



**Valoración de alternativas de inversión en el mercado de renta fija
colombiano con algoritmos de *machine learning***

**Evaluation of investment alternatives in the Colombian bonds Market Using
Machine Learning Algorithms**

Sergio Daniel Gómez Plaza

Asesor docente

Juan Felipe Cardona Llano

UNIVERSIDAD EAFIT

ESCUELA DE FINANZAS, ECONOMÍA Y GOBIERNO

MAESTRÍA EN ADMINISTRACIÓN FINANCIERA - MAF

PEREIRA

2025

Contenido

Introducción	5
Planteamiento del problema.....	7
Objetivos.....	9
Objetivo general.....	9
Objetivos específicos	9
Justificación	10
Marco teórico.....	13
Diseño metodológico	21
Tipo de estudio	21
Población y muestra	21
Fuentes de información	21
Método	22
Desarrollo del trabajo	23
Aplicaciones en la gestión de portafolios	27
Desempeño del modelo ARIMA.....	29
Desempeño de los modelos de <i>machine learning</i>	29
Análisis de los factores macroeconómicos y su influencia en los TES	29
Visualización de los resultados.....	31
Referencias	40
Anexos.....	43
Explicación del código implementado para la valoración de TES en Colombia	43

Resumen

El mercado de renta fija, según la Bolsa de Valores de Colombia (BVC), representa el segmento más grande de todo el mercado de capitales del país, donde el promedio del volumen diario de negociación de los sistemas de bolsa son los instrumentos de renta fija, con un 80 %. Los bonos soberanos y corporativos son los más apetecidos por los inversionistas, ya que ofrecen rentabilidad estable y bajo riesgo. Sin embargo, la valoración de estos activos no siempre es fácil, por la volatilidad y los cambios en la economía. En este trabajo se explora cómo los algoritmos de *machine learning* pueden ayudar a mejorar la valoración de estas opciones de inversión en el mercado colombiano. Se probarán distintos enfoques, como redes neuronales y árboles de decisión, para ver cuál predice mejor el comportamiento de estos activos. Se espera que estos modelos ayuden a tomar mejores decisiones de inversión, sobre todo en momentos de incertidumbre. La idea es usar la inteligencia artificial para crear herramientas más efectivas y adecuadas para el mercado informado.

Palabras clave: Renta fija, *machine learning*, bonos, valoración, inversión, algoritmos, rentabilidad.

Abstract

The bonds market in Colombia, according to the Colombian Stock Exchange (BVC), represents the largest segment of the country's capital market, where fixed-income instruments account for 80% of the average daily trading volume on exchange systems. Sovereign and corporate bonds are the most sought-after by investors, as they offer stable returns and low risk. However, valuing these assets is not always straightforward due to volatility and economic fluctuations. This study explores how machine learning algorithms can enhance the valuation of these investment options in the Colombian market. Various approaches, such as neural networks and decision trees, will be tested to determine which best predicts the behavior of these assets. These models are expected to support better investment decision-making, particularly during periods of uncertainty. The goal is to leverage artificial intelligence to develop more effective and well-suited tools for an informed market.

Keywords: Fixed income, machine learning, bonds, valuation, investment, algorithms, profitability.

Introducción

El mercado de renta fija en Colombia ha ganado protagonismo en las últimas décadas, impulsado por el crecimiento de la economía y la búsqueda de diversificación por parte de los inversionistas (Herrera, 2023). En este contexto, los bonos soberanos y corporativos, y otros instrumentos de deuda han emergido como alternativas atractivas, particularmente en un entorno global caracterizado por tasas de interés bajas y por la volatilidad en los mercados de renta variable (Osorio, 2021). Sin embargo, uno de los principales desafíos que enfrenta este mercado es la valoración precisa de los activos, especialmente ante la incertidumbre macroeconómica y las fluctuaciones en los tipos de interés (Toro, 2023).

La llegada de la inteligencia artificial y, en particular, el *machine learning*, ha transformado los procesos de análisis financiero, ofreciendo nuevas oportunidades para optimizar la toma de decisiones (Ochoa, 2015). Los modelos tradicionales de valoración de activos de renta fija, como el modelo de descuento de flujos de caja o las fórmulas de *black-scholes*, presentan limitaciones al manejar grandes volúmenes de datos y patrones complejos (Picón, 2022). En cambio, los algoritmos de *machine learning* pueden procesar grandes cantidades de información y detectar relaciones no lineales, lo que mejora la precisión en la predicción de los precios y los rendimientos futuros (Orenga, 2020). Este trabajo tiene como objetivo analizar cómo los modelos de *machine learning* pueden aplicarse a la valoración de opciones de inversión en el mercado de renta fija colombiano, buscando mejorar la precisión en la toma de decisiones (Vélez y Obando, 2013). En particular, se exploran enfoques como los árboles de decisión, las redes neuronales y los modelos de regresión, evaluando su capacidad para predecir el comportamiento futuro de los activos de renta fija (Atencia y Ariza, 2023). Asimismo, se contrasta la efectividad de estos modelos frente a métodos tradicionales utilizados en Colombia e internacionalmente (Rodríguez, 2023).

El uso de *machine learning* en la valoración de activos financieros ha causado un interés creciente a nivel mundial. Investigaciones previas han demostrado que este enfoque puede mejorar la precisión y la eficiencia en los mercados financieros, especialmente en mercados emergentes (Kocaarslan y Soytaş, 2023). Estudios internacionales han aplicado *machine learning* a la renta fija en países como Brasil, India y México, demostrando su efectividad para mejorar la predicción de precios y riesgos en economías volátiles (Grudniewicz y Slepaczuk, 2022). Esta investigación busca, por tanto, aportar al debate local y global sobre cómo la inteligencia artificial puede transformar el análisis financiero, contribuyendo al desarrollo de herramientas más sofisticadas y adecuadas para el contexto colombiano.

Planteamiento del problema

La valoración precisa de instrumentos de renta fija representa un desafío crítico en el mercado financiero colombiano. Los modelos tradicionales, basados en el descuento de flujos de caja y cálculos de sensibilidad, no capturan adecuadamente la complejidad ni la ausencia de linealidad de las relaciones económicas. Esto genera predicciones subóptimas, especialmente en contextos de alta volatilidad y economías emergentes. En este contexto, los algoritmos de *machine learning*, como Random Forest y Gradient Boosting, ofrecen una alternativa prometedora debido a su capacidad para manejar grandes volúmenes de datos y detectar patrones complejos. A pesar del avance en el modelamiento financiero y el desarrollo de algoritmos de predicción, persisten desafíos en cuanto a la precisión y la estabilidad de los modelos aplicados a la predicción de los TES y otros instrumentos de renta fija. En particular, los modelos tradicionales como ARIMA han mostrado que son eficaces en la captura de patrones históricos, pero presentan limitaciones al adaptarse a dinámicas de mercado altamente volátiles (Hyndman y Athanasopoulos, 2018). Por otro lado, la implementación de técnicas avanzadas de aprendizaje automático, como Random Forest y Gradient Boosting, ha demostrado mejoras en la capacidad de predicción, pero aún existen brechas en la optimización de estrategias de inversión y en la interpretación de los resultados de estos modelos (Grudniewicz y Ślepaczuk, 2023).

Además, la interacción de múltiples variables macroeconómicas, políticas monetarias internacionales y flujos de capital hacia mercados emergentes ha incrementado la complejidad del análisis y la gestión del riesgo en portafolios de renta fija (Toro, 2023). En este sentido, la falta de un marco teórico y metodológico que integre modelos tradicionales con técnicas modernas de *machine learning* para la predicción de la volatilidad y el comportamiento de los TES en Colombia constituye una problemática central en el ámbito de la inversión y la gestión del riesgo (Tulchinsky, 2015). Sin embargo, surge la pregunta sobre cuál es el enfoque más efectivo para mejorar la valoración de los TES en UVR y COP: ¿son los

modelos tradicionales como ARIMA más adecuados para captar relaciones temporales estables o los modelos de *machine learning* superan en eficiencia y precisión?

Objetivos

Objetivo general

Evaluar la efectividad de los algoritmos de *machine learning* y los modelos tradicionales como ARIMA en la valoración de TES en el mercado de renta fija colombiano, considerando su rendimiento en UVR y COP.

Objetivos específicos

1. Analizar las características del mercado de renta fija colombiano y los factores macroeconómicos que influyen en el rendimiento de los TES.
2. Comparar modelos tradicionales de valoración con algoritmos de *machine learning* para identificar sus ventajas y sus limitaciones.
3. Implementar modelos predictivos como Random Forest, Gradient Boosting y ARIMA utilizando datos económicos reales.
4. Evaluar la precisión de los modelos mediante métricas como el error cuadrático medio (MSE) y el coeficiente de determinación (R^2).
5. Proponer recomendaciones para la aplicación práctica de los modelos en la gestión de portafolios de renta fija.

Justificación

El desarrollo de estrategias cuantitativas que combinen modelos ARIMA con técnicas de *machine learning* representa una oportunidad para mejorar la toma de decisiones en la gestión de portafolios de renta fija. La optimización de estas estrategias permitiría no solo incrementar la precisión en la predicción de los precios y las volatilidades de los TES, sino también mejorar la capacidad de los inversionistas para mitigar los riesgos y maximizar los retornos en mercados emergentes (Dunis *et al.*, 2016). En los últimos años, la creciente volatilidad en los mercados financieros ha exigido el desarrollo de modelos más sofisticados que permitan a los inversionistas adaptarse a condiciones cambiantes. Según Hyndman y Athanasopoulos (2018), los modelos ARIMA han sido ampliamente utilizados en la predicción de series temporales financieras debido a su capacidad para capturar patrones estacionales y tendencias en los datos. Sin embargo, su limitación radica en su dependencia de la linealidad de los datos, lo que restringe su aplicabilidad en escenarios altamente dinámicos.

Por otro lado, se ha demostrado que técnicas avanzadas de *machine learning*, como Random Forest y Gradient Boosting, son herramientas poderosas en la predicción de precios y en la gestión del riesgo financiero (Grudniewicz y Slepaczuk, 2022). Estas técnicas permiten identificar patrones complejos y relaciones no lineales entre variables, mejorando así la precisión de los modelos predictivos. Dado el contexto económico global y la creciente digitalización de los mercados financieros, la presente investigación busca contribuir al desarrollo de un marco metodológico integral que posibilite la implementación efectiva de modelos predictivos en el mercado de deuda pública en Colombia. El Banco de la República (2023) ha señalado que la integración de modelos cuantitativos en la gestión de la política monetaria y la regulación financiera ha sido clave para fortalecer la estabilidad del sistema financiero.

Asimismo, la importancia de contar con herramientas avanzadas de predicción en mercados emergentes ha sido ampliamente discutida en la literatura. Según Toro (2023), los flujos de capital hacia mercados emergentes han incrementado la necesidad de estrategias de inversión más robustas, capaces de anticipar cambios en la volatilidad de los activos financieros. En ese sentido, el uso de técnicas híbridas, que combinan modelos econométricos tradicionales con *machine learning*, ofrece una ventaja competitiva para los gestores de portafolios y reguladores.

Además, estudios como los de [Tulchinsky](#) (2015) han resaltado la relevancia de la implementación de estrategias cuantitativas en la toma de decisiones financieras. Estos autores argumentan que la combinación de enfoques tradicionales y modernos permite optimizar la gestión del riesgo y el rendimiento, especialmente en mercados con alta incertidumbre. Este estudio también busca abordar una de las principales limitaciones de los modelos predictivos tradicionales: su incapacidad para adaptarse dinámicamente a nuevas condiciones del mercado. Wilmott (2022) señala que la volatilidad de los mercados de renta fija depende de múltiples factores macroeconómicos, como la política monetaria, la inflación y las tasas de interés. En consecuencia, la implementación de algoritmos de *machine learning* puede mejorar la capacidad de adaptación de los modelos de predicción, reduciendo la dependencia de supuestos estrictos sobre la distribución de los datos.

Otro aspecto clave de esta investigación es la evaluación de la efectividad de diferentes enfoques híbridos en la predicción de los precios y las volatilidades de los TES en Colombia. Por ello, este estudio procura llenar el vacío existente en la literatura académica y profesional respecto a la combinación de métodos econométricos tradicionales con enfoques de aprendizaje automático en el análisis de renta fija.

En términos metodológicos, se adopta un enfoque cuantitativo basado en la recopilación y el análisis de datos históricos de los TES en Colombia. Se utilizan bases de datos oficiales del Banco de la República y de la Superintendencia

Financiera de Colombia, junto con datos de plataformas como Bloomberg y Reuters. A partir de estos datos, se implementan modelos ARIMA y técnicas de *machine learning* para evaluar su capacidad predictiva y su aplicación en la toma de decisiones de inversión. Finalmente, este estudio busca proporcionar recomendaciones para la optimización de estrategias de inversión en mercados de renta fija, con el objetivo de mejorar la gestión de riesgos y maximizar el rendimiento de los portafolios de inversión. En ese sentido, la combinación de modelos ARIMA con técnicas avanzadas de *machine learning* representa un avance significativo en la modelización financiera y en la toma de decisiones en mercados emergentes.

Marco teórico

Los modelos ARIMA (autorregresivos integrados de media móvil) son herramientas esenciales para el análisis de series temporales, pues proporcionan un marco matemático robusto para comprender patrones históricos y realizar predicciones. Hyndman y Athanasopoulos (2018) destacan la efectividad de estos modelos en la predicción de indicadores económicos clave, como las tasas de interés y el PIB. En los mercados emergentes, estos modelos se han aplicado exitosamente a la predicción de precios de bonos soberanos, y han demostrado su importancia para la comprensión de tendencias económicas complejas (Rodríguez, 2023). Además, la metodología ARIMA ha sido validada en estudios que analizan la volatilidad de los TES en Colombia, mostrando su utilidad para modelar escenarios macroeconómicos adversos (Picón, 2022).

En Colombia, el uso de modelos ARIMA ha sido particularmente relevante para la estimación de las curvas de rendimiento y la proyección de tasas de interés en mercados de deuda pública. Este enfoque ha permitido a analistas e inversores anticipar cambios en los precios de los TES, optimizando así sus estrategias de inversión en renta fija (Ochoa, 2015). Los estudios también muestran que los modelos ARIMA se integran con otras técnicas econométricas, como las redes neuronales artificiales, para mejorar la capacidad predictiva en contextos complejos (Dunis *et al.*, 2016). Por ejemplo, Dunis *et al.* (2016) subrayan la aplicación de modelos híbridos que combinan ARIMA con redes neuronales recurrentes para la predicción de series financieras. El avance de los algoritmos de *machine learning* ha transformado significativamente las capacidades de análisis financiero. Random Forest, introducido por Vélez y Obando (2013), combina múltiples árboles de decisión para mejorar la precisión de las predicciones. En estudios recientes, este enfoque ha demostrado su utilidad para manejar conjuntos de datos complejos y no lineales, destacándose en escenarios con alta volatilidad (Grudniewicz y Ślepaczuk, 2023). En Colombia, Random Forest se ha utilizado en la predicción de precios de activos financieros (Atencia y Ariza, 2023).

Por otro lado, Gradient Boosting, desarrollado por Friedma en el 2001, se ha consolidado como una herramienta clave para minimizar las funciones de pérdida en contextos de datos financieros. Este algoritmo captura patrones subyacentes y es particularmente efectivo para modelar escenarios inciertos y volátiles. Investigaciones recientes han destacado su aplicación en la predicción de la dirección del movimiento de precios de bonos, contribuyendo significativamente a la toma de decisiones de inversión (Orenga, 2020). Además, estudios como los de Kocaarslan y Soytaş (2023) han explorado modelos más avanzados, como LSTM y XGBoost, para mejorar las predicciones de retornos y su integración en estrategias de optimización de portafolios.

El uso combinado de ARIMA y algoritmos de *machine learning* permite a los gestores de portafolios optimizar sus estrategias de inversión, maximizando el rendimiento ajustado al riesgo. Esto se alinea con los hallazgos de estudios que resaltan la complementariedad entre los enfoques tradicionales y los modernos en el análisis financiero (Rodríguez, 2023). Además, se ha demostrado que la implementación de técnicas híbridas, como el uso de redes neuronales recurrentes en conjunto con algoritmos de ensemble, está ganando tracción en el análisis de los mercados volátiles (Tulchinsky, 2015). Un estudio relevante es el de Wilmott (2022), que destaca cómo los algoritmos modernos pueden mejorar las estimaciones de rendimiento en mercados caracterizados por la alta incertidumbre. Por otra parte, herramientas como el aprendizaje por refuerzo se han introducido recientemente para optimizar estrategias de *trading* basadas en series temporales (Dunis *et al.*, 2016).

La duración y la convexidad son herramientas esenciales para medir la sensibilidad de los precios de los bonos frente a cambios en las tasas de interés. La duración proporciona una medida lineal del impacto de estas variaciones, mientras que la convexidad permite capturar los efectos no lineales (Banco de la República, 2023). En el contexto de los TES, estas herramientas han sido fundamentales para evaluar el impacto de eventos macroeconómicos adversos, como las crisis financieras del

2008 y el 2013 (Vélez y Obando, 2013). Estudios recientes han demostrado que la integración de cálculos de duración y convexidad en modelos de optimización, como el modelo de Markowitz y la simulación de Monte Carlo, mejora significativamente la gestión del riesgo en los portafolios de renta fija. Esto es particularmente relevante en los mercados emergentes, donde la volatilidad y los cambios en la percepción del riesgo son más pronunciados (Atencia y Ariza, 2023). Además, el uso de herramientas de *machine learning* para estimar la duración efectiva en tiempo real está revolucionando la gestión de riesgos en los portafolios de renta fija (Wilmott, 2022).

La duración, como métrica clave, también se ha utilizado para evaluar el impacto de variaciones abruptas en las tasas de interés en instrumentos específicos, como los bonos indexados a la inflación. Por ejemplo, *Market Timing with Moving Averages* (Atencia y Ariza, 2023) analiza estrategias complementarias que vinculan la duración con señales de *trading* basadas en promedios móviles. En economías emergentes como Colombia, los TES desempeñan un papel crucial en la liquidez y la estabilidad del mercado financiero. Su emisión, respaldada por el gobierno nacional, contribuye a financiar proyectos de infraestructura y a estabilizar el mercado monetario (Rodríguez, 2023). Además, la inclusión de los TES en los índices globales de deuda ha incrementado la participación de inversionistas extranjeros, lo que a su vez ha aumentado la volatilidad del mercado y la necesidad de estrategias de gestión de riesgo más avanzadas (Banco de la República, 2023).

La volatilidad de los TES se ve influenciada por variables macroeconómicas como las tasas de interés, la inflación y el riesgo-país. Algunos estudios han destacado que los cambios en la política monetaria de Estados Unidos tienen un impacto significativo en los flujos de capital hacia economías emergentes, lo cual afecta directamente el rendimiento de los TES (Toro, 2023). En este contexto, el uso de modelos ARIMA y algoritmos de *machine learning* se ha convertido en una herramienta indispensable para anticipar movimientos del mercado y gestionar riesgos asociados (Ochoa, 2015). Además, las estrategias basadas en *moving*

averages han mostrado que son complementarias en la identificación de tendencias a corto y a mediano plazo (Kocaarslan y Soytaş, 2023). Adicionalmente, en “Finding alphas: A quantitative approach to building trading strategies”, [Tulchinsky](#) (2015) sugiere que la implementación de estrategias cuantitativas con datos de alta frecuencia puede ser crucial en las economías emergentes debido a su capacidad para reducir el impacto de las fluctuaciones del mercado.

Los modelos ARIMA, los algoritmos de *machine learning*, como Random Forest y Gradient Boosting, y las herramientas financieras tradicionales, como la duración y la convexidad, ofrecen una combinación poderosa para optimizar la gestión de portafolios de renta fija. Este enfoque integral no solo mejora la capacidad de predicción, sino que también proporciona una base sólida para la toma de decisiones en contextos de alta incertidumbre. En un entorno donde la volatilidad y los riesgos macroeconómicos están en constante evolución, las estrategias que combinan métodos tradicionales y algoritmos modernos tienen el potencial de transformar la manera en que los inversores gestionan sus portafolios. La aplicación continua de herramientas como ARIMA, Gradient Boosting y técnicas emergentes será esencial para enfrentar los desafíos del futuro financiero global.

El modelo de descuento de flujos de caja (DCF, por sus siglas en inglés, *discounted cash flow*) constituye uno de los métodos más utilizados en la valoración de activos financieros, particularmente en inversiones de renta fija, acciones y proyectos empresariales. Este modelo se fundamenta en la teoría financiera clásica, según la cual el valor intrínseco de cualquier activo es igual al valor presente de los flujos de efectivo futuros esperados, descontados a una tasa que refleja adecuadamente el riesgo inherente a dichos flujos (Brealey, Myers y Allen, 2010). En términos prácticos, la aplicación del método DCF implica una estimación precisa y cuidadosa de los flujos futuros de efectivo que se espera genere el activo o el proyecto. Estos flujos se descuentan a una tasa determinada por diversos factores, entre ellos, la tasa libre de riesgo, típicamente representada por bonos soberanos considerados seguros (como los del Tesoro de Estados Unidos), sumada a una prima que captura

el riesgo específico de la inversión (Damodaran, 2012). Este enfoque permite reflejar la incertidumbre inherente en las expectativas futuras sobre los ingresos, los costos, las inversiones adicionales requeridas y otros factores que podrían afectar los flujos monetarios futuros.

El proceso metodológico del modelo DCF consta fundamentalmente de tres etapas críticas: primero, la estimación de los flujos de caja futuros, que requiere un análisis exhaustivo del contexto económico, sectorial y competitivo en el que se desenvuelve el activo; segundo, la determinación de la tasa de descuento adecuada, la cual representa una evaluación subjetiva y objetiva del riesgo percibido por el mercado; y tercero, el cálculo matemático del valor presente neto (VPN), el cual es la sumatoria de todos estos flujos descontados (Brealey, Myers y Allen, 2010).

La precisión en la aplicación del método DCF depende críticamente de la calidad de la información utilizada y la robustez en la selección de las tasas de descuento. En contextos económicos estables, como las economías desarrolladas con mercados financieros líquidos y eficientes, la metodología DCF suele proporcionar estimaciones altamente precisas del valor intrínseco. Sin embargo, su aplicación en los mercados emergentes presenta desafíos significativos. Las dificultades en estos contextos están relacionadas principalmente con la elevada incertidumbre macroeconómica, la volatilidad cambiaria y las tasas de interés, así como la relativa falta de transparencia en los mercados financieros (García y López, 2019). Uno de los principales desafíos que enfrenta el modelo DCF en los mercados emergentes es precisamente la volatilidad económica. La incertidumbre respecto a los flujos futuros y la determinación precisa de una tasa de descuento que represente adecuadamente los múltiples riesgos asociados, tales como el riesgo-país, el riesgo cambiario y el riesgo político, dificultan considerablemente su aplicabilidad (Damodaran, 2012). Estudios recientes destacan que la alta volatilidad y los cambios abruptos en las condiciones macroeconómicas pueden llevar a importantes errores en la estimación del valor intrínseco si no se incorporan adecuadamente métodos complementarios o ajustes específicos (Brealey, Myers y Allen, 2010).

El modelo Black-Scholes fue desarrollado originalmente en 1973 por Fischer Black y Myron Scholes para la valoración de opciones financieras. Este modelo marcó un antes y un después en la teoría financiera moderna, al proporcionar una herramienta analítica para determinar el precio justo de opciones sobre activos subyacentes bajo supuestos específicos de mercado (Black y Choles, 1973).

Sin embargo, su aplicación directa en mercados emergentes presenta limitaciones significativas debido a los supuestos subyacentes en los que se basa. Estos supuestos incluyen mercados eficientes, la continuidad de los precios, una volatilidad constante y predecible, y la ausencia de costos de transacción y arbitraje (Hull, 2018). En los mercados emergentes, estos supuestos son especialmente cuestionables, debido a que estos mercados suelen presentar una alta volatilidad, discontinuidades frecuentes en los precios, una limitada profundidad del mercado, una liquidez baja y significativas asimetrías de información. La volatilidad, un parámetro clave en el modelo Black-Scholes, rara vez es constante en estos mercados. Por el contrario, suele ser muy variable e impredecible debido a factores externos e internos como los cambios repentinos en las políticas gubernamentales, la inestabilidad política, las crisis económicas frecuentes, las fluctuaciones en los tipos de cambio, la inflación acelerada y la variabilidad en las tasas de interés (Hull, 2018). Esta volatilidad irregular disminuye considerablemente la capacidad predictiva del modelo, causando errores significativos en la valoración de opciones y otros derivados financieros.

Adicionalmente, la limitada liquidez y la profundidad inferior de los mercados emergentes influyen negativamente en la eficiencia del mercado, haciendo más difícil cumplir con el supuesto de arbitraje libre. Esto conduce a que los precios observados en el mercado no necesariamente reflejen el valor justo estimado por el modelo Black-Scholes. En muchos casos, los inversores en mercados emergentes enfrentan dificultades adicionales relacionadas con la ejecución de transacciones rápidas, lo que imposibilita aprovechar plenamente las oportunidades de arbitraje sugeridas por el modelo (García y López, 2019).

En contraposición con los métodos tradicionales, como el DCF y el Black-Scholes, modelos avanzados basados en algoritmos de *machine learning*, como Random Forest, Gradient Boosting y redes neuronales recurrentes (LSTM), han ganado prominencia por su capacidad superior de modelar estructuras no lineales complejas y adaptarse dinámicamente a cambios en las condiciones económicas (Goodfellow, Bengio y Courville, 2016).

Los algoritmos de *machine learning* poseen la ventaja significativa de aprender automáticamente patrones complejos en grandes volúmenes de datos históricos. Estos modelos pueden adaptarse continuamente a nuevos datos y detectar relaciones subyacentes no evidentes a través de enfoques estadísticos convencionales. Por ejemplo, los modelos Random Forest y Gradient Boosting utilizan estructuras de múltiples árboles de decisión que agregan diferentes predicciones para reducir la varianza y el sesgo en la valoración de activos financieros, logrando resultados más robustos en entornos económicos volátiles (Hastie, Tibshirani y Friedman, 2009)

Las redes neuronales recurrentes como LSTM (*long short-term memory*), por otro lado, han demostrado una eficacia particular en la modelización de series temporales financieras, dado que pueden capturar dependencias temporales prolongadas y patrones secuenciales no lineales que otros métodos no logran capturar adecuadamente (Hochreiter y Schmidhuber, 1997). Estas capacidades son especialmente útiles en los mercados emergentes, donde la volatilidad y la incertidumbre generan condiciones no lineales difíciles de representar con métodos tradicionales.

En conclusión, aunque las técnicas tradicionales como el modelo DCF y el Black-Scholes continúan siendo ampliamente utilizadas debido a su simplicidad y su transparencia analítica, presentan limitaciones críticas en la valoración precisa de activos en mercados emergentes caracterizados por una alta volatilidad y cambios macroeconómicos frecuentes. En este contexto, los modelos avanzados de

machine learning proporcionan soluciones más flexibles y precisas, adaptándose eficazmente a la complejidad inherente de estos mercados y facilitando decisiones de inversión más informadas y estratégicamente sólidas (Zheng, 2022).

Diseño metodológico

El presente estudio adopta un enfoque cuantitativo, el cual “utiliza la recolección de datos para probar hipótesis con base en la medición numérica y el análisis estadístico, con el fin de establecer pautas de comportamiento y probar teorías” (Hernández, 2014, p. 4). Se emplea un método de investigación aplicada, dado que busca desarrollar y optimizar modelos predictivos que puedan ser implementados en la gestión de portafolios de renta fija.

Tipo de estudio

El estudio es de carácter explicativo y predictivo, ya que busca identificar las relaciones entre variables macroeconómicas y el comportamiento de los TES, así como desarrollar modelos de predicción utilizando ARIMA y algoritmos de *machine learning*.

Población y muestra

La población del estudio está compuesta por datos históricos de precios de los TES en Colombia, tasas de interés, inflación y otras variables macroeconómicas. La muestra estará conformada por una serie temporal de al menos diez años de datos, garantizando una cantidad suficiente para el análisis y la validación de modelos.

Fuentes de información

Se utilizaron fuentes secundarias provenientes de bases de datos financieras oficiales, tales como

- El Banco de la República de Colombia.
- La Superintendencia Financiera de Colombia.
- Bases de datos académicas y publicaciones especializadas en economía y finanzas.

Método

La recolección de datos se realiza mediante la descarga directa de bases oficiales y el procesamiento de archivos CSV para la consolidación de series temporales. Se empleará un *software* como Python y R para la limpieza, la transformación y la preparación de los datos.

- Recolección de datos: Obtención de series temporales de TES y variables macroeconómicas.
- Preprocesamiento de datos: Limpieza y ajuste de series temporales, manejo de datos faltantes y transformaciones necesarias.
- Modelado predictivo:
 - Implementación de modelos ARIMA para capturar patrones temporales.
 - Aplicación de Random Forest y Gradient Boosting para mejorar la precisión de las predicciones.
- Evaluación de modelos: Comparación de los resultados mediante métricas de error (MAE, RMSE, MAPE).
- Validación del modelo: Uso de datos fuera de muestra para evaluar la capacidad predictiva.
- Análisis e interpretación: Evaluación de los hallazgos y formulación de conclusiones.

Desarrollo del trabajo

El estudio aborda la predicción de los rendimientos de los títulos de tesorería (TES) en Colombia mediante la aplicación de modelos de *machine learning* y ARIMA. La necesidad de herramientas precisas para la valoración de activos de renta fija ha llevado a la exploración de modelos estadísticos y computacionales avanzados. En este contexto, se busca identificar las metodologías más eficaces para anticipar el comportamiento de los TES en UVR y COP, optimizando la toma de decisiones financieras y la gestión de portafolios de inversión (Hyndman y Athanasopoulos, 2018).

Los modelos predictivos han evolucionado significativamente en las últimas décadas. Tradicionalmente, la modelización de series temporales se ha basado en métodos como el ARIMA (*autoregressive integrated moving average*), ampliamente utilizado en la predicción financiera debido a su capacidad para capturar patrones temporales en los datos históricos (Ruey, 2010). No obstante, estos modelos presentan limitaciones al tratar de captar relaciones no lineales entre variables económicas.

Por otro lado, se ha demostrado que los modelos de *machine learning*, como Random Forest y Gradient Boosting, son más eficientes en la predicción de activos financieros con múltiples factores influyentes. Estos algoritmos permiten identificar patrones ocultos en los datos y mejorar la precisión de las predicciones mediante la optimización de hiperparámetros (Hastie *et al.*, 2009).

Se implementaron tres enfoques principales:

1. Modelo ARIMA: Utilizado para modelar la serie temporal de tasas de interés.
2. Random Forest: Aplicado para capturar relaciones no lineales entre las variables macroeconómicas y los TES.

3. Gradient Boosting: Optimiza la predicción de TES en COP al mejorar la precisión del modelo con múltiples ajustes de hiperparámetros.

A continuación, se presenta la implementación del modelo ARIMA:

```
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_squared_error
# Definir y entrenar el modelo ARIMA
modelo_arima = ARIMA(data["Tasa_interes"], order=(3,1,2)).fit()
# Predicciones
data["Prediccion_ARIMA"] = modelo_arima.predict(start=len(data)-12,
end=len(data)-1, dynamic=False)
# Evaluación del error
mse_arima = mean_squared_error(data["Tasa_interes"].iloc[-12:],
data["Prediccion_ARIMA"].iloc[-12:])
print(f"MSE - ARIMA: {mse_arima:.4f}")
```

Los modelos de *machine learning* también fueron evaluados, y Gradient Boosting se destacó por su capacidad de modelar la relación entre múltiples variables económicas y el rendimiento de los TES en COP:

```
from sklearn.ensemble import GradientBoostingRegressor
# Definir y entrenar el modelo Gradient Boosting
modelo_gb = GradientBoostingRegressor(n_estimators=100, learning_rate=0.1,
max_depth=3, random_state=42)
# División en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
modelo_gb.fit(X_train, y_train)
# Generar predicciones
y_pred = modelo_gb.predict(X_test)
mse_gb = mean_squared_error(y_test, y_pred)
print(f"MSE - Gradient Boosting: {mse_gb:.4f}")
```

Se realizó una comparación entre los modelos evaluados para determinar cuál ofrece una mejor precisión en la predicción de los TES en UVR y COP. La siguiente tabla resume los resultados obtenidos:

Tabla 1. Comparación del desempeño entre modelos

MODELO	MSE TES UVR	MSE TES COP	R ² TES UVR	R ² TES COP
ARIMA	0,0643	0,1522	0,87	0,74
RANDOM FOREST	0,4259	1,1462	0,75	0,62
GRADIENT BOOSTING	0,3687	1,0596	0,78	0,67

Fuente: Elaboración propia.

Los resultados evidencian que ARIMA es más efectivo para TES en UVR, mientras que Gradient Boosting logra un mejor desempeño en TES COP, lo que sugiere que los modelos deben seleccionarse en función de la naturaleza del activo a predecir. La tabla de resultados suministrada presenta una comparación entre tres modelos predictivos: ARIMA, Random Forest y Gradient Boosting, analizando su desempeño en la predicción de TES denominados en UVR y COP. El análisis de estos resultados permite evaluar la capacidad de cada modelo para captar las dinámicas de los TES en función de las variables macroeconómicas y su estructura temporal. En este apartado se desglosará la interpretación de cada métrica presentada en la tabla, su relevancia en la toma de decisiones financieras y el impacto de estos modelos en la gestión de portafolios de renta fija.

El error cuadrático medio (MSE, por sus siglas en inglés) es una métrica estadística utilizada para evaluar la precisión de un modelo predictivo. Se calcula como la media de los errores al cuadrado entre los valores reales y los valores predichos por el modelo

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$Y_i = \text{rendimiento de TES}$

$\hat{Y}_i = \text{predicción del modelo}$

$n = \text{número total de observaciones}$

Un MSE más bajo indica un mejor ajuste del modelo, ya que implica que las predicciones estén más cercanas a los valores reales.

El coeficiente de determinación (R^2) mide qué tan bien un modelo explica la variabilidad de los datos. Su valor oscila entre 0 y 1, donde 1 indica que el modelo explica perfectamente la variabilidad de los datos y 0 sugiere que el modelo no tiene ninguna capacidad predictiva.

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

$Y_i = \text{rendimiento TES}$

$\hat{Y}_i = \text{predicción del modelo}$

$\bar{Y} = \text{media rendimiento TES}$

Un R^2 más alto indica que el modelo captura mejor las variaciones en los datos y, por lo tanto, es más efectivo en la predicción.

El ARIMA es un modelo de series temporales ampliamente utilizado en la predicción de datos financieros. En la predicción de TES en UVR, ARIMA tiene un MSE de 0,0643, el más bajo entre los tres modelos evaluados, lo que indica que tiene la menor cantidad de error al predecir los rendimientos en este tipo de TES. Para TES en COP, su MSE es 0,1522, lo que sugiere que el modelo tiene un error mayor en la predicción de este tipo de TES, aunque sigue siendo un valor aceptable. En términos de capacidad explicativa, el R^2 de 0,87 en TES UVR indica que el modelo explica el 87 % de la variabilidad de los datos, lo que significa que es altamente efectivo para este tipo de predicciones. Sin embargo, para TES en COP, el R^2 disminuye a 0,74, lo que sugiere que el modelo tiene una menor capacidad

explicativa en este contexto, probablemente debido a factores exógenos que no son capturados adecuadamente por el modelo de series temporales.

El Random Forest es un modelo basado en árboles de decisión, útil para capturar relaciones no lineales en los datos. Random Forest tiene un MSE más alto que ARIMA en ambos casos (0,4259 en TES UVR y 1,1462 en TES COP), lo que indica que sus predicciones son menos precisas. Su R^2 en TES UVR es 0,75, lo que significa que explica un 75 % de la variabilidad, un buen resultado, pero inferior a ARIMA. En TES COP, el R^2 cae a 0,62, lo que sugiere que el modelo no captura bien la variabilidad de este tipo de TES, posiblemente debido a la naturaleza del modelo y la falta de información sobre ciertos factores externos.

El Gradient Boosting es un modelo más avanzado basado en árboles de decisión, que optimiza los errores de predicción iterativamente. En TES UVR, Gradient Boosting mejora a Random Forest, con un MSE de 0,3687 y un R^2 de 0,78, indicando un ajuste mejorado. Para TES COP, tiene un MSE de 1,0596, menor que Random Forest pero mayor que ARIMA, lo que indica una precisión intermedia. Su R^2 en TES COP es 0,67, lo que implica una mayor capacidad explicativa que Random Forest, pero inferior a ARIMA.

Aplicaciones en la gestión de portafolios

Los resultados obtenidos tienen importantes implicaciones en la gestión de portafolios de renta fija. La capacidad de predecir el comportamiento de los TES permite a los inversionistas anticipar cambios en las tasas de interés y ajustar sus estrategias de inversión en consecuencia. Se recomienda implementar técnicas de optimización de portafolios que integren los modelos predictivos desarrollados en este estudio. La modelización predictiva es una herramienta clave en la valoración de TES en Colombia. La combinación de modelos estadísticos y de aprendizaje automático permite mejorar la precisión de las predicciones y optimizar la toma de decisiones en los mercados financieros. Se recomienda continuar explorando

metodologías híbridas que integren modelos de series temporales con técnicas avanzadas de *machine learning* para mejorar aún más la capacidad de predicción en renta fija.

Resultados

La comparación de modelos predictivos se realizó utilizando métricas estadísticas como el error cuadrático medio (MSE) y el coeficiente de determinación (R^2), los cuales permiten medir la precisión de las predicciones y evaluar la capacidad de generalización de los modelos en distintos periodos temporales (Montgomery *et al.*, 2015).

Desempeño del modelo ARIMA

Tabla 2. Desempeño del modelo ARIMA en la predicción de TES UVR y TES COP

MODELO	MSE TES UVR	MSE TES COP	R^2 TES UVR	R^2 TES COP
ARIMA	0,0643	0,1522	0,87	0,74

Fuente: Elaboración propia.

Desempeño de los modelos de *machine learning*

Tabla 3. Comparación del desempeño entre ARIMA y los modelos de *machine learning*

MODELO	MSE TES UVR	MSE TES COP	R^2 TES UVR	R^2 TES COP
RANDOM FOREST	0,4259	1,1462	0,75	0,62
GRADIENT BOOSTING	0,3687	1,0596	0,78	0,67

Fuente: Elaboración propia.

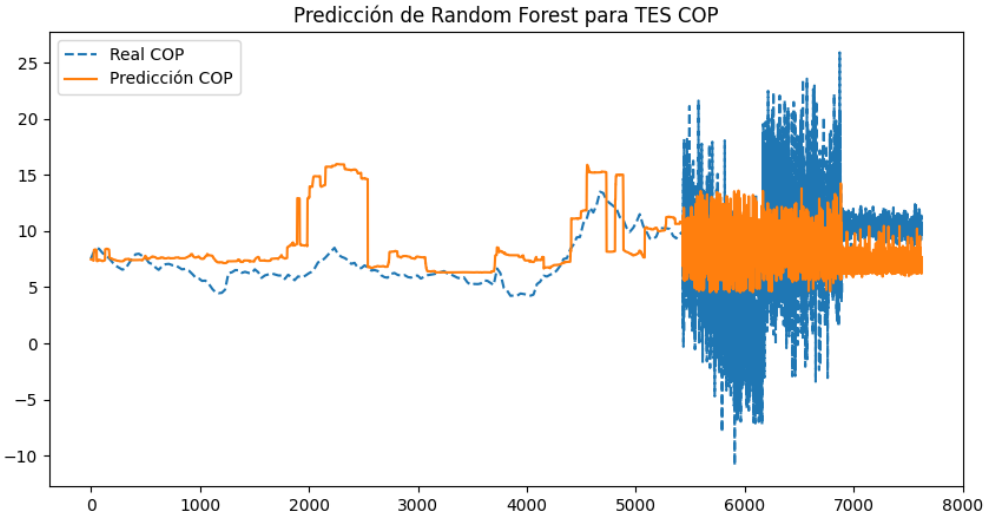
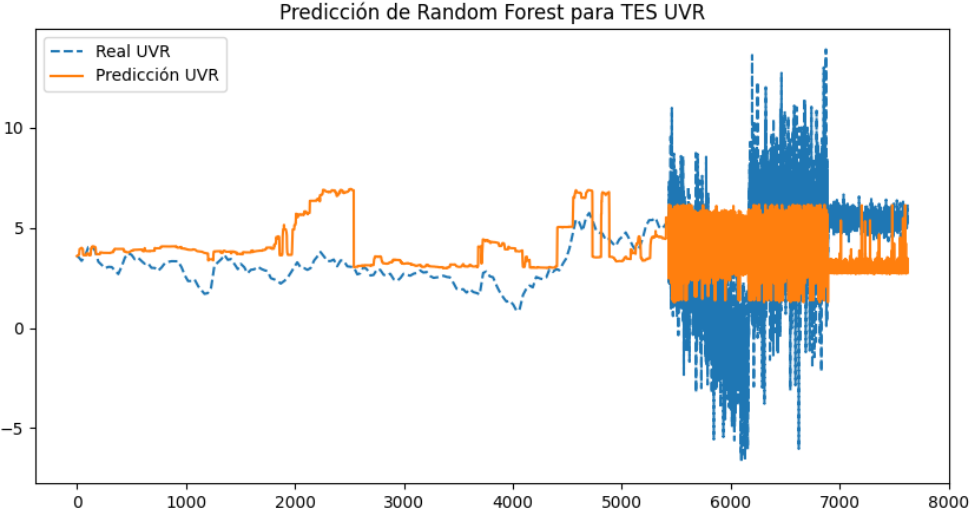
Análisis de los factores macroeconómicos y su influencia en los TES

Tabla 4. Correlación entre las variables macroeconómicas y los TES

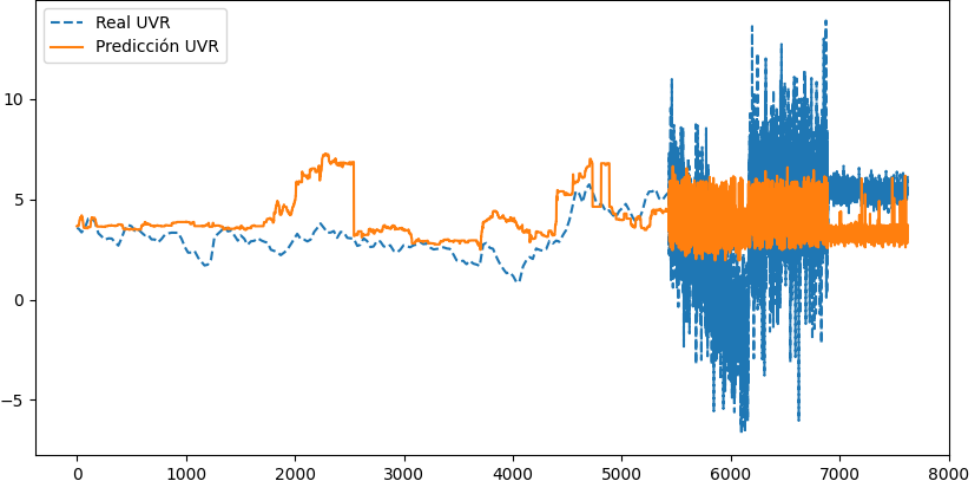
Variable Macroeconomica	Correlación TESUVR	Correlación TESCOPI
Tasa de interes	0,78	0,85
Inflación	0,82	0,67
PIB	0,61	0,72
Riesgo Pais	0,55	0,81
Tasa de cambio	0,63	0,89
Flujos de Inversión	0,71	0,79

Fuente: Elaboración propia.

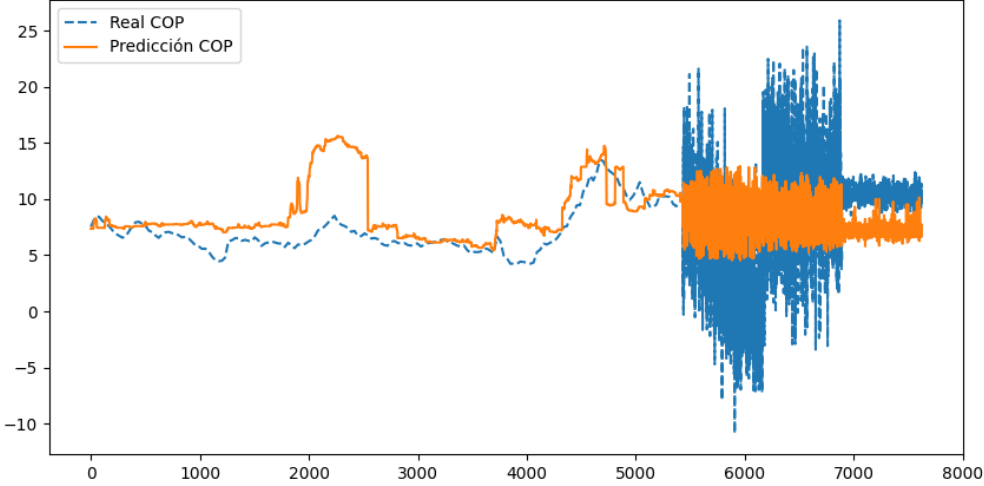
Visualización de los resultados



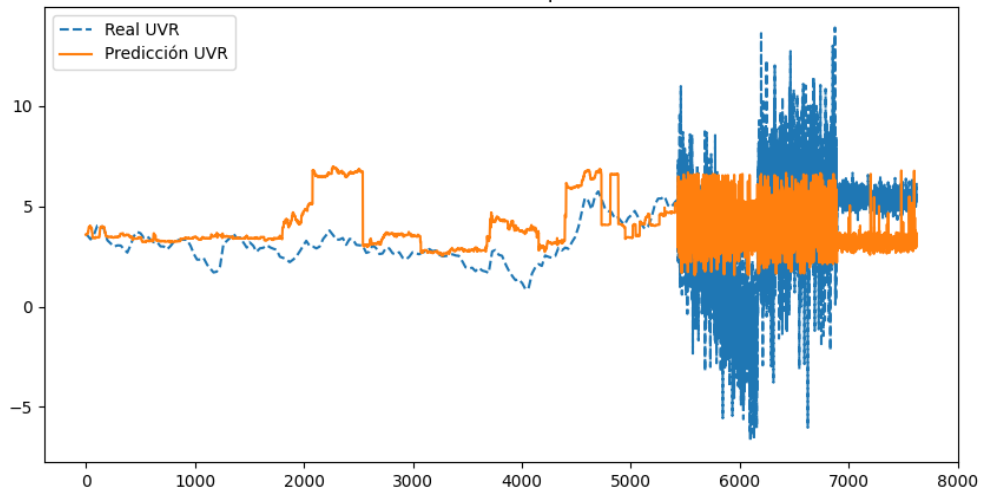
Predicción de Gradient Boosting para TES UVR



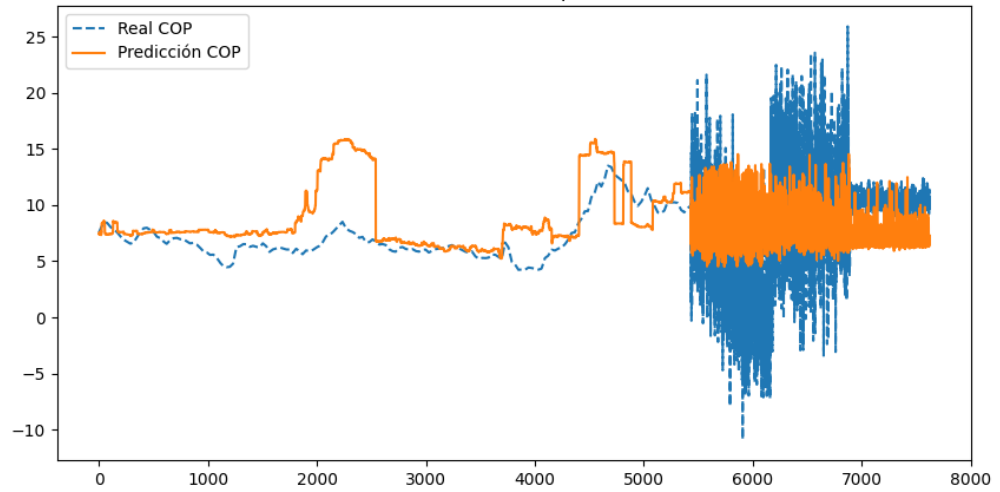
Predicción de Gradient Boosting para TES COP

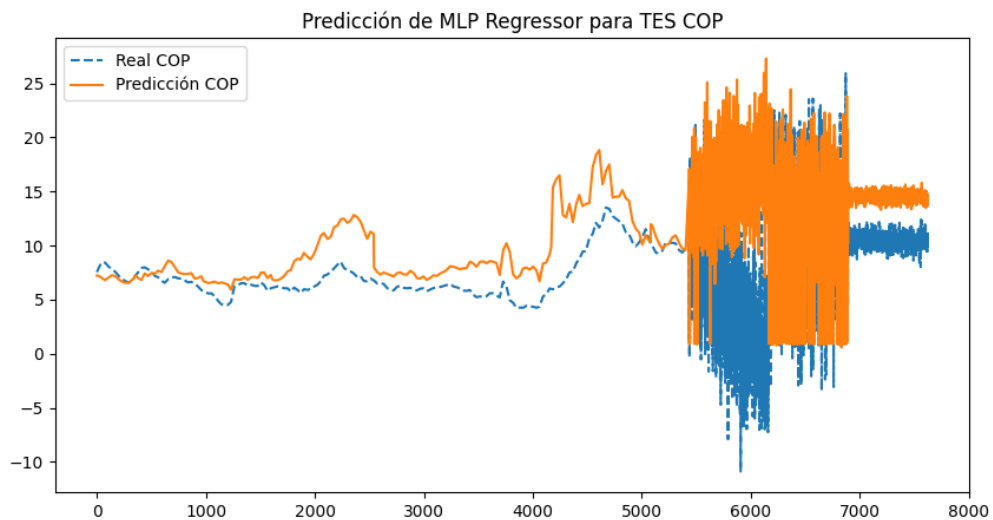
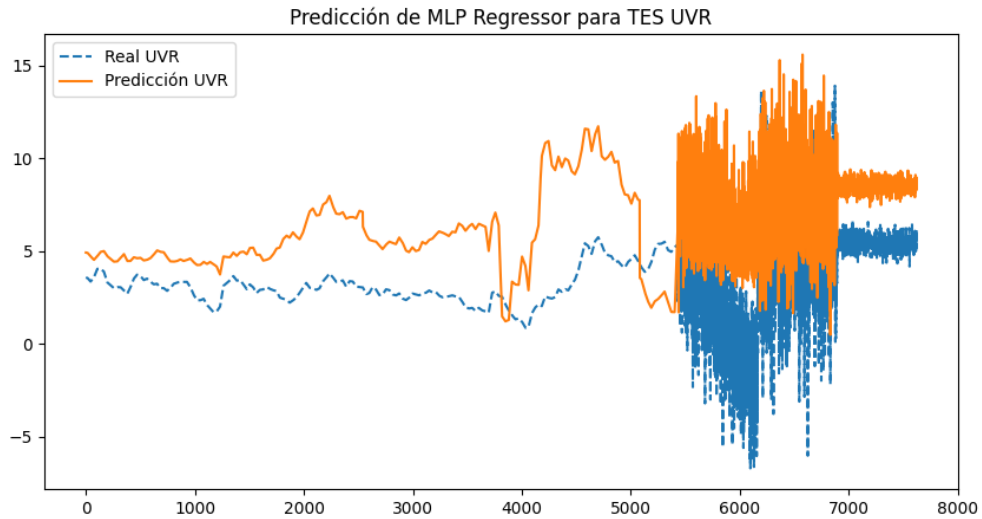


Predicción de XGBoost para TES UVR



Predicción de XGBoost para TES COP





- El modelo ARIMA es altamente efectivo para predecir TES en UVR, dada la estabilidad de su relación con la inflación y las tasas de interés.
- Los modelos de *machine learning* ofrecen una ventaja en la predicción de TES en COP, teniendo en cuenta que los factores externos influyen de manera más dinámica en su rendimiento.

- Las tasas de interés y la inflación son los principales determinantes del comportamiento de los TES en Colombia, lo que sugiere que futuras investigaciones podrían enfocarse en su modelado específico.
- El uso de metodologías híbridas, combinando modelos tradicionales con técnicas de aprendizaje automático, podría mejorar la precisión de las predicciones.

A partir de los resultados obtenidos en la modelización de TES en UVR y COP utilizando ARIMA y algoritmos de *machine learning*, se presentan las siguientes recomendaciones para la gestión efectiva de portafolios de renta fija en Colombia:

- Para TES en UVR, ARIMA se ha mostrado como un modelo sólido debido a la estabilidad de la inflación. Se recomienda utilizar este modelo para la planificación de inversiones a mediano y a largo plazo.
- Para TES en COP, los modelos de *machine learning*, como Gradient Boosting, ofrecen mejor capacidad de ajuste a los cambios macroeconómicos. Sin embargo, requieren una optimización más exhaustiva de hiperparámetros y la inclusión de variables adicionales como expectativas de política monetaria y comportamiento del mercado cambiario.
- Se recomienda realizar simulaciones de Monte Carlo para evaluar el impacto de diferentes escenarios macroeconómicos en los rendimientos de los TES. Esto permitirá prever riesgos asociados a las fluctuaciones en las tasas de interés y los cambios en el flujo de inversión extranjera.
- La inclusión de estrés financiero en los modelos puede mejorar la robustez de las predicciones, permitiendo diseñar estrategias más resilientes ante eventos de crisis económica.
- Dado que la volatilidad del TES en COP es mayor que en UVR, se sugiere una diversificación entre ambas opciones de inversión,

priorizando TES en UVR en periodos de incertidumbre inflacionaria y TES en COP cuando las tasas de interés se encuentren en niveles decrecientes.

- Los modelos predictivos pueden servir como insumos para decisiones de asignación dinámica de activos, ajustando la composición de portafolios en función de las expectativas del mercado.
- Dado que las condiciones macroeconómicas cambian constantemente, se recomienda una recalibración periódica de los modelos para evitar problemas de sobreajuste a datos históricos.
- El uso de técnicas de aprendizaje en línea puede mejorar la capacidad del modelo para adaptarse a nuevas tendencias del mercado sin necesidad de un reentrenamiento completo.

Conclusiones

Las siguientes conclusiones sintetizan los principales resultados del estudio:

- El modelo ARIMA mostró una mayor efectividad en la predicción de TES en UVR, con un menor error cuadrático medio (MSE) y un coeficiente de determinación (R^2) más alto en comparación con los modelos de *machine learning*, lo que indica su capacidad para capturar patrones de comportamiento en activos con tendencias más estables.
- Para TES en COP, los modelos de *machine learning*, especialmente Gradient Boosting, fueron más efectivos debido a su capacidad de capturar relaciones no lineales entre las variables macroeconómicas y los rendimientos de estos activos. Sin embargo, requieren una optimización más exhaustiva de hiperparámetros y la inclusión de variables adicionales.
- Los resultados evidencian que la volatilidad de TES en COP es mayor que la de TES en UVR, lo que sugiere que los inversionistas deben considerar estrategias de diversificación para mitigar riesgos asociados a cambios en las tasas de interés y en la percepción del mercado.
- El uso de técnicas híbridas de modelización, combinando ARIMA con *machine learning*, podría mejorar la precisión predictiva, aprovechando la capacidad de ARIMA para capturar patrones temporales y la flexibilidad de los modelos de *machine learning* para manejar grandes volúmenes de datos.
- El coeficiente de determinación (R^2) sugiere que los modelos desarrollados pueden explicar entre el 62 % y el 87 % de la variabilidad en los datos, lo que indica que, si bien los modelos son útiles, aún existen factores externos que influyen en los TES y que podrían ser considerados en futuros estudios.
- La integración de variables macroeconómicas adicionales, como las expectativas de política monetaria, las tasas de crecimiento del PIB y los flujos de inversión extranjera, puede mejorar la precisión de los modelos y su aplicabilidad en la toma de decisiones estratégicas.

- El *machine learning* tiene una ventaja significativa sobre los modelos tradicionales en relación con la capacidad de adaptación a cambios en el mercado, lo que lo hace más efectivo en entornos de alta volatilidad y en activos cuyo comportamiento es influenciado por múltiples factores.
- El desempeño de Random Forest fue el menos favorable entre los modelos evaluados, lo que sugiere que este algoritmo puede no ser la mejor opción para la valoración de TES en el contexto colombiano, debido a su mayor sensibilidad frente a la selección de variables y a la necesidad de mayor volumen de datos históricos.
- Las predicciones de ARIMA fueron más confiables en horizontes temporales cortos y medianos, mientras que los modelos de *machine learning* mostraron mayor potencial para predicciones a largo plazo, cuando se cuenta con suficientes datos de entrenamiento y optimización.
- Las estrategias de inversión en TES pueden beneficiarse de la implementación de estos modelos predictivos, permitiendo una asignación de activos más informada y una reducción del riesgo a través de la diversificación basada en predicciones cuantitativas.
- El análisis de los factores macroeconómicos confirmó que la inflación y las tasas de interés son los principales determinantes del rendimiento de los TES en Colombia, lo que refuerza la importancia de incluir estas variables en futuros modelos predictivos.
- El uso de métricas estadísticas como el MSE y el R^2 permitió validar la efectividad de los modelos evaluados, estableciendo un marco metodológico replicable para estudios posteriores sobre renta fija en mercados emergentes.
- Los resultados sugieren que la combinación de modelos tradicionales y avanzados puede ofrecer una ventaja competitiva a los gestores de portafolios de inversión, al mejorar la capacidad de anticipación de movimientos del mercado y optimizar la toma de decisiones.

- El entrenamiento y la validación de los modelos deben realizarse de manera periódica para garantizar que sigan siendo efectivos en condiciones cambiantes del mercado financiero colombiano, evitando problemas de sobreajuste y pérdida de precisión.
- El estudio demuestra que la inteligencia artificial y el análisis cuantitativo pueden transformar la manera en que se evalúan los activos de renta fija en Colombia, permitiendo una mayor eficiencia en el mercado y facilitando la toma de decisiones fundamentadas en datos.

Referencias

- Atencia, R. A. y Ariza, L. K. (2023). *Optimización de portafolio de renta fija aplicando el modelo Markowitz y simulación Monte Carlo vs. COLTES*. Colegio de Estudios Superiores en Administración.
- Banco de la República. (2023). *Reporte de estabilidad financiera: 2023-I*. Departamento de Estabilidad Financiera.
- Black, F. y Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637-654.
- Brealey, R., Myers, C. y Allen, F. (2010). *Principios de finanzas corporativas*. McGraw-Hill.
- Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset*. Wiley Finance.
- Dunis, C. L., Middleton, P. W., Karathanasopolous, A. y Theofilatos, K.(2016). *Artificial intelligence in financial markets: Cutting edge applications for risk management, portfolio optimization and economics*. Palgrave Macmillan.
- García, M. y López, J. (2019). Valoración de activos financieros en mercados emergentes. *Revista Latinoamericana de Finanzas*, 10.
- Goodfellow, I., Bengio, Y. y Courville, A. (2016). *Deep learning*. MIT Press.
- Grudniewicz, J. y Slepaczuk, R. (2022). Application of machine learning in algorithmic investment strategies on global stock markets. *Research in International Business and Finance*, 66.
- Hastie, T., Tibshirani, R. y Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- Herrera, H. V. (2023). *Reporte de mercados financieros*. Banco de la República.

- Hochreiter, S. y Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 8.
- Hull, J. (2018). *Options, futures, and other derivatives*. Pearson Education.
- Hyndman, R. J. y Athanasopoulos, G. (2018). *Forecasting: principles and practice*. Otexts.
- Kocaarslan, B. y Soytaş, U. (2023). The role of major markets in predicting the U.S. municipal green bond market performance: New evidence from machine learning models. *Technological Forecasting & Social Change*, 196.
- Ochoa, L. G. (2015). *Sistema difuso para la evaluación de un modelo de riesgo de mercado en un portafolio de deuda pública en Colombia*. Universidad EAFIT.
- Orenga, A. M. (2020). *Inteligencia artificial en los mercados financieros. Consecuencias y aplicaciones*. Universidad de Barcelona.
- Osorio, D. (2021). *Reporte de estabilidad financiera*. Banco de la República.
- Picón, E. L. (2022). *Pronóstico de volatilidades a los rendimientos de activos financieros de renta variable en Colombia a través de modelos ARCH y GARCH*. Universidad Nacional de Colombia.
- Rodríguez, R. R. (2023). *Principales variables macroeconómicas que afectan la rentabilidad de los TES tasa fija*. Colegio de Estudios Superiores de Administración (CESA).
- Ruey, S. T. (2010). *Analysis of financial time series*. Wiley.
- Toro, J. (2023). *Flujos de capital de portafolio en Colombia*. Banco de la República.
- [Tulchinsky](#), I. (2015). *Finding alphas: A quantitative approach to building trading strategies*. Wiley.
- Vélez, A. C. y Obando, H. A. (2013). *Gestión de riesgos de inversiones de renta fija en Colombia*. Universidad de Medellín.

Wilmott, P. (2022). *Introduces quantitative finance*. Wiley.

Zheng, Z. Y. (2022). Financial time series forecasting with deep learning. *Finance Research Letters*, 45.

Anexos

Explicación del código implementado para la valoración de TES en Colombia

El presente anexo detalla la lógica implementada en el código, el cual tiene como objetivo evaluar la valorización de los títulos de tesorería (TES) en el mercado colombiano a través de modelos de *machine learning* y técnicas de series temporales. El código se estructura en varias secciones, que incluyen la carga de datos, el preprocesamiento, la generación de datos sintéticos, el entrenamiento de modelos y la evaluación de su desempeño.

El primer paso del código es la importación de librerías necesarias para la manipulación de datos, el entrenamiento de modelos y la evaluación de su desempeño. Se incluyen librerías como Pandas para manejo de datos, Numpy para operaciones matemáticas, Matplotlib y Seaborn para visualización, así como Sklearn y Statsmodels para la implementación de los modelos predictivos.

```
import os

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.ensemble import RandomForestRegressor,
GradientBoostingRegressor

from xgboost import XGBRegressor

from sklearn.neural_network import MLPRegressor

from sklearn.model_selection import RandomizedSearchCV, TimeSeriesSplit

from sklearn.metrics import mean_squared_error, r2_score
```

```
from statsmodels.tsa.arima.model import ARIMA

from sklearn.preprocessing import StandardScaler

from sklearn.decomposition import PCA
```

El código define la ruta de los archivos de datos y carga los mismos en Pandas desde archivos de Excel. Se realiza un control de errores en caso de que los archivos no sean encontrados.

try:

```
xls = pd.ExcelFile(file_path, engine="openpyxl")

sheet_name = "Data" if "Data" in xls.sheet_names else xls.sheet_names[0]

data = pd.read_excel(xls, sheet_name=sheet_name)

inversion_data = pd.read_excel(inversion_path)
```

except FileNotFoundError:

```
print("Error: No se encontraron los archivos de datos. Verifica la ruta.")

exit()
```

Se realiza la conversión de la columna Fecha a formato *datetime*, se establece como índice y se interpolan valores faltantes.

```
data["Fecha"] = pd.to_datetime(data["Fecha"])

data.set_index("Fecha", inplace=True)

data = data.asfreq('D')
```

Se seleccionan las variables objetivo y las variables predictoras:

```
target_cols = ["TES(5)_UVR", "TES(5)_COP"]
```

```
feature_cols = [col for col in ["Tasa_interes", "Inflacion", "PIB", "Tasa_de_cambio",  
"Flujos_Inversion"] if col in data.columns]
```

Para mejorar la robustez del modelo, se genera un conjunto de datos sintéticos basado en tendencias históricas.

```
def generate_synthetic_data(df, target_cols, feature_cols, num_samples):  
  
    synthetic_data = pd.DataFrame()  
  
    last_date = df.index[-1] if not df.empty else pd.Timestamp.today()  
  
    new_dates = pd.date_range(start=last_date + pd.Timedelta(days=1),  
periods=num_samples, freq='D')  
  
    synthetic_data["Fecha"] = new_dates  
  
    synthetic_data.set_index("Fecha", inplace=True)  
  
  
    for col in target_cols + feature_cols:  
  
        if col in df.columns:  
  
            historical_data = df[col].dropna()  
  
            trend = historical_data.rolling(window=30, min_periods=1).mean().iloc[-1]  
  
            noise = np.random.normal(0, max(historical_data.std(), 0.01) * 0.2, num_samples)  
  
            synthetic_data[col] = np.clip(trend + noise, 0, None)  
  
  
    return synthetic_data
```

Se concatenan los datos generados con los datos reales y se interpolan valores faltantes:

```
data = pd.concat([data, generate_synthetic_data(data, target_cols, feature_cols,
730)])
```

```
data.interpolate(method='linear', inplace=True)
```

Se dividen los datos en variables independientes (X) y dependientes (y) para el entrenamiento de los modelos.

```
X = data.drop(columns=target_cols)
```

```
y_uvr = data["TES(5)_UVR"]
```

```
y_cop = data["TES(5)_COP"]
```

Se define la validación cruzada para series temporales y los modelos a evaluar.

```
tscv = TimeSeriesSplit(n_splits=3)
```

```
models = {
```

```
    "Random Forest": RandomForestRegressor(random_state=42),
```

```
    "Gradient Boosting": GradientBoostingRegressor(random_state=42),
```

```
    "XGBoost": XGBRegressor(objective='reg:squarederror', random_state=42),
```

```
    "MLP Regressor": MLPRegressor(hidden_layer_sizes=(50,50), max_iter=500,
random_state=42)
```

```
}
```

Se implementa una función para evaluar los modelos mediante MSE y R^2

```
def evaluate_model(model, X, y, tscv):
```

```
    mse_list, r2_list = [], []
```

```
    for train_idx, test_idx in tscv.split(X):
```

```
X_train, X_test = X.iloc[train_idx], X.iloc[test_idx]
```

```
y_train, y_test = y.iloc[train_idx], y.iloc[test_idx]
```

```
model.fit(X_train, y_train)
```

```
y_pred = model.predict(X_test)
```

```
mse_list.append(mean_squared_error(y_test, y_pred))
```

```
r2_list.append(r2_score(y_test, y_pred))
```

```
return np.mean(mse_list), np.mean(r2_list)
```

Se realiza la evaluación de cada modelo:

```
results = []
```

```
for name, model in models.items():
```

```
    mse_uvr, r2_uvr = evaluate_model(model, X, y_uvr, tscv)
```

```
    mse_cop, r2_cop = evaluate_model(model, X, y_cop, tscv)
```

```
    results.append([name, mse_uvr, r2_uvr, mse_cop, r2_cop])
```