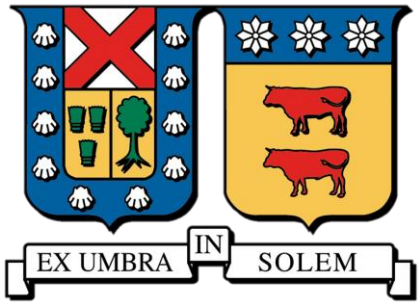


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARIA

DEPARTAMENTO DE INGENIERÍA CIVIL INDUSTRIAL

OPTIMIZACIÓN DE CARTERA BASADA EN PREDICCIÓN DEL VALOR EN RIESGO UTILIZANDO ALGORITMOS DE REGRESIÓN DE MACHINE LEARNING PARA MERCADOS BURSÁTILES.

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

AUTOR

STEVE IGNACIO CALDERÓN BEIZA

PROFESOR GUIA

WERNER KRISTJANPOLLER

PROFESOR CORREFERENTE

DIEGO CANESSA

SANTIAGO DE CHILE, 07 DE OCTUBRE, 2026

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): X Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Optimización de cartera basada en predicción del valor en riesgo utilizando algoritmos de regresión de Machine Learning para mercados bursátiles.

Nombre del candidato(a): Steve Ignacio Calderón Beiza

Carrera / Grado: Ingeniero Civil Industrial

Campus: Vitacura **Departamento: Ingeniería Civil Industrial**

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Werner Kristjanpoller Rodríguez, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

X El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por **(marcar una opción)**:

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

_____

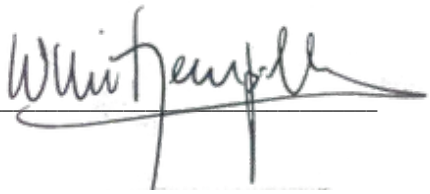
_____

_____

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 12/11/2026_____ **Firma:** _____



Estudiante o Candidato(a):

Fecha: 12/11/2026_____ **Firma:** _____



Agradecimientos

Culminar esta etapa es el resultado de la perseverancia, y no habría sido posible sin el apoyo de personas fundamentales.

A mis padres, por nunca dudar de mí, por su apoyo incondicional y por darme el impulso y la fortaleza necesarios para retomar este desafío y cerrar este importante ciclo profesional y personal. De manera muy especial, dedico este logro a la memoria de mi abuela, quien fue un pilar fundamental en mi decisión de estudiar esta carrera. Aunque no alcanzó a verme terminar este proceso, su apoyo y su profundo deseo de verme cumplir esta meta me acompañaron en cada paso hasta el final.

Agradezco a mi profesor guía, Werner Kristjanpoller, por su paciencia, su confianza y su valiosa orientación a lo largo de todo este proyecto. Su disposición fue fundamental para llevar a cabo y concluir con éxito esta investigación.

Finalmente, a toda la comunidad de la Universidad Técnica Federico Santa María: a mis compañeros de ruta más cercanos, con quienes compartí y que hicieron de las clases y trabajos una experiencia inolvidable; a los profesores, por haberme entregado las herramientas y el temple necesarios para resolver problemas; y al personal administrativo, por su constante apoyo en las gestiones del día a día.

Índice de Contenidos

Resumen Ejecutivo 1
Introducción 3
Problema de Investigación 4
Objetivos 6
  Objetivo General 6
  Objetivos Específicos 6
Marco Teórico 7
  Random Forest (RF) 8
  Support Vector Machine Regression (SVR) 9
  Adaptive Boosting (AdaBoost) 9
  Extreme Gradient Boosting (XGBoost) 10
  K-Nearest Neighbors (KNN) 11
  Artificial Neural Network (ANN) 12
  Modelo de Optimización de Portafolio Mean-VaR 13
Metodología 16
  Enfoque y Diseño 16
  Recolección y Procesamiento de Datos 17
  Modelamiento Predictivo de Machine Learning 18
  Optimización de Portafolios por Mean-Value-at-Risk 22
Alcance 24
Resultados 25
  Preprocesamiento y Consolidación de los Datos 25
  Desempeño Predictivo de los Modelos de Machine Learning 27
  Optimización de Portafolios (MVaR) 29
  Validación Dinámica *Walk-Forward* (2024-2025) 34
Conclusiones y Recomendaciones 38
Discusión 39
Limitaciones 40
Referencias 42
Anexos 44

Resumen Ejecutivo

El presente trabajo de investigación desarrolla y evalúa un sistema de optimización de portafolios aplicado al mercado bursátil estadounidense, específicamente sobre las empresas que componen el índice S&P 500. El objetivo principal es construir carteras de inversión que maximicen la rentabilidad esperada mientras se controla el riesgo de pérdidas. Para lograrlo, el sistema integra la predicción de seis algoritmos de regresión de Machine Learning con el modelo de optimización matemática Mean-Value-at-Risk (MVaR).

La metodología contempló tres etapas. El procesamiento inicial de datos que extrajo, limpió y transformó el historial de cotizaciones de 466 acciones durante el período 2016-2025. Sobre esta base estacionaria, se entrenaron los modelos: Random Forest (RF), Support Vector Regression (SVR), AdaBoost, XGBoost, K-Nearest Neighbors (KNN) y Artificial Neural Networks (ANN), usando variables desfasadas (lags) para proyectar retornos mensuales. Para la asignación de capital y validación, se implementó una simulación Walk-Forward Validation con recalibración trimestral durante el período 2024-2025. Entrenada inicialmente con datos de 2016 a 2023, esta estrategia dinámica adapta los modelos a las rotaciones del mercado y mitiga la fuga de datos (data leakage). Las predicciones Out-of-Sample generadas trimestralmente alimentaron al algoritmo MVaR, el cual distribuyó el capital maximizando el Ratio de Sharpe bajo la restricción de inversión total y la prohibición de ventas en corto (long-only).

Los resultados demostraron que el uso de un filtro basado en el Coeficiente de Determinación (R^2) resulta determinante para depurar el universo de inversión y descartar activos altamente impredecibles. Al evaluar el rendimiento predictivo, XGBoost y SVR registraron los menores errores cuadráticos medios (RMSE de 0,0748 y 0,0749, respectivamente), consolidándose como las arquitecturas con mayor capacidad de generalización para aislar el ruido del mercado.

Sin embargo, durante la validación dinámica (Walk-Forward) de las carteras en el bienio 2024-2025, el benchmark (S&P 500) registró un rendimiento acumulado del 41,29%,

un Ratio de Sharpe de 1,46 y un Maximum Drawdown (caída máxima) de -4,70%. Frente a esto, el portafolio ANN obtuvo el mayor retorno acumulado (109,41%) y riesgo ajustado (Sharpe de 2,37) con un Drawdown de apenas -0,39%; aunque su estrategia reflejó una alta concentración de capital. SVR demostró un desempeño estructuralmente robusto, alcanzando un rendimiento del 86,53% (Sharpe de 1,47; Drawdown de -9,80%). En el segmento medio, tanto KNN como XGBoost se acoplaron al mercado con retorno de 42,82% y 39,89%, respectivamente, aunque con perfiles de riesgo dispares: KNN mantuvo una mayor eficiencia defensiva (Sharpe de 1,51; Drawdown de -2,92%), mientras que XGBoost asumió una mayor exposición con un Drawdown de -17,33%. Finalmente, Random Forest y AdaBoost exhibieron un rendimiento inferior al índice de referencia; si bien Random Forest cerró el período con un 17,48%, AdaBoost presentó métricas indicativas de sobreajuste severo (overfitting), finalizando con un 4,01% y un Ratio de Sharpe negativo (-0,15).

Se concluye que la precisión predictiva no garantiza rendimientos óptimos sin una correcta gestión del riesgo. Para mercados de alta eficiencia y liquidez como el S&P 500, las arquitecturas que no se basan en árboles de decisión (ANN, SVR y KNN) demostraron superioridad financiera al estructurar carteras eficientes con un Ratio de Sharpe superior a 1. Dentro de los ensambles basados en árboles, únicamente XGBoost logró un desempeño respetable frente al mercado, aunque exponiendo el capital a caídas más pronunciadas. Estos resultados confirman que la integración de algoritmos no lineales continuos y basados en distancias, respaldados por un esquema de validación de ventana móvil (Walk-Forward), constituye el enfoque más robusto para la selección de carteras en mercados dinámicos.

Introducción

La selección óptima de portafolios es uno de los mayores desafíos financieros. En mercados de alta eficiencia y liquidez como el estadounidense, los precios están sujetos a múltiples factores dinámicos y a un alto nivel de ruido. Históricamente, la asignación de capital se fundamentó en modelos paramétricos estáticos como la frontera eficiente de Markowitz; sin embargo, su vulnerabilidad frente a caídas asimétricas ha evidenciado la necesidad de incorporar métricas de riesgo más robustas como el Valor en Riesgo (VaR).

Actualmente, la ciencia de datos ofrece herramientas capaces de modelar el comportamiento no lineal de las series de tiempo financieras. El uso de modelos de Machine Learning permite analizar grandes volúmenes de datos históricos para identificar patrones y proyectar la rentabilidad de los instrumentos. No obstante, aún existe debate sobre qué modelo resulta más eficiente al aplicarse en un índice tan complejo como el S&P 500.

Este trabajo aborda dicho desafío mediante el diseño, implementación y validación de un sistema computacional de optimización de carteras. La investigación busca determinar qué algoritmos logran aislar mejor el ruido del mercado para integrar sus proyecciones en un modelo de asignación de capital centrado en minimizar el riesgo (MVaR). Para garantizar la validez del sistema, se implementó una arquitectura de validación dinámica (Walk-Forward Validation). Este diseño somete a los portafolios a un reentrenamiento y recalibración trimestral, adaptando secuencialmente las decisiones de inversión a las rotaciones del mercado y mitigando por completo la fuga de datos (data leakage).

El documento inicia exponiendo el problema de investigación y los objetivos del estudio. A continuación, el marco teórico sienta las bases matemáticas de los algoritmos seleccionados y la teoría de la optimización MVaR. Con estos fundamentos, la metodología detalla el tratamiento de los datos y el proceso de entrenamiento. El trabajo finaliza con el análisis de los resultados, donde se evalúa tanto la precisión predictiva como el rendimiento financiero real de las carteras simuladas durante el bienio 2024-2025, dando paso a las conclusiones y recomendaciones para futuras líneas de investigación.

Problema de Investigación

El mercado financiero exige estrategias orientadas a maximizar la rentabilidad y mitigar la exposición al riesgo. La optimización de portafolios aborda este desafío mediante la formulación de modelos matemáticos que determinan la asignación eficiente del capital entre diversos instrumentos.

La teoría de portafolios, propuesta por Markowitz (1952), evalúa la rentabilidad esperada del portafolio mediante la media, el riesgo mediante la varianza y la actitud del inversor frente al riesgo. Desde entonces, la varianza ha sido un indicador de riesgo popular y el modelo media-varianza se ha extendido ampliamente, aplicándose variaciones como lo propuesto por Konno y Yamazaki (1991), quienes introdujeron un enfoque de selección de cartera lineal definiendo el riesgo como la desviación absoluta.

Los modelos de media-varianza y media-desviación absoluta han demostrado eficacia cuando los rendimientos presentan distribuciones simétricas; sin embargo, resultan insuficientes ante asimetrías, ya que penalizan matemáticamente tanto las desviaciones positivas (ganancias inusuales) como las negativas (pérdidas). Ante esta limitación, se ha integrado el modelo de Valor en Riesgo (VaR), métrica que cuantifica la exposición al riesgo estableciendo la pérdida máxima esperada que puede experimentar una inversión en un horizonte temporal, dado un nivel de confianza determinado (generalmente entre 70% y 99% en el caso de aseguradoras).

Dado que el modelo VaR se centra en la contención del riesgo sin evaluar la calidad de los instrumentos, la construcción óptima de un portafolio requiere abordar la selección de capital desde dos dimensiones. Primero, la idoneidad del activo según el perfil y la tolerancia al riesgo de cada inversor. Esto afecta directamente a los agentes del mercado fiduciario —tales como gestores de fondos mutuos, inversores institucionales y hedge funds—, quienes requieren herramientas dinámicas que protejan el patrimonio frente a caídas asimétricas sin sacrificar su competitividad. Segundo, la estimación del comportamiento futuro de estos activos, lo cual constituye una tarea compleja debido a la naturaleza estocástica del mercado bursátil. Las series financieras exhiben dinámicas no lineales, ruidosas y caóticas, donde los

precios son impactados continuamente por múltiples variables políticas, macroeconómicas y conductuales (Wang et al., 2011). Ante esta incertidumbre predictiva, los modelos econométricos tradicionales resultan insuficientes, requiriendo arquitecturas de Machine Learning capaces de aislar el ruido y descifrar patrones subyacentes.

Para abordar esta dinámica, la literatura reciente emplea algoritmos predictivos basados en datos históricos. Este trabajo se enfoca en modelos de regresión como Random Forest (RF), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Support Vector Machine Regression (SVR), K-Nearest Neighbors (KNN) y Artificial Neural Network (ANN), los cuales han demostrado superioridad para procesar series no estacionarias y no lineales (Song et al., 2004; Boulesteix y Schmid, 2014; Zhang et al., 2018). Estudios de frontera (Behera et al. 2023) han propuesto la integración de modelos de optimización, reportando precisiones predictivas destacadas utilizando AdaBoost en mercados asiáticos.

Si bien la integración de aprendizaje automático y optimización paramétrica resuelve el problema de asignación en entornos específicos, su aplicación directa a mercados de alta eficiencia, liquidez y capitalización, como el índice estadounidense S&P 500, plantea nuevos desafíos algorítmicos. La eficiencia informacional de este mercado eleva el riesgo de sobreajuste (overfitting) al modelar patrones históricos. Adicionalmente, la evaluación estática de las carteras en un único período resulta insuficiente para representar la realidad financiera, ya que omite las rotaciones sectoriales y la degradación predictiva de los modelos a lo largo del tiempo.

Por consiguiente, surge la necesidad de investigar si estas arquitecturas mantienen su superioridad predictiva y logran evitar el sobreajuste en el mercado estadounidense bajo condiciones operativas continuas. En este contexto, cabe preguntarse: ¿qué modelo entrega mejores resultados para predecir el valor futuro y gestionar el capital en el mercado estadounidense? Para dar respuesta, este trabajo resolverá el problema de selección de portafolios mediante un ensamble de los seis algoritmos de regresión mencionados y el modelo Mean-VaR. Para garantizar la validez, se incorporará una arquitectura de validación dinámica (Walk-Forward Validation).

Objetivos

Objetivo General

Diseñar e implementar un algoritmo de optimización de carteras que integre diferentes modelos de regresión de Machine Learning para la predicción de la rentabilidad de las acciones del mercado bursátil estadounidense (S&P500), acoplado a un modelo de programación matemática Mean-Value-at-Risk (MVaR), con el fin de seleccionar y gestionar dinámicamente combinaciones de activos (portafolios) que maximicen el retorno ajustado del riesgo.

Objetivos Específicos

1. Implementar los seis modelos de regresión de Machine Learning (RF, AdaBoost, XGBoost, SVR, KNN y ANN) utilizando series temporales del índice S&P 500 para proyectar la rentabilidad futura de los activos, proveyendo de insumo a la construcción de portafolios de inversión.
2. Examinar el rendimiento de los seis modelos de Machine Learning antes de la creación de cartera para garantizar la selección de los mejores activos.
3. Formular y aplicar el modelo de optimización MVaR para resolver el problema de selección de capital.
4. Analizar y comparar los resultados obtenidos de cada modelo para concluir sobre el uso de los modelos de regresión como herramienta de apoyo para la toma de decisiones financieras.

Marco Teórico

El mercado bursátil estadounidense opera como el principal centro de liquidez a nivel mundial, concentrando sistemáticamente más del 40% de la capitalización financiera global y superando los 40 billones de dólares en volumen transaccional corporativo (Securities Industry and Financial Markets Association, 2026). Dada esta magnitud, sus fluctuaciones y ciclos macroeconómicos generan repercusiones directas sobre las economías internacionales (Balli, Hasan, & Ozer-Balli, 2021), consolidándolo como el escenario primario para el diseño, estrés y validación empírica de modelos de optimización cuantitativa de carteras.

Dentro del mercado estadounidense, el índice Standard & Poor's 500 (S&P500) es el indicador financiero más representativo. Desde 1957, agrupa a las 500 empresas de mayor capitalización bursátil que cotizan en la Bolsa de Valores de Nueva York (NYSE) y la National Association of Securities Dealers Automated Quotations (NASDAQ). Este índice abarca aproximadamente el 80% del valor total disponible en EE.UU. (S&P Dow Jones Indices, 2026). Estas compañías, seleccionadas intencionadamente de diversos sectores estratégicos del mercado, tienen asignada una ponderación por capitalización flotante, por lo que existe un amplio consenso académico en que refleja fielmente la evolución estructural de la economía estadounidense (Parnes, 2020).

A causa de su alta eficiencia informacional y volumen transaccional continuo, el S&P500 presenta un comportamiento dinámico, ruidoso y altamente complejo de modelar. En la actualidad, el desarrollo de la ciencia de datos ha demostrado que las herramientas computacionales avanzadas logran procesar la no linealidad inherente de estas series financieras con mayor eficacia que los enfoques estadísticos clásicos. En esta investigación, se implementarán modelos predictivos de Machine Learning para analizar el comportamiento histórico de los activos e identificar patrones subyacentes. Estas predicciones algorítmicas actúan como un mecanismo de filtrado para depurar el universo de inversión, proveyendo estimaciones de rentabilidad que posteriormente alimentan a un motor de optimización convexa (MVaR) para la asignación eficiente del capital.

A continuación, se describen los fundamentos teóricos y matemáticos de las seis arquitecturas algorítmicas a evaluar. Esta selección metodológica abarca un espectro técnico diverso, transitando desde ensambles robustos basados en árboles de decisión hasta modelos predictivos fundamentados en el cálculo de distancias y redes neuronales continuas:

Random Forest (RF)

El Random Forest es una técnica de Machine Learning introducida por Ho (1995), que es utilizada para resolver problemas de clasificación y regresión a través de la combinación de múltiples salidas de árboles de decisión que llega a un único resultado. Cada árbol se construye usando una muestra bootstrap, que se forma a partir del conjunto de partida de manera aleatoria y con reemplazo. Aproximadamente un tercio de los datos no se incluye en el entrenamiento y se reserva para pruebas, conformando el conjunto Out-Of-Bag (OOB). Este proceso garantiza que cada árbol sea diferente, lo que reduce la varianza general del modelo y previene el sobreajuste frente a datos ruidosos.

La determinación de la predicción variará en función del tipo de problema, en el caso de la clasificación, los árboles de decisión del modelo de predicción inducido por RF se organizan en un grupo y el resultado se determina por las predicciones realizadas por cada árbol, como un voto por mayoría, mientras que, para el caso de regresión, la media de las predicciones de los árboles constituye la predicción final, es decir, cuando RF obtiene un vector de entrada (x) que contiene los valores de las muchas características evidentes que se han examinado para una determinada área de entrenamiento genera m árboles de regresión y promedia los resultados. La ecuación de predicción de RF después de m árboles $\{T(x)\}_1^M$ es:

$$f_{rf}^M(x) = \frac{1}{M}\sum_{m=1}^{M} T(x)$$

Support Vector Machine Regression (SVR)

El método de Support Vector Machine (SVM) fue desarrollado en la década de los 90 e introducido por primera vez por Cortes y Vapnik (1995). Es un algoritmo utilizado en diversos sectores para resolver problemas de regresión y clasificación. Para el caso de clasificación, el SVM tiene un enfoque que se basa en el concepto de hiperplano y de mayor margen para ofrecer la máxima separación entre las diferentes clases. Las instancias más cercanas al hiperplano de margen máximo para un problema de aprendizaje dado se denominan vectores de soporte.

Para el caso de regresión, el modelo es una ampliación denominada Support Vector Machine Regression (SVR), introducida por Drucker et al. (1996). El SVR usa técnicas de optimización convexa para transformar complejos problemas de regresión no lineal en lineales. En la práctica, el modelo crea un "tubo" de tolerancia (margen épsilon, ϵ) alrededor de los datos, ignorando los errores menores que caen dentro de él y penalizando únicamente las desviaciones significativas. Mediante esta conversión, se logra un ajuste óptimo entre el modelo y la precisión predictiva. En SVR, la relación no lineal entre las variables aleatorias X e Y es expresada por la función lineal:

$$f(x) = \omega\phi(x) + b$$

Adaptive Boosting (AdaBoost)

Adaptive Boosting fue propuesto por Freund y Schapire (1997) como el primer enfoque de refuerzo. Este algoritmo utiliza un proceso iterativo que consiste en entrenar secuencialmente varios predictores débiles (o sencillos). En cada paso, el algoritmo se vuelve adaptativo porque obliga al modelo a prestar mayor atención a las acciones que fueron más difíciles de predecir en la iteración anterior ajustando sus pesos. Este proceso se repite hasta alcanzar T, que representa el límite de iteraciones del algoritmo básico.

El método de Schapire (1999) para AdaBoost se ejecuta en un conjunto de entrenamiento (x, y) con k observaciones formando vectores como (x_1, y_1), …, (x_k, y_k), donde x_k es un elemento perteneciente al dominio de X e y_k es un elemento del conjunto Y. Los pesos de la distribución correspondiente al caso i, es denotado como $D_t(i)$, i = 1, 2, …, k y son determinados en cada iteración t. El peso se establece uniformemente como $D_1(i) = \frac{1}{k}$, para t=1. Después se construye un clasificador $h_t(x) \rightarrow X$ (0,1) donde se obtiene un clasificador débil. El error del aprendizaje débil $h_t(x)$ es determinado utilizando $\epsilon_t = \sum_{i=1}^{k} D_t(i)\delta_t$, donde:

$$\delta_t \begin{cases} 0 & si\ h_t(x_i) = y_i \\ 1 & en\ caso\ contrario \end{cases}$$

$$D_{t+1}(i) = D_t(i)\frac{F_i}{Z_t}, \qquad actualiza\ el\ peso\ en\ cada\ iteración\ donde:$$

$$F_i \begin{cases} e^{-\alpha_t} & si\ h_t(x_i) = y_i \\ e^{\alpha_t} & si\ h_t(x_i) \neq y_{i,} \end{cases}$$

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right),$$

Z_t = factor de normalización para que D_t suma uno.

Extreme Gradient Boosting (XGBoost)

El Extreme Gradient Boosting es un algoritmo de código abierto basado en árboles de decisión que se construyen de forma secuencial. En lugar de generar predicciones aisladas, el modelo minimiza los errores residuales de los árboles anteriores mediante la optimización por descenso de gradiente. Este modelo de Machine Learning es empleado para problemas de clasificación y regresión, y fue diseñado para mejorar la velocidad y el rendimiento de modelos anteriores como RF o SVR (Tarwidi et al., 2023).

La naturaleza arbórea de XGBoost lo hace insensible a los valores atípicos y es resistente al sobreajuste como otros modelos boosting, lo que facilita la selección del modelo (Alabdullah et al., 2022; Pasayat et al., 2022). La siguiente ecuación representa el modelo regularizado de XGBoost en el objetivo t-ésimo del paso de entrenamiento, donde

$l\,(y_{pred}^{(t)}, y_{truth})$ denota la pérdida, que se refiere al cálculo de la diferencia entre el valor simulado $y_{pred}^{(t)}$ y la verdad fundamental asociada y_{truth}

$$L^{(t)} = \sum_{i} l\,\left(y_{pred}^{(t)}, y_{truth}\right) + \sum_{k} \Omega(f_k)$$

Donde $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda||\omega||^2$ del k-ésimo árbol, en el cual T significa el número de hojas y $||\omega||^2$ especifica el ℓ_2 norma de todas las puntuaciones de hoja para ejemplos de entrenamiento. Al buscar en el árbol, los parámetros γ y λ regulan el grado de conservadurismo.

K-Nearest Neighbors (KNN)

El K-Nearest Neighbors es un algoritmo de aprendizaje supervisado del tipo no paramétrico que utiliza la proximidad para resolver problemas tanto de clasificación como de predicción (regresión) sobre la agrupación de un punto de datos individual, donde "K" significa la cantidad de "puntos vecinos" que tenemos en cuenta en las cercanías para clasificar los grupos o clases. En el caso de clasificaciones, se asigna una etiqueta de clase basado en el voto mayoritario, es decir, se utilizan las etiquetas que están representadas con mayor frecuencia alrededor de un punto de "consulta" de datos determinados.

En el caso de la regresión, el KNN se basa en localizar un grupo de puntos dentro de un conjunto de datos que esté más cerca del punto de "consulta" medido por alguna métrica de distancia (como Euclidiana, Minkowski o Manhattan) (Lin et al., 2022). Para un valor específico de K:

$$f(x) = \frac{1}{K} \sum_{x_i \in N(x)} y_i$$

La elección del hiperparámetro K representa un equilibrio entre el sesgo y la varianza: un valor pequeño genera un modelo con bajo sesgo y alta varianza, mientras que un valor mayor invierte esta relación. Si p = (p1, …, pn) y q = (q1, …, qn) son vectores, la distancia métrica es determinada como:

Distancia Euclidiana = $(\sum_{i=1}^{n} |p_i - q_i|^2)^{\frac{1}{2}}$

Distancia Minkowski = $(\sum_{i=1}^{n} |p_i - q_i|^{\alpha})^{\frac{1}{\alpha}}$

Distancia Manhattan = $\sum_{i=1}^{n} |p_i - q_i|$

Artificial Neural Network (ANN)

El término de Red Neuronal Artificial (ANN) hace referencia a un subcampo de inteligencia artificial que está inspirada a partir del modelo de redes neuronales biológicas del cerebro, imitando la forma en que las neuronas se envían señales entre sí, donde el núcleo de las neuronas se conoce como nodo dentro de ANN.

Las ANN son de las técnicas más fundamentales del Deep learning, estructuradas mediante un flujo de información hacia adelante (*Feedforward*). Están formadas por numerosas capas ocultas de nodos, donde cada capa está conectada a la capa anterior (que provee los datos de entrada) y a la capa siguiente (que recibe el resultado procesado). Al tomar en consideración redes con capas crecientes e interconectadas, se puede construir cualquier función general.

La combinación lineal ponderada de entradas es procesada por cada capa y transformada utilizando una función de activación. La salida de la j-ésima unidad cuando w_{ji}^{1} es el peso que conecta con la j-ésima unidad de la i-ésima unidad en la capa 1 es:

$$Z_j = g\left(\sum_{i=0}^{d} w_{ji}^{1} x_i\right)$$

Entre las opciones de funciones de activación, en este trabajo se utiliza la unidad lineal rectificada (ReLU). Las ANN son capaces de aprender la característica por sí solas y se entrenan mediante el método de aprendizaje de redes neuronales muy apreciado a pesar de que la solución no es global (Qiu et al., 2021). Sin embargo, la ANN tiene algunos inconvenientes, ya que requieren mucha potencia computacional, los modelos son difíciles de explicar y requieren de una cuidadosa atención en los datos, además de una gran cantidad de datos para su entrenamiento.

Los seis algoritmos de aprendizaje automático expuestos previamente cumplen la función exclusiva de descifrar patrones históricos para proyectar el comportamiento futuro de los precios. Sin embargo, la mera estimación direccional de retornos no constituye una estrategia de inversión integral. Para gestionar el capital con rigor financiero, las predicciones individuales generadas por estos modelos (y_{pred}) se agrupan para conformar el vector de retornos esperados (μ). Este vector actúa como el insumo matemático primordial que, junto con la matriz de covarianzas, alimenta a los motores de optimización. Ante este escenario, el presente trabajo acopla dichas proyecciones de Machine Learning al modelo de programación matemática Mean-Value-at-Risk (MVaR), transformando las estimaciones algorítmicas en ponderaciones exactas de capital para minimizar la exposición a caídas.

Modelo de Optimización de Portafolio Mean-VaR

El modelo de selección de portafolios, propuesto inicialmente por Markowitz (1952), define la cartera como una combinación de múltiples activos financieros evaluada bajo 2 dimensiones: la rentabilidad esperada (medida a través de la media) y el riesgo (medido mediante la varianza). El objetivo principal es identificar aquella composición de activos que maximice la rentabilidad para un nivel de riesgo determinado, o que minimice el riesgo para una rentabilidad esperada dada. El modelo clásico se fundamenta en las siguientes hipótesis:

1. La rentabilidad de cualquier activo financiero es una variable aleatoria. La media (o esperanza matemática) de dicha variable se acepta como la métrica representativa de la rentabilidad de la inversión.
2. Se establece como medida de riesgo la dispersión de la variable aleatoria que describe la rentabilidad, cuantificada a través de la varianza o la desviación estándar.
3. La conducta racional del inversor lo llevará invariablemente a preferir aquellos portafolios que ofrezcan una mayor rentabilidad frente a un menor riesgo.

Para capturar de forma más estricta las caídas extremas del mercado, la definición clásica evoluciona hacia el uso del Valor en Riesgo (VaR). Según Alexander y Baptista (2002, 2004), Sea α* el cuantil α de la distribución normal estándar, α ∈ (0,5, 1], el VaR paramétrico de un portafolio para un período fijo de tiempo se expresa de la siguiente manera:

$$VaR = \alpha * \sigma_p - \mu_p = \alpha * \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}x_i x_j} - \sum_{i=1}^{n}\mu_i x_i$$

En esta investigación se aplica una adaptación del modelo de optimización de Media-VaR propuesto por Sheng et al. (2012). El problema de optimización central de este sistema se define como:

$$\min VaR = \alpha * \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}x_i x_j} - \sum_{i=1}^{n}\mu_i x_i$$

$$subject\ to: \sum_{i=1}^{n}\mu_i x_i = \mu_{fix}$$

$$\sum_{i=1}^{n} x_i = 1\,, \qquad 0 \le x_i < 1, \qquad \forall\, i = 1,2,..,n$$

Este modelo híbrido minimiza el VaR de la cartera (función objetivo) en lugar de la varianza tradicional. Sujeto a las restricciones matemáticas del sistema, el algoritmo busca la configuración de menor riesgo posible exigiendo que el rendimiento esperado del portafolio alcance exactamente un valor objetivo preestablecido (μ_{fix}), asegurando simultáneamente que la suma de las proporciones de inversión (x_i) sea exactamente 1 (100% del capital asignado) y prohibiendo las posiciones en corto ($x_i \geq 0$).

Metodología

Enfoque y Diseño

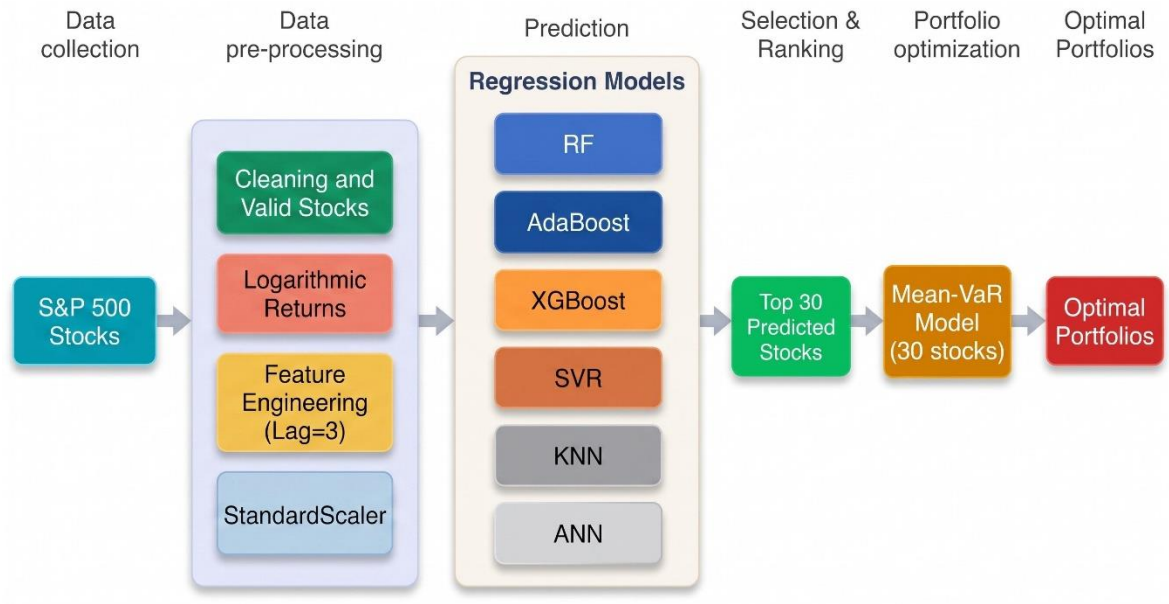
Este estudio tiene un enfoque cuantitativo y un diseño empírico-experimental. Utiliza datos históricos reales del mercado financiero para entrenar, evaluar y simular un sistema computacional de optimización de portafolios. El alcance del trabajo es de tipo correlacional y predictivo, partiendo de una premisa financiera básica: la rentabilidad de las acciones está vinculada a su comportamiento en el tiempo, por lo que las empresas más estables presentan patrones subyacentes que son más fáciles de predecir que aquellas expuestas a un alto nivel de ruido en el mercado.

Para ordenar el trabajo, el diseño metodológico se estructuró en un flujo secuencial de tres etapas:

1. Recolección y Procesamiento de Datos
2. Modelamiento Predictivo, Selección y Ranking
3. Optimización de Portafolios por MVaR

Figura 1

Esquema general de procesamiento, predicción y optimización MVaR



Esta secuencia metodológica, junto con el flujo de datos y los modelos algorítmicos aplicados en cada fase del sistema, se ilustra en la Figura 1.

Todo el entrenamiento, modelamiento predictivo y posterior optimización convexa fueron programados utilizando el lenguaje Python, apoyado principalmente en la librería de aprendizaje automático Scikit-learn.

Recolección y Procesamiento de Datos

El universo de estudio abarca a las empresas que componen el índice S&P 500 de Estados Unidos. La recolección de los datos se hizo de forma automatizada mediante la API de Yahoo Finance, extrayendo el historial de cotizaciones para el período comprendido entre el 1 de enero de 2016 y el 31 de diciembre de 2025.

Para evitar sesgos o errores por datos faltantes, se programó un filtro de supervivencia. Este script evaluó las 503 acciones nominales del índice y descartó automáticamente a 38 instrumentos que no tenían un historial continuo desde la fecha de inicio (principalmente debido a aperturas en bolsa recientes o reestructuraciones corporativas). Como resultado, el sistema consolidó un universo válido de 465 acciones.

Posteriormente, se pasó a la etapa de transformación temporal. Dado que los datos originales venían en frecuencia diaria, el sistema ejecutó una agrupación mensual para aislar y extraer específicamente el último Precio de Cierre Ajustado (Adj Close) registrado al final de cada mes.

Para cumplir con los supuestos de estacionariedad exigidos por los modelos predictivos, se prescindió del uso de estos precios nominales para el entrenamiento. En su lugar, el sistema calculó los retornos logarítmicos mensuales a partir del Precio de Cierre Ajustado (Adj Close), el cual incorpora internamente el pago de dividendos y los splits de acciones.

La fórmula aplicada para el cálculo del retorno mensual (r_t) fue la siguiente:

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

Donde P_t es el precio ajustado en el mes actual y P_{t-1} es el precio ajustado en el mes inmediatamente anterior. Este procesamiento dio como resultado una matriz de datos estacionaria, lista para ser utilizada en la fase de entrenamiento de Machine Learning.

Modelamiento Predictivo de Machine Learning

Los inversionistas en el mercado financiero buscan, por lo general, implementar estrategias que les ayuden a construir un portafolio de inversión óptimo, el cual ofrezca un potencial retorno significativo con el menor riesgo posible. Bajo esta premisa, el objetivo principal de esta etapa se enfoca en la predicción y selección de acciones con alto potencial de rendimiento.

Para la proyección de los retornos futuros, se implementó un ensamble comparativo de seis algoritmos de Machine Learning: Random Forest (RF), Support Vector Regression (SVR), AdaBoost, XGBoost, K-Nearest Neighbors (KNN) y Artificial Neural Networks (ANN).

Dado que la mayoría de los algoritmos de regresión estándar son estáticos y no manejan secuencias temporales de forma nativa, se implementó un proceso de Feature Engineering para dotar al sistema de memoria. Para cada acción, se generaron variables desfasadas temporalmente (lags). Se estableció una ventana de $lag = 3$, lo que significa que el algoritmo utilizó como variables predictoras o características (X) los retornos de los tres meses anteriores (el último trimestre cerrado) para predecir el retorno del mes actual como variable objetivo (y). Este desfase estructural asegura que el modelo identifique patrones de tendencia basándose estrictamente en el pasado, evitando la fuga de datos (Data Leakage) al momento de la predicción. Los valores nulos generados en los primeros meses de la serie a causa de este desfase fueron eliminados de la matriz.

Para garantizar la validez prospectiva de los modelos y evitar el fenómeno de fuga de datos (Data Leakage), el conjunto histórico se dividió cronológicamente. Los datos comprendidos entre los años 2016 y 2023 se utilizaron exclusivamente como conjunto de entrenamiento (Train), mientras que el bienio 2024-2025 conformó el conjunto de prueba (Out-of-Sample).

Dado que algoritmos basados en cálculo de distancias (como KNN y SVR) son altamente sensibles a la magnitud de las variables, se aplicó un proceso de estandarización. Para mantener el aislamiento de los datos, el escalador estadístico (StandardScaler) fue ajustado únicamente con la media y varianza del conjunto de entrenamiento.

Con el objetivo de maximizar la capacidad predictiva, se ejecutó una búsqueda automatizada de hiperparámetros (Grid Search). A través de este proceso iterativo, se definieron las configuraciones óptimas para cada algoritmo (detalladas en la Tabla 1), incorporando transversalmente parámetros operativos de ejecución en paralelo (n_jobs=-1) y semillas de control de aleatoriedad (random_state=42) para garantizar la reproducibilidad del experimento.

Tabla 1

Parámetros para la Predicción de los Modelos

Modelo	RF	SVR	Adaboost	XGboost	KNN	ANN
Parámetro (Valor)	n_estimators (300)	Kernel (rbf)	Estimator(max_depth=1)	n_estimators (75)	N_neighbors (20)	Hidden-layer_sizes (5,3)
	Max_depth (1)	C (0.05)	n_estimators (80)	Leaning_rate (0.01)	Weights (uniform)	Activation (relu)
	Max_features (sqrt)	Epsilon (0.05)	Leaning_rate (0.005)	Max_depth (2)	P (2)	Solver (lbfgs)
	min_samples_split (2)	Gamma (0.01)	Loss (square)	Subsample (1.0)		Alpha (0.1)
	min_samples_leaf (5)			colsample_bytree (0.8)		Learning_rate_init (0.001)

El ajuste reveló que, frente a la estocasticidad y el ruido inherente a la volatilidad bursátil, los modelos requerían configuraciones estrictamente regularizadas para prevenir el sobreajuste (overfitting). En el caso de los ensambles basados en árboles, Random Forest se restringió a estimadores de una sola división (max_depth=1), promediando 300 árboles sobre un subconjunto reducido de variables (max_features='sqrt') para estabilizar su varianza. De manera similar, AdaBoost utilizó como estimador base árboles de un solo nivel

(max_depth=1) acotados a 80 iteraciones, aplicando una tasa de aprendizaje conservadora (learning_rate=0,005) y una función de pérdida cuadrática (loss='square') para forzar una convergencia gradual sin memorizar anomalías mensuales. Por su parte, XGBoost se estructuró con 75 árboles de profundidad 2 (max_depth=2), implementando un muestreo aleatorio del 80% de las variables por árbol (colsample_bytree=0.8) junto con penalizaciones matemáticas L1 y L2 (reg_alpha=1.0, reg_lambda=1.0) para aislar la tendencia estructural subyacente.

En cuanto a las arquitecturas basadas en distancias, vectores y nodos, el modelo SVR convergió hacia un margen de insensibilidad estricto (épsilon=0.01), una regularización fuerte (C=0.05) y un alcance de influencia amplio (gamma=0.01). El uso de un kernel radial (rbf) le permitió proyectar patrones no lineales descartando las fluctuaciones espurias menores al 5% mensual. El algoritmo KNN requirió agrupar 20 vecinos (n_neighbors=20) con una ponderación equitativa (weights='uniform') y calculados mediante distancia Euclidiana, lo que diluyó el impacto de meses con retornos extremos al promediar un mayor volumen de observaciones históricas. Finalmente, la Red Neuronal Artificial (ANN) se estructuró mediante una topología de "cuello de botella" con dos capas ocultas de 5 y 3 neuronas, utilizando la función de activación ReLU y una tasa de aprendizaje inicial baja (learning_rate_init=0,001). Dado el tamaño acotado de las observaciones mensuales, se seleccionó el algoritmo de optimización cuasi-Newton (solver='lbfgs'), el cual garantizó una convergencia matemática más estable y precisa.

Para comprobar la precisión de los modelos durante esta fase, se utilizaron cuatro métricas estadísticas estandarizadas:

- **Error cuadrático medio (MSE):** Es una medida de qué tan cerca está una línea de regresión de un grupo de puntos (Behera et al., 2022). Se resuelve con la diferencia al cuadrado entre los valores reales y predichos. Cuanto menor sea el MSE, la predicción será más precisa.

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2$$

donde y_j = valores reales, $\hat{y}_j$ = valores predichos.

- **Raíz del error cuadrático medio (RMSE):** Representa la desviación estándar de los errores de predicción (residuales) (Ho et al., 2021). Proporciona la medida de cuán dispersos están los residuales alrededor de la línea de mejor ajuste, penalizando severamente las predicciones atípicas (*outliers*), siendo crítica para entornos de alta volatilidad

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2}$$

donde y_j = valores reales, $\hat{y}_j$ = valores predichos.

- **Error absoluto medio (MAE):** Se estima promediando la diferencia absoluta entre los valores reales y previstos en todo el conjunto de datos, es decir, es el promedio aritmético del error absoluto (Elavarasan et al., 2018). Mide la magnitud de los errores y sin importar la dirección, cuanto más bajo es el valor del MAE, más preciso es el modelo.

$$\text{MAE} = \frac{1}{N}\sum_{j=1}^{N}\left|y_j - \hat{y}_j\right|$$

donde y_j = valores reales, $\hat{y}_j$ = valores predichos.

- **Coeficiente de Determinación (R^2):** Indica la proporción de la varianza en el retorno real que es predecible por el modelo.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

donde y_i = valores reales, $\hat{y}_i$ = valores predichos; $\bar{y}$ = promedio de valores reales.

Tras validar los modelos y revisar sus márgenes de error, el siguiente paso fue generar las proyecciones reales para armar las carteras del 2025 (el período $t + 1$). Para lograrlo, el sistema tomó los datos del último mes disponible (diciembre de 2024) y estimó el retorno esperado de cada instrumento.

Antes de asignar el capital, se aplicó un filtro para limpiar el universo de inversión. El sistema evaluó el Coeficiente de Determinación (R^2) histórico de cada acción y descartó directamente aquellas que resultaron ser demasiado impredecibles (es decir, las que obtuvieron un $R^2 < -0.5$). Finalmente, de los activos que pasaron este corte, el algoritmo seleccionó a los 30 con mayor rentabilidad proyectada. Este "Top 30" fue el subconjunto definitivo que se exportó hacia la fase de optimización.

Para garantizar la validez prospectiva de los modelos y evitar el fenómeno de fuga de datos (data leakage), el conjunto histórico se estructuró mediante una arquitectura de validación dinámica (Walk-Forward Validation). Inicialmente, los datos comprendidos entre 2016 y 2023 se utilizaron exclusivamente como conjunto de entrenamiento base (Train). A partir de enero de 2024, el sistema implementó una recalibración trimestral: los modelos predicen el retorno de los activos para el trimestre siguiente (Out-of-Sample). Al finalizar dicho período, las observaciones reales del mercado se incorporan al bloque de entrenamiento para reajustar los pesos e hiperparámetros antes de proyectar el trimestre posterior. Esta mecánica de ventana móvil avanzó de forma secuencial hasta abarcar la totalidad de los años 2024 y 2025.

Optimización de Portafolios por Mean-Value-at-Risk

La fase final del sistema estructuró la asignación de capital mediante el modelo de programación no lineal Mean-Value-at-Risk (MVaR). A diferencia de la optimización clásica de media-varianza de Markowitz, que penaliza de manera simétrica cualquier desviación estándar, el enfoque MVaR aísla y penaliza específicamente el riesgo de pérdidas extremas. Para ello, emplea como medida de riesgo el Valor en Riesgo (VaR) paramétrico, calibrado en este estudio a un nivel de confianza del 70%.

Operativamente, la ponderación porcentual de capital asignada a cada activo (x_i) se recalculó de forma secuencial al inicio de cada trimestre. En cada iteración temporal, el motor de optimización recibió dos parámetros de entrada: el vector de retornos esperados (μ), derivado de las proyecciones individuales de cada modelo de Machine Learning, y la matriz de covarianzas empírica (Σ), calculada con la historia de precios hasta la fecha de corte. La resolución matemática de este problema no lineal se llevó a cabo mediante el algoritmo de optimización convexa Sequential Least Squares Programming (SLSQP).

Para sortear las inestabilidades matemáticas propias del cálculo del VaR ante proyecciones algorítmicas alcistas —donde una varianza proyectada cercana a cero induce un denominador marginal y distorsiona el ratio de eficiencia—, la función objetivo se configuró para maximizar el diferencial de retorno ajustado por riesgo (una adaptación estructural del Ratio de Sharpe). Esta formulación matemática permite que el algoritmo numérico converja de manera estable hacia el portafolio de tangencia sobre la frontera eficiente, garantizando la máxima compensación de rentabilidad por cada unidad de riesgo paramétrico asumido.

Toda la optimización trimestral se ejecutó respetando dos restricciones financieras estrictas:

- **Inversión total del capital disponible:** La suma de los pesos asignados a los activos debe ser exactamente igual al 100% del capital disponible ($\sum x_i = 1$).
- **Restricción de venta en corto (Long-only):** Se excluyó la posibilidad de operar con apalancamiento o ventas en corto, restringiendo el dominio de asignación para cada activo a un valor nulo o positivo con un límite máximo del total de la cartera ($0 \leq x_i \leq 1$).

Las ponderaciones óptimas derivadas de este proceso establecieron la composición exacta de los portafolios al inicio de cada subperíodo. Estas asignaciones estáticas trimestrales alimentaron posteriormente el motor de validación dinámica (*Walk-Forward Validation*), permitiendo cuantificar la viabilidad real y el rendimiento acumulado de las estrategias integradas durante el bienio 2024-2025.

Alcance

La investigación se delimita al mercado bursátil estadounidense, focalizándose en las empresas de renta variable del índice Standard & Poor's 500 (S&P 500). El horizonte temporal abarca ininterrumpidamente desde enero de 2016 hasta diciembre de 2025, estructurado metodológicamente en una fase de entrenamiento algorítmico base (2016-2023) y una validación dinámica mediante ventana móvil (Walk-Forward Validation) a lo largo del bienio 2024-2025. Para mitigar sesgos de supervivencia, el universo de inversión se restringe a un subconjunto depurado de 465 acciones con un historial de cotización continuo durante todo el período de evaluación.

La asignación de capital se rige por las restricciones del modelo de programación matemática Mean-Value-at-Risk (MVaR), el cual exige una inversión total del patrimonio disponible y prohíbe las posiciones de venta en corto (estrategia long-only). Asimismo, la evaluación de las carteras asume un entorno de fricción transaccional nula durante los procesos de rebalanceo periódico por trimestre. Bajo esta dinámica, la simulación excluye el impacto financiero de comisiones de intermediación (brokers), deslizamiento de precios (slippage), diferencial de liquidez (bid-ask spread) y obligaciones tributarias, con el propósito de aislar y cuantificar estrictamente el desempeño técnico predictivo de las arquitecturas algorítmicas frente al mercado.

Resultados

Preprocesamiento y Consolidación de los Datos

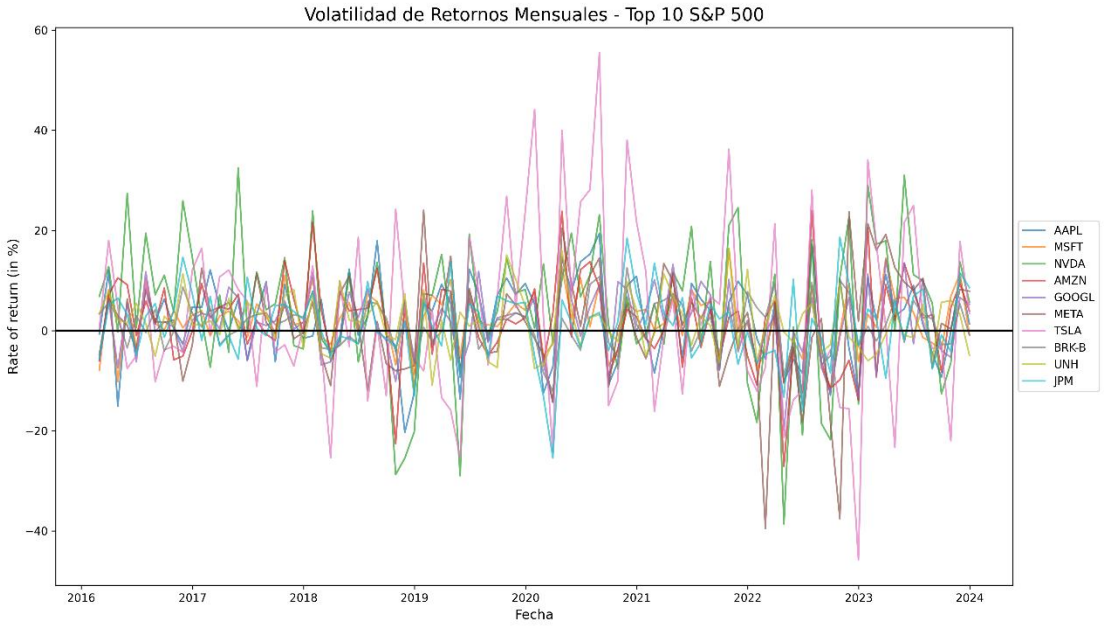
La fase inicial del procesamiento de datos consistió en la extracción automatizada del historial de cotizaciones mediante la API de Yahoo Finance. De las 503 acciones que componen el índice S&P500, se aplicó un filtro para verificar la continuidad operativa de cada instrumento desde enero de 2016 hasta diciembre de 2025. Este procedimiento de depuración consolidó exitosamente 465 activos válidos, excluyendo a 38 empresas que presentaban datos faltantes por aperturas bursátiles recientes o reestructuraciones corporativas. Posteriormente, los precios de cierre ajustados (Adj Close) se transformaron en una matriz de retornos logarítmicos mensuales, garantizando la estacionariedad de las series temporales y asimilando el efecto de eventos como el pago de dividendos o splits accionarios.

Al analizar los datos, ilustrada en la Figura 2 para las diez empresas de mayor capitalización del mercado, evidencia un comportamiento estocástico, no lineal y de alta dispersión. Los retornos mensuales observados fluctúan en rangos amplios que alcanzan contracciones del -40% y expansiones cercanas al +60%. Esta extrema volatilidad y ausencia de patrones simples evidencian las limitaciones de los modelos lineales tradicionales (como ARIMA), estableciendo el escenario base para el uso de Machine Learning, herramientas capaces de descifrar patrones subyacentes en series financieras complejas y el ruido inherente a estas series financieras.

Adicionalmente, el análisis visual captura la exposición del mercado tanto a riesgos sistemáticos como idiosincráticos. En primer lugar, se identifica una contracción generalizada y simultánea durante el primer trimestre de 2020, producto de la crisis sanitaria global (COVID-19). La inclusión de estos valores atípicos (outliers) de alta severidad dentro del conjunto de entrenamiento resulta vital, ya que obliga a los algoritmos a modelar el comportamiento del mercado bajo escenarios de estrés.

Figura 2

Volatilidad de retornos mensuales de las 10 acciones más representativas del S&P500



En segundo lugar, se constata una alta volatilidad individual y descorrelacionada en activos específicos; por ejemplo, mientras el mercado general asimilaba el impacto de 2020, activos de alta ponderación como Tesla registraron repuntes alcistas cercanos al 60% a finales de ese mismo año, seguidos de correcciones superiores al -40% hacia finales de 2022. Una dinámica oscilatoria similar se observa en NVIDIA a partir de 2022.

Estas divergencias individuales demuestran el peligro propio de cada empresa. Este riesgo justifica la integración del motor de programación matemática Mean-Value-at-Risk (MVaR) en la fase final del flujo, cuya función es distribuir el capital para mitigar la probabilidad de que el colapso abrupto de un activo particular comprometa la viabilidad del portafolio. Con el universo de inversión consolidado en un formato estacionario, la ejecución del sistema avanza hacia la evaluación estadística del desempeño de regresión, con el fin de cuantificar qué algoritmo posee la mayor capacidad técnica para aislar el ruido y generar estimaciones de rentabilidad precisas.

Desempeño Predictivo de los Modelos de Machine Learning

Tras el preprocesamiento de las series temporales, se evaluó la capacidad de los seis modelos para aislar el ruido estocástico y proyectar la tendencia del mercado. Para garantizar la validez estadística y evitar el sesgo de sobreajuste (overfitting), el cálculo del error predictivo se ejecutó sobre el primer bloque del conjunto de prueba (Out-of-Sample), correspondiente al año 2024. Este diseño permitió asegurar que la evaluación se realice frente a datos estrictamente no observados durante la fase de entrenamiento base (2016-2023). La Tabla 2 consolida el promedio poblacional ($\bar{x}$), la desviación estándar (σ) y la varianza (σ^2) de las métricas de desempeño —Error Cuadrático Medio (MSE), Raíz del Error Cuadrático Medio (RMSE) y Error Absoluto Medio (MAE)— junto con el Coeficiente de Determinación (R^2) obtenidos para todas las acciones analizadas.

Tabla 2

Rendimiento Predictivo Out-of-Sample de los Diferentes Modelos (Año 2024)

Modelo	Estadístico	MSE	RMSE	MAE	R^2
AdaBoost	$\bar{x}$	0,00649	0,07568	0,06056	-0,114385
	σ	0,00627	0,02771	0,02158	0,168821
	σ^2	0,0000393	0,000768	0,000465	0,028500
ANN	$\bar{x}$	0,00639	0,07506	0,05998	-0,090142
	σ	0,00605	0,02752	0,02108	0,130245
	σ^2	0,0000367	0,000757	0,000444	0,016963
KNN	$\bar{x}$	0,00658	0,07640	0,06123	-0,141537
	σ	0,00604	0,02725	0,02114	0,182427
	σ^2	0,0000364	0,000743	0,000447	0,033279
RandomForest	$\bar{x}$	0,00639	0,07509	0,05995	-0,093381
	σ	0,00615	0,02742	0,02105	0,133747
	σ^2	0,0000379	0,000752	0,000443	0,017888
SVR	$\bar{x}$	0,00635	0,07488	0,05981	-0,088251
	σ	0,00604	0,02735	0,02088	0,140792
	σ^2	0,0000365	0,000748	0,000436	0,019822
XGBoost	$\bar{x}$	0,00633	0,07481	0,05967	-0,083697
	σ	0,00600	0,02718	0,02068	0,112160
	σ^2	0,0000360	0,000739	0,000427	0,012579

Los resultados indican que, para el mercado estadounidense, XGBoost logró el mejor desempeño predictivo general. El modelo obtuvo el menor Error Cuadrático Medio (MSE:

0,00633), el menor RMSE (0,07481) y el menor Error Absoluto Medio (MAE: 0,05967). En términos financieros, valores reducidos de MSE y RMSE implican desviaciones marginales respecto a los retornos reales observados. Específicamente, un MAE cercano a 0,059 señala que la magnitud promedio del error del modelo es de 5,9% mensual por acción. Considerando que los activos de renta variable pueden presentar variaciones atípicas de hasta un 50% mensual, concentrar el margen de error poblacional en torno a un 0,07481 demuestra una alta eficacia para filtrar la volatilidad del mercado. Adicionalmente, el modelo SVR exhibió un desempeño estadísticamente equivalente, posicionándose como una arquitectura competitiva con un RMSE de 0,07488 y un MAE de 0,05981, seguido por la red neuronal ANN con un RMSE: 0,07507. Al evaluar la varianza del error (σ^2), XGBoost presenta la métrica más baja en MSE (0,000036) y RMSE (0,000739), seguido por KNN (RMSE 0,000743) y SVR (RMSE 0,000748). Esta combinación de bajo error promedio y alta estabilidad confirma que estos algoritmos logran mitigar la memorización de anomalías al modelar el riesgo idiosincrático de cada empresa.

Respecto al Coeficiente de Determinación (R^2), la evaluación revela una media poblacional negativa transversal a todos los modelos, con el registro menos adverso liderado por XGBoost (-0,083697). En el modelado de retornos bursátiles de alta liquidez, la proporción de ruido frente a la señal real es predominantemente alta, dado que el mercado exhibe una mayor cantidad de ruido aleatorio que patrones estructurales predecibles. Un R^2 negativo a nivel poblacional indica que predecir el retorno exacto de un activo individual genera, en promedio, mayor dispersión que utilizar simplemente su media histórica. No obstante, el objetivo en esta fase no radica en la exactitud del valor absoluto proyectado, sino en la generación de un ordenamiento relativo (ranking) para identificar los instrumentos con mayor expectativa comparativa de rentabilidad. Para controlar esta variabilidad intrínseca y mitigar el riesgo de selección, se aplicó un filtro de exclusión ($R^2 \geq -0{,}5$) que descarta directamente los activos altamente impredecibles, conservando exclusivamente el subconjunto donde los modelos detectaron una señal paramétrica consistente.

La aplicación de este filtro permitió aislar un universo depurado, seleccionando las 30 acciones con mayor rentabilidad esperada mensual para cada modelo predictivo. Sin embargo, la mera estimación de retornos resulta insuficiente para la toma de decisiones corporativas si no se integra una estrategia integral de inversión, este subconjunto de activos opera como el insumo matricial que alimentará el motor de programación matemática no lineal Mean-Value-at-Risk (MVaR). La fase de optimización subsiguiente tendrá como objetivo distribuir el capital disponible estructurando carteras que maximizan la rentabilidad proyectada, sujetas a la minimización estricta del riesgo extremo frente a posibles correcciones del mercado.

Optimización de Portafolios (MVaR)

Al procesar las proyecciones de rentabilidad del subconjunto de activos seleccionados (Top 30) en el motor de programación matemática no lineal Mean-Value-at-Risk (MVaR) para el primer período de la simulación (Out-of-Sample correspondiente a enero de 2024), se estructuraron las carteras óptimas maximizando la esperanza de retorno mensual. Tal como se detalla en la Tabla 3, se observan diferencias sustanciales en los perfiles de riesgo gestionados por cada modelo. Para evaluar la viabilidad financiera de estas configuraciones teóricas, el análisis se centra en dos métricas de eficiencia estructural. Primero, el Valor en Riesgo al 70% (VaR), que cuantifica la pérdida máxima de capital esperada en un escenario adverso bajo dicho nivel de confianza. Segundo, el ratio Retorno/VaR, que dimensiona la eficiencia matemática del capital asignado frente a la exposición a pérdidas; un valor alto indica una rentabilidad robusta frente al riesgo asumido, mientras que un valor bajo o nulo advierte una compensación deficiente o un sobreajuste algorítmico.

Los resultados de esta calibración inicial evidencian perfiles de asignación diferenciados, caracterizados transversalmente por un fuerte optimismo predictivo. En la formulación paramétrica del MVaR, la obtención de un Valor en Riesgo negativo indica que la esperanza de retorno (μ) del portafolio supera matemáticamente a la penalización por volatilidad paramétrica al nivel de confianza del 70%. En términos financieros, esto significa

que los algoritmos proyectaron un impulso alcista suficientemente sólido como para neutralizar la expectativa de pérdida nominal en el límite del escenario adverso evaluado.

Tabla 3

*Resultados de la Esperanza del Optimizador por Modelo (*Enero de 2024*)*

Modelo	Retorno mensual esperado	VaR 70%	Retorno/VaR
Adaboost	5,69%	-2,18%	2,6118
ANN	3,71%	-1,09%	3,4144
KNN	3,71%	-1,77%	2,0927
RF	3,07%	-0,63%	4,8635
SVR	3,58%	-1,12%	3,1889
XGBoost	2,55%	-0,03%	91,4222

Bajo este contexto, el modelo AdaBoost estructuró la cartera con mayor sesgo de crecimiento, proyectando el retorno mensual más alto de la muestra (5,69%), lo cual derivó en un VaR de -2,18% y un ratio Retorno/VaR de 2,6118. Una postura similar se observó en ANN y KNN; ambos modelos igualaron su expectativa de retorno en 3,71%, aunque KNN asumió una volatilidad estructural ligeramente mayor, reflejada en un VaR de -1,77% frente al -1,09% de ANN. En una franja de moderación paramétrica, SVR y Random Forest lograron configuraciones balanceadas. SVR estimó un retorno de 3,58% limitando el VaR a -1,12% (ratio de 3,1889), mientras que Random Forest proyectó un 3,07% de retorno con un VaR contenido en -0,63%, logrando el ratio estructural más eficiente entre los ensambles convencionales (4,8635).

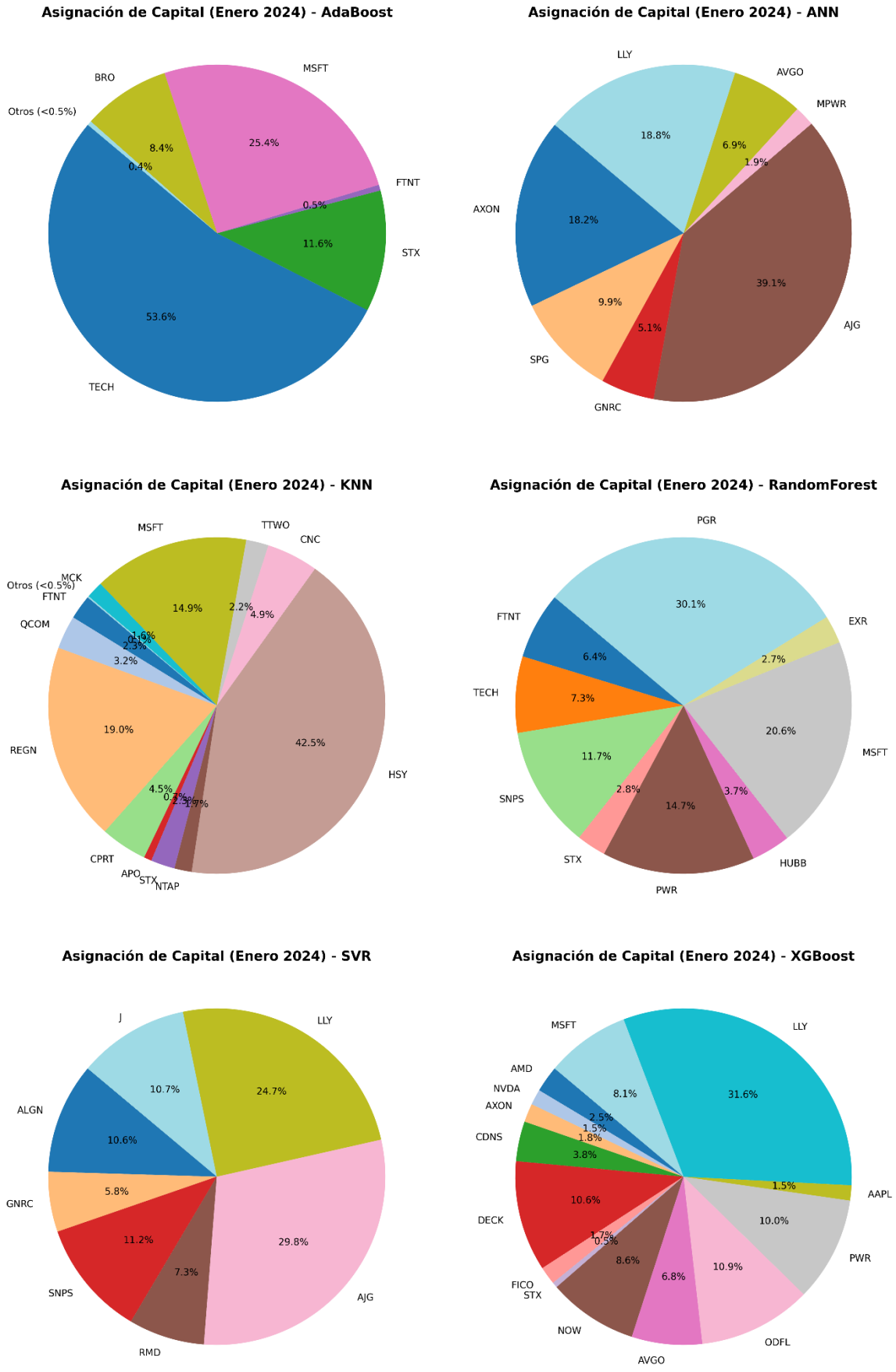
Finalmente, XGBoost generó la asignación de mayor aversión al riesgo: estructuró la rentabilidad esperada más baja (2,55%) mediante la selección de activos altamente defensivos, comprimiendo el riesgo del portafolio hasta un VaR estadísticamente tendiente a cero (-0,03%). La división de un retorno estándar por este denominador casi nulo disparó el ratio a 91,4222, un fenómeno matemático que confirma la priorización de la protección del capital por parte de este algoritmo frente a la incertidumbre del mercado.

Al analizar la composición interna de los portafolios óptimos generados para enero de 2024 (Figura 3), las divergencias en los niveles de riesgo se materializan en estrategias de asignación sectorial muy distintas. Al observar las tres tenencias principales que anclan la rentabilidad esperada de cada cartera, la concentración del riesgo se manifiesta agudamente en el modelo AdaBoost, el cual destinó el 53,57% del capital a Bio-Techne Corp (TECH), un 25,41% a Microsoft (MSFT) y un 11,63% a Seagate Technology (STX), absorbiendo más del 90% del portafolio en solo tres instrumentos. De manera análoga, ANN y KNN anclaron posiciones dominantes de forma focalizada: ANN asignó un 39,12% a Arthur J. Gallagher & Co. (AJG), un 18,84% a Eli Lilly (LLY) y un 18,20% a Axon Enterprise (AXON), mientras que KNN destinó el 42,48% de su patrimonio a The Hershey Company (HSY), secundado por Regeneron Pharmaceuticals (REGN, 19,02%) y Microsoft (MSFT, 14,93%).

En contraste, los modelos que lograron mayor balance estadístico optaron por una diversificación ponderada sobre empresas de alta resiliencia corporativa. XGBoost lideró su cartera con Eli Lilly (LLY, 31,61%), complementada por participaciones en industria y consumo discrecional con Old Dominion Freight Line (ODFL, 10,89%) y Deckers Brands (DECK, 10,62%). Por su parte, Random Forest y SVR distribuyeron el riesgo en una combinación de aseguradoras, tecnología, infraestructura y salud; Random Forest concentró el capital en Progressive Corp (PGR, 30,08%), Microsoft (MSFT, 20,57%) y Quanta Services (PWR, 14,67%), mientras que SVR ancló su capital en Arthur J. Gallagher & Co. (AJG, 29,76%), Eli Lilly (LLY, 24,69%) y Synopsys (SNPS, 11,19%). El detalle exhaustivo con las 30 acciones seleccionadas y las ponderaciones exactas para cada uno de los seis modelos se encuentra disponible en la sección de anexos (Tablas A1 a A6).

Figura 3

Asignación de portafolios por modelo



Al concluir esta fase de optimización, cabe precisar que las ponderaciones de capital estructuradas por el motor MVaR se fundamentan en estimaciones ex ante; específicamente, en la esperanza de rentabilidad proyectada para el mes inmediatamente siguiente (enero de 2024). Si bien las métricas paramétricas demuestran la capacidad matemática del sistema para equilibrar rentabilidad y riesgo en un entorno aislado, la validez financiera de estas estrategias depende de su solidez a través del tiempo. En este contexto, contrastar la rentabilidad proyectada individual de los activos frente a su rendimiento efectivo durante el primer mes de cotización permite cuantificar la brecha de error predictivo inicial.

Los reportes detallados en la sección de anexos (Tablas A7 a A12) evidencian desviaciones significativas entre la teoría algorítmica y la estocasticidad práctica del mercado. Por un lado, se registraron severas sobreestimaciones en empresas que proyectaban retornos moderados, pero experimentaron contracciones reales de dos dígitos; el caso más extremo fue Tesla (TSLA), que sufrió una caída mensual del -28,27%, generando una brecha de error predictivo superior al 30%, seguido por ON Semiconductor (ON) con un diferencial de error cercano al 18,5%. Por otro lado, los modelos subestimaron fuertemente los impulsos alcistas agresivos del sector tecnológico, como el caso de NVIDIA (NVDA), cuyo rendimiento real alcanzó el 21,71% frente a proyecciones conservadoras cercanas al 4%, resultando en un sesgo de subestimación superior al 17%.

Esta divergencia inherente entre el rendimiento exigido por el optimizador y la volatilidad real del mercado confirma que un análisis de rentabilidad estático, acotado a un solo mes, resulta insuficiente. La fluctuación evidenciada en los activos que componen las carteras requiere someter a los algoritmos a una evaluación continua para cuantificar si la gestión del riesgo (VaR) logró absorber estas anomalías tempranas y generar un exceso de retorno genuino. Por consiguiente, la siguiente sección evalúa la viabilidad técnica y financiera de estos portafolios mediante una validación dinámica de ventana móvil (Walk-Forward Validation), ejecutando reentrenamientos y recalibraciones trimestrales ininterrumpidas a lo largo del bienio 2024-2025.

Validación Dinámica *Walk-Forward* (2024-2025)

Para estructurar una validación medible y cuantificable de las asignaciones de capital, se ejecutó una simulación prospectiva dinámica (Walk-Forward Validation) utilizando los datos fuera de la muestra (out-of-sample) correspondientes al bienio 2024-2025. Esta metodología implementa una ventana temporal móvil que reentrena los modelos predictivos y recalibra las ponderaciones del portafolio (vía el optimizador MVaR) al inicio de cada trimestre. Operativamente, el sistema proyecta los retornos, ajusta las posiciones de inversión y acumula el rendimiento real obtenido durante los meses subsiguientes, repitiendo este ciclo iterativamente para asimilar las rotaciones macroeconómicas. La evolución del patrimonio se determinó mediante la capitalización geométrica de los retornos, aplicando la formulación del interés compuesto:

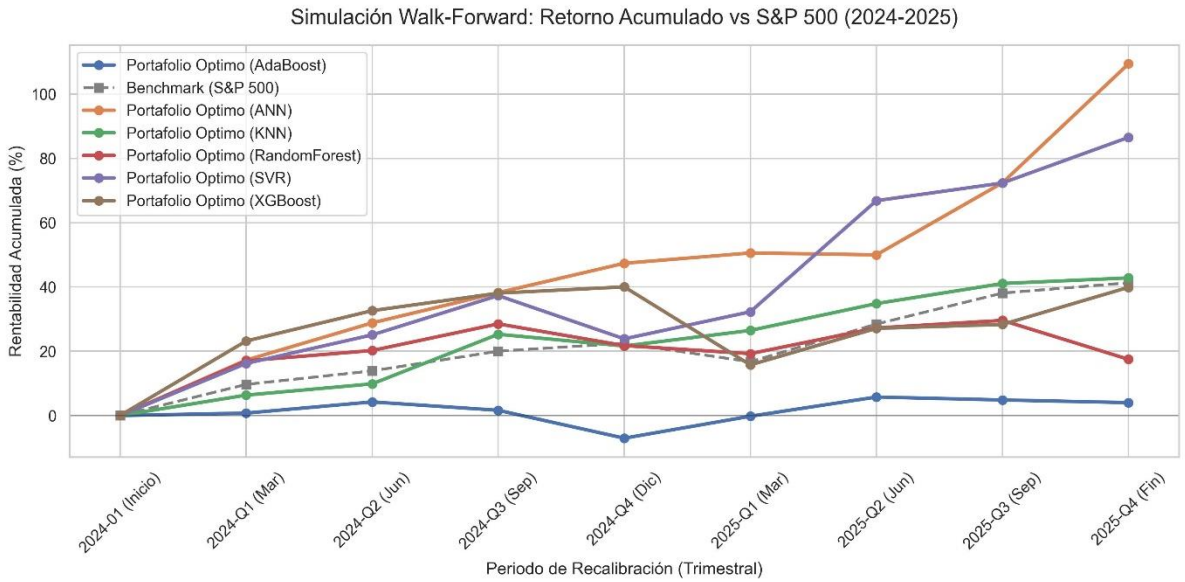
$$R_{acum} = \prod_{t=1}^{T} \left(1 + R_{port,t}\right) - 1$$

donde $R_{port,t}$ representa el rendimiento consolidado de la cartera en el mes t. Como línea base de rendimiento (benchmark), se incorporó la evolución del índice S&P 500, permitiendo evaluar si la gestión activa de los algoritmos logró generar un exceso de retorno (alfa) frente al mercado general.

Las trayectorias de rendimiento acumulado (Figura 4) y las métricas de riesgo resultantes (Tabla 4) revelan que la recalibración trimestral generó disparidades profundas entre los modelos, fuertemente condicionadas por sus rotaciones sectoriales en períodos críticos. La arquitectura basada en redes neuronales (ANN) exhibió un desempeño marcadamente superior al resto, capturando una tendencia alcista sostenida a partir del segundo semestre de 2024 y acelerando drásticamente en 2025 para finalizar el bienio con un retorno acumulado de 109,41%.

Figura 4

Rendimiento de los modelos para el periodo de2024-2025



Esta aceleración se fundamenta en su rebalanceo estratégico a inicios de 2025; en la recalibración de enero, ANN estructuró una asignación focalizada en activos de alta tracción, destinando sus mayores ponderaciones a Arthur J. Gallagher & Co. (AJG, 33,43%), Southern Company (SO, 26,33%) y Axon Enterprise (AXON, 25,42%). Desde el punto de vista del riesgo, su eficiencia resultó excepcional: ANN registró el mayor Ratio de Sharpe de la prueba (2,37) y un Max Drawdown estadísticamente marginal de -0,39%, demostrando que la rotación trimestral logró capturar la prima del mercado sin exponer el patrimonio a retrocesos severos.

Tabla 4

Metricas de Riesgo por Modelo

Modelo / Indice	Retorno Total	Volatilidad Anual	Sharpe Ratio	Max Drawdown
S&P 500	41,29%	9,65%	1,46	-4,70%
AdaBoost	4,01%	10,14%	-0,15	-10,83%
ANN	109,41%	15,04%	2,37	-0,39%
KNN	42,82%	9,75%	1,51	-2,92%
RandomForest	17,48%	16,37%	0,33	-9,35%
SVR	86,53%	20,53%	1,47	-9,80%
XGBoost	39,89%	22,74%	0,68	-17,33%

Por su parte, el modelo SVR se consolidó como la segunda arquitectura más robusta (86,53%). Aunque experimentó una compresión temporal de retornos entre el tercer trimestre de 2024 y el cuarto trimestre de 2024, el rebalanceo logró reestructurar la cartera para capitalizar el fuerte impulso de los últimos nueve meses, finalizando con un Sharpe sólido de 1,47 y un drawdown acotado a -9,80%.

En el segmento de riesgo medio, el modelo KNN logró replicar de cerca la inercia del índice de referencia, superando por un margen estrecho al S&P 500 al cierre del período (42,82% frente a 41,29% del benchmark). Su gestión de exposición fue altamente eficiente, registrando un Ratio de Sharpe de 1,51 y conteniendo su máxima pérdida a solo un -2,92%, demostrando ser una estrategia conservadora pero estadísticamente preferible a la indexación pasiva. En contraste, XGBoost presentó la mayor inestabilidad paramétrica de la prueba. A pesar de su alta precisión en fases previas, la simulación expuso la cartera a una volatilidad anualizada del 22,74%. Entre el tercer trimestre de 2024 y el primer trimestre de 2025, el algoritmo sufrió una contracción severa evidenciada por un Max Drawdown de -17,33%. Esta caída se explica al auditar su composición interna; durante el inicio de este período de estrés (octubre de 2024), el modelo mantuvo una alta concentración en activos sujetos a correcciones sectoriales temporales, liderados por Eli Lilly (LLY, 33,10%), Arista Networks (ANET, 14,53%) y Quanta Services (PWR, 9,02%). Aunque logró recuperar posiciones en el último tramo para cerrar en 39,89%, su Ratio de Sharpe cayó a 0,68, indicando que el retorno nominal final no compensó adecuadamente el estrés financiero asumido.

En el cuartil inferior, los modelos Random Forest y AdaBoost evidenciaron deficiencias estructurales severas para adaptarse al horizonte de dos años. Random Forest inició con rendimientos alineados al mercado, pero perdió capacidad predictiva a partir de mediados de 2024, estancándose de forma progresiva hasta cerrar con un rendimiento subóptimo de 17,48% (Sharpe de 0,33). El comportamiento de AdaBoost fue el más deficiente de la investigación: la cartera fue incapaz de generar tracción a lo largo de los ocho trimestres, cerrando el ciclo con un rendimiento cuasi nulo del 4,01% y un Ratio de Sharpe negativo (-0,15). Al evaluar sus rotaciones, como la efectuada en julio de 2024, se constata una asignación

ineficiente y descorrelacionada de los vectores de crecimiento del mercado, priorizando posiciones estáticas en Aptiv (APTV, 44,42%), FedEx (FDX, 27,77%) y Merck & Co. (MRK, 23,02%). Esta degradación confirma que el sobreoptimismo detectado en la fase de optimización inicial (overfitting) invalidó por completo su capacidad para navegar en escenarios variables, anulando la generación de valor. Para revisar el desglose histórico completo de las asignaciones de capital estructuradas en cada trimestre de la simulación, referirse a las Tablas A13 a A18 en la sección de anexos.

La simulación dinámica confirma que combinar proyecciones de Machine Learning con un optimizador MVaR es una metodología práctica y efectiva para la gestión de portafolios, aunque su éxito depende directamente de la calidad del modelo predictivo. Los resultados demuestran que algoritmos flexibles y adaptables, como ANN y SVR, logran procesar la nueva información cada trimestre, descartar inversiones desfavorables y construir carteras que superan claramente al mercado general. Por el contrario, los modelos que tienden a confundir las variaciones aleatorias de los precios con patrones reales sufren fuertes caídas en su rentabilidad. Esto evidencia que, para proteger el capital en el mercado bursátil, no basta con ajustar las posiciones periódicamente; es indispensable que el sistema esté respaldado por algoritmos robustos que logren generalizar el comportamiento del mercado y eviten el sobreajuste a datos pasados.

Conclusiones y Recomendaciones

El presente trabajo de investigación desarrolló y validó un sistema de optimización de portafolios para el mercado bursátil estadounidense, demostrando que la integración de modelos predictivos de Machine Learning con la programación no lineal Mean-Value-at-Risk (MVaR) constituye una metodología viable para la gestión de capital en entornos de alta volatilidad y liquidez como el índice S&P 500. Un aspecto metodológico determinante para la estabilidad del sistema fue la aplicación de un filtro de exclusión basado en el Coeficiente de Determinación ($R^2 \geq -0{,}5$). Descartar los activos altamente estocásticos permitió concentrar el capital exclusivamente en instrumentos con señales estructurales identificables, facilitando la convergencia del optimizador hacia carteras eficientes y reduciendo la exposición al ruido aleatorio inherente a las series financieras de alta liquidez.

La validación mediante la simulación dinámica de ventana móvil (Walk-Forward Validation) para el bienio 2024-2025 confirmó que la recalibración trimestral incide de manera crítica en la viabilidad financiera de los modelos. La arquitectura basada en redes neuronales (ANN) demostró la mayor capacidad matemática para procesar nueva información, rotando su exposición sectorial para registrar un rendimiento acumulado de 109,41%, el mayor Ratio de Sharpe de la evaluación (2,37) y una exposición mínima a caídas severas (Max Drawdown de -0,39%). De manera complementaria, el modelo SVR evidenció una alta solidez estructural, capitalizando la inercia del mercado para alcanzar un retorno de 86,53%. En contraste, arquitecturas de ensamble que mostraron alta precisión en regresión estática, como XGBoost, exhibieron vulnerabilidades ante la dinámica temporal real, sufriendo inestabilidad paramétrica significativa (volatilidad de 22,74% y Max Drawdown de -17,33%). Asimismo, el desempeño casi nulo de AdaBoost (4,01%) confirmó empíricamente que el sobreajuste (overfitting) a anomalías históricas anula por completo la capacidad algorítmica para generar valor sostenido en el mercado.

Para futuras líneas de investigación, se propone profundizar en la modelación de la microestructura del mercado. Dado que la presente simulación asumió un entorno de fricción cero, resulta pertinente incorporar variables de costos operacionales de transacción,

diferenciales de cotización (bid-ask spread) y estimaciones de impacto en el precio de ejecución (slippage), con el objetivo de cuantificar la rentabilidad neta exigible tras cada rebalanceo trimestral. Adicionalmente, la expansión de la matriz de características predictivas mediante la inclusión de variables macroeconómicas exógenas —tales como la inflación, el crecimiento del PIB o las variaciones en las tasas de interés referenciales de la Reserva Federal— permitiría a los algoritmos asimilar los ciclos económicos subyacentes, aislando el riesgo sistémico de manera más eficiente durante períodos de estrés financiero.

Discusión

Los hallazgos de este estudio contrastan directamente con la literatura reciente, en particular con la investigación de referencia desarrollada por Behera et al. (2023). Al analizar mercados bursátiles asiáticos, dichos autores concluyeron que AdaBoost constituía el algoritmo con mayor precisión predictiva y eficiencia financiera. No obstante, la validación dinámica aplicada al índice S&P 500 evidencia que este comportamiento algorítmico no es replicable y depende intrínsecamente de la microestructura del mercado. Al operar en el mercado estadounidense, AdaBoost resultó sensible a anomalías direccionales históricas. A pesar de aplicar restricciones matemáticas sobre sus hiperparámetros, el modelo sufrió de sobreajuste (overfitting), estructurando asignaciones que fueron incapaces de adaptarse a las rotaciones macroeconómicas trimestrales, lo que derivó en un rendimiento acumulado deficiente (4,01%) y un Ratio de Sharpe negativo (-0,15).

Esta divergencia demuestra que un mercado de alta capitalización y eficiencia exige herramientas predictivas más estables y adaptables. Al evaluar la red neuronal (ANN), los resultados comprobaron que esta arquitectura posee la flexibilidad matemática necesaria para ajustar sus ponderaciones de forma dinámica a lo largo del tiempo. Gracias a su capacidad de reentrenamiento trimestral, ANN logró capturar el crecimiento del mercado sin exponer el capital a caídas estructurales, reflejando una pérdida máxima (Max Drawdown) de apenas -0,39% frente a retornos superiores al 100%. De manera análoga, el modelo SVR validó su

utilidad práctica, demostrando que sus penalizaciones internas ayudan a construir carteras seguras y resilientes durante los dos años de simulación.

En contrapartida, las metodologías de ensamble basadas en árboles de decisión revelaron limitaciones operativas bajo este régimen dinámico. Específicamente, XGBoost, que suele liderar las métricas de regresión transversal estática, evidenció una fuerte inestabilidad paramétrica durante la simulación secuencial. Su alta sensibilidad a fluctuaciones de corto plazo lo llevó a reestructuraciones deficientes durante las transiciones trimestrales, exponiendo la cartera a una volatilidad anualizada superior al 22% y a contracciones de capital severas (-17,33%).

En consecuencia, este estudio demuestra que no existe un único algoritmo de Machine Learning que sea el mejor para todos los escenarios financieros. La elección y ajuste del modelo debe adaptarse siempre a la naturaleza y volatilidad específica del mercado donde se invierte. Asimismo, los resultados confirman que evaluar un portafolio en un único período de tiempo no es suficiente. Para comprobar si una estrategia de inversión es realmente segura y rentable, es indispensable utilizar una validación dinámica continua (Walk-Forward). Este enfoque obliga a los modelos a procesar información nueva constantemente, separando las tendencias reales de las fluctuaciones aleatorias del mercado, lo que resulta fundamental para sostener y proteger el capital a largo plazo.

Limitaciones

El diseño y ejecución de la presente investigación contemplan restricciones metodológicas estructurales que deben ser consideradas al interpretar los niveles de rentabilidad reportados. En primer lugar, el alto costo computacional derivado del procesamiento de matrices de datos a gran escala obligó a estandarizar la calibración de hiperparámetros. En lugar de optimizar la arquitectura algorítmica de manera individual para cada empresa, se implementó una búsqueda exhaustiva (*Grid Search*) parametrizada de forma generalizada para todo el universo de acciones. Esta homogeneización analítica implica que

los modelos predictivos pueden ceder precisión local o limitar su capacidad para capturar la ciclicidad idiosincrática de activos financieros con comportamientos corporativos atípicos.

En segundo lugar, la evaluación del rendimiento prospectivo mediante la simulación dinámica (*Walk-Forward Validation*) para el bienio 2024-2025 se ejecutó asumiendo un entorno de fricción transaccional nula. La metodología analítica excluyó explícitamente variables operativas críticas del mercado real, tales como las comisiones de intermediación institucional (*brokers*), los costos por deslizamiento de precios en la ejecución de órdenes (*slippage*), el impacto del diferencial entre los precios de compra y venta (*bid-ask spread*) y las eventuales cargas impositivas sobre la realización de ganancias de capital de corto plazo.

Bajo la recalibración continua implementada, el motor de optimización MVaR exige ejecutar la reestructuración de instrumentos al inicio de cada trimestre para actualizar las ponderaciones del portafolio. En un escenario de gestión fiduciaria real, esta rotación periódica de activos acumularía costos de transacción sostenidos que no se encuentran reflejados en el cálculo del interés compuesto del simulador. Bajo esta dinámica, la incorporación de estas obligaciones operativas y tributarias reduciría de manera directa el margen de ganancia neto frente a los retornos brutos obtenidos en la evaluación fuera de la muestra, impactando financieramente con mayor severidad a aquellos modelos predictivos que exijan una alta rotación estructural del capital.

Referencias

Alabdullah, A. A., Iqbal, M., Zahid, M., Khan, K., Amin, M. N., & Jalal, F. E. (2022). Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Construction and Building Materials, 345*, 128-296. https://doi.org/10.1016/j.conbuildmat.2022.128296

Alexander, G. J., & Baptista, A. M. (2002). Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis. *Journal of economic dynamics and control, 26*(7-8), 1159-1193. https://doi.org/10.1016/S0165-1889(01)00041-0

Alexander, G. J., & Baptista, A. M. (2004). A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science, 50*(9), 1261-1273. https://doi.org/10.1287/mnsc.1040.0201

Balli, F., Hasan, M., & Ozer-Balli, H. G.-A. (2021). Why do U.S. uncertainties drive stock market spillovers? International evidence. *International Review of Economics & Finance, 76*, 288-301. https://doi.org/10.1016/j.iref.2021.06.015

Behera, J., Pasayat, A. K., & Behera, H. (2022). COVID-19 Vaccination Effect on Stock Market and Death Rate in India. *Asia-Pac Financ Markets*, 651-673. https://doi.org/10.1007/s10690-022-09364-w

Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence, 120*, 105843. https://doi.org/10.1016/j.engappai.2023.105843

Boulesteix, A.-L., & Schmid, M. (2014). Machine learning versus statistical modeling. *Biometrical Journal, 56*(4), 588-593. https://doi.org/10.1002/bimj.201300226

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Springer link, 20*, 273-297. https://link.springer.com/article/10.1007/bf00994018

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support Vector Regression Machines. *NIPS'96: Proceedings of the 9th International Conference on Neural Information Processing Systems*, 155-161. https://dl.acm.org/doi/proceedings/10.5555/2998981

Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture, 155*, 257-282. https://doi.org/10.1016/j.compag.2018.10.024

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences, 55*(1), 119-139. https://doi.org/10.1006/jcss.1997.1504

Freund, Y., & Schapire, R. E. (1999). A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence, 14*(5), 771-780. https://cseweb.ucsd.edu/~yfreund/papers/IntroToBoosting.pdf

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition, 1*, 278-282. https://doi.org/10.1109/ICDAR.1995.598994

Ho, W. K., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research, 38*, 48-70. https://doi.org/10.1080/09599916.2020.1832558

Konno, H., & Yamazaki, H. (1991). Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market. *Management Science, 37*(5), 519-531. https://doi.org/10.1287/mnsc.37.5.519

Lin, G., Lin, A., & Danlei, G. (2022). Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. *Information Sciences, 608*, 517-531. https://doi.org/10.1016/j.ins.2022.06.090

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91. https://doi.org/10.2307/2975974

Parnes, D. (2020). Exploring economic anomalies in the S&P500 index. *The Quarterly Review of Economics and Finance, 76*, 292-309. https://doi.org/10.1016/j.qref.2019.09.012

Pasayat, A. K., Mitra, A., & Bhowmick, B. (2022). Determination of essential features for predicting start-up success: an empirical approach using machine learning. *Technology Analysis & Strategic Management*, 1-19. https://doi.org/10.1080/09537325.2022.2116569

Qiu, Y., Garg, D., Kim, S.-M., Mudawar, I., & Kharangate, C. R. (2021). Machine learning algorithms to predict flow boiling pressure drop in mini/micro-channels based on universal consolidated data. *178*, 121-607. https://doi.org/10.1016/j.ijheatmasstransfer.2021.121607

S&P Dow Jones Indices. (julio de 2026). *S&P U.S. Indices Methodology*. S&P Global: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf

Safari, A. (2024). Hybrid Emerging Model Predictive Data-Driven Forecasting of Three-Phase Electrical Signals of Photovoltaic Systems using GBR-LSTM. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, 100549. https://doi.org/10.1016/j.prime.2024.100549

Securities Industry and Financial Markets Association. (13 de agosto de 2026). *Capital Markets Fact Book*. SIFMA: https://www.sifma.org/research/statistics/fact-book

Sheng, Z., Benshan, S., & Zhongping, W. (2012). Analysis of mean-VaR model for financial risk control. *Systems Engineering Procedia, 4*, 40-45. https://doi.org/10.1016/j.sepro.2011.11.047

Song, X., Mitnitski, A., Cox, J., & Rockwood, K. (2004). Comparison of Machine Learning Techniques with Classical Statistical Models in Predicting Health Outcomes. *IOS Press Ebooks, 107*, 736-740. https://doi.org/10.3233/978-1-60750-949-3-736

Tarwidi, D., Pudjaprasetya, S. R., Adytia, D., & Apri, M. (2023). An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach. *MethodsX, 10*, 102-119. https://doi.org/10.1016/j.mex.2023.102119

Wang, J.-Z., Wang, J.-J., & Zhang, Z.-G. (2011). Forecasting stock indices with back propagation neural network. *Expert Systems with Applications, 38*(11), 14346-14355. https://doi.org/10.1016/j.eswa.2011.04.222

Zhang, J., Li, Z., Pu, Z., & Xu, C. (2018). Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods. *Institute of Electrical and Electronics Engineers, 6*, 60079-60087. https://doi.org/10.1109/ACCESS.2018.2874979

Anexos

Tabla A1

Composición Óptima del Portafolio MVaR (Enero 2024) – Modelo Adaptive Boosting (AdaBoost)

N°	Stock	Nombre	Sector	Peso (%)
1	TECH	Bio-Techne Corp	Salud	53,57
2	MSFT	Microsoft Corp	Tecnología	25,41
3	STX	Seagate Technology	Tecnología	11,63
4	BRO	Brown & Brown, Inc.	Financiero	8,41
5	FTNT	Fortinet Inc.	Tecnología	0,55
6	LOW	Lowe's Companies, Inc.	Consumo Discrecional	0,43
7	ALGN	Align Technology	Salud	0
8	NVDA	NVIDIA Corp	Tecnología	0
9	URI	United Rentals	Industrial	0
10	GNRC	Generac Holdings	Industrial	0
11	LRCX	Lam Research	Tecnología	0
12	SNPS	Synopsys Inc.	Tecnología	0
13	CPRT	Copart Inc.	Industrial	0
14	CDNS	Cadence Design Systems	Tecnología	0
15	NFLX	Netflix Inc.	Servicios de Comunicación	0
16	KLAC	KLA Corp	Tecnología	0
17	AMAT	Applied Materials	Tecnología	0
18	ANET	Arista Networks	Tecnología	0
19	PHM	PulteGroup	Consumo Discrecional	0
20	TER	Teradyne	Tecnología	0
21	NTAP	NetApp Inc.	Tecnología	0
22	PWR	Quanta Services	Industrial	0
23	ADBE	Adobe Inc.	Tecnología	0
24	FCX	Freeport-McMoRan	Materiales	0
25	SCHW	Charles Schwab Corp	Financiero	0
26	IDXX	IDEXX Laboratories	Salud	0
27	AVGO	Broadcom Inc.	Tecnología	0
28	BLDR	Builders FirstSource	Industrial	0
29	VEEV	Veeva Systems	Salud	0
30	QCOM	QUALCOMM Inc.	Tecnología	0

Tabla A2

Composición Óptima del Portafolio MVaR (Enero 2024) – Modelo Red Neuronal Artificial (ANN)

N°	Stock	Nombre	Sector	Peso (%)
1	AJG	Arthur J. Gallagher & Co.	Financiero	39,12
2	LLY	Eli Lilly and Co.	Salud	18,84
3	AXON	Axon Enterprise	Industrial	18,20
4	SPG	Simon Property Group	Bienes Raíces	9,90
5	AVGO	Broadcom Inc.	Tecnología	6,86
6	GNRC	Generac Holdings	Industrial	5,15
7	MPWR	Monolithic Power Systems	Tecnología	1,94
8	NVDA	NVIDIA Corp	Tecnología	0
9	AMD	Advanced Micro Devices	Tecnología	0
10	ARES	Ares Management	Financiero	0
11	TSLA	Tesla Inc.	Consumo Discrecional	0
12	BLDR	Builders FirstSource	Industrial	0
13	CDNS	Cadence Design Systems	Tecnología	0
14	DECK	Deckers Brands	Consumo Discrecional	0
15	LRCX	Lam Research	Tecnología	0
16	TPL	Texas Pacific Land Corp	Energía	0
17	FICO	Fair Isaac Corp	Tecnología	0
18	SNPS	Synopsys Inc.	Tecnología	0
19	SMCI	Super Micro Computer	Tecnología	0
20	NOW	ServiceNow	Tecnología	0
21	FANG	Diamondback Energy	Energía	0
22	KLAC	KLA Corp	Tecnología	0
23	ODFL	Old Dominion Freight Line	Industrial	0
24	PWR	Quanta Services	Industrial	0
25	CPRT	Copart Inc.	Industrial	0
26	FTNT	Fortinet Inc.	Tecnología	0
27	AAPL	Apple Inc.	Tecnología	0
28	ON	ON Semiconductor	Tecnología	0
29	AMAT	Applied Materials	Tecnología	0
30	URI	United Rentals	Industrial	0

Tabla A3

Composición Óptima del Portafolio MVaR (Enero 2024) – Modelo K-Nearest Neighbors (KNN)

N°	Stock	Nombre	Sector	Peso (%)
1	HSY	The Hershey Company	Consumo Básico	42,48
2	REGN	Regeneron Pharmaceuticals	Salud	19,02
3	MSFT	Microsoft Corp	Tecnología	14,93
4	CNC	Centene Corporation	Salud	4,95
5	CPRT	Copart Inc.	Industrial	4,46
6	QCOM	QUALCOMM Inc.	Tecnología	3,20
7	FTNT	Fortinet Inc.	Tecnología	2,34
8	STX	Seagate Technology	Tecnología	2,31
9	TTWO	Take-Two Interactive	Servicios de Comunicación	2,15
10	NTAP	NetApp Inc.	Tecnología	1,66
11	MCK	McKesson Corp	Salud	1,62
12	APO	Apollo Global Management	Financiero	0,74
13	NVDA	NVIDIA Corp	Tecnología	0,12
14	ALGN	Align Technology	Salud	0
15	PHM	PulteGroup	Consumo Discrecional	0
16	FCX	Freeport-McMoRan	Materiales	0
17	ADBE	Adobe Inc.	Tecnología	0
18	TECH	Bio-Techne Corp	Salud	0
19	IDXX	IDEXX Laboratories	Salud	0
20	TER	Teradyne	Tecnología	0
21	ANET	Arista Networks	Tecnología	0
22	SNPS	Synopsys Inc.	Tecnología	0
23	MRVL	Marvell Technology	Tecnología	0
24	AES	The AES Corp	Servicios Públicos	0
25	A	Agilent Technologies	Salud	0
26	HUM	Humana Inc.	Salud	0
27	CDNS	Cadence Design Systems	Tecnología	0
28	CRL	Charles River Laboratories	Salud	0
29	CTAS	Cintas Corporation	Industrial	0
30	KR	The Kroger Co.	Consumo Básico	0

Tabla A4

Composición Óptima del Portafolio MVaR (Enero 2024) – Modelo Random Forest (RF)

N°	Stock	Nombre	Sector	Peso (%)
1	PGR	Progressive Corp	Financiero	30,08
2	MSFT	Microsoft Corp	Tecnología	20,57
3	PWR	Quanta Services	Industrial	14,67
4	SNPS	Synopsys Inc.	Tecnología	11,73
5	TECH	Bio-Techne Corp	Salud	7,34
6	FTNT	Fortinet Inc.	Tecnología	6,39
7	HUBB	Hubbell Inc.	Industrial	3,73
8	STX	Seagate Technology	Tecnología	2,82
9	EXR	Extra Space Storage	Bienes Raíces	2,68
10	AMD	Advanced Micro Devices	Tecnología	0
11	NVDA	NVIDIA Corp	Tecnología	0
12	ALGN	Align Technology	Salud	0
13	TSLA	Tesla Inc.	Consumo Discrecional	0
14	CDNS	Cadence Design Systems	Tecnología	0
15	AMAT	Applied Materials	Tecnología	0
16	AXON	Axon Enterprise	Industrial	0
17	QCOM	QUALCOMM Inc.	Tecnología	0
18	BLDR	Builders FirstSource	Industrial	0
19	NFLX	Netflix Inc.	Servicios de Comunicación	0
20	MRVL	Marvell Technology	Tecnología	0
21	URI	United Rentals	Industrial	0
22	MSCI	MSCI Inc.	Financiero	0
23	LOW	Lowe's Companies, Inc.	Consumo Discrecional	0
24	ARES	Ares Management	Financiero	0
25	GNRC	Generac Holdings	Industrial	0
26	LII	Lennox International	Industrial	0
27	IDXX	IDEXX Laboratories	Salud	0
28	KLAC	KLA Corp	Tecnología	0
29	APO	Apollo Global Management	Financiero	0
30	AES	The AES Corp	Servicios Públicos	0

Tabla A5

Composición Óptima del Portafolio MVaR (Enero 2024) – Modelo Support Vector Machine Regression (SVR)

N°	Stock	Nombre	Sector	Peso (%)
1	AJG	Arthur J. Gallagher & Co.	Financiero	29,76
2	LLY	Eli Lilly and Co.	Salud	24,69
3	SNPS	Synopsys Inc.	Tecnología	11,19
4	J	Jacobs Solutions Inc.	Industrial	10,66
5	ALGN	Align Technology	Salud	10,64
6	RMD	ResMed Inc.	Salud	7,29
7	GNRC	Generac Holdings	Industrial	5,77
8	NVDA	NVIDIA Corp	Tecnología	0
9	APO	Apollo Global Management	Financiero	0
10	FTNT	Fortinet Inc.	Tecnología	0
11	AMD	Advanced Micro Devices	Tecnología	0
12	LRCX	Lam Research	Tecnología	0
13	MSCI	MSCI Inc.	Financiero	0
14	ADBE	Adobe Inc.	Tecnología	0
15	NFLX	Netflix Inc.	Servicios de Comunicación	0
16	ODFL	Old Dominion Freight Line	Industrial	0
17	MRVL	Marvell Technology	Tecnología	0
18	MPC	Marathon Petroleum	Energía	0
19	IDXX	IDEXX Laboratories	Salud	0
20	WST	West Pharmaceutical Services	Salud	0
21	NTAP	NetApp Inc.	Tecnología	0
22	NVR	NVR, Inc.	Consumo Discrecional	0
23	BLDR	Builders FirstSource	Industrial	0
24	TECH	Bio-Techne Corp	Salud	0
25	TRGP	Targa Resources	Energía	0
26	CDNS	Cadence Design Systems	Tecnología	0
27	TER	Teradyne	Tecnología	0
28	URI	United Rentals	Industrial	0
29	LII	Lennox International	Industrial	0
30	ARES	Ares Management	Financiero	0

Tabla A6

Composición Óptima del Portafolio MVaR (Enero 2024) – Modelo Extreme Gradient Boosting (XGBoost)

N°	Stock	Nombre	Sector	Peso (%)
1	LLY	Eli Lilly and Co.	Salud	31,61
2	ODFL	Old Dominion Freight Line	Industrial	10,89
3	DECK	Deckers Brands	Consumo Discrecional	10,62
4	PWR	Quanta Services	Industrial	10,05
5	NOW	ServiceNow	Tecnología	8,63
6	MSFT	Microsoft Corp	Tecnología	8,10
7	AVGO	Broadcom Inc.	Tecnología	6,77
8	CDNS	Cadence Design Systems	Tecnología	3,83
9	AMD	Advanced Micro Devices	Tecnología	2,53
10	AXON	Axon Enterprise	Industrial	1,78
11	FICO	Fair Isaac Corp	Tecnología	1,72
12	NVDA	NVIDIA Corp	Tecnología	1,50
13	AAPL	Apple Inc.	Tecnología	1,46
14	STX	Seagate Technology	Tecnología	0,53
15	BLDR	Builders FirstSource	Industrial	0
16	SNPS	Synopsys Inc.	Tecnología	0
17	KLAC	KLA Corp	Tecnología	0
18	MPWR	Monolithic Power Systems	Tecnología	0
19	LRCX	Lam Research	Tecnología	0
20	TSLA	Tesla Inc.	Consumo Discrecional	0
21	ON	ON Semiconductor	Tecnología	0
22	CPRT	Copart Inc.	Industrial	0
23	URI	United Rentals	Industrial	0
24	FTNT	Fortinet Inc.	Tecnología	0
25	AMAT	Applied Materials	Tecnología	0
26	ALGN	Align Technology	Salud	0
27	MSCI	MSCI Inc.	Financiero	0
28	MU	Micron Technology	Tecnología	0
29	NFLX	Netflix Inc.	Servicios de Comunicación	0
30	JBL	Jabil Inc.	Tecnología	0

Tabla A7

Contraste de Rentabilidad Proyectada vs. Real (Enero 2024) – Modelo Adaptive Boosting (AdaBoost)

Stock	t+1 (%)	Retorno real (%)	Diff (%)
TECH	7,34	-9,28	16,63
ALGN	6,07	-2,47	8,54
STX	5,45	0,36	5,09
NVDA	4,74	21,71	-16,97
FTNT	4,28	9,70	-5,42
URI	3,62	8,68	-5,06
GNRC	3,60	-12,84	16,44
LOW	3,57	-3,96	7,53
LRCX	3,52	5,21	-1,69
SNPS	3,38	3,52	-0,14
MSFT	3,36	5,57	-2,21
CPRT	3,22	-1,98	5,20
CDNS	3,19	5,74	-2,55
NFLX	3,10	14,72	-11,62
KLAC	3,01	2,17	0,84

Stock	t+1 (%)	Retorno real (%)	Diff (%)
AMAT	2,92	1,37	1,55
ANET	2,90	9,38	-6,48
PHM	2,87	1,29	1,58
BRO	2,67	8,68	-6,01
TER	2,64	-11,65	14,29
NTAP	2,62	-0,51	3,14
PWR	2,62	-10,62	13,25
ADBE	2,58	3,49	-0,91
FCX	2,53	-6,64	9,17
SCHW	2,51	-8,93	11,44
IDXX	2,50	-7,47	9,98
AVGO	2,43	5,55	-3,12
BLDR	2,38	3,99	-1,61
VEEV	2,34	7,45	-5,11
QCOM	2,34	2,65	-0,31

Tabla A8

Contraste de Rentabilidad Proyectada vs. Real (Enero 2024) – Modelo Red Neuronal Artificial (ANN)

Stock	t+1 (%)	Retorno real (%)	Diff (%)
AXON	6,75	-3,65	10,41
SPG	5,45	-2,87	8,32
NVDA	4,40	21,71	-17,31
AMD	4,06	12,89	-8,83
GNRC	3,30	-12,84	16,14
ARES	3,08	2,13	0,95
TSLA	2,98	-28,27	31,25
BLDR	2,95	3,99	-1,04
AJG	2,81	3,19	-0,38
CDNS	2,67	5,74	-3,06
DECK	2,67	12,01	-9,34
LRCX	2,66	5,21	-2,56
TPL	2,64	-7,33	9,97
FICO	2,60	2,95	-0,35
MPWR	2,59	-4,55	7,14

Stock	t+1 (%)	Retorno real (%)	Diff (%)
SNPS	2,58	3,52	-0,94
SMCI	2,56	62,23	-59,66
NOW	2,49	8,01	-5,52
FANG	2,48	-0,87	3,35
AVGO	2,48	5,55	-3,07
KLAC	2,47	2,17	0,30
ODFL	2,44	-3,59	6,04
PWR	2,42	-10,62	13,04
CPRT	2,41	-1,98	4,39
LLY	2,39	10,21	-7,83
FTNT	2,38	9,70	-7,32
AAPL	2,38	-4,31	6,70
ON	2,37	-16,07	18,44
AMAT	2,36	1,37	0,99
URI	2,35	8,68	-6,33

Tabla A9

Contraste de Rentabilidad Proyectada vs. Real (Enero 2024) – Modelo K-Nearest Neighbors (KNN)

Stock	t+1 (%)	Retorno real (%)	Diff (%)
NVDA	6,19	21,71	-15,51
FTNT	4,66	9,70	-5,04
QCOM	4,48	2,65	1,83
REGN	4,42	7,09	-2,66
ALGN	4,12	-2,47	6,59
PHM	3,91	1,29	2,62
CPRT	3,86	-1,98	5,84
FCX	3,82	-6,64	10,46
APO	3,81	7,45	-3,65
STX	3,78	0,36	3,42
ADBE	3,52	3,49	0,03
NTAP	3,48	-0,51	4,00
HSY	3,47	3,74	-0,27
CNC	3,46	1,47	1,98
TTWO	3,44	2,44	1,00

Stock	t+1 (%)	Retorno real (%)	Diff (%)
TECH	3,40	-9,28	12,68
IDXX	3,37	-7,47	10,85
TER	3,37	-11,65	15,01
ANET	3,35	9,38	-6,03
MSFT	3,35	5,57	-2,22
SNPS	3,30	3,52	-0,22
MRVL	3,27	11,66	-8,40
AES	3,13	-13,31	16,44
A	3,13	-6,64	9,77
HUM	3,12	-19,14	22,26
CDNS	2,96	5,74	-2,78
MCK	2,92	7,67	-4,75
CRL	2,81	-8,90	11,70
CTAS	2,80	0,32	2,48
KR	2,77	0,94	1,83

Tabla A10

Contraste de Rentabilidad Proyectada vs. Real (Enero 2024) – Modelo Random Forest (RF)

Stock	t+1 (%)	Retorno real (%)	Diff (%)
AMD	4,34	12,89	-8,55
NVDA	4,24	21,71	-17,47
FTNT	3,98	9,70	-5,72
TECH	3,88	-9,28	13,16
ALGN	3,81	-2,47	6,28
SNPS	3,69	3,52	0,17
STX	3,68	0,36	3,32
PWR	3,50	-10,62	14,12
TSLA	3,49	-28,27	31,76
CDNS	3,42	5,74	-2,32
AMAT	3,12	1,37	1,75
AXON	3,09	-3,65	6,75
HUBB	3,01	2,00	1,01
QCOM	2,99	2,65	0,34
BLDR	2,86	3,99	-1,13

Stock	t+1 (%)	Retorno real (%)	Diff (%)
MSFT	2,84	5,57	-2,73
NFLX	2,73	14,72	-11,99
MRVL	2,71	11,66	-8,95
URI	2,64	8,68	-6,04
MSCI	2,55	5,67	-3,11
LOW	2,53	-3,96	6,49
ARES	2,46	2,13	0,33
EXR	2,45	-10,44	12,89
GNRC	2,43	-12,84	15,26
LII	2,42	-4,42	6,84
IDXX	2,42	-7,47	9,89
PGR	2,41	11,75	-9,35
KLAC	2,40	2,17	0,23
APO	2,38	7,45	-5,07
AES	2,38	-13,31	15,68

Tabla A11

Contraste de Rentabilidad Proyectada vs. Real (Enero 2024) – Modelo Support Vector Machine Regression (SVR)

Stock	t+1 (%)	Retorno real (%)	Diff (%)
ALGN	7,64	-2,47	10,11
NVDA	5,57	21,71	-16,13
GNRC	5,19	-12,84	18,03
SNPS	3,97	3,52	0,45
APO	3,82	7,45	-3,64
FTNT	3,68	9,70	-6,02
AMD	3,64	12,89	-9,25
LRCX	3,37	5,21	-1,84
MSCI	3,09	5,67	-2,57
ADBE	3,09	3,49	-0,40
NFLX	3,09	14,72	-11,64
ODFL	3,04	-3,59	6,63
RMD	3,03	10,05	-7,01
MRVL	2,99	11,66	-8,68
MPC	2,98	10,99	-8,01

Stock	t+1 (%)	Retorno real (%)	Diff (%)
IDXX	2,95	-7,47	10,43
AJG	2,85	3,19	-0,34
WST	2,85	5,82	-2,98
NTAP	2,81	-0,51	3,33
NVR	2,80	1,06	1,74
BLDR	2,77	3,99	-1,22
TECH	2,75	-9,28	12,03
TRGP	2,74	-1,64	4,39
CDNS	2,74	5,74	-3,00
TER	2,74	-11,65	14,38
URI	2,74	8,68	-5,94
LLY	2,72	10,21	-7,49
LII	2,66	-4,42	7,08
ARES	2,66	2,13	0,53
J	2,63	3,76	-1,13

Tabla A12

Contraste de Rentabilidad Proyectada vs. Real (Enero 2024) – Modelo Extreme Gradient Boosting (XGBoost)

Stock	t+1 (%)	Retorno real (%)	Diff (%)
AMD	4,69	12,89	-8,20
NVDA	4,39	21,71	-17,31
BLDR	2,95	3,99	-1,04
AXON	2,89	-3,65	6,55
CDNS	2,68	5,74	-3,06
DECK	2,66	12,01	-9,35
FICO	2,60	2,95	-0,35
SNPS	2,59	3,52	-0,93
KLAC	2,58	2,17	0,41
MPWR	2,57	-4,55	7,12
STX	2,57	0,36	2,20
LRCX	2,55	5,21	-2,67
TSLA	2,50	-28,27	30,77
NOW	2,49	8,01	-5,52
ON	2,49	-16,07	18,56

Stock	t+1 (%)	Retorno real (%)	Diff (%)
AVGO	2,48	5,55	-3,08
ODFL	2,44	-3,59	6,04
PWR	2,42	-10,62	13,04
CPRT	2,41	-1,98	4,38
URI	2,39	8,68	-6,28
FTNT	2,39	9,70	-7,31
AAPL	2,39	-4,31	6,70
LLY	2,39	10,21	-7,83
AMAT	2,35	1,37	0,98
ALGN	2,33	-2,47	4,79
MSFT	2,32	5,57	-3,25
MSCI	2,27	5,67	-3,39
MU	2,27	0,48	1,79
NFLX	2,24	14,72	-12,48
JBL	2,24	-1,67	3,91

Tabla A13

Evolución Histórica de las Asignaciones Trimestrales (Walk-Forward 2024-2025) – Modelo Adaptive Boosting (AdaBoost)

Trimestre de Recalibración	Stock	Peso Asignado (%)
2024-Q1 (Enero)	TECH	56,38
	MSFT	19,63
	STX	11,69
	BRO	11,43
	LOW	0,87
2024-Q2 (Abril)	XEL	38,56
	PANW	27,77
	KDP	17,69
	SMCI	12,38
	VEEV	2,72
	DECK	0,88
2024-Q3 (Julio)	APTV	44,42
	FDX	27,77
	MRK	23,02
	MKC	4,79
2024-Q4 (Octubre)	LLY	27,53
	ERIE	26,62
	BRO	13,70
	ZBRA	9,67
	FTNT	6,06
	SCHW	4,90
	DHR	4,12
	GDDY	3,72
	NVDA	2,28
	SNPS	1,40
2025-Q1 (Enero)	CHTR	29,01
	CL	25,15
	STLD	20,21
	ELV	13,54
	COR	5,97
	PWR	4,49
	AXON	1,62
2025-Q2 (Abril)	BX	56,13
	WSM	32,01
	CMI	11,86
2025-Q3 (Julio)	CMS	53,15
	MKC	12,79
	AXON	11,48
	BMY	8,24
	NFLX	4,73
	WMT	3,59
	GDDY	2,78
	NVDA	2,02
	BRO	1,20
2025-Q4 (Octubre)	TXN	56,20
	SO	16,96
	LLY	10,33
	STZ	8,15
	FTNT	5,46

	DLR	2,90

Tabla A14

Evolución Histórica de las Asignaciones Trimestrales (Walk-Forward 2024-2025) – Modelo Red Neuronal Artificial (ANN)

Trimestre de Recalibración	Stock	Peso Asignado (%)
2024-Q1 (Enero)	AJG	41,32
	LLY	21,31
	AXON	19,76
	SPG	8,94
	GNRC	6,03
	SMCI	1,43
	CDNS	1,21
2024-Q2 (Abril)	LLY	28,48
	PODD	23,49
	TECH	21,46
	DECK	9,39
	AMAT	6,33
	LITE	5,43
	MA	1,92
	SMCI	1,52
	PCG	1,36
	NVDA	0,62
2024-Q3 (Julio)	FSLR	28,84
	LLY	24,26
	EG	15,18
	STX	14,30
	TPL	6,45
	FISV	6,13
	COHR	3,59
	URI	0,66
	GE	0,58
2024-Q4 (Octubre)	LLY	30,93
	ANET	15,91
	DECK	15,67
	FIX	14,60
	GM	12,61
	BKR	7,25
	NOW	2,61
	GE	0,41
2025-Q1 (Enero)	AJG	33,43
	SO	26,33
	AXON	25,42
	FSLR	8,53
	AVGO	5,32
	ES	0,98
2025-Q2 (Abril)	BBY	42,74
	LLY	23,05
	FICO	21,68
	MCHP	9,07
	ANET	2,70
	IBKR	0,75
2025-Q3 (Julio)	AXON	35,04

	STX	32,95
	GNRC	17,92
	COST	6,23
	FSLR	4,40
	META	3,46
2025-Q4 (Octubre)	ECHO	74,16
	NOW	16,71
	AXON	4,99
	CDNS	4,13

Tabla A15

Evolución Histórica de las Asignaciones Trimestrales (Walk-Forward 2024-2025) – Modelo K-Nearest Neighbors (KNN)

Trimestre de Recalibración	Stock	Peso Asignado (%)
2024-Q1 (Enero)	HSY	41,75
	REGN	20,66
	MSFT	14,08
	CNC	5,98
	HUM	5,09
	QCOM	3,80
	NTAP	2,85
	STX	2,82
	CPRT	2,79
	AES	0,18
2024-Q2 (Abril)	CNC	20,32
	META	15,43
	PODD	14,93
	IBKR	9,60
	ANET	9,59
	ZTS	9,82
	CIEN	4,08
	NRG	3,96
	FIX	2,26
	CAT	1,37
	TPL	0,17
2024-Q3 (Julio)	ABT	24,28
	FISV	18,25
	FIX	14,81
	FSLR	8,73
	FTNT	8,18
	RMD	6,79
	TMO	6,61
	TPL	4,49
	DXCM	4,08
	LDOS	3,79
2024-Q4 (Octubre)	HUM	30,23
	DHR	16,63
	DECK	16,09
	NVR	9,97
	EME	9,52
	LULU	6,46
	IBKR	5,44
	SNPS	2,30

	AVGO	2,15
	BR	0,75
	AXON	0,44
	AMD	0,03
2025-Q1 (Enero)	PGR	43,12
	RMD	15,33
	CAT	13,40
	MPWR	12,13
	TSLA	3,88
	EW	3,85
	GEN	3,44
	DXCM	2,28
	EOG	1,38
	AXON	0,85
	MGM	0,36
2025-Q2 (Abril)	LLY	34,75
	TDY	25,48
	PODD	13,45
	CDNS	9,31
	DECK	4,87
	ANET	4,65
	AVGO	3,96
	GE	2,25
	GNRC	1,28
2025-Q3 (Julio)	WMT	33,44
	ABBV	16,33
	AXON	15,47
	DECK	13,49
	ACGL	9,00
	GNRC	7,95
	TTWO	4,33
2025-Q4 (Octubre)	FTNT	23,68
	ADM	16,34
	LDOS	14,22
	MTD	12,73
	DECK	10,23
	NVR	10,24
	MU	5,29
	SNPS	3,24
	TSLA	2,70
	ODFL	1,34

Tabla A16

Evolución Histórica de las Asignaciones Trimestrales (Walk-Forward 2024-2025) – Modelo Random Forest (RF)

Trimestre de Recalibración	Stock	Peso Asignado (%)
2024-Q1 (Enero)	PGR	32,49
	PWR	14,86
	MSFT	14,53
	SNPS	12,25
	HUBB	6,83
	FTNT	6,61
	EXR	4,73

	STX	2,96
	TECH	2,79
	AXON	1,95
2024-Q2 (Abril)	EME	16,50
	HIG	16,07
	PODD	12,71
	SNPS	9,54
	META	8,83
	KLAC	8,25
	VEEV	6,10
	AXON	5,66
	PANW	5,59
	PWR	5,47
	SMCI	3,83
	AMD	1,17
	TSLA	0,29
2024-Q3 (Julio)	LLY	35,42
	MSCI	22,27
	PWR	13,79
	MSFT	6,35
	QCOM	5,79
	FTNT	4,88
	APTV	4,66
	TPL	3,00
	STX	2,31
	FSLR	0,95
	HPQ	0,58
2024-Q4 (Octubre)	LLY	27,86
	SNPS	24,01
	VRTX	12,31
	EME	10,68
	ERIE	8,25
	RF	6,43
	DECK	5,39
	LULU	2,36
	AXON	2,12
	DHR	0,59
2025-Q1 (Enero)	PGR	27,19
	PWR	16,82
	VRTX	16,51
	LDOS	13,99
	AVGO	12,46
	AXON	6,07
	HUBB	3,02
	DE	2,09
	NVDA	1,85
2025-Q2 (Abril)	LLY	39,22
	BX	14,25
	TXN	13,74
	ANET	10,70
	PH	8,16
	META	6,33
	ODFL	5,19
	CMI	2,41
2025-Q3 (Julio)	PGR	33,05

	ABBV	16,54
	AJG	13,38
	PWR	9,87
	AXON	6,89
	MSFT	5,92
	CPRT	4,72
	STX	2,92
	TTWO	2,76
	NOW	2,24
	COST	1,72
2025-Q4 (Octubre)	TXN	37,49
	ORCL	26,84
	ECHO	12,25
	FTNT	7,75
	CDNS	4,65
	PTC	3,45
	AXON	2,57
	FICO	2,55
	NOW	1,48
	TSLA	0,97

Tabla A17

Evolución Histórica de las Asignaciones Trimestrales (Walk-Forward 2024-2025) – Modelo Support Vector Machine Regression (SVR)

Trimestre de Recalibración	Stock	Peso Asignado (%)
2024-Q1 (Enero)	AJG	37,97
	LLY	26,81
	SNPS	12,74
	ALGN	11,45
	GNRC	5,43
	RMD	5,07
	NVDA	0,53
2024-Q2 (Abril)	NVDA	18,01
	PODD	14,64
	MSCI	14,45
	ARES	12,79
	AVGO	12,17
	META	10,01
	SNPS	9,62
	JBL	7,38
	NOW	0,92
2024-Q3 (Julio)	J	24,47
	COO	17,15
	MSFT	17,03
	GPN	11,87
	PANW	8,61
	FSLR	4,42
	HUBB	4,34
	PWR	4,20
	MSCI	3,41
	NVDA	2,65
	ARES	1,87
2024-Q4 (Octubre)	LLY	32,99

	PGR	31,57
	DG	19,43
	ON	5,99
	NVDA	4,78
	FERG	3,67
	MPWR	1,55
2025-Q1 (Enero)	PGR	30,35
	UNH	25,01
	LLY	17,85
	STE	7,50
	MPWR	6,06
	PWR	5,81
	HII	3,79
	NVDA	3,25
	AXON	0,38
2025-Q2 (Abril)	ARES	29,54
	ADI	26,10
	ANET	11,58
	AVGO	10,72
	ORCL	8,91
	JBL	8,31
	DECK	3,03
	HLT	1,81
2025-Q3 (Julio)	WMT	28,84
	PGR	16,76
	AJG	14,01
	ABBV	12,76
	LLY	11,54
	GNRC	6,28
	CPRT	4,23
	NOW	2,93
	NVDA	2,65
2025-Q4 (Octubre)	LLY	25,26
	ARES	18,78
	UNH	17,33
	DE	10,32
	INTU	7,91
	NDAQ	5,50
	NVDA	5,43
	PANW	5,15
	EME	1,93
	TXN	1,32
	AVGO	1,07

Tabla A18

Evolución Histórica de las Asignaciones Trimestrales (Walk-Forward 2024-2025) – Modelo Extreme Gradient Boosting (XGBoost)

Trimestre de Recalibración	Stock	Peso Asignado (%)
2024-Q1 (Enero)	LLY	35,67
	DECK	14,31
	PWR	10,91
	ODFL	10,77
	MSFT	6,43

	CDNS	5,24
	NOW	4,72
	FICO	4,35
	NVDA	3,77
	AMD	1,94
	AXON	1,03
	STX	0,87
2024-Q2 (Abril)	LLY	37,57
	DECK	17,11
	PWR	8,66
	ODFL	7,61
	NVDA	5,64
	AMD	4,78
	CDNS	4,72
	NOW	4,04
	TPL	3,85
	PANW	3,51
	FICO	1,55
	STX	0,96
2024-Q3 (Julio)	LLY	38,75
	DECK	11,50
	NVDA	9,92
	PWR	9,76
	ARES	6,70
	TPL	5,45
	MSFT	4,24
	FTNT	3,78
	NOW	3,75
	CDNS	3,42
	FICO	2,06
	STX	0,67
2024-Q4 (Octubre)	LLY	33,10
	ANET	14,53
	PWR	9,02
	DECK	8,56
	FTNT	8,30
	APA	7,60
	NVDA	6,65
	FICO	5,04
	NOW	3,92
	MPWR	3,30
2025-Q1 (Enero)	LLY	29,12
	AXON	15,27
	AVGO	14,76
	PWR	11,01
	DECK	9,13
	NOW	8,75
	NVDA	6,71
	FIX	2,70
	FICO	2,54
2025-Q2 (Abril)	PGR	36,59
	LLY	23,46
	ARES	9,72
	PWR	8,22
	AVGO	7,84

	ANET	6,01
	NVDA	5,67
	NOW	1,51
	TKO	0,98
2025-Q3 (Julio)	LLY	30,52
	AXON	19,01
	AVGO	12,08
	PWR	12,01
	MSFT	10,14
	NVDA	8,10
	FICO	3,30
	CTAS	2,80
	EME	0,81
	NOW	0,71
	NRG	0,53
2025-Q4 (Octubre)	LLY	33,63
	TSLA	13,75
	EME	9,31
	ARES	8,33
	AVGO	7,75
	AXON	7,39
	NVDA	6,73
	STX	4,39
	SMCI	3,44
	PWR	2,94
	TKO	1,68
	CDNS	0,66