

Plantilla sugerida para la elaboración de trabajos escritos

Esta plantilla, elaborada por el Centro Cultural Biblioteca Luis Echavarría Villegas de la Universidad EAFIT, busca orientar a la comunidad universitaria en la realización de trabajos escritos. Este documento presenta la estructura básica de los contenidos bajo la norma Icontec NTC **5613 referencias bibliográficas, contenido, forma y estructura**. Además, presenta una breve explicación sobre cómo se debe desarrollar cada ítem.

Cabe señalar que de los ítems contemplados solo es obligatorio la portada. Es posible añadir y modificar el orden o eliminar alguno

si es necesario.

Para una correcta utilización de este documento le recomendamos crear una copia para editar este archivo, eliminar esta presentación y organizar uniformemente los espacios de la portada y otros elementos de la plantilla¹.

Centro Cultural Biblioteca Luis Echavarría Villegas.



Esta obra está bajo una [licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).

¹ Fuentes consultadas para la elaboración de esta plantilla:

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. (2014). *Metodología de la investigación*. McGraw-Hill.
Instituto Colombiano de Normas Técnicas. (2018). *Documentación: Presentación de trabajos académicos* (NTC 1486).



COMPARACIÓN DE LA EFICIENCIA DE PORTAFOLIOS: UN ANÁLISIS ENTRE
LA TEORÍA DE MARKOWITZ Y UN MODELO BASADO EN RANDOM FOREST

Comparison of portfolio efficiency: An analysis between the Markowitz Theory and
a model based on Random Forest

ESTEFANÍA VARGAS GALVIS

Tesis

Asesor:

Juan Carlos Botero Ramírez

UNIVERSIDAD EAFIT
ESCUELA DE FINANZAS, ECONOMÍA Y GOBIERNO
MAESTRÍA EN ADMINISTRACIÓN FINANCIERA - MAF
BOGOTÁ
2025

CONTENIDO

RESUMEN.....	7
ABSTRACT.....	8
INTRODUCCIÓN.....	9
PLANTEAMIENTO DEL PROBLEMA.....	11
JUSTIFICACIÓN.....	12
OBJETIVOS.....	13
OBJETIVO GENERAL.....	13
OBJETIVOS ESPECÍFICOS.....	13
MARCO CONCEPTUAL.....	14
TEORÍA DE MARKOWITZ Y LA OPTIMIZACIÓN DE PORTAFOLIOS.....	14
ML Y LA OPTIMIZACIÓN DE PORTAFOLIOS.....	15
ALGORITMOS DE ÁRBOLES DE DECISIÓN Y RF.....	15
APLICACIÓN DE ML EN LA OPTIMIZACIÓN DE PORTAFOLIOS.....	16
LIMITACIONES.....	16
DISEÑO METODOLÓGICO.....	18
SELECCIÓN DE ACTIVOS Y BASE DE DATOS.....	19
PROCEDIMIENTO DE EVALUACIÓN Y COMPARACIÓN DE MODELOS.....	20
ANÁLISIS DE LA FRONTERA EFICIENTE Y PESOS ÓPTIMOS.....	20
ANÁLISIS DE ESTABILIDAD DEL SHARPE RATIO.....	20
RESULTADOS.....	22
DESEMPEÑO PREDICTIVO DE LOS MODELOS.....	22

FRONTERA EFICIENTE Y LÍNEA DEL MERCADO DE CAPITALES (CML)	24
PESOS ÓPTIMOS DE INVERSIÓN	26
ANÁLISIS DE ESTABILIDAD DEL SHARPE RATIO	29
CONCLUSIONES	31
EFICIENCIA PREDICTIVA.....	31
DISTRIBUCIÓN DE PESOS	31
ROBUSTEZ Y SUPOSICIONES	32
INTERPRETABILIDAD Y APLICACIÓN.....	32
CONSIDERACIONES FINALES	33
REFERENCIAS	34

LISTA DE FIGURAS

Figura 1. Comparativa entre predicciones y retornos reales enero 2025.....	23
Figura 2. Frontera eficiente y Capital Market Line	25
Figura 3. Comparación de pesos óptimos: Markowitz vs Random Forest (RF)	26
Figura 4. Estabilidad del radio de Sharpe	29

LISTA DE TABLAS

Tabla 1. Pesos óptimos de los modelos28

RESUMEN

Este estudio compara la eficiencia en la construcción de portafolios de inversión, utilizando dos enfoques: la clásica Teoría de Markowitz y un modelo basado en Random Forest. De esta manera se puede decir que mientras Markowitz optimiza el portafolio bajo supuestos paramétricos y lineales, Random Forest aprovecha la capacidad de capturar patrones no lineales y dinámicos en datos financieros complejos, especialmente en mercados volátiles.

La investigación se fundamenta en datos reales de las 30 acciones más líquidas del S&P 500, entre 2020 y 2024, evaluando el desempeño de ambos modelos a través de métricas clave como el Sharpe Ratio, la varianza y la capacidad predictiva de retornos en enero de 2025. Los resultados obtenidos evidenciarán si el modelo de Random Forest podrá igualar o superar la eficiencia de Markowitz en la relación riesgo-retorno, ofreciendo una alternativa adaptativa y robusta para la gestión de portafolios en entornos financieros actuales.

Palabras clave:

Portafolio eficiente, Teoría de Markowitz, Random Forest, Machine Learning, S&P500.

ABSTRACT

This study compares the efficiency of investment portfolio construction using two approaches: the classic Markowitz Theory and a model based on Random Forest. Thus, it can be said that while Markowitz optimizes the portfolio under parametric and linear assumptions, Random Forest leverages the ability to capture nonlinear and dynamic patterns in complex financial data, especially in volatile markets.

The research is based on real data from the 30 most liquid stocks in the S&P 500, between 2020 and 2024, evaluating the performance of both models through key metrics such as the Sharpe Ratio, variance, and return predictive capacity in January 2025. The results obtained will demonstrate whether the Random Forest model can match or exceed Markowitz's efficiency in the risk-return relationship, offering an adaptive and robust alternative for portfolio management in current financial environments.

Keywords:

Efficient portfolio, Markowitz Theory, Random Forest, Machine Learning, S&P500.

INTRODUCCIÓN

En la construcción de portafolios de inversión, la Teoría de Markowitz (1952) ha sido predominante, debido a su robusta propuesta matemática para maximizar el rendimiento esperado, a través de la diversificación eficiente del riesgo. Sin embargo, los supuestos de esta teoría limitan su aplicación en mercados volátiles y complejos (Merton, 1972). Ante este escenario, surge el interés por explorar metodologías no paramétricas y adaptativas, como los algoritmos de Machine Learning (en adelante ML), los cuales permiten adecuarse mejor a patrones complejos y no lineales, presentes en los datos financieros actuales (Bennett & Lemoine, 2017).

La relevancia de este estudio radica en la creciente evidencia de que los métodos tradicionales de optimización de portafolios pueden ser superados por técnicas de ML, especialmente en entornos de alta volatilidad e incertidumbre (Gu *et al.*, 2020). A diferencia de los métodos clásicos, los algoritmos de Random Forest (en adelante RF) o asumen relaciones lineales y pueden manejar mejor la alta complejidad de los datos, lo que podría traducirse en mejores decisiones de inversión (Varian, 2014).

El alcance del estudio incluye la construcción de dos modelos: primero, un portafolio basado en la Teoría de Markowitz, que sigue un enfoque paramétrico y optimiza la varianza como medida de riesgo y, segundo, un portafolio construido mediante un modelo de RF, el cual aprovecha la capacidad del ML para identificar patrones no lineales en los datos históricos.

A su vez, el enfoque metodológico es cuantitativo y comparativo, empleando métricas de desempeño como el Sharpe Ratio, el Alpha de Jensen y la varianza del portafolio, para evaluar la eficiencia de los modelos propuestos. El documento está estructurado de la siguiente manera: en primera instancia, se revisa la literatura relevante sobre la teoría de portafolios y el uso de técnicas de ML en finanzas; luego se describen los datos y la metodología empleada; posteriormente, se presentan los resultados del análisis comparativo; y, finalmente, se ofrecen conclusiones y recomendaciones basadas en los hallazgos del estudio.

Este trabajo busca aportar nuevas evidencias empíricas sobre el uso de técnicas de ML, para la optimización de portafolios de inversión, ofreciendo una alternativa a los métodos tradicionales y evaluando su aplicabilidad en contextos de alta volatilidad. Se espera que los resultados de esta investigación contribuyan al debate académico y práctico sobre el uso efectivo de RF, para la construcción de portafolios diversificados y eficientes.

PLANTEAMIENTO DEL PROBLEMA

La Teoría Moderna de Portafolios, propuesta por Harry Markowitz en 1952, ha sido, durante décadas, el pilar de la gestión de inversiones, al ofrecer un marco matemático para maximizar el rendimiento esperado de un portafolio, dado un nivel de riesgo (Markowitz, 1952). Sin embargo, en la práctica, esta teoría enfrenta críticas sustanciales debido a supuestos restrictivos como la normalidad en los retornos, la estabilidad de las varianzas y covarianzas en el tiempo, y la perfecta racionalidad de los inversores (Best & Grauer, 1991; Michaud, 1989). Estos supuestos rara vez se cumplen en los mercados financieros reales, lo que puede comprometer la efectividad de las carteras optimizadas bajo este enfoque.

En los últimos años, el auge del aprendizaje automático (ML) ha generado nuevas expectativas en la construcción de portafolios, al permitir modelos más flexibles, capaces de captar relaciones no lineales y dinámicas complejas entre variables financieras (Breiman, 2001; Gu *et al.*, 2020). En particular, el algoritmo RF ha demostrado ser una herramienta eficaz en la predicción de retornos y clasificación de activos, gracias a su capacidad para manejar grandes volúmenes de datos, su resistencia al sobreajuste y su interpretación intuitiva (Henrique *et al.*, 2019; Krauss *et al.*, 2017).

A pesar de estos avances, aún existe una brecha en los estudios, respecto a comparaciones directas entre modelos tradicionales como el de Markowitz y enfoques basados en algoritmos de ML, en términos de eficiencia y desempeño predictivo y en la conformación de portafolios reales (Huerta & Salas, 2022). Por ello surge la necesidad de analizar si un modelo basado en RF puede generar portafolios igual de eficientes o incluso más, que aquellos construidos bajo la Teoría de Markowitz, en especial en un contexto como el del mercado accionario estadounidense.

JUSTIFICACIÓN

La presente investigación tiene como propósito comparar la eficiencia de portafolios contruidos con base en la Teoría de Markowitz y aquellos generados mediante un modelo de RF, evaluando su desempeño sobre datos reales del S&P500 en el periodo de estudio (Niaki & Hoseinzade, 2013). Esta comparación no solo responde a una necesidad académica identificada en la literatura (Huerta & Salas, 2022), sino también a una inquietud práctica de gestores de inversión y analistas financieros que buscan modelos más robustos, adaptables y eficientes, ante condiciones de mercado cambiantes.

El enfoque clásico de optimización, aunque matemáticamente sólido, es altamente sensible a errores de estimación en los insumos (Best & Grauer, 1991; Michaud, 1989). En contraste, los algoritmos de aprendizaje automático, como RF, ofrecen una alternativa basada en el descubrimiento de patrones históricos, sin necesidad de supuestos rígidos sobre la distribución de los retornos (Breiman, 2001; Gu *et al.*, 2020). Además, su capacidad de generalización permite evaluar el desempeño ex post de los portafolios y estimar su eficiencia predictiva (Henrique *et al.*, 2019).

Este estudio aporta valor en tres frentes: primero, evalúa empíricamente el rendimiento y riesgo de ambos modelos; segundo, ofrece visualizaciones comparativas que permiten al lector identificar las diferencias en eficiencia y, tercero, genera evidencia útil para tomadores de decisiones en gestión de portafolios. De esta manera, se espera contribuir al debate académico y práctico sobre el uso de técnicas de ML en finanzas, particularmente en el diseño de estrategias de inversión más precisas y adaptativas.

OBJETIVOS

OBJETIVO GENERAL

Comparar la eficiencia de portafolios de inversión, contruidos mediante la Teoría de Markowitz con aquellos basados RF, evaluando cuál de los dos enfoques optimiza mejor la relación riesgo-retorno, la diversificación y la maximización de rendimientos en una muestra representativa de acciones en el mercado accionario estadounidense.

OBJETIVOS ESPECÍFICOS

- Desarrollar el modelo basado en la Teoría de Markowitz, analizando el comportamiento de los portafolios contruidos en términos de riesgo-retorno y diversificación, utilizando datos de una muestra representativa del S&P 500 para los últimos 5 años.
- Desarrollar el modelo de RF para la construcción de portafolios de inversión, aprovechando la capacidad de la técnica de ML para identificar patrones no lineales en los datos financieros, con la misma base de datos histórica usada como base en la Teoría de Markowitz.
- Evaluar los resultados obtenidos de cada modelo y determinar cuál metodología (Teoría de Markowitz o RF) es más eficiente para gestionar portafolios de inversión, teniendo en cuenta los datos del 2020 al 2024.

MARCO CONCEPTUAL

El presente apartado busca fundamentar la explicación de los conceptos clave, relacionados con la optimización de portafolios de inversión, mediante técnicas tradicionales y modernas de ML. A partir de la revisión de literatura, se abordarán las bases teóricas de la Teoría de Markowitz, los algoritmos de RF y su relación en la aplicación en el análisis financiero.

TEORÍA DE MARKOWITZ Y LA OPTIMIZACIÓN DE PORTAFOLIOS

La Teoría de Markowitz sentó las bases del análisis moderno de portafolios, al introducir el concepto de diversificación como estrategia para minimizar el riesgo, sin sacrificar el rendimiento esperado. El autor planteó que un portafolio eficiente es aquel que, para un nivel dado de riesgo, maximiza el retorno esperado (Markowitz, 1952). Su enfoque utiliza medidas estadísticas como la media (retorno esperado) y la varianza (riesgo), para evaluar y construir portafolios, asumiendo que los inversionistas prefieren un menor riesgo para un mismo nivel de retorno. A su vez, la frontera eficiente se entiende como el conjunto de portafolios óptimos que proporcionan la mejor relación riesgo-retorno.

Sin embargo, los supuestos de esta teoría presentan limitaciones significativas:

- **Distribución normal de los retornos:** la teoría asume que los retornos siguen una distribución normal, lo cual no siempre se cumple en los mercados financieros (Merton, 1972).
- **Estabilidad de las covarianzas:** se presupone que la matriz de covarianzas entre activos es estable, lo que puede ser irrealista en entornos de alta volatilidad o crisis económicas (Chopra & Ziemba, 1993).

Algunos estudios recientes han cuestionado la aplicabilidad de la Teoría de Markowitz en escenarios complejos y dinámicos, sugiriendo que los métodos tradicionales de optimización de portafolios pueden ser superados por enfoques adaptativos y no paramétricos (Gu *et al.*, 2020).

ML Y LA OPTIMIZACIÓN DE PORTAFOLIOS

En las últimas dos décadas, los avances en técnicas de ML han transformado el análisis financiero, permitiendo una mejor modelación de datos complejos y la identificación de patrones no lineales, que los métodos tradicionales no captan (Varian, 2014). El ML utiliza algoritmos capaces de aprender de los datos históricos para hacer predicciones y recomendaciones, ofreciendo un enfoque flexible y dinámico para la optimización de portafolios.

Entre las técnicas más destacadas de ML para la construcción de portafolios se encuentran los algoritmos de ensemble, que combinan múltiples modelos para mejorar la precisión y robustez de las predicciones. Dentro de esta categoría, el algoritmo RF ha ganado popularidad, debido a su capacidad para manejar datos de alta dimensionalidad y capturar relaciones complejas entre variables (Breiman, 2001).

ALGORITMOS DE ÁRBOLES DE DECISIÓN Y RF

Los árboles de decisión son algoritmos que dividen los datos en subconjuntos más pequeños, basándose en reglas de decisión derivadas de las características de los datos. Cada división representa una decisión basada en un umbral específico, lo que permite modelar relaciones no lineales y complejas entre las variables (Bennett & Lemoine, 2017). Aunque los árboles de decisión son intuitivos y fáciles de interpretar, su principal desventaja es que pueden ser propensos al sobreajuste, especialmente cuando se entrenan con datos ruidosos o de gran tamaño.

De otro lado, el Random Forest (RF), introducido por Breiman (2001), es un algoritmo de ensemble que construye múltiples árboles de decisión y combina sus predicciones para obtener un resultado más preciso y estable. Utiliza técnicas de bagging (Bootstrap Aggregating) para crear subconjuntos aleatorios del conjunto de datos y entrenar cada árbol de decisión, de manera independiente. La predicción final se obtiene mediante un promedio (regresión) o a través de un voto mayoritario (clasificación).

Los beneficios del RF incluyen:

- Reducción del sobreajuste: al combinar múltiples árboles, se reduce el riesgo de que el modelo se ajuste excesivamente a los datos de entrenamiento.
- Manejo de alta dimensionalidad: el RF puede manejar conjuntos de datos con muchas características y es efectivo para identificar interacciones complejas entre variables.
- Robustez y precisión: el uso de múltiples modelos permite obtener predicciones más precisas y robustas frente a datos con alta volatilidad.

APLICACIÓN DE ML EN LA OPTIMIZACIÓN DE PORTAFOLIOS

Los estudios que han explorado el uso de técnicas de ML, para la optimización de portafolios, han mostrado algunos resultados prometedores. Por ejemplo, Gu et al. (2020), encontraron que los modelos basados en ML superan a los métodos tradicionales en términos de predicción de retornos y reducción de riesgo, especialmente en mercados caracterizados por alta volatilidad. Por su parte Varian (2014), destacó que los enfoques basados en Big Data y ML pueden capturar mejor los cambios rápidos en los mercados financieros y adaptarse más rápidamente a nuevos patrones.

La aplicación de RF para la construcción de portafolios se justifica por su capacidad para manejar relaciones no lineales y complejas en los datos financieros, lo que permite capturar patrones que no son evidentes, a través de métodos lineales como la Teoría de Markowitz (Bennett & Lemoine, 2017; Breiman, 2001). Además, el RF no requiere supuestos sobre la distribución de los retornos, lo que lo hace más adecuado para mercados volátiles y datos financieros reales.

LIMITACIONES

Aunque los algoritmos de ML, como el RF, ofrecen ventajas significativas, también presentan desafíos. El principal reto es la interpretabilidad de los modelos, pues al combinar múltiples árboles, el modelo final puede volverse difícil de entender para los analistas financieros. Además, el rendimiento del modelo depende de la calidad

de los datos históricos, lo que implica que eventos inesperados o cambios estructurales en el mercado, pueden afectar su precisión (Chopra & Ziemba, 1993).

Otra crítica es que, aunque los modelos de ML pueden identificar patrones complejos en los datos, no siempre proporcionan una comprensión clara de los factores económicos subyacentes que impulsan estos patrones. Esto puede limitar su aplicabilidad para decisiones de inversión basadas en fundamentos económicos (Varian, 2014).

DISEÑO METODOLÓGICO

Para abordar el problema esta investigación se ha planteado un estudio comparativo, cuyo objetivo es analizar la eficiencia relativa de dos enfoques distintos para la construcción de portafolios de inversión: la Teoría de Markowitz y un modelo de RF. La comparación busca determinar cuál de los dos métodos presenta mayor precisión, al predecir retornos diarios y construir portafolios más eficientes en términos de rendimiento ajustado por riesgo.

Al respecto se puede afirmar lo siguiente:

- El modelo clásico de Markowitz, basado en la teoría media-varianza (Markowitz, 1952), busca construir portafolios eficientes, maximizando el rendimiento esperado y minimizando la varianza del portafolio. En él se asume una distribución normal de los retornos y se emplean estimaciones de rendimiento medio y matriz de covarianzas entre activos, para resolver el problema de optimización. Además, el portafolio óptimo es aquel que maximiza el Ratio de Sharpe, utilizando una tasa libre de riesgo (para este caso constante del 2 %).
- El algoritmo de RF, propuesto por Breiman (2001), es un modelo de aprendizaje supervisado, basado en un conjunto de árboles de decisión (ensemble). Este modelo fue elegido por su capacidad para capturar relaciones no lineales y patrones complejos, en series temporales financieras. En este estudio, el RF se entrena sobre una ventana histórica de retornos, para predecir los rendimientos esperados del siguiente día. A partir de estas predicciones, se asignan pesos óptimos al portafolio en función proporcional al retorno esperado.

Para hacer el análisis comparativo entre ambos modelos, se realizaron los siguientes pasos:

SELECCIÓN DE ACTIVOS Y BASE DE DATOS

La muestra de activos que se usa en el estudio corresponde a las 30 acciones más líquidas del índice S&P 500, seleccionadas en función de su puntaje de liquidez (Liquidity Assessment Score, LQA), provisto por Bloomberg durante los últimos 12 meses. Esta métrica considera múltiples factores para medir el costo de transacción de un activo:

- Volumen: frecuencia y cantidad de acciones negociadas
- Costo de ejecución: incluye el diferencial bid-ask y otros costos implícitos.
- Tiempo de ejecución: rapidez para ejecutar una operación, sin afectar significativamente el precio.

Las acciones seleccionadas corresponden al percentil 100 del LQA, es decir, aquellas con los menores costos de liquidez.

Teniendo en cuenta la información histórica seleccionada, esta se almacenó en los siguientes dos archivos:

- BASE_FINAL12M.xlsx: contiene los precios diarios del periodo comprendido entre el 1 de enero de 2020 y el 31 de diciembre de 2024.
- BASE_FINAL12M_25.xlsx: incluye los precios diarios del mes de enero de 2025, utilizados exclusivamente para validar las predicciones generadas por ambos modelos.

Los precios se transformaron a retornos diarios, mediante el cálculo de cambio porcentual entre días consecutivos. Para el tratamiento de datos faltantes, se aplicó la técnica de imputación por llenado hacia adelante (*forward fill*), seguida de la eliminación de cualquier fila con valores ausentes restantes.

PROCEDIMIENTO DE EVALUACIÓN Y COMPARACIÓN DE MODELOS

PREDICCIÓN DE RETORNOS EN ENERO 2025:

Ambos modelos fueron entrenados sobre la información histórica disponible, hasta el día anterior a cada fecha de predicción, durante enero de 2025. Para cada día hábil del mes, se generaron predicciones de retorno del portafolio utilizando:

- Pesos optimizados mediante la Teoría de Markowitz.
- Pesos derivados del modelo RF.

La eficiencia predictiva de cada modelo se evaluó mediante el Error Absoluto Promedio (MAE, por sus siglas en inglés) entre el retorno proyectado del portafolio de cada modelo y el retorno real promedio observado para ese día.

ANÁLISIS DE LA FRONTERA EFICIENTE Y PESOS ÓPTIMOS

Se simularon 5,000 portafolios aleatorios, usando combinaciones de pesos aplicados a los activos seleccionados. Para cada portafolio se calculó su retorno esperado, riesgo (volatilidad) y ratio de Sharpe. De igual forma, la frontera eficiente fue trazada a partir de estos puntos y se identificó el portafolio con máxima razón de Sharpe. A partir de este, se construyó la Capital Market Line (CML). Luego se compararon visualmente los portafolios óptimos, derivados de ambos modelos (Markowitz y RF) con respecto a dicha frontera.

ANÁLISIS DE ESTABILIDAD DEL SHARPE RATIO

Para complementar la evaluación de la eficiencia de los portafolios, se llevó a cabo un análisis de la estabilidad del Sharpe Ratio, para cada uno de los métodos. Además de encontrar los portafolios óptimos con la Teoría de Markowitz y el modelo de RF, se analizó qué tan estable era el ratio de Sharpe de cada portafolio, a lo

largo del tiempo. Esto significa que, una vez se identifica un portafolio óptimo en una fecha determinada, con un radio de Sharpe específico, se evaluará cómo se comporta ese mismo portafolio en periodos futuros, comparando el radio de Sharpe, proyectado con el que se obtuvo inicialmente. La idea es saber si el portafolio mantiene un buen equilibrio entre riesgo y rendimiento o si su eficiencia cambia mucho con el tiempo. Este análisis ayudará a determinar cuál método genera portafolios que no solo son óptimos en el momento, sino que también son más estables y confiables en el futuro.

RESULTADOS

DESEMPEÑO PREDICTIVO DE LOS MODELOS

Con el objetivo de evaluar la eficiencia predictiva de los modelos de optimización de portafolios (Markowitz y RF), se realizaron predicciones diarias para el mes de enero de 2025. De esta manera, para cada día del mes se utilizó una ventana móvil de retornos históricos (incluyendo hasta el día anterior a cada fecha de predicción), con el fin de estimar los pesos óptimos de inversión, según cada modelo. Posteriormente, estos pesos fueron aplicados a los retornos reales del día correspondiente, para calcular el rendimiento proyectado de cada portafolio.

A su vez, la eficiencia de las predicciones se evaluó a través del error absoluto promedio (MAE), entre los retornos proyectados por cada modelo y el retorno real promedio del mercado ese día.

Ahora, para evaluar la capacidad predictiva de los modelos implementados, se calcularon los errores absolutos promedio entre los retornos proyectados y los rendimientos reales observados en cada fecha del mes de enero de 2025. Los resultados fueron los siguientes:

- Modelo de Markowitz: error absoluto promedio de 0.4252 %
- Modelo RF: error absoluto promedio de 0.5941 %

Estos valores sugieren que el modelo de optimización tradicional, basado en la teoría media-varianza, logró una mayor precisión en la estimación de retornos diarios, superando al modelo basado en RF durante el periodo de evaluación.

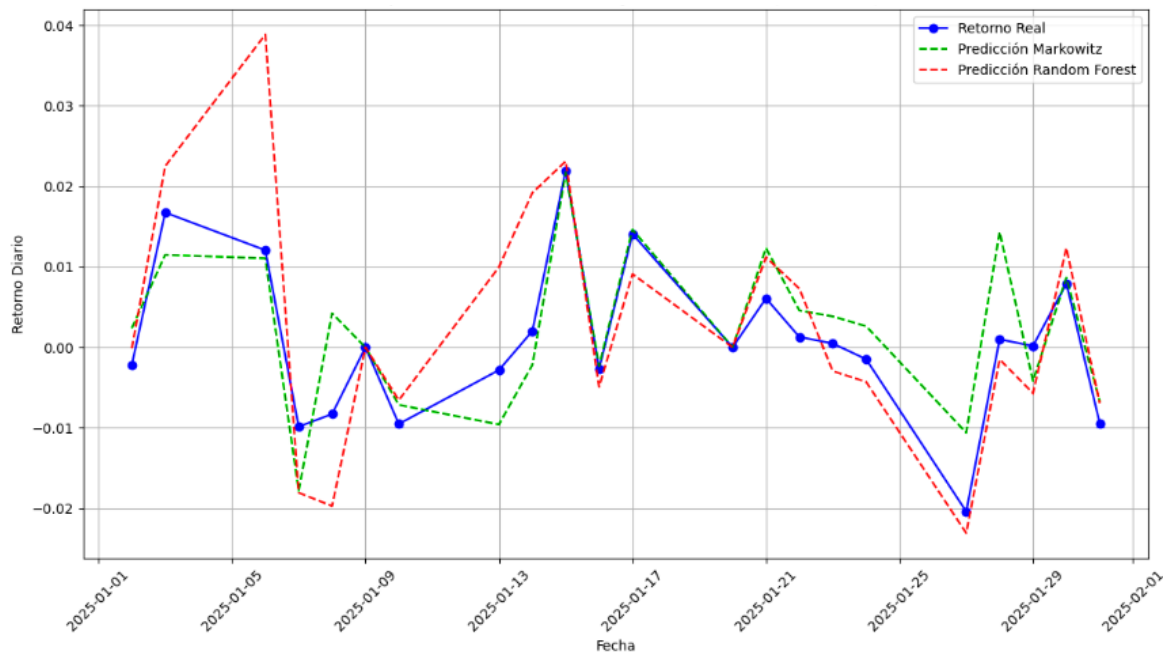
Si bien se esperaba que el modelo RF captara patrones no lineales de forma más eficiente, sus predicciones resultaron menos precisas en este caso. Esto podría

atribuirse a la naturaleza del periodo evaluado (enero de 2025), a la limitada cantidad de datos recientes para predicción diaria, o a posibles condiciones de mercado, donde las relaciones lineales predominan.

Por lo tanto, bajo el criterio de menor error absoluto, en la predicción de retornos, se concluye que el modelo de Markowitz mostró un mejor desempeño predictivo en este análisis puntual.

La figura 1 muestra una comparación visual entre los retornos reales del mercado y las predicciones realizadas por cada modelo:

Figura 1. Comparativa entre predicciones y retornos reales enero 2025



Fuente: Elaboración propia código Python.

Descripción:

Esta gráfica muestra la evolución diaria de los retornos reales del mercado en enero de 2025, comparada con las predicciones realizadas por los modelos de Markowitz y RF.

Las líneas representan:

- Línea azul: retorno real diario promedio de las acciones.
- Línea verde discontinua: predicciones del modelo de Markowitz.
- Línea roja discontinua: predicciones del modelo RF.

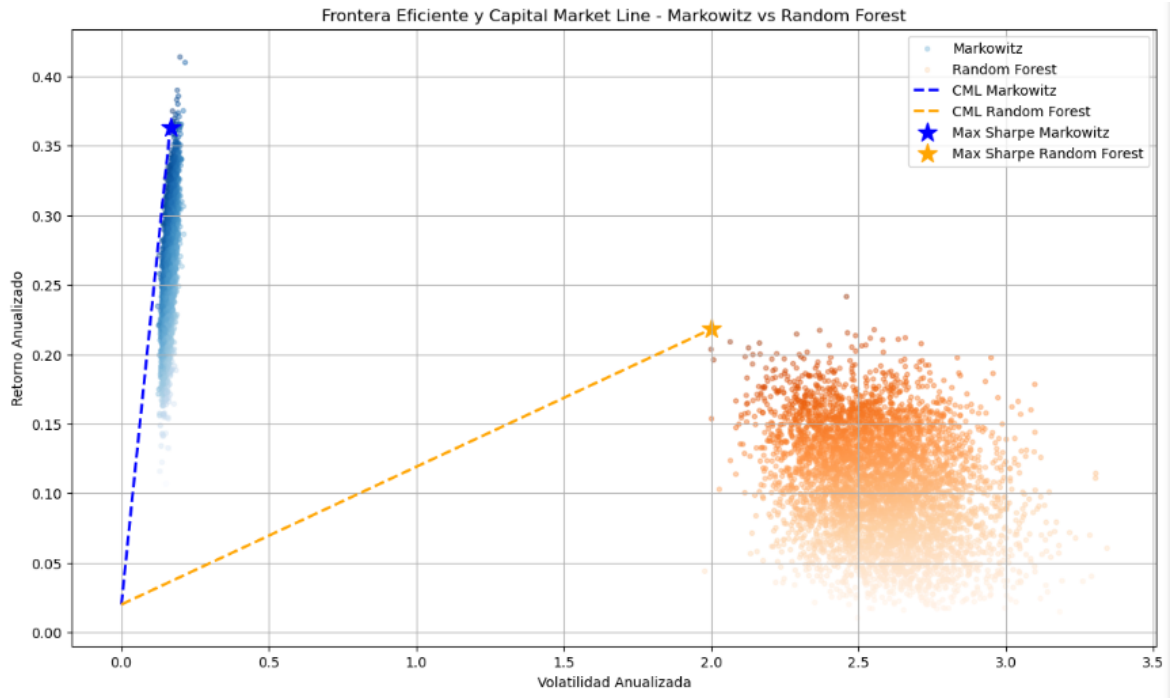
FRONTERA EFICIENTE Y LÍNEA DEL MERCADO DE CAPITALES (CML)

Se simularon 5,000 portafolios aleatorios, utilizando los retornos históricos desde enero de 2020 hasta diciembre de 2024. Para cada portafolio se calcularon el retorno esperado, la volatilidad y el radio de Sharpe. A partir de estos datos, se construyó la frontera eficiente, identificando el portafolio con el mayor radio de Sharpe (portafolio óptimo). Además, se graficó la Capital Market Line (CML), partiendo de una tasa libre de riesgo del 2 %.

Se considera la tasa del 2 %, ya que la base de datos abarca retornos desde enero de 2020 hasta diciembre de 2024. Durante este período, especialmente tras la pandemia y en los años posteriores, la tasa de los bonos del Tesoro estadounidense a un año fluctuó entre valores cercanos a cero (2020-2021) y valores por encima del 4-5 % (2022-2023). Por tanto, una tasa promedio del 2 % puede considerarse una simplificación razonable, para cubrir todo el período.

Como se observa en la figura 2, el portafolio con el mayor Sharpe ratio se encuentra en el punto de tangencia entre la CML y la frontera eficiente del modelo de Markowitz, que lo convierte en el portafolio teóricamente óptimo.

Figura 2. Frontera eficiente y Capital Market Line



Fuente: Elaboración propia código Python.

Descripción:

La gráfica muestra la Frontera Eficiente generada para dos modelos de construcción de portafolios: el clásico basado en la teoría de Markowitz y un modelo alternativo basado en RF. A su vez, cada punto en la nube representa un portafolio simulado con distintas combinaciones de pesos, mostrando su retorno anualizado, esperado en función de su volatilidad anualizada.

Además, se incluyen las respectivas Capital Market Lines (CML) para cada modelo, las cuales representan la relación óptima entre riesgo y retorno, cuando se combina el portafolio de máxima eficiencia (máximo Sharpe Ratio) con un activo libre de riesgo. A su vez, los portafolios con la mejor relación riesgo-retorno están indicados con estrellas, destacando la diferencia en la eficiencia y riesgo asumido entre ambos enfoques.

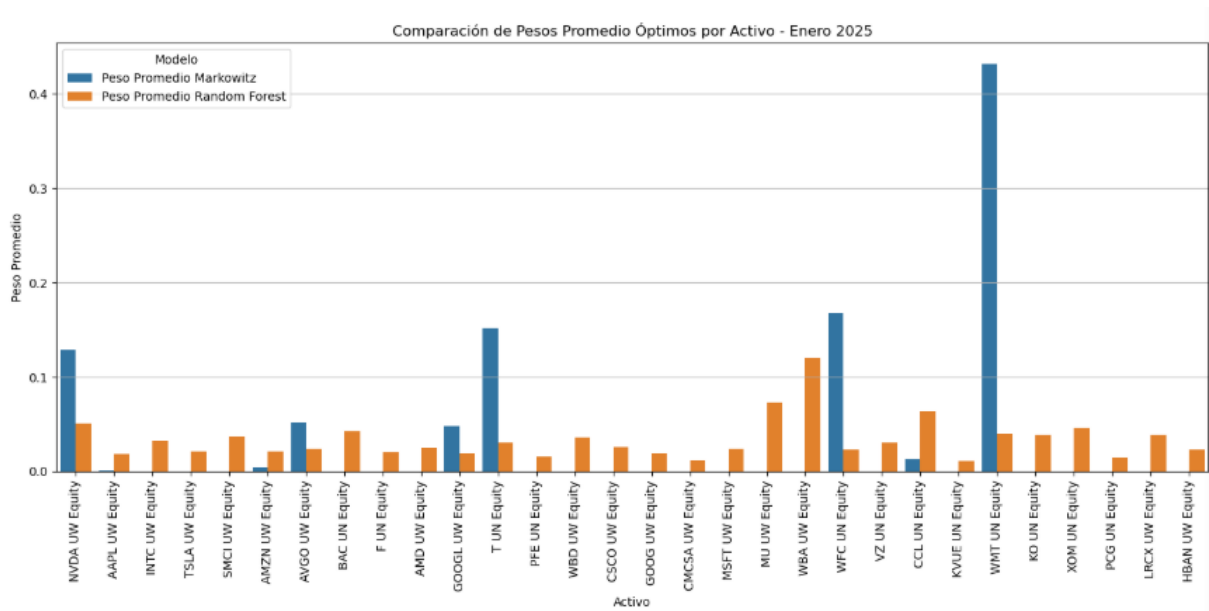
Esta comparación visual permite apreciar cómo cada modelo optimiza el trade-off entre riesgo y retorno, facilitando una evaluación directa de su desempeño relativo en términos de eficiencia financiera.

PESOS ÓPTIMOS DE INVERSIÓN

Se calcularon los pesos óptimos, asignados por cada modelo (Markowitz y RF) a los activos del portafolio, al cierre de diciembre de 2024. Tanto la figura 3 como la tabla 1 muestran la distribución de pesos sugerida por cada enfoque.

Se observa que el modelo de Markowitz tiende a diversificar más homogéneamente entre los activos, mientras que el modelo RF concentra más peso en aquellos activos que, según sus predicciones, ofrecen mayor rendimiento esperado.

Figura 3. Comparación de pesos óptimos: Markowitz vs Random Forest (RF)



Fuente: Elaboración propia código Python.

Descripción:

Gráfica de barras que visualiza, para cada acción, los pesos asignados por los modelos de Markowitz y RF.

- Barras azules: pesos asignados por Markowitz.
- Barras naranjas: pesos asignados por RF.

Se observa que el modelo de Markowitz tiende a distribuir los pesos entre un mayor número de activos, asignando participaciones relevantes a WMT (41.55 %), T (16.66 %), WFC (15.61 %) y NVDA (13.26 %). Aunque también incluye, en menor proporción, a otras acciones como AAPL y AVGO, su enfoque busca un equilibrio entre riesgo y retorno, optimizando toda la frontera eficiente.

Por otro lado, el modelo RF asigna pesos significativos únicamente a un grupo reducido de acciones, destacándose XOM (60.93 %), VZ (25.74 %), PCG (5.17 %), WBD (2.32 %), PFE (2.19 %) y T (3.65 %). Esta concentración sugiere que el modelo identificó patrones de retorno futuros, particularmente favorables en estos activos, posiblemente como resultado de su naturaleza no lineal y predictiva.

Esta diferencia en la asignación de pesos también refleja la lógica subyacente de ambos modelos:

- Markowitz busca la diversificación óptima, bajo la suposición de normalidad en los retornos y correlación lineal entre activos.
- RF actúa como modelo predictivo, priorizando activos que, con base en sus patrones pasados, pueden tener mayores retornos esperados, lo cual puede llevar a concentraciones más extremas.

Tabla 1. Pesos óptimos de los modelos

	Acciones	Pesos Markowitz	Pesos Random Forest
0	NVDA UW Equity	1.326377e-01	0.000000
1	AAPL UW Equity	1.601820e-02	0.000000
2	INTC UW Equity	0.000000e+00	0.000000
3	TSLA UW Equity	3.200917e-15	0.000000
4	SMCI UW Equity	1.489399e-15	0.000000
5	AMZN UW Equity	2.613684e-04	0.000000
6	AVGO UW Equity	5.654052e-02	0.000000
7	BAC UN Equity	0.000000e+00	0.000000
8	F UN Equity	0.000000e+00	0.000000
9	AMD UW Equity	8.072814e-17	0.000000
10	GOOGL UW Equity	4.394807e-02	0.000000
11	T UN Equity	1.666063e-01	0.036485
12	PFE UN Equity	0.000000e+00	0.021898
13	WBD UW Equity	6.471213e-16	0.023265
14	CSCO UW Equity	2.161391e-16	0.000000
15	GOOG UW Equity	1.087871e-15	0.000000
16	CMCSA UW Equity	0.000000e+00	0.000000
17	MSFT UW Equity	8.531858e-16	0.000000
18	MU UW Equity	0.000000e+00	0.000000
19	WBA UW Equity	8.934121e-16	0.000000
20	WFC UN Equity	1.560584e-01	0.000000
21	VZ UN Equity	7.971935e-16	0.257428
22	CCL UN Equity	1.239244e-02	0.000000
23	KVUE UN Equity	9.242473e-16	0.000000
24	WMT UN Equity	4.155370e-01	0.000000
25	KO UN Equity	9.934985e-16	0.000000
26	XOM UN Equity	3.216455e-16	0.609259
27	PCG UN Equity	1.581353e-15	0.051665
28	LRCX UW Equity	0.000000e+00	0.000000
29	HBAN UW Equity	4.828732e-16	0.000000

Fuente: Elaboración propia código Python.

Descripción:

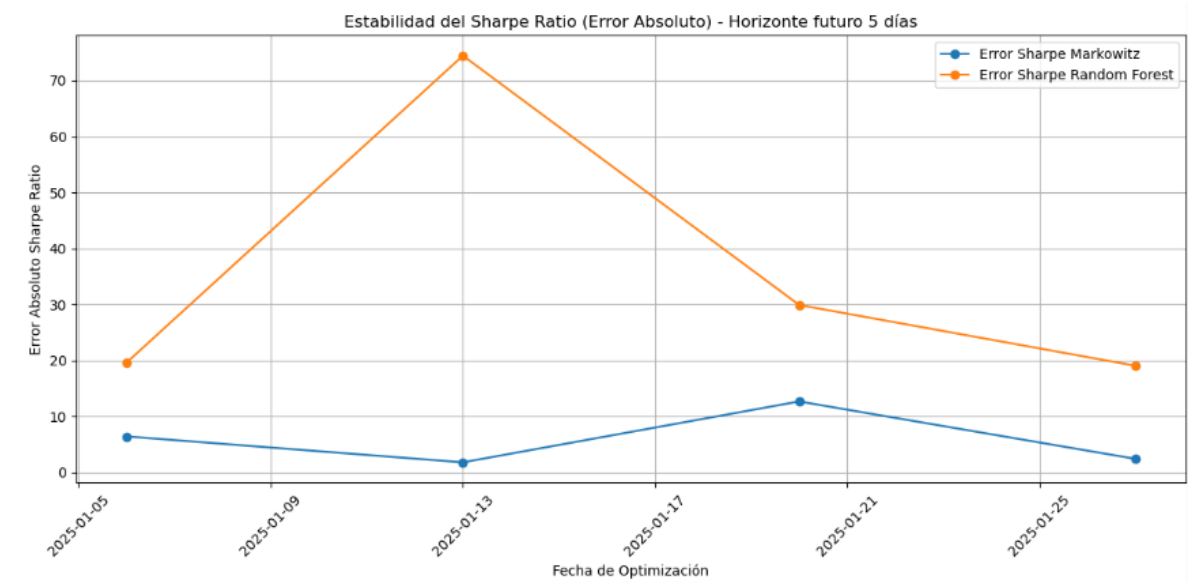
Tabla que compara los pesos asignados a cada acción por los dos modelos optimizadores:

- Columna “Pesos Markowitz”: distribución óptima basada en maximización del Sharpe Ratio.
- Columna “Pesos Random Forest”: distribución derivada de predicciones de retornos realizadas por el modelo de RF.

ANÁLISIS DE ESTABILIDAD DEL SHARPE RATIO

Para evaluar la estabilidad de los modelos de optimización de portafolio, se analizó el error absoluto entre el Sharpe Ratio estimado, en el momento de la optimización y el Sharpe Ratio realizado durante un horizonte futuro de 5 días hábiles. Este análisis se aplicó tanto al modelo clásico de Markowitz, como al modelo basado en RF, durante el primer trimestre de 2025.

Figura 4. Estabilidad del radio de Sharpe



Fuente: Elaboración propia código Python.

Como los resultados que se muestran en la figura 4, la manera consistente del modelo de RF presentó errores absolutos más altos en todas las fechas evaluadas. Es decir, la diferencia entre el Sharpe Ratio predicho y el observado 5 días después, fue sistemáticamente mayor para el modelo de RF en comparación con el de Markowitz.

Este patrón indica que, aunque RF puede capturar ciertas dinámicas no lineales en los retornos, su estabilidad predictiva a corto plazo es menor. En contraste, los portafolios optimizados, bajo el modelo de Markowitz, resultaron más robustos frente a los cambios reales del mercado, en ese mismo horizonte.

Una posible explicación es que el modelo de Markowitz, al basarse en estadísticas agregadas como medias y covarianzas, produce asignaciones menos volátiles. Por el contrario, el enfoque de RF, más flexible pero también más complejo, puede ser más sensible al ruido en los datos históricos, lo que se traduce en mayor variabilidad entre lo esperado y lo observado.

Estabilidad promedio del Sharpe Ratio (error absoluto medio):

Markowitz: 5.8493 RF: 35.7649

CONCLUSIONES

Esta investigación comparó la eficiencia de dos enfoques para la construcción de portafolios de inversión: el modelo tradicional de optimización media-varianza propuesto por Markowitz (1952) y un modelo predictivo basado en ML, específicamente el algoritmo RF. La comparación se centró en la capacidad de cada modelo para generar portafolios eficientes, evaluados según su capacidad predictiva durante el mes de enero de 2025.

EFICIENCIA PREDICTIVA

El modelo de Markowitz presentó un error absoluto promedio de 0.4252 %, mientras que el modelo RF mostró un error mayor, de 0.5941 %, en la predicción de los retornos diarios. Esta diferencia sugiere que, bajo las condiciones y datos utilizados, el modelo de Markowitz fue más eficiente al estimar el comportamiento futuro del mercado, lo cual concuerda con su uso extendido en la literatura financiera, como punto de referencia para evaluaciones de eficiencia (Elton *et al.*, 2014).

DISTRIBUCIÓN DE PESOS

Los portafolios óptimos generados por cada modelo exhiben diferencias significativas en la asignación de activos:

- El modelo de Markowitz distribuyó los pesos de forma más diversificada, asignando participación a un mayor número de acciones. Entre las más representativas se encuentran WMT, T, WFC y NVDA.
- En contraste, el modelo RF concentró sus asignaciones en pocos activos, destacando XOM, VZ y PCG, lo que sugiere un enfoque más agresivo y orientado hacia activos específicos que luego el modelo identificó como altamente prometedores (Breiman, 2001).

ROBUSTEZ Y SUPOSICIONES

El mejor desempeño del modelo de Markowitz podría atribuirse a la estabilidad histórica de las métricas de retorno y riesgo durante el período analizado. Además, este modelo se beneficia de una formulación matemática clara y de supuestos bien definidos —como la distribución normal de los retornos y la estabilidad de las covarianzas— que lo hacen consistente bajo condiciones normales de mercado (Fabozzi *et al.*, 2002; Markowitz, 1952).

Por su parte, el modelo RF, aunque potente en contextos con relaciones no lineales o estructuras complejas, puede requerir más datos o una arquitectura distinta para superar consistentemente a modelos estadísticos tradicionales, en entornos financieros (Gu *et al.*, 2020).

Los resultados mostraron que el modelo de Markowitz mantiene una mayor estabilidad en sus proyecciones de desempeño, reflejada en un menor error absoluto entre el Sharpe Ratio estimado y el realizado a lo largo del horizonte futuro de 5 días. Esta característica lo convierte en una herramienta más confiable para entornos donde la consistencia y la capacidad de anticipación son fundamentales, como en la gestión activa de portafolios.

Por el contrario, aunque el modelo de RF tiene la ventaja de incorporar relaciones no lineales y puede captar patrones complejos del mercado, su desempeño fue más volátil y menos predecible. La mayor sensibilidad de este modelo, al ruido de los datos históricos, afecta negativamente su capacidad para mantener una buena relación entre rentabilidad y riesgo en el corto plazo (DeMiguel, Garlappi, & Uppal, 2009).

INTERPRETABILIDAD Y APLICACIÓN

El modelo de Markowitz ofrece una mayor interpretabilidad y transparencia en la toma de decisiones, lo cual es valioso en contextos institucionales (Elton *et al.*, 2014). A su vez, el modelo RF, si bien más flexible y potencialmente más adaptativo,

introduce desafíos en la explicación de decisiones y control de riesgos, especialmente cuando realiza asignaciones extremas a pocos activos (Rasouli & Heidarzadeh, 2021).

CONSIDERACIONES FINALES

Esta comparación evidencia que los modelos tradicionales aún conservan ventajas competitivas importantes en entornos controlados y con buenos datos históricos.

Sin embargo, los modelos de ML ofrecen un camino prometedor, especialmente si se combinan con técnicas de validación más avanzadas, ingeniería de características o información alternativa (como datos macroeconómicos o del sentimiento del mercado) (Gu *et al.*, 2020; Nguyen *et al.*, 2021).

Futuros trabajos podrían incluir pruebas de robustez en distintos periodos de mercado, evaluar la estabilidad de los pesos a lo largo del tiempo o combinar ambos enfoques en modelos híbridos (Chen *et al.*, 2021).

REFERENCIAS

- Basak, S., Pavlidis, E., & Kim, K. H. (2019). An Empirical Comparison of Random Forest-Based Portfolio Optimization with Traditional Mean-Variance Frameworks. *Journal of Quantitative Finance*, 19(4), 615–635. <https://doi.org/10.1080/14697688.2019.1631931>
- Bennett, C. J., & Lemoine, G. J. (2017). What's Really Driving Stock Returns? Machine Learning Provides New Insights. *Journal of Financial Economics*, 125(2), 239-258.
- Bernard, S., Heutte, L., & Adam, S. (2012). On the selection of decision trees in Random Forests. *Machine Learning*, 81(3), 317–342. <https://doi.org/10.1109/IJCNN.2009.5178693>
- Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, 4(2), 315–342. <https://doi.org/10.1093/rfs/4.2.315>
- Bianchi, D., & Tamoni, A. (2019). Machine Learning and Asset Allocation. *Financial Analysts Journal*, 75(4), 14-28.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Chopra, V. K., & Ziemba, W. T. (1993). The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *Journal of Portfolio Management*, 19(2), 6-11. https://people.duke.edu/~charvey/Teaching/BA453_2006/Chopra_The_effect_of_1993.pdf
- Chen, A., Pelger, M., & Zhu, J. (2021). Deep learning in asset pricing. *The Review of Financial Studies*, 34(11), 5149–5203. <https://doi.org/10.1093/rfs/hhab083>
- De Miguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5), 1915–1953. <https://doi.org/10.1093/rfs/hhm075>
- Dey, S. R., & Saha, S. (2018). Optimal Portfolio Selection Using Random Forest Algorithm. *Quantitative Finance and Economics*, 2(2), 305–322. <https://doi.org/10.3934/qfe.2018.2.305>

- Ding, J. D. (2024). AI-Driven financial modeling techniques: transforming investment strategies. *The Journal of Applied Business and Economics*, 26(4), 63-74. <https://doi.org/10.33423/jabe.v26i4.718>
- Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2014). *Modern portfolio theory and investment analysis* (9th ed.). Wiley.
- Fabozzi, F. J., Gupta, F., & Markowitz, H. M. (2002). *The legacy of modern portfolio theory*. *The Journal of Investing*, 11(3), 7–22. <https://doi.org/10.3905/joi.2002.319510>
- Fernandes, L., Kulkarni, M., & Pande, M. B. (2023). A Systematic Literature Review of Classical and Quantum Machine Learning Approaches for Mutual Fund Portfolio Optimization. In *2023 IEEE Pune Section International Conference (PuneCon)* (pp. 1-6). IEEE
- Franco-Arbeláez, L. C., Avendaño-Rúa, C. T., & Barbutín-Díaz, H. (2011). Modelo de Markowitz y Modelo de Black-Litterman en la optimización de portafolios de inversión. *TecnoLógicas*, (26), 71-88. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-77992011000100005
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 28(2), 337–374. <https://doi.org/10.1214/aos/1016218223>
- Garmaise, M. J. (2015). Borrower Misreporting and Loan Performance. *The Journal of Finance*, 70(1), 449–484. <https://doi.org/10.1111/jofi.12156>
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *Review of Financial Studies*, 33(5), 2223-2273. <https://dachxiu.chicagobooth.edu/download/ML.pdf>
- Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124, 226–251. <https://doi.org/10.1016/j.eswa.2019.01.012>
- Huang, Z. (James), & Xu, L. (2021). Negative conversion premium. *The Journal of Finance and Data Science*, 7, 1–21. <https://doi.org/10.1016/j.jfds.2020.11.001>
- Huerta, R., & Salas, R. (2022). Machine learning in portfolio optimization: A systematic literature review. *Expert Systems with Applications*, 190, 116210. <https://doi.org/10.1016/j.eswa.2021.116210>

- Ishwaran, H., & Kogalur, U. B., Blackstone, E. H. & Lauer, M. S. (2008). Random Survival Forests. *The Annals of Applied Statistics*, 2(3), 841–860. <https://doi.org/10.1214/08-AOAS169>
- Kotsiantis, S. B., & Pintelas, P. E. (2004). Recent advances in decision trees. *Artificial Intelligence Review*, 21(3), 137–178. <https://doi.org/10.1023/B:AIRE.0000037132.72525.73>
- Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689–702. <https://doi.org/10.1016/j.ejor.2016.10.031>
- López de Prado, M. (2018). *Advances in Financial Machine Learning*. Wiley.
- Marengo Orozco, J. M. (s.f.). *Optimización de inversión en plataformas de créditos P2P*.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.2307/2975974>
- Markowitz, H. (1952). The utility of wealth. *Journal of political Economy*, 60(2), 151-158. <https://www.jstor.org/stable/1825964>
- Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, 7(4), 1851-1872. <https://www.jstor.org/stable/2329621>
- Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is “Optimized” Optimal? *Financial Analysts Journal*, 45(1), 31–42. <https://doi.org/10.2469/faj.v45.n1.31>
- Nguyen, T. T., Nguyen, Q. V., & Nguyen, T. A. (2021). A comparison of machine learning algorithms for stock price prediction. *Computational Economics*, 58(1), 1–25. <https://doi.org/10.1007/s10614-020-10050-y>
- Niaki, S. T. A., & Hoseinzade, S. (2013). Forecasting S&P 500 index using artificial neural networks and design of experiments. *Journal of Industrial Engineering International*, 9(1), 1–9. <https://doi.org/10.1186/2251-712X-9-1>
- Rapach, D., Strauss, J., & Zhou, G. (2010). Out-of-sample equity premium prediction: combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2), 821–862. <https://doi.org/10.1093/rfs/hhp063>
- Rasouli, S., & Heidarzadeh, H. (2021). Random forest applications in finance: A review. *Journal of Risk and Financial Management*, 14(10), 497. <https://doi.org/10.3390/jrfm14100497>

- Sugadev, T., Hameed, N. S., Vijayakumar, S., Tamilarasan, P., & Islam, M. S. (2023). Portfolio Optimization Using Machine Learning Techniques. In *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)* (pp. 1-7). IEEE.
- Sundaramoorthi, D., & Dong, L. (2024). Machine Learning and Optimization-Based Decision-Support Tool for Seed Variety Selection. *Annals of Operations Research*, *341*(1), 5-39. <https://doi.org/10.1007/s10479-022-04995-8>
- Szymański, B., & Szałowski, A. (2020). Machine Learning approaches for portfolio optimization: a comparative study. *Journal of Finance and Data Science*, *6*(4), 329–342. <https://doi.org/10.1016/j.jfds.2020.11.001>
- Varian, H. R. (2014). Big Data: new tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3-28. <https://www.aeaweb.org/articles?id=10.1257/jep.28.2.3>
- Yeung, H. W. F., Zhou, M., Chung, Y. Y., Moule, G., Thompson, W., Ouyang, W., Cai, W., & Bennamoun, M. (2022). Deep-learning-based solution for data deficient satellite image segmentation. *Expert Systems with Applications*, *191*, 116210. <https://doi.org/10.1016/j.eswa.2021.116210>
- Zhang, L., & Wei, Q. (2018). Machine learning for financial portfolio optimization: A survey and future directions. *Journal of Financial Research*, *41*(2), 261–288.
- Zhu, L., & Xie, W. (2023). A comparative study of machine learning models for portfolio optimization. *Computational Economics*, *64*(1), 45-68. <https://doi.org/10.1007/s10614-023-10384-5>