validación con datos reales del mercado

Para complementar la evaluación de los modelos, se realizó una validación utilizando precios históricos reales de las acciones entre 2014 y 2024. Con los pesos obtenidos para cada portafolio (tradicional, ML optimizado y ML estático), se calcularon sus retornos mensuales reales y se estimó el Sharpe Ratio de cada uno. Además, se utilizó una técnica estadística (*bootstrap*) para calcular un margen de confianza del 95%, indicando el rango dentro del cual es probable que se ubique el verdadero valor del Sharpe Ratio en el largo plazo.

Este análisis sirve para verificar si los buenos resultados proyectados por los modelos se mantienen con datos reales. También se calculó el rendimiento ajustado al riesgo del índice S&P 500 como referencia externa. La siguiente tabla resume los resultados:

Tabla 12. Comparación entre portafolios óptimos por método

Portafolio	Sharpe Real	Rango de confianza 95%
Tradicional	0.36	[0.18, 0.53]
ML Optimizado	0.34	[0.09, 0.59]
ML Estático	0.14	[-0.09, 0.39]
S&P 500 (benchmark)	0.18	[-0.02, 0.38]

Fuente: Código 8. Intervalos de confianza
Elaboración: Propia

Se observa que tanto el portafolio tradicional como el optimizado por IA superan al índice de referencia. Sin embargo, dado que sus rangos de confianza se superponen, no se puede afirmar con certeza que uno sea claramente superior al otro en todos los escenarios.

CAPÍTULO 4:

4.1 Conclusiones

Los resultados obtenidos en esta investigación demuestran que los modelos de machine learning, en particular el Random Forest Regressor, pueden ser utilizados como una alternativa eficaz para la construcción de portafolios de inversión, compitiendo directamente con enfoques tradicionales como el modelo de media-varianza de Markowitz. A partir de datos históricos de precios y volumen de acciones del S&P 500, se logró proyectar retornos esperados que, utilizados como insumo para la optimización de portafolios, ofrecieron resultados sólidos y superiores en algunos casos.

En la simulación basada en las cinco acciones seleccionadas por criterios ESG y desempeño histórico, el modelo ML ofreció un rendimiento acumulado cercano al del enfoque tradicional, con un Sharpe Ratio competitivo. Sin embargo, fue en la aplicación extendida sobre el universo completo de acciones donde el modelo ML demostró su mayor fortaleza, al generar un portafolio estático con un rendimiento acumulado de 59,71 % y un Sharpe Ratio de 2,45. Esto evidenció una relación riesgo-rentabilidad significativamente más eficiente que la del portafolio tradicional (1,40).

Además de los resultados proyectados, se validó el desempeño de los portafolios utilizando datos reales del mercado durante el mismo período. Esta evaluación confirmó que tanto el portafolio tradicional como el optimizado mediante machine learning superaron al índice S&P 500 en términos de rendimiento ajustado por riesgo. Sin embargo, al analizar los márgenes de confianza estadística, se observó que sus rangos se superponen, lo cual indica que las diferencias no son concluyentes desde un punto de vista estadístico. Esto refuerza la

necesidad de interpretar los resultados con prudencia, especialmente al considerar su aplicación en escenarios reales.

A pesar de estas ventajas, es importante destacar que los modelos de machine learning presentan desafíos inherentes de interpretabilidad y explicabilidad. A diferencia del modelo de Markowitz, donde la lógica detrás de las asignaciones de pesos es clara y derivada de estadísticas observables como la varianza y la covarianza, los modelos ML operan como cajas negras, dificultando la comprensión detallada de por qué se escogen ciertas acciones o asignaciones. Esto puede limitar su adopción por parte de gestores institucionales que necesitan justificar cada decisión ante clientes o reguladores.

Otro aspecto crítico es el posible sesgo de selección en la muestra de empresas elegidas. Las cinco acciones seleccionadas inicialmente presentaban un buen desempeño sostenido durante el periodo 2014–2024, lo cual puede haber favorecido tanto al modelo tradicional como al predictivo. En contextos reales, la selección ex ante de activos con rentabilidad futura es precisamente uno de los retos más complejos, y cualquier sesgo retrospectivo debe ser considerado con cautela al interpretar los resultados.

Finalmente, si bien se logró una mejora en la eficiencia del portafolio ML libre, este modelo fue estático y no consideró rebalanceos dinámicos ni ajustes periódicos en función de nuevas condiciones del mercado. A pesar de esta limitación, el desempeño obtenido frente al portafolio tradicional resulta relevante, ya que valida el potencial de los algoritmos de machine learning para predecir retornos y construir estrategias de inversión altamente eficientes. Esto demuestra que, incluso en un esquema sin ajustes continuos, es posible obtener resultados sólidos si se cuenta con un modelo bien entrenado y datos de calidad. No obstante, para fortalecer su aplicabilidad práctica, estos enfoques deben complementarse con

criterios adicionales de evaluación, mecanismos de validación cruzada y controles rigurosos para mitigar el riesgo de sobreajuste o interpretaciones erróneas en contextos reales.

Además, este trabajo contribuye al avance de enfoques más sostenibles e innovadores en la gestión financiera, al integrar tecnologías emergentes en la construcción de portafolios y considerar empresas seleccionadas bajo criterios de desempeño ambiental, social y de gobernanza (ESG). Esta combinación no solo refuerza el valor de la diversificación, sino que también apunta hacia una toma de decisiones más alineada con prácticas responsables y orientadas al largo plazo en los mercados financieros.

4.2 Limitaciones del estudio

A pesar de los resultados positivos obtenidos, el estudio presenta limitaciones que abren espacio para mejoras futuras. En primer lugar, el análisis se centró exclusivamente en variables técnicas (precio y volumen), dejando fuera elementos fundamentales y contextuales que influyen en la dinámica bursátil, como estados financieros, indicadores macroeconómicos o eventos externos no reflejados en los datos históricos.

Asimismo, no se incorporaron variables relacionadas con criterios ESG (ambientales, sociales y de gobernanza) en la modelización, a pesar de que fueron considerados de forma cualitativa en la etapa de selección de empresas. Su inclusión como datos cuantificables podría enriquecer el modelo, permitiendo capturar factores de sostenibilidad relevantes para el comportamiento de los activos.

En tercer lugar, el portafolio optimizado fue evaluado en una estructura estática, sin considerar rebalanceos ni costos de transacción. Esto difiere de los escenarios reales, donde los inversionistas ajustan sus posiciones con el tiempo y enfrentan fricciones de mercado.

También se debe considerar que la muestra de empresas utilizadas, aunque extensa, fue filtrada en función de disponibilidad completa de datos, lo que puede generar sesgos de supervivencia o excluir empresas relevantes por falta de información continua.

Finalmente, aunque el modelo de machine learning demostró buena capacidad predictiva, su implementación práctica enfrenta obstáculos relacionados con la explicabilidad y justificación regulatoria, factores clave para su adopción por parte de asesores financieros o instituciones.

4.3 Recomendaciones

Con base en los resultados obtenidos, se plantean las siguientes recomendaciones para futuras investigaciones y aplicaciones prácticas:

- Ampliar el conjunto de variables explicativas del modelo ML, incorporando datos fundamentales (ratios financieros, flujo de caja, etc.), indicadores macroeconómicos y variables alternativas como sentimiento del mercado o información climática. Esto podría mejorar la precisión predictiva y reducir el riesgo de decisiones basadas en ruido estadístico.
- Explorar algoritmos adicionales, como XGBoost, LightGBM o redes neuronales profundas, que podrían superar al Random Forest en capacidad predictiva o velocidad de ejecución, especialmente si se aplican sobre ventanas móviles de entrenamiento.
- Implementar rebalanceos periódicos, simulando decisiones de inversión más realistas. Esto permitiría observar el desempeño del modelo ML bajo condiciones dinámicas y responder a eventos de alta volatilidad, como los experimentados durante la pandemia o crisis geopolíticas.

- Fomentar la explicabilidad del modelo mediante herramientas como SHAP (SHapley Additive exPlanations) o LIME (Local Interpretable Model-agnostic Explanations), para que los resultados del modelo ML puedan ser comunicados de forma clara a gestores y clientes.
- Aplicaciones prácticas para gestores y asesores financieros: el modelo desarrollado puede ser utilizado como un complemento para detectar combinaciones óptimas de activos, especialmente en fases exploratorias o de construcción de escenarios.

Aunque su aplicación directa puede estar limitada por su falta de explicabilidad, su capacidad de cribar oportunidades entre cientos de acciones puede aportar valor agregado a las decisiones estratégicas de inversión.

5. Referencias Bibliográficas

Amat Rodrigo, J. (2020, octubre). *Random Forest en Python: fundamentos teóricos y aplicación práctica*. Ciencia de Datos. Recuperado el 4 de diciembre de 2024, de https://cienciadedatos.net/documentos/py08_random_forest_python

RankiaPro. (2023, abril 11). *Machine Learning: Qué es y cómo puede usarse en la gestión de inversiones*. RankiaPro. Recuperado el 26 de noviembre de 2024, de <https://rankiapro.com/latam/insights/machine-learning-que-es-usarse-gestion-inversiones/>

Bloomberg. (2024). *La IA en las terminales Bloomberg: un cambio de juego*. <https://www.bloomberg.com/latam/blog/el-financiero-la-ia-en-las-terminales-bloomberg-un-cambio-de-juego>

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91.

Fabozzi, F. J., Gupta, F., & Markowitz, H. M. (2002). The legacy of modern portfolio theory. *The Journal of Investing*, 11(3), 7-22.

Fabozzi, F. J., Gupta, F., & Markowitz, H. M. (2002). *The Theory and Practice of Investment Management*.

Naciones Unidas. (2015). *Objetivos de desarrollo sostenible: 17 objetivos para transformar nuestro mundo*. <https://www.un.org/sustainabledevelopment/es/>

CFA Institute. (2022, November 16). *How Goals-Based Portfolio Theory Came to Be*. CFA Institute. Recuperado de <https://blogs.cfainstitute.org/investor/tag/modern-portfolio-theory/>

CFA Institute. (2020). *Artificial intelligence in asset management*. CFA Institute. Recuperado de <https://www.cfainstitute.org/sites/default/files/-/media/documents/book/rf-lit-review/2020/rflr-artificial-intelligence-in-asset-management.pdf>

- SIFMA. (2024). *Equity market structure*. Securities Industry and Financial Markets Association. Recuperado de <https://www.sifma.org/explore-issues/equity-market-structure/>
- Ross, S. A., Westerfield, R. W., & Jordan, B. D. (2016). *Finanzas Corporativas (9na ed.)*. McGraw-Hill Education.
- Gu, S., Kelly, B. T., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273. <https://doi.org/10.1093/rfs/hhaa009>
- Bodie, Z., Kane, A., & Marcus, A. J. (2018). *Investments* (11th ed.). McGraw-Hill Education.
- Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning in finance. *Annual Review of Financial Economics*, 9, 145-165.
- Jensen, M. C., Meckling, W. H., & Schmidt, R. (2021). Alternative data and investment decision making. *Journal of Financial Economics*, 141(3), 785-808. <https://doi.org/10.1016/j.jfineco.2021.03.012>
- Kolanovic, M., & Krishnamachari, R. (2017). Big data and AI strategies: Machine learning and alternative data approach to investing. J.P. Morgan Global Quantitative and Derivatives Strategy Report.
- Nasdaq. (s.f.). North American Markets, Trading Solutions & Services. Recuperado de <https://www.nasdaq.com/solutions/north-american-markets>
- New York Stock Exchange. (s.f.). The NYSE Market Model. Recuperado de <https://www.nyse.com/market-model>
- Securities and Exchange Commission. (n.d.). *What We Do*. SEC.gov. Recuperado el 7 de abril de 2025, de <https://www.sec.gov/about/what-we-do>
- Securities and Exchange Commission. (n.d.). *Laws and Rules*. SEC.gov. Recuperado el 7 de abril de 2025, de <https://www.sec.gov/about/laws>

S&P Dow Jones Indices. (s.f.). *S&P 500*. Recuperado el 7 de abril de 2025, de <https://www.spglobal.com/spdji/en/indices/equity/sp-500/>

Pinelis, M., & Ruppert, D. (2022). Machine learning portfolio allocation. *The Journal of Finance and Data Science*, 8, 35–54. <https://doi.org/10.1016/j.jfds.2021.12.001>

MSCI. (2024). ESG Ratings Methodology. MSCI Inc. <https://www.msci.com/our-solutions/esg-investing/esg-ratings>

Sustainalytics. (2024). Company ESG Risk Ratings Reports. Morningstar Sustainalytics. <https://www.sustainalytics.com/esg-ratings>

6. Anexos

Código 1. Descarga base S&P 500 de YH Finance

```
import yfinance as yf

start_date = "2014-01-01"
end_date = "2024-12-31"

sp500_data = yf.download(tickers, start=start_date, end=end_date,
interval="1mo", group_by="ticker")

print(sp500_data)
file_path = "SP500_Historical_Prices.xlsx"
sp500_data.to_excel('/Users/aaron97/Desktop/Python tesis/Empresa
SP500.xlsx', engine="openpyxl")

print(f"Datos guardados en {'/Users/aaron97/Desktop/Python
tesis/Empresa SP500.xlsx'}")
```

Código 2. Modelo Rendimiento MAE R^2

```
from sklearn.metrics import mean_absolute_error, r2_score

metricas = []
for ticker in tickers_todos:
    try:
        close_col = f"{ticker}_Close"
        vol_col = f"{ticker}_Volume"

        df_t = df[['Date', close_col, vol_col]].dropna().copy()
        df_t.columns = ['Date', 'Close', 'Volume']
        df_t['Ticker'] = ticker

        df_t['Return'] = df_t['Close'].pct_change()
        df_t['Vol_Change'] = df_t['Volume'].pct_change()
        df_t['MA_3'] = df_t['Close'].rolling(window=3).mean()
        df_t['MA_6'] = df_t['Close'].rolling(window=6).mean()
        df_t['Target'] = df_t['Close'].pct_change().shift(-1)
        df_t = df_t.dropna()

        df_t['Date'] = pd.to_datetime(df_t['Date'])
        train_mask = df_t['Date'] <= '2019-12-31'

        X_train = df_t.loc[train_mask, ['Return', 'Vol_Change',
'MA_3', 'MA_6']]
        y_train = df_t.loc[train_mask, 'Target']

        if X_train.empty or y_train.empty:
            continue
```

```

        model = RandomForestRegressor(n_estimators=100,
random_state=42)
        model.fit(X_train, y_train)
        y_pred = model.predict(X_train)

        mae = mean_absolute_error(y_train, y_pred)
        r2 = r2_score(y_train, y_pred)

        metricas.append({'Ticker': ticker, 'MAE': mae, 'R2': r2})

    except:
        continue

df_metricas = pd.DataFrame(metricas)
df_metricas = df_metricas.sort_values(by='R2', ascending=False)
print(df_metricas.head(10)) # las mejores 10 acciones modeladas
df_metricas.to_excel("/Users/aaroncastro/Desktop/metricas_modelo_regresion_RF.xlsx", index=False)

```

Código 3. Rendimiento Portafolio con ML

```

import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, r2_score

ruta = "/Users/aaroncastro/Desktop/Python
tesis/Precios_CierreYVolumen_SP500_Limpio.xlsx"
archivo = pd.ExcelFile(ruta)
df = archivo.parse(archivo.sheet_names[0])

#Rendimiento portafolio con ML
acciones_tradicionales = ['AAPL', 'MSFT', 'MOS', 'LIN', 'SBUX']
predicciones_5 = []

for ticker in acciones_tradicionales:
    try:
        close_col = f"{ticker}_Close"
        vol_col = f"{ticker}_Volume"

        df_t = df[['Date', close_col, vol_col]].dropna().copy()
        df_t.columns = ['Date', 'Close', 'Volume']
        df_t['Ticker'] = ticker

        # Features
        df_t['Return'] = df_t['Close'].pct_change()
        df_t['Vol_Change'] = df_t['Volume'].pct_change()
        df_t['MA_3'] = df_t['Close'].rolling(window=3).mean()
        df_t['MA_6'] = df_t['Close'].rolling(window=6).mean()
        df_t['Target'] = df_t['Close'].pct_change().shift(-1)
        df_t = df_t.dropna()

        df_t['Date'] = pd.to_datetime(df_t['Date'])
        train_mask = df_t['Date'] <= '2019-12-31'
        test_mask = df_t['Date'] > '2019-12-31'
    
```

```

        X_train = df_t.loc[train_mask, ['Return', 'Vol_Change', 'MA_3',
'MA_6']]
        y_train = df_t.loc[train_mask, 'Target']
        X_test = df_t.loc[test_mask, ['Return', 'Vol_Change', 'MA_3',
'MA_6']]
        df_test = df_t.loc[test_mask, ['Date', 'Ticker']].copy()

        if X_train.empty or y_train.empty or X_test.empty:
            print(f"⚠️ {ticker}: datos insuficientes, se omite.")
            continue

        model = RandomForestRegressor(n_estimators=100, random_state=42)
        model.fit(X_train, y_train)

        df_test['Predicted_Return'] = model.predict(X_test)
        predicciones_5.append(df_test)

    except Exception as e:
        print(f"Error con {ticker}: {e}")

df_predicciones_5 = pd.concat(predicciones_5).reset_index(drop=True)
df_predicciones_5.to_excel("/Users/aaroncastro/Desktop/predicciones_5acciones_RF.xlsx", index=False)
print(f"✅ Archivo guardado como predicciones_5acciones_RF.xlsx en el escritorio.")

```

Código 4. Pesos Portafolio con ML

```

!pip install git+https://github.com/robertmartin8/PyPortfolioOpt.git

# Paso 1: Subir el archivo .xlsx
from google.colab import files
uploaded = files.upload()

# Paso 2: Cargar predicciones
import pandas as pd

df_pred = pd.read_excel("predicciones_5acciones_RF.xlsx")
df_pred['Date'] = pd.to_datetime(df_pred['Date'])

pivot = df_pred.pivot(index='Date', columns='Ticker',
values='Predicted_Return').dropna()

# Paso 3: Optimización con PyPortfolioOpt (ya viene instalado en Colab)
from pypfopt.efficient_frontier import EfficientFrontier

mu = pivot.mean()
S = pivot.cov()

ef = EfficientFrontier(mu, S)
weights = ef.max_sharpe()
cleaned_weights = ef.clean_weights()

# Paso 4: Mostrar resultados y exportar a Excel

```

```

pesos = pd.DataFrame.from_dict(cleaned_weights, orient='index',
columns=['Peso'])
pesos = pesos[pesos['Peso'] > 0]
print(pesos)

output_file = "pesos_optimos_ml_markowitz.xlsx"
pesos.to_excel(output_file)

from google.colab import files
files.download(output_file)

```

Código 5. MonteCarlo portafolios tradicional y ML

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

ruta = '/Users/aaroncastro/Desktop/Python
tesis/Precios_CierreYVolumen_SP500_Limpio.xlsx'

df['Date'] = pd.to_datetime(df['Date'], errors='coerce')
df = df.dropna(subset=['Date'])
df = df[(df['Date'] >= '2020-01-01') & (df['Date'] <= '2024-12-31')]

acciones = ['AAPL', 'MSFT', 'LIN', 'SBUX']
columnas = [f"{ticker}_Close" for ticker in acciones]
df = df[['Date'] + columnas].copy().set_index('Date')

retornos = df.pct_change().dropna()
mean_returns = retornos.mean() * 12
cov_matrix = retornos.cov() * 12

n_portfolios = 10000
np.random.seed(42)

resultados = {'retorno': [], 'riesgo': [], 'sharpe': [], 'pesos': []}
for _ in range(n_portfolios):
    weights = np.random.dirichlet(np.ones(len(acciones)))
    retorno = np.dot(weights, mean_returns)
    riesgo = np.sqrt(np.dot(weights.T, np.dot(cov_matrix, weights)))
    sharpe = retorno / riesgo
    resultados['retorno'].append(retorno)
    resultados['riesgo'].append(riesgo)
    resultados['sharpe'].append(sharpe)
    resultados['pesos'].append(weights)

df_frontera = pd.DataFrame(resultados)

ret_trad = 0.2992 # rendimiento del modelo tradicional
risk_trad = 0.1982
ret_ml = 0.2507 # rendimiento estimado por el modelo ML
risk_ml = 0.2164

plt.figure(figsize=(10, 6))
scatter = plt.scatter(df_frontera['riesgo'], df_frontera['retorno'],
c=df_frontera['sharpe'], cmap='viridis', alpha=0.5)

```

```

plt.colorbar(scatter, label='Sharpe Ratio')
plt.scatter(risk_trad, ret_trad, color='red', marker='X', s=100,
label='Portafolio Tradicional')
plt.scatter(risk_ml, ret_ml, color='blue', marker='D', s=100,
label='Portafolio ML')
plt.title('Frontera eficiente y portafolios optimizados')
plt.xlabel('Riesgo (Desviación estándar)')
plt.ylabel('Retorno esperado')
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.savefig("/Users/aaroncastro/Desktop/frontera_eficiente.png", dpi=300)
plt.show()

df_frontera.to_excel("simulaciones_frontera.xlsx", index=False)
df_frontera.to_excel("/Users/aaroncastro/Desktop/simulaciones_frontera.xlsx", index=False)

```

Código 6. Modelo Todas las acciones

```

import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestRegressor

ruta = "/Users/aaroncastro/Desktop/Python
tesis/Precios_CierreYVolumen_SP500_Limpio.xlsx"
archivo = pd.ExcelFile(ruta)
df = archivo.parse(archivo.sheet_names[0])

close_columns = [col for col in df.columns if col.endswith('_Close')]
volume_columns = [col for col in df.columns if col.endswith('_Volume')]
tickers_todos = list(set([c.replace('_Close', '') for c in close_columns
if f"{c.replace('_Close', '')}_Volume" in volume_columns]))
resultados_rf_regresion_todos = []

for ticker in tickers_todos:
    try:
        close_col = f"{ticker}_Close"
        vol_col = f"{ticker}_Volume"

        df_t = df[['Date', close_col, vol_col]].dropna().copy()
        df_t.columns = ['Date', 'Close', 'Volume']
        df_t['Ticker'] = ticker

        # Variables predictoras
        df_t['Return'] = df_t['Close'].pct_change()
        df_t['Vol_Change'] = df_t['Volume'].pct_change()
        df_t['MA_3'] = df_t['Close'].rolling(window=3).mean()
        df_t['MA_6'] = df_t['Close'].rolling(window=6).mean()
        df_t['Target'] = df_t['Close'].pct_change().shift(-1)
        df_t = df_t.dropna()

        df_t['Date'] = pd.to_datetime(df_t['Date'])
        train_mask = df_t['Date'] <= '2019-12-31'
        test_mask = df_t['Date'] > '2019-12-31'
    
```

```

        X_train = df_t.loc[train_mask, ['Return', 'Vol_Change', 'MA_3',
'MA_6']]
        y_train = df_t.loc[train_mask, 'Target']
        X_test = df_t.loc[test_mask, ['Return', 'Vol_Change', 'MA_3',
'MA_6']]
        df_test = df_t.loc[test_mask, ['Date', 'Ticker']].copy()

        if X_train.empty or y_train.empty or X_test.empty:
            print(f"⚠️ {ticker}: datos insuficientes, se omite.")
            continue

        model = RandomForestRegressor(n_estimators=100, random_state=42)
        model.fit(X_train, y_train)
        df_test['Predicted_Return'] = model.predict(X_test)
        resultados_rf_regresion_todos.append(df_test)

    except Exception as e:
        print(f"Error con {ticker}: {e}")

df_predicciones_todas =
pd.concat(resultados_rf_regresion_todos).reset_index(drop=True)
df_predicciones_todas.to_excel("/Users/aaroncastro/Desktop/predicciones_r
f_regresion_todas.xlsx", index=False)

print(df_predicciones_todas.head())

resumen_mensual =
df_predicciones_todas.groupby('Date')['Predicted_Return'].mean().reset_in
dex()
print(resumen_mensual.tail())

top_5_mensual = df_predicciones_todas.sort_values(by='Predicted_Return',
ascending=False).groupby('Date').head(5)
print(top_5_mensual)

```

Código 7. Modelo ML estatico, todas las acciones

```

!pip install PyPortfolioOpt

import pandas as pd
import numpy as np
from pypfopt.efficient_frontier import EfficientFrontier
from pypfopt.risk_models import sample_cov
import matplotlib.pyplot as plt

from google.colab import files
uploaded = files.upload()

from google.colab import files
uploaded = files.upload()

df_pred = pd.read_excel("predicciones_rf_regresion_todas.xlsx")
df_metricas = pd.read_excel("metricas_modelo_regresion_RF.xlsx")

```

```

df_vol = pd.ExcelFile("Copia de
Precios_CierreYVolumen_SP500_Limpio.xlsx")

df_raw = df_vol.parse(df_vol.sheet_names[0], header=None)
df_header = df_raw.iloc[0]
df_vol_data = df_raw[1:].copy()
df_vol_data.columns = df_header
df_vol_data = df_vol_data.reset_index(drop=True)
df_vol_data['Date'] = pd.to_datetime(df_vol_data['Date'])

tickers_final_ml = ['INVH', 'XOM', 'PYPL', 'VST', 'DD', 'FOXA', 'CRM',
'UNH', 'JNPR', 'CCL']

df_pred_top10 = df_pred[df_pred['Ticker'].isin(tickers_final_ml)]
pivot_pred = df_pred_top10.pivot(index='Date', columns='Ticker',
values='Predicted_Return').dropna()

pred_mean = pivot_pred.mean()

valid_close_columns = [f"{ticker}_Close" for ticker in pred_mean.index if
f"{ticker}_Close" in df_vol_data.columns]
df_real_close = df_vol_data[['Date'] + valid_close_columns].dropna()
df_real_close['Date'] = pd.to_datetime(df_real_close['Date'])
df_real_close = df_real_close.set_index('Date')
df_real_close.columns = [col.replace('_Close', '') for col in
df_real_close.columns]

fecha_inicio = '2022-01-01'
df_real_close = df_real_close[df_real_close.index >= fecha_inicio]
pivot_pred = pivot_pred[pivot_pred.index >= fecha_inicio]
ret_real = df_real_close.pct_change().dropna()

mu = pred_mean[ret_real.columns]
S = sample_cov(df_real_close)

ef = EfficientFrontier(mu, S)
weights = ef.max_sharpe()
cleaned_weights = ef.clean_weights()

ret_port =
ret_real[list(cleaned_weights.keys())].dot(np.array(list(cleaned_weights.
values()))))
valor_port = (1 + ret_port).cumprod() * 100
sharpe_static = (ret_port.mean() / ret_port.std()) * np.sqrt(252)

rf = 0.02 / 252 # tasa diaria
ret_mean = ret_port.mean()
ret_std = ret_port.std()

sharpe_ajustado_rf = ((ret_mean - rf) / ret_std) * np.sqrt(252)
print(f"Sharpe Ratio sin tasa libre de riesgo: {sharpe_static:.2f}")
print(f"Sharpe Ratio ajustado con tasa libre de riesgo (2% anual):
{sharpe_ajustado_rf:.2f}")

plt.figure(figsize=(10,6))
plt.plot(valor_port.index, valor_port, label='Portafolio ML Estático',
linewidth=2)

```

```

plt.title("Evolución Portafolio ML Estático con Predicciones Promedio")
plt.xlabel("Fecha")
plt.ylabel("Valor Base 100")
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()

print("Sharpe Ratio Portafolio ML Estático:", round(sharpe_static, 2))
print("Pesos Óptimos:", cleaned_weights)

df_pesos = pd.DataFrame.from_dict(cleaned_weights, orient='index',
columns=['Peso'])
df_pesos = df_pesos[df_pesos['Peso'] > 0]
df_pesos = df_pesos.sort_values(by='Peso', ascending=False)
df_pesos.to_excel("pesos_portafolio_ML.xlsx")
df_valor = valor_port.reset_index()
df_valor.columns = ['Fecha', 'Valor Base 100']
df_valor.to_excel("evolucion_portafolio_ML.xlsx", index=False)

from google.colab import files
files.download("pesos_portafolio_ML.xlsx")
files.download("evolucion_portafolio_ML.xlsx")

valor_port.iloc[[0, -1]]

```

Código 8. Intervalos de confianza

```

import pandas as pd

import numpy as np

def sharpe_ratio(r, rf):

    return (r - rf).mean() / np.std(r, ddof=1)

def bootstrap_sharpe(r, rf, n=1000, alpha=0.05):

    s = [sharpe_ratio(np.random.choice(r, size=len(r), replace=True),
rf) for _ in range(n)]

    return np.mean(s), np.percentile(s, alpha/2*100), np.percentile(s,
(1 - alpha/2)*100), s

```



```

rf_mensual = 0.00165 # Equivalente a 2% anual

df_sp500 = pd.read_excel('/Users/aaroncastro/Desktop/Python
tesis/Primer descarga yh/SP500_YH.xlsx', sheet_name="Sheet1")

df_sp500['Date'] = pd.to_datetime(df_sp500['Date'])
df_sp500 = df_sp500.set_index('Date').sort_index()

ret_sp500 = df_sp500['Close'].pct_change().dropna()
sr_sp500, ci_low, ci_up, _ = bootstrap_sharpe(ret_sp500.to_numpy(),
rf_mensual)

print("=== SHARPE RATIO DEL ÍNDICE S&P 500 ===")
print(f"Sharpe Ratio: {sr_sp500:.3f}")
print(f"IC 95%: [{ci_low:.3f}, {ci_up:.3f}]")

# === Funciones auxiliares ===
def sharpe_ratio(r, rf):
    return (r - rf).mean() / np.std(r, ddof=1)

def bootstrap_sharpe(r, rf, n=1000, alpha=0.05):
    s = [sharpe_ratio(np.random.choice(r, size=len(r), replace=True),
rf) for _ in range(n)]
    return np.mean(s), np.percentile(s, alpha/2*100), np.percentile(s,
(1 - alpha/2)*100), s

rf_mensual = 0.00165 # Equivalente a 2% anual

```

```

df_sp500      =      pd.read_excel('/Users/aaroncastro/Desktop/Python
tesis/Primer descarga yh/SP500_YH.xlsx', sheet_name="Sheet1")
df_sp500['Date'] = pd.to_datetime(df_sp500['Date'])
df_sp500 = df_sp500.set_index('Date').sort_index()
ret_sp500 = df_sp500['Close'].pct_change().dropna()
sr_sp500, ci_low, ci_up, _ = bootstrap_sharpe(ret_sp500.to_numpy(),
rf_mensual)
print("=== SHARPE RATIO DEL ÍNDICE S&P 500 ===")
print(f"Sharpe Ratio: {sr_sp500:.3f}")
print(f"IC 95%: [{ci_low:.3f}, {ci_up:.3f}]")

```