



Vigilada Mineducación

**OPTIMIZACIÓN DE UN PORTAFOLIO DE INVERSIÓN CON ACCIONES DEL
COLCAP APLICANDO TÉCNICAS DE *MACHINE LEARNING***

*Optimization of an investment portfolio with Colcap shares applying machine learning
techniques*

Maribel Osorio Buitrón

Juan David Rico

Trabajo de grado

Asesor, docente

Bryan Ricardo Rojas Ormaza MSc

UNIVERSIDAD EAFIT

ESCUELA DE ECONOMÍA Y FINANZAS

MAESTRÍA EN ADMINISTRACIÓN FINANCIERA - MAF

CALI

2024

Resumen

A partir de los incrementos en los volúmenes de información histórica de las acciones de empresas públicas, surge el interrogante sobre si es posible conformar portafolios mejor optimizados que los generados a partir de la teoría tradicional, haciendo uso de las recientes innovaciones en inteligencia artificial y análisis de datos de la última década. Por esta razón, la presente investigación pretende comparar la teoría tradicional de optimización de portafolio con las recientes metodologías de análisis de datos aplicadas al Colcap. La metodología utilizada se basa en doce investigaciones previas las cuales han testeado y demostrado el rendimiento de diferentes modelos de machine learning en bolsas de todo el mundo, tales como S&P500, NASDAQ, DAX, SET y Colcap. De aquí se seleccionaron los mejores prospectos y se aplicaron a las acciones del Colcap para predecir el movimiento futuro en el precio de la acción a partir del comportamiento histórico de ciertas variables significativas para después ser comparados con la metodología tradicional. Se encontró que el mejor modelo de predicción en el movimiento del precio aplicado al Colcap es el *random forest*, y que las variables que mejor explican los cambios futuros en el precio de las acciones de esta bolsa son el precio de cierre de la acción, el índice TRM y el índice Colcap. Además, los modelos de *machine learning* lograron optimizar portafolios con menor número de acciones y mayores rentabilidades respaldadas en la información histórica.

Palabras clave: bolsa de valores, inteligencia artificial, análisis de datos, KNIME

Abstract

Based on the increases in the volumes of historical information on the shares of public companies, the question arises as to whether it is possible to create better optimized portfolios than those generated from traditional theory, making use of recent innovations in artificial intelligence and analysis of data from the last decade. For this reason, the present research aims to compare the traditional theory of portfolio optimization with the recent data analysis methodologies applied to Colcap. The methodology used is based on twelve previous investigations which had tested and demonstrated the performance of different machine learning models on stock exchanges around the world, such as S&P500, NASDAQ, DAX, SET and Colcap. From here, the best prospects were selected and applied to Colcap shares, to predict the future movement in the share price based on the historical behavior of certain significant variables and then compared with the traditional methodology. It was found that the best prediction model in the price movement applied to Colcap is the Random Forest, and the variables that best explain the future changes in the price of the shares of this exchange are the closing price of the share, the TRM index and the Colcap index. In addition, machine learning models managed to optimize portfolios with a smaller number of shares and higher returns backed by historical information.

Keywords: stock market, artificial intelligence, data analytics, KNIME

Agradecimientos

Agradecida con Dios, por darme la bendición de seguirme formando; a mi mamá y a mi hermana, por siempre estar conmigo en mis procesos. A mi compañero, Juan David, por su inteligencia, admiración y paciencia. Gracias.

Maribel Osorio Buitrón

A mi papá, por preguntarme incansablemente: “¿Cómo va con la tesis?”; a mi mamá, por alentarme a terminar este trabajo, y a mi *fiancé*, por acompañarme durante todo este proceso. A Maribel, infinitas gracias por permitirme trabajar y aprender con ella durante estos tres años de academia.

Juan David Rico Villarreal

Tabla de contenido

Introducción	10
1. Situación en estudio – problema	13
1.2. Pregunta de investigación	14
1.3. Objetivo general	14
1.4. Objetivos específicos	14
2. Marco de referencia conceptual	15
2.1. Teoría portafolio	15
2.2. Método Markowitz.....	15
2.3. Riesgos financieros	16
2.4. Colcap.....	17
2.5. <i>Machine learning</i>	18
2.6. Etiquetas.....	18
2.7. Tipos de aprendizajes	18
2.8. Árboles de decisión (DT)	19
2.9. <i>Random forest</i> (RF).....	21
2.10. Regresión logística (R-LOG).....	22
2.11. Indicador Kappa	23
3. Metodología.....	25
3.1. Antecedentes	25
3.2. Conjunto de datos	35
3.3. Procesamiento de datos	37
3.4. Lenguaje de programación y librerías	37

3.5. Métodos de predicción	37
3.6. Variable dependiente	38
3.7. Aplicación de los métodos seleccionados	39
3.8. Evaluación de los métodos seleccionados	40
3.9. Teoría de portafolio	42
3.10. Retornos por activo	44
3.11. Supuestos	44
4. Resultados	45
4.1. Análisis descriptivo.....	45
4.2. <i>Machine learning</i>	48
4.3. Estrategia.....	50
4.3.1. Elección de acciones.....	50
4.3.2. Acción Ecopetrol	52
4.3.3. Acción Bancolombia.....	56
4.3.4. Acción Celsia	59
4.4. Señales de compra y venta	62
4.5. Teoría de portafolio	64
5. Conclusiones	68
5.1. Inteligencia artificial, modelación, simulación	68
5.2. Rentabilidad	69
6. Recomendaciones	71
Referencias bibliográficas	73
Anexos.....	80

Anexo 1 Estadísticas después de la simulación de los tres modelos <i>machine learning</i> ,	80
Anexo 2 Comportamiento del precio de la acción de Ecopetrol en el período 2018- 2022	81
Anexo 3 Comportamiento del precio de la acción de Celsia en el período 2018-2022	82
Anexo 4 Comportamiento del precio de la acción de Bancolombia en el período 2018-2022	83

Lista de tablas

Tabla 1 Matriz de antecedentes	26
Tabla 2 Ejemplificación de una variable dependiente	38
Tabla 3 Período de análisis para entrenar y evaluar la predicción de la dirección del precio	40
Tabla 4 Estrategia para portafolio	51
Tabla 5 Desempeño de los modelos de <i>machine learning</i> con la acción de Ecopetrol	54
Tabla 6 Desempeño de los modelos de <i>machine learning</i> con la acción de Bancolombia	57
Tabla 7 Desempeño de los modelos de <i>machine learning</i> con la acción de Celsia	60
Tabla 8 Porcentaje de precisión balanceada de los modelos de <i>machine learning</i>	62
Tabla 9 Predicción y decisión de posición para las acciones analizadas	63
Tabla 10 Volatilidad media por sector Colcap	64
Tabla 11 Portafolios modelados	65

Lista de figuras

Figura 1 Algunas contribuciones a la teoría de carteras en las últimas décadas	12
Figura 2 Ejemplo de un árbol de decisión	20
Figura 3 Ejemplificación del modelo <i>random forest</i>	22
Figura 4 Resultados de una regresión logística	23
Figura 5 Métodos de aprendizaje automático hallados en la literatura en el campo financiero	34
Figura 6 Matriz de confusión	41
Figura 7 Esquema metodológico propuesto	45
Figura 8 Composición por sectores económicos de las acciones estudiadas	45
Figura 9 Rendimientos de las acciones durante el período estudiado (2018-2022)	46
Figura 10 Comportamiento histórico retornos de los índices Colcap y S&P500	47
Figura 11 Comportamiento de la variable de respuesta sin predicción	49
Figura 12 Aplicación de inteligencia artificial a la teoría de portafolio	50
Figura 13 Retorno por acción de la estrategia <i>machine learning</i> en el período 2018-2022	52
Figura 14 Retornos de Ecopetrol, Colcap y TRM en el período 2018-2022	56
Figura 15 Retornos de Bancolombia, Colcap y TRM en el período 2018-2022	58
Figura 16 Retornos de Celsia, Colcap y TRM en el período 2018-2022	61
Figura 17 Conformación portafolios optimizados Colcap en el período 2018-2022	66
Figura 18 Diagrama de los portafolios obtenidos y frontera eficiente de una de las iteraciones estocásticas	67

Introducción

Nos encontramos en una actualidad en la cual, la inteligencia artificial ha adquirido tal capacidad de automatizar actividades, tomar decisiones y resolver problemas de aprendizaje en una ínfima fracción de tiempo del que un ser humano común y corriente requeriría para realizar la misma tarea.

Este fenómeno mundial se ha proliferado a diferentes ramas de aplicación como las decisiones de inversión, en bolsas de valores que manejan altos volúmenes de negociación en países como Estados Unidos y Reino Unido, quienes figuran en la literatura como pioneros en investigaciones acerca del tema. Es a raíz de este panorama que se ha formulado este trabajo de investigación con el objetivo de establecer las herramientas de inteligencia artificial que permitan el análisis de grandes cantidades de información en el mercado accionario colombiano y comparar sus resultados en materia de rentabilidad y riesgo con la teoría de portafolio tradicional.

La teoría y la práctica financiera han sido profundamente influenciadas por el enfoque tradicional para la selección de portafolios de inversión creado por Markowitz (1952), enfoque bajo el cual se supone un agente racional que desea minimizar el riesgo sujeto a un nivel de retorno mínimo esperado, o maximizar la rentabilidad sujeto a un máximo de riesgo deseado.

Posteriormente, aparecen modelos basados en factores como el de CAPM (Sharpe, 1964), que utiliza un único factor Beta que relaciona la rentabilidad del activo respecto

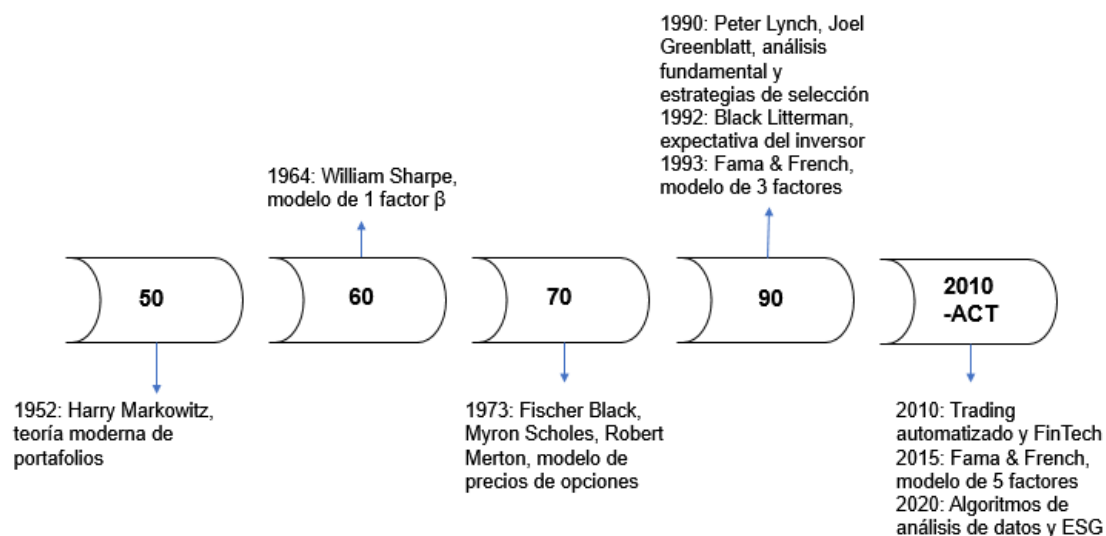
al mercado. Por otro lado, Fama y French (1993) introducen un modelo de tres factores en el que, además del Beta, tienen en cuenta el tamaño y el crecimiento, mientras que el modelo Black-Litterman permite incluir las expectativas del inversor y la confianza sobre determinado activo (Black & Litterman, 1992). Años más tarde, Fama y French (2015) le añaden a su modelo de conformación de portafolio dos factores adicionales que tienen en cuenta la rentabilidad y los patrones de inversión.

La modelación financiera basada en la teoría inicial y con apoyo de la econometría, ha permitido generar modelos que explican fenómenos económicos a partir de la estadística. Los avances tecnológicos en la actualidad hablan de *machine learning* como una rama de la inteligencia artificial que a partir de técnicas estadísticas posibilita la creación de sistemas capaces de reconocer patrones y predicciones. Su aplicación al mercado de valores ha demostrado resultados positivos en el mercado de renta fija como lo demuestra (Coronado, 2022), fondos de pensiones voluntarias y FIC en (Castro y Guerrero, 2021) y en mercados de renta variable Echeverri (2021)

En la figura 1, se resumen algunas de las contribuciones a la teoría de carteras, desde Markowitz (1952) hasta la actualidad.

Figura 1

Algunas contribuciones a la teoría de carteras en las últimas décadas



Nota. Elaboración propia con base en las fuentes consultadas.

Con lo anterior, en la presente investigación se pretende aplicarle al mercado accionario algunas de las mejores técnicas de *machine learning* encontradas en la literatura, a partir del Colcap, el cual refleja las variaciones de los precios de las acciones más líquidas de la Bolsa de Valores de Colombia (BVC, s. f.).

1. Situación en estudio – problema

En razón de las coyunturas económicas actuales, la globalización y otros factores, los mercados de valores presentan fluctuaciones pronunciadas, las cuales conllevan un rendimiento y una eficacia no significativos en las estrategias de inversión tradicional (Chen y otros, 2020). Lo anterior ha repercutido negativamente en la toma de decisiones en el mercado de valores, dando cabida a mercados ineficientes en los que el precio de un activo en un momento determinado no representa su valor total en el mercado (Saetia & Yokrattanasak, 2022).

A medida que se ha incrementado la información sobre empresas y mercados, se ha vuelto cada vez más fundamental reducir esta gran cantidad de datos, que ya no podrían ser procesados mediante métodos convencionales (Albeda, 2021), para convertirlos en información útil para la toma de decisiones (Caparrini, 2021). Adicionalmente, Echeverri (2021) afirma que durante los procesos de negociación se presentan errores humanos que afectan negativamente las rentabilidades esperadas.

Una solución a este diagnóstico, es la aplicación de modelos de *machine learning* e inteligencia artificial a los mercados de valores. Este es un campo ampliamente investigado en países tales como Estados Unidos y en Reino Unido, los cuales en efecto hacen uso de estas herramientas en la mayoría de los volúmenes de negociación de activos, caso diametralmente opuesto al de Colombia Echeverri (2021).

Por lo anterior, esta investigación busca sentar un precedente en el estudio de

herramientas de análisis de datos aplicados al mercado de renta variable en el contexto colombiano, con el ánimo de generar conocimiento y de proporcionar herramientas para que los académicos e inversores para que puedan hacer una eficiente toma de decisiones en el mercado de renta variable.

1.2. Pregunta de investigación

¿Qué modelo de *machine learning* le permite al inversionista alcanzar la mayor rentabilidad en la conformación de un portafolio para inversión en activos del Colcap?

1.3. Objetivo general

Determinar el modelo que le permita al inversionista la selección del portafolio óptimo para activos financieros de renta variable en Colombia.

1.4. Objetivos específicos

- Seleccionar los activos pertenecientes al grupo Colcap, que sean adecuados para incluir en un portafolio de inversión.
- Aplicar las metodologías de *machine learning* a los activos seleccionados.
- Comparar los resultados obtenidos por las metodologías de *machine learning* y analizar los resultados obtenidos.

2. Marco de referencia conceptual

2.1. Teoría portafolio

Un portafolio consiste en un conjunto de instrumentos o activos financieros, cuyo objetivo es alcanzar el mayor rendimiento posible minimizando riesgos. La técnica utilizada para seleccionar la composición de estos recibe el nombre de teoría moderna de portafolios Markowitz (1952), la cual se propone conseguir el mejor portafolio posible, el cual deberá satisfacer las necesidades de los inversionistas en lo que respecta a riesgo y tiempo. En este sentido, es responsabilidad del administrador del portafolio incrementar el rendimiento de conformidad al riesgo asumido.

2.2. Método Markowitz

Markowitz publicó en 1952 un artículo basado en su tesis doctoral, titulado *Portfolio Selection*, y en 1959 publica el libro *Portfolio Selection, Efficient Diversification of Investments* (Markowitz, 1959), en los cuales desarrolla su modelo sobre la base del comportamiento racional del inversor; es decir, en el que el inversor desea la rentabilidad y rechaza el riesgo. Por lo tanto, para Markowitz una cartera será eficiente si proporciona la máxima rentabilidad posible para un riesgo dado o, de forma equivalente, si presenta el menor riesgo posible para un nivel determinado de rentabilidad (Mangram, 2013).

Los principales supuestos de la teoría de Markowitz (1952) son los siguientes:

- Los inversionistas son razonables al tratar de maximizar sus rendimientos minimizando el riesgo.
- Los inversionistas están dispuestos a asumir un mayor riesgo si son compensados con mayores rendimientos.
- Los inversionistas tienen la información pertinente y oportunamente al momento de tomar su decisión de inversión.
- Los inversionistas pueden pedir prestado o prestar capital ilimitado a una tasa libre de riesgo.
- Los mercados son perfectamente eficientes.
- Los mercados no incluyen costos de transacción ni impuestos.
- Es posible seleccionar acciones cuyo rendimiento es independiente de otras carteras.

Algunos detractores del método de Markowitz (1952) argumentan que este presenta algunos problemas, entre los cuales se destacan: su complejidad matemática, debido a la naturaleza cuadrática paramétrica; el número elevado de estimaciones en rendimientos esperados, varianzas y covarianza; la ausencia de evaluación de actitud ante el riesgo y la deducción de función de utilidad, y las decisiones subjetivas por parte de los gestores (Mendizábal y otros, 2002). Con la aplicación de las técnicas de *machine learning* se pretende dar alternativa a estos obstáculos.

2.3. Riesgos financieros

La medición de los riesgos financieros ha contribuido a tomar mejores decisiones en beneficio de su administración dentro de la economía del país. Por lo tanto, a partir de

1988 el Comité de Supervisión Bancaria de Basilea 1 dictaminó una serie de pautas en términos de regulación, lo que suponía un estándar internacional de la gestión del riesgo en los mercados financieros. En este se indica que el riesgo financiero proviene de tres fuentes: mercado, crédito y operación (De Lara, 2005).

En Colombia, la entidad encargada de inspeccionar, vigilar y controlar que se cumplan las especificaciones de Basilea es la Superintendencia Financiera de Colombia (Ministerio Hacienda, 2005). Su objetivo radica en garantizar que las instituciones financieras desarrollen e implementen un sistema de administración de riesgos. Entre ellos, el riesgo de mercado (SARM), el sistema de administración de riesgo de crédito (SARC) y el sistema de administración del riesgo operativo (SARO). Dichos sistemas son libres de realizar sus cálculos, siempre y cuando guarden correspondencia con los niveles de exposición y estimen el monto de capital adecuado siguiendo las normas establecidas.

2.4. Colcap

El Colcap es un índice de capitalización que refleja las variaciones de los precios de las acciones más líquidas de la BVC (s. f.). La participación de cada acción en el índice está determinada por el correspondiente valor de la capitalización bursátil ajustada (Banco de la República, s. f.).

2.5. Machine learning

Machine learning, en español aprendizaje automático, es una rama de la inteligencia artificial. El proceso de aprendizaje automático es similar al de la minería de datos, ya que ambos sistemas buscan encontrar patrones entre los datos; sin embargo, en lugar de extraer los datos para la comprensión humana, como es el caso de las aplicaciones de minería de datos, el aprendizaje automático utiliza esos datos para detectar patrones y ajustar las acciones del programa.

2.6. Etiquetas

A lo largo de esta investigación se menciona reiteradamente el término etiqueta. Las etiquetas identifican los vectores de datos apropiados que se van a extraer para el entrenamiento de modelos, para posteriormente aprender a realizar las mejores predicciones (IBM, s. f.); es decir, permiten asignarles un nombre a los datos con los que se está modelando.

2.7. Tipos de aprendizajes

Los algoritmos de aprendizaje automático presentan diferentes tipos de aprendizajes, en función de qué conjunto de datos se utilizan para realizar el entrenamiento, y a menudo se clasifican como supervisados o no supervisados.

En el caso del aprendizaje supervisado, los modelos parten de datos etiquetados y se entrenan con dichos datos, para inferir patrones y asignarles una clase a nuevos datos

sin etiquetar (IBM, s. f.). Los algoritmos de aprendizaje supervisado analizan los datos de entrenamiento y crean una función que se puede utilizar para mapear nuevos ejemplos. Este es el entrenamiento que se utiliza en la presente investigación.

En el caso de los algoritmos no supervisados, se parte de datos que no han sido previamente etiquetados. Mediante estas técnicas, los datos se agrupan en algunos conjuntos en función de sus características, siendo los elementos que pertenecen a un mismo grupo más parecidos entre ellos que los que se encuentran en grupos distintos (Caparrini, 2021). Estos algoritmos son de utilidad cuando hay datos de entrenamiento que no contienen ninguna información sobre el resultado deseado, y lo que se pretende es comprender los datos obteniendo la estructura básica de los mismos.

2.8. Árboles de decisión (DT)

El objetivo de esta técnica es predecir la etiqueta de clase de ciertas entradas por clasificar, a partir de información de otros atributos que previamente han entrenado al algoritmo para predecir (Botía, 2007).

En los modelos de clasificación, el objetivo es predecir un valor específico de una variable al agrupar la información en categorías basadas en otras variables. Asimismo, en los modelos de regresión, la meta es predecir el valor de una variable en función de otras variables independientes. Investigaciones tales como la de Echeverri (2021) han usado este método para predecir si el precio de determinada acción sube o baja.

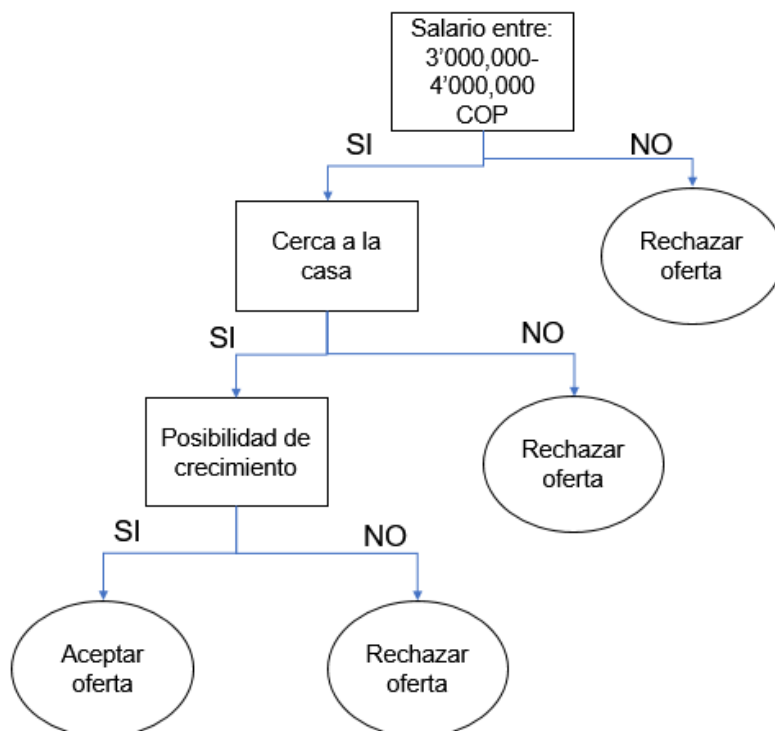
A continuación, se describe la estructura de un árbol de decisión. según Unir (2021):

- Los nodos internos representan cada una de las características a considerar para tomar una decisión.
- Las ramas representan la decisión en función de una determinada condición como la probabilidad de ocurrencia.
- Los nodos finales representan el resultado de la decisión. (párr. 4)

Un ejemplo práctico de los árboles de decisión se muestra a continuación en la figura 2, en el cual como resultado final se busca saber si se debe o no aceptar una oferta laboral en función de unos atributos especificados en cada nodo.

Figura 2

Ejemplo de un árbol de decisión



El interés del presente trabajo es aplicarle el concepto a la optimización de portafolios de inversión en acciones del Colcap, a partir de ciertas variables de entrada definidas más adelante.

2.9. *Random forest* (RF)

El algoritmo *random forest*, o bosque aleatorio, es una técnica de clasificación que crea múltiples árboles de decisión a partir de submuestras aleatorias de datos, cada uno con la capacidad de producir resultados de predicción. Estos múltiples árboles se combinan para lograr una alta precisión y reducir el riesgo de sobreajuste en comparación con otros modelos (Koehrsen, 2017).

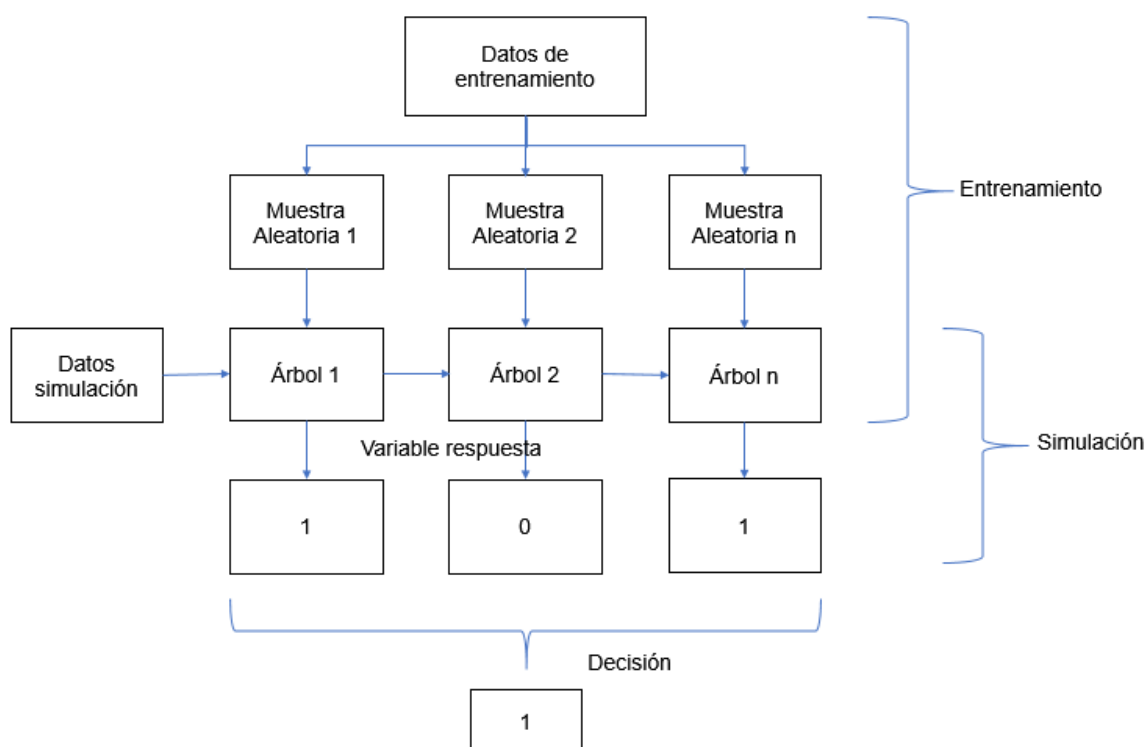
En el caso de una regresión, se toma la media de los valores resultantes de todos los árboles, y en la clasificación se selecciona la clase más común entre los árboles de decisión. El *random forest* resulta muy acertado en la predicción a partir de datos en formato tabla, y su objetivo es dividir los conjuntos de datos analizados en grupos homogéneos; es decir, dada una variable exógena Y , y unos atributos X , un árbol divide el conjunto en grupos que en Y sean tan homogéneos como sea posible (Madeeh & Abdullah, 2020).

Las divisiones se realizan en función de una variable, entre el conjunto de sus características. En caso de una variable numérica, entonces el árbol será un árbol de regresión, mientras que, si la variable es categórica, el árbol será de clasificación (Coronado, 2022).

En el ejemplo de los árboles de decisiones, para saber si se debe aceptar o no la oferta de empleo se aplica *random forest*, que agrupa de manera estratégica los datos disponibles y elabora un número óptimo de árboles de decisión que permitan tomar una decisión con base en la mayoría de las respuestas obtenidas por medio de los árboles. En la figura 2 se muestra el funcionamiento de la metodología *random forest*.

Figura 3

Ejemplificación del modelo random forest



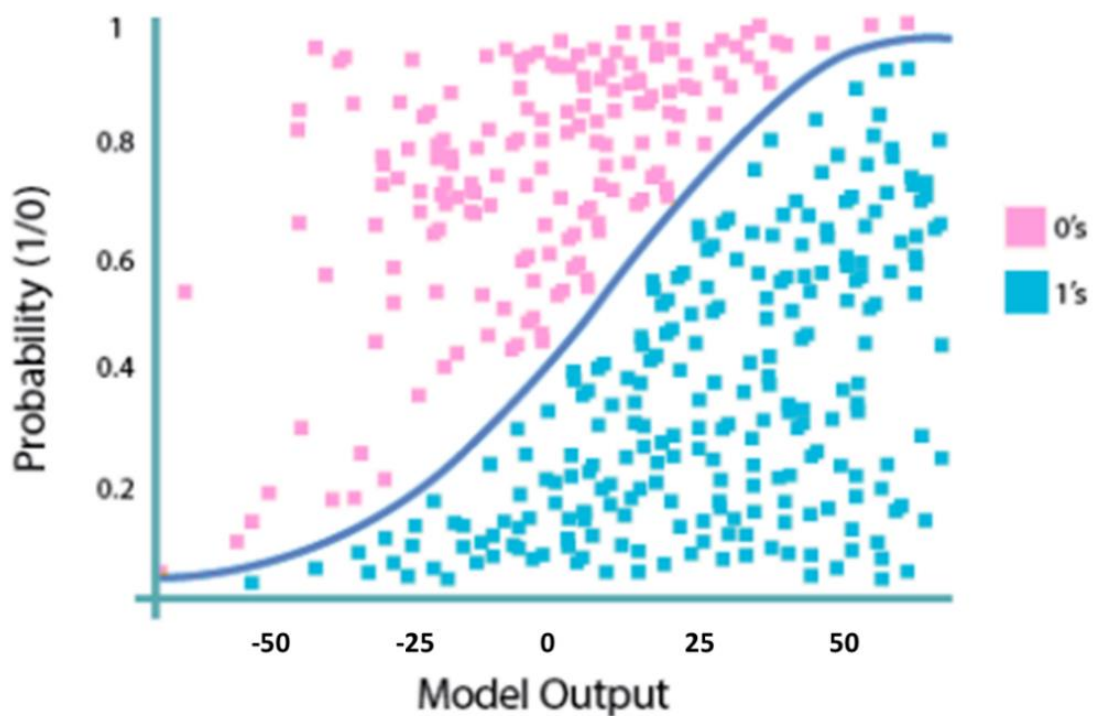
2.10. Regresión logística (R-LOG)

La regresión logística es clasificada como un modelo de aprendizaje supervisado, que se utiliza para predecir la probabilidad de que ocurra un evento de estudio, a partir de un conjunto de datos de variables independientes y de algoritmos basados en ecuaciones matemáticas (Saetia & Yokrattanasak, 2022).

Las ecuaciones matemáticas son utilizadas para encontrar las relaciones entre dos factores de datos. Posteriormente, utiliza esta relación para predecir el valor de uno de esos factores basándose en el otro. Normalmente, la predicción tiene un número finito de resultados, como un sí o un no, y permiten hacer predicciones y descubrir patrones en los datos para toma de decisiones (Saetia & Yokrattanasak, 2022). En la figura 3 se aprecia una salida de Python a una aplicación de esta metodología, en donde los resultados son dicotómicos.

Figura 4

Resultados de una regresión logística



2.11. Indicador kappa

El indicador kappa corresponde a una medida de concordancia que se utiliza para evaluar en nivel de proximidad de las predicciones realizadas con los modelos de

machine learning y los datos a posteriori. Rubiales (2018) sugiere que el resultado de Kappa se interpreta de la siguiente manera: valores ≤ 0 indican que no hay acuerdo; 0,01-0,20 son ninguno o leve; 0,21-0,40 son regulares; 0,41-0,60 son moderados; 0,61-0,80 son sustanciales, y 0,81-1,00 son casi acuerdo perfecto.

3. Metodología

3.1. Antecedentes

La tabla 1, que se presenta a continuación, se elaboró a partir de la búsqueda en bases de datos sobre publicaciones previas relacionadas con la aplicación de las técnicas de *machine learning* a los mercados financieros, con algunas de las investigaciones más relevantes en el campo.

En esta tabla se destaca que las aplicaciones de las técnicas de aprendizaje automático son transversales a los mercados de renta fija, renta variable, productos de banca patrimonial y bonos del Tesoro. Para cada mercado, el desempeño de los modelos es diferente, y así también sus resultados.

De manera general, se destaca la necesidad de disponer de grandes volúmenes de datos tanto para modelos supervisados como no supervisados, divididos en un conjunto de datos para entrenamiento, y otro, para simulación.

Tabla 1*Matriz de antecedentes*

Autor	Tipo de publicación	Metodología aplicada	Activo financiero	Hallazgos
Chen y otros (2020)	<i>A Quantitative Investment Model Based on Random Forest and Sentiment Analysis</i>	* Árboles de decisión (DT) * <i>Random forest</i> (RF) * Regresión logística * <i>SVM support vector machine</i>	10 acciones del mercado de valores chino	* La mayor precisión promedio se obtuvo del modelo RF seguida del DT. * El mejor rendimiento se obtuvo mediante un modelo RF-SA (<i>random forest - sentiment analysis</i>).
Madeeh y Abdullah (2020)	<i>An Efficient Prediction Model based on Machine Learning Techniques for Prediction of the Stock Market</i>	* KNN vecinos más cercanos * <i>Random forest</i> (RF)	12 empresas del mercado accionario NYSE	Los resultados de los experimentos demostraron que el modelo RF logró la mayor precisión de predicción en precisión, respectivamente, en

Autor	Tipo de publicación	Metodología aplicada	Activo financiero	Hallazgos
				comparación con el modelo K-NN. Por lo tanto, emplear el modelo de predicción <i>random forest</i> (RF) en el mercado de valores puede proporcionarles más ganancias a los inversores y reducir el riesgo de pérdidas tanto como sea posible.
Saetia y	<i>Stock Movement Prediction</i>	* Regresión logística	100 principales	El modelo XGBoost funciona
Yokrattanasak (2022)	<i>Using Machine Learning Based on Technical Indicators and Google Trend Searches in Thailand</i>	* <i>Random forest</i> (RF) * Aumento de gradiente extremo (XGBoost)	acciones de la bolsa de Tailandia	significativamente mejor durante períodos de extrema volatilidad del mercado, tales como la pandemia de covid-19.
Surayagari y otros (2021)	<i>Stock market predictions using machine learning</i>	* Redes neuronales	7 acciones más populares en	El objetivo principal de la tesis era establecer la conexión

Autor	Tipo de publicación	Metodología aplicada	Activo financiero	Hallazgos
		* <i>Long Short Term Memory</i> (LSTM)	internet, pertenecientes al Nasdaq: Microsoft, Google, Netflix, Apple, Nvidia, AMD, Amazon	entre las noticias y el movimiento de las acciones. Aunque los resultados mostraron que era posible clasificar qué noticias pueden impactar en las acciones, no se tradujo de manera efectiva en predicciones de precios de acciones.
Echeverri (2021)	Conformación automática de portafolios de inversión usando analítica financiera	* Regresión logística * <i>Random forest</i> (RF) * Árbol de decisión * KNN vecinos más ceranos	Todas las acciones del Índice Colcap	La regresión logística y el método de vecinos más ceranos se destacan como las metodologías más efectivas para predecir el movimiento de

Autor	Tipo de publicación	Metodología aplicada	Activo financiero	Hallazgos
		* Red neuronal recurrente LSTM		precios en una cantidad significativa de acciones. Se recomienda utilizar el modelo de IA que mejor se desempeñe para cada acción, en lugar de un enfoque único para todas las acciones.
Castro y Guerrero (2021)	Modelo predictivo de propensión de ahorro e inversión en productos de banca patrimonial	* Regresión logística * K vecinos más cercanos * Máquinas vectores de soporte * Naive Bayes * Redes neuronales * <i>Gradient boosting</i>	Primera muestra: 2371 clientes. Segunda muestra: 116500 clientes.	Los modelos de ensamble <i>random forest</i> mostraron el mejor rendimiento tanto para la identificación de propensión como para la estimación de saldos potenciales. Los resultados del modelo de propensión se alinean con la

Autor	Tipo de publicación	Metodología aplicada	Activo financiero	Hallazgos
		* <i>Loss distribution approach</i> (LDA) * Árboles de clasificación * <i>Random forest</i>		distribución de la población de clientes del banco y sugieren un potencial de crecimiento significativo en el segmento de banca patrimonial.
Caparrini (2021)	Optimización de cartera de activos financieros aplicando aprendizaje automático	* Árboles de decisión * <i>Random forest</i> * XGBoost	Todas las empresas que han pertenecido al SP500 entre 1997 y 2019, hasta seleccionar 15.	Los resultados mostraron que utilizando tanto factores de la literatura como el algoritmo XGBoost, se pueden lograr retornos superiores al índice de referencia, con un aumento del 5% anual en los rendimientos a lo largo de 16 años.
Albeda (2021)	Composición de cartera réplica del índice bursátil	* Regularización LASSO	30 acciones más representativas	El <i>autoencoder</i> con tres capas intermedias se destacó como la

Autor	Tipo de publicación	Metodología aplicada	Activo financiero	Hallazgos
	alemán DAX utilizando técnicas pertenecientes al <i>machine learning</i> : comparativa entre <i>lasso</i> , <i>random forest</i> y <i>autoencoder</i>	* <i>Random forest</i> (RF) * Redes neuronales (<i>autoencoders</i>)	del DAX a lo largo de la historia reciente.	metodología más fiable para replicar el comportamiento del índice bursátil DAX durante el período de estudio, seguido de <i>random forest</i> .
Coronado (2022)	Aplicación del machine learning al factor investing en renta fija corporativa.	* Modelos de regresión penalizada lineales * Árboles de decisión * Redes neuronales y máquinas de vectores de soporte (SVM)	15 bonos: 3 nominales, 12 ordinales discretos	<i>Random forest</i> (RF) se muestra como la alternativa más relevante entre las estudiadas.
Wei y otros (2021)	<i>Exploring the financial indicators to improve the pattern recognition of</i>	* Redes neuronales (<i>backpropagation</i>)	Acciones de varias bolsas	Los algoritmos KNN y VSM pueden capturar eficazmente las tendencias cambiantes de

Autor	Tipo de publicación	Metodología aplicada	Activo financiero	Hallazgos
	<i>economic data based on machine learning</i>		alrededor del mundo.	indicadores financieros y predecir con precisión la realización de operaciones de financiamiento.
Castro (2020)	<i>Machine learning investment strategies and the liquidity premium for the Colombian yield curve</i>	* Árboles de decisión * <i>Random forest</i>	Títulos del Tesoro colombiano	En general, para todos los períodos analizados se concluye que es posible utilizar cualquiera de los métodos propuestos para predecir el beneficio de la estrategia, ya que todos arrojan buenos resultados. Además, el uso de métodos de conjunto tales como <i>random forest</i> y <i>extra trees</i> muestra mejores

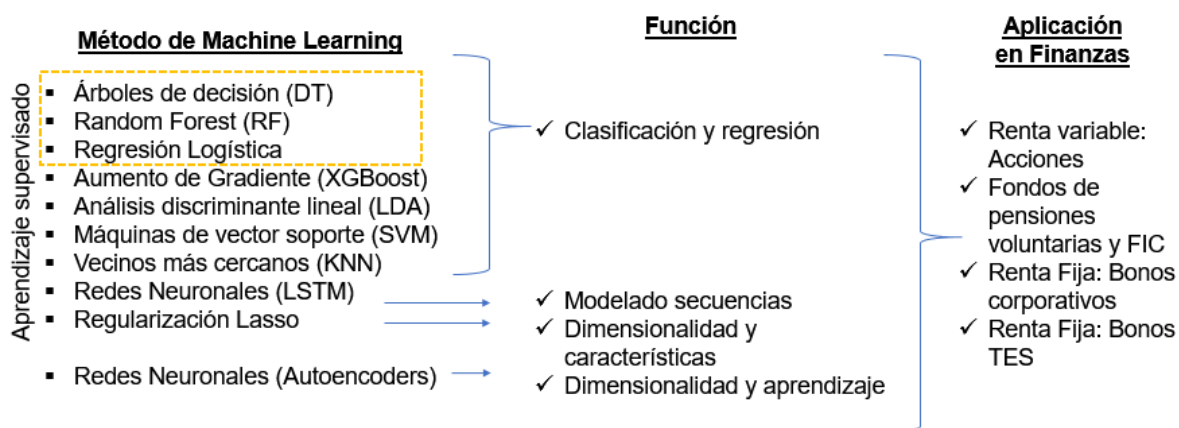
Autor	Tipo de publicación	Metodología aplicada	Activo financiero	Hallazgos
				resultados para el indicador de error cuadrático medio (MSE).
Mendizábal y otros (2002)	El modelo de Markowitz en la gestión de carteras	* Método Markowitz	IBEX-35	La aplicación del modelo de Markowitz proporciona carteras eficientes, que superan en rentabilidad y presentan menor riesgo en comparación con las carteras representadas por los índices de referencia IBEX-35 e IGBM.

En referencia al mercado de activos de renta variable, se encontró una tendencia, independiente de la bolsa estudiada, al uso de determinadas técnicas de *machine learning*. Es así como índices tales como Nasdaq, S&P500, DAX, SET y Colcap demostraron resultados favorables en la aplicación de técnicas tales como árboles de decisión, *random forest* y regresión logística, cuando lo que se buscaba era una predicción en el movimiento futuro del precio del activo en cuestión, basándose en datos históricos de investigaciones desarrolladas en contextos de analítica de datos e ingeniería de sistemas.

En la figura 5 se resume el uso de las metodologías encontradas en el campo financiero.

Figura 5

Métodos de aprendizaje automático hallados en la literatura en el campo financiero



Estos algoritmos presentan la particularidad de que permiten hacer inferencias no lineales en los conjuntos de datos estudiados a partir de patrones en la información, y se presentan como las opciones óptimas en temas de computación y programación.

Por lo anterior, la presente investigación, se orientará hacia la aplicación de las técnicas árboles de decisión, *random forest* y regresión logística para optimizar un portafolio de inversión en activos del Colcap.

3.2. Conjunto de datos

La intención principal del presente trabajo es aplicar las herramientas del *machine learning* para la construcción de un portafolio que maximice la rentabilidad invirtiendo en acciones del Colcap. Por esta razón, la *data set* se conforma con los datos diarios durante el período 2018-2022 de los precios de apertura, cierre, mínimo, máximo y volumen de negociación de cada acción, tal y como lo proponen Saetia y Yokrattanasak (2022).

Adicionalmente, se consideran el índice Colcap, el ETF del S&P500 y la TRM, tal como recomienda Echeverri (2021), quien encontró que las variaciones en los precios de los activos de renta variable colombiana se encuentran estadísticamente relacionados con las variaciones en los índices anteriormente mencionados. Las definiciones de las variables independientes se extrajeron de la BVC (s. f.) y de Nickolas (2023).

- Cantidad: número de unidades del respectivo título o papel negociados en un día.
- Volumen: cantidad total de la actividad comercial del día.
- Precio cierre: precio de cotización al cierre del día.
- Precio mayor: precio máximo de negociación del día.

- Precio menor: precio mínimo de negociación del día.
- Colcap: es un indicador que refleja las variaciones de los precios de las 20 acciones más líquidas de la BVC.
- S&P500: sigue el índice Standard & Poor's (S&P) 500, que comprende 500 acciones estadounidenses de gran capitalización. Funciona como *proxy* de la salud financiera y la estabilidad de la economía en Estados Unidos.
- TRM: la tasa de cambio representativa del mercado es la cantidad de pesos colombianos por un dólar estadounidense.

Adicionalmente se cuenta con la serie histórica de la TRM, con el fin de medir, para el período de análisis, el impacto sobre la volatilidad con base en recomendaciones de expertos. En ese orden de ideas, se conforma una base de datos que recopila las variables descritas para cada acción y el comportamiento histórico de los índices Colcap y S&P500.

La data ha sido recolectada de los servicios de información financiera Yahoo Finance e Investing.com. La información relacionada con los rendimientos e indicadores financieros se extrajeron de la base de datos EMIS.

Se incluyeron las siguientes acciones: Bcolombia, Bogotá, Celsia, Cemargos, Cnec, Corficolcf, Ecopetrol, Geb, Grubolivar, Grupoargos, Gruposura, Isa, Mineros, Nutresa, Pfaval, Pfbcolom, Pfcemargos, Pfcorficol, Pfdavvnda, Pfgrupoarg, Pfgrupsura, Promigas y Terpel. Del análisis se excluyeron las acciones de ETB, por pérdida de valor en las últimas décadas, y las de BVC, por falta de información financiera.

3.3. Procesamiento de datos

El preprocesamiento de datos incluye las etapas de limpieza de datos y normalización (Madeeh & Abdullah, 2020). En la limpieza de datos, los valores faltantes se abordan mediante la estrategia de reemplazar dichos valores por el promedio de ese atributo. La normalización de datos se realiza para escalar los atributos en un rango específico, lo que facilita las técnicas de aprendizaje automático; además, se eliminan valores atípicos.

3.4. Lenguaje de programación y librerías

El *software* usado para la modelación de la estrategia es la herramienta de análisis de minería de datos KNIME (por sus siglas en alemán *konstanz information miner*), y para el proceso de normalización y limpieza de los datos, es SQL.

3.5. Métodos de predicción

El objetivo se enfoca en predecir la dirección en la que se moverá el precio de cada acción al cierre de la jornada bursátil del siguiente día, y así conocer algún patrón relacionado tal que permita tomar decisiones de compra o venta, garantizando algún grado de rentabilidad.

Para esto se prueban los modelos: *random forest*, regresión logística y árboles de decisión.

3.6. Variable dependiente

Los modelos generarán una variable dependiente con una respuesta en forma binaria: 1 si el precio de la acción subirá al día siguiente y 0 si el precio de la acción tiende a la baja utilizando la ecuación 1.

$$Si (P_t < P_{t+1}, 1,0) \quad (1)$$

En donde







P_t : precio de cierre del día actual

P_{t+1} : precio de cierre del día siguiente

De esta manera, cada algoritmo se aplica para obtener sus respectivos resultados de rentabilidad y riesgo con las acciones cuya simulación arroje los mejores resultados para el portafolio. En la tabla 2 se ejemplifica cómo funciona la predicción en la caída o subida en el precio de la acción al día siguiente.

Tabla 2

Ejemplificación de una variable dependiente

Stock	Fecha	Precio Cierre	Tendencia Precio
CEMARGOS	21/12/2022	3560	0 
CEMARGOS	22/12/2022	3540	0 
CEMARGOS	23/12/2022	3540	1 
CEMARGOS	26/12/2022	3690	1 
CEMARGOS	27/12/2022	3727	0 
CEMARGOS	28/12/2022	3600	0 

Tal como se aprecia en la tabla anterior, por ejemplo, para el 21 de diciembre de 2022, la acción Cemargos cerró a \$3560 por acción, y el modelo predice que, para el 22 de diciembre, el precio irá a la baja. Al revisar el precio de cierre de la misma

acción el 22 de diciembre, se encuentra que se estableció en los \$3540, indicando una predicción acertada.

3.7. Aplicación de los métodos seleccionados

Para el entrenamiento del modelo, se crea primero una *data set* de fechas tomando todos los días de bolsa desde el 2017 para cada una de las acciones del índice. La medición de la volatilidad no se realiza sobre los precios *per se*, sino sobre los rendimientos diarios, que no son más que el cambio diario en el precio del activo, haciendo uso de los retornos logarítmicos, debido a las ventajas que tienen cuando se analizan múltiples períodos (Miskolczi, 2017).

Una vez obtenido la *data set*, se emplea la simulación con los métodos *random forest*, regresión y árboles de decisión. Estas predicciones encontradas se traducen a señales de compra y(o) venta de las acciones que conformen un portafolio óptimo, esperando una rentabilidad arrojada por cada modelo.

Para predecir los movimientos en el precio de las acciones a partir de los modelos discutidos, se requiere hacer una partición en la data por períodos tal y como se muestra en la tabla 3.

Tabla 3

Período de análisis para entrenar y evaluar la predicción de la dirección del precio

Entrenamiento			Simulación	
2018	2019	2020	2021	2022

Se toma un período de entrenamiento para los modelos de *machine learning* y otro período para evaluar la predicción de estos, siguiendo el principio de aprendizaje supervisado.

Adicionalmente, se definen iterativamente los hiperparámetros del modelo, tales como el número de árboles, el criterio de división y la profundidad máxima de los árboles, para optimizar los modelos en función de las restricciones computacionales. Por otra parte, la medición de la rentabilidad se da por tenencia *long-only* para los portafolios, no guardando continuidad del precio del día siguiente (Leung y otros, 2003; Choudhry y Garg, 2008).

3.8. Evaluación de los métodos seleccionados

Con el fin de comparar la precisión en la predicción de los algoritmos propuestos, se utiliza la matriz de confusión para cada modelo, que permite ver qué tipos de aciertos y errores está teniendo el modelo durante el proceso de aprendizaje con los datos (Barrios, 2023).

La matriz de confusión permite identificar los siguientes errores y aciertos en un modelo de predicción:

- Verdadero positivo (VP): el valor real es positivo y la prueba predijo también que era positivo.
- Verdadero negativo (VN): el valor real es negativo y la prueba predijo también que el resultado era negativo.
- Falso negativo (FN): el valor real es positivo, y la prueba predijo que el resultado es negativo. Conocido en estadística como error tipo II.
- Falso positivo (FP): el valor real es negativo, y la prueba predijo que el resultado es positivo. Conocido en estadística como error tipo I.

Una vez conocidos estos valores se reúnen en la matriz, tal y como se ilustra en la figura 6, para de esta manera obtener la precisión equilibrada, utilizada para estimar el rendimiento del modelo, a partir de la sensibilidad y especificidad del modelo.

Figura 6

Matriz de confusión

		Realidad		Predicción-Simulación
		Verdadero Positivo (VP)	Falso Positivo (FP)	
Predicción-Simulación	Verdadero Positivo (VP)	Verdaderos Positivos (VP)	Falsos Positivos (FP)	
	Falso Negativo (FN)	Falsos Negativos (FN)	Verdaderos Negativos (VN)	

La sensibilidad cubre la proporción de verdaderos positivos mientras que la especificidad hace lo mismo para los verdaderos negativos, siguiendo las ecuaciones que se presentan a continuación (2, 3 y 4).

$$\text{Especificidad} = \left(\frac{VN}{VN+FP} \right) \quad (2)$$

$$\text{Sensibilidad} = \left(\frac{VP}{VP+FN} \right) \quad (3)$$

$$\text{Precisión balanceada} = \frac{\text{Especificidad} + \text{Sensibilidad}}{2} \quad (4)$$

Asimismo, se comparan los resultados obtenidos con la teoría de optimización de portafolio. En este caso, con el método Markowitz, o de mínima varianza.

3.9. Teoría de portafolio

Uno de los objetivos de la presente investigación es pretende determinar las diferencias en resultados que se puede obtener entre la teoría de portafolio tradicional y las nuevas técnicas de *machine learning*. En este caso, corresponde describir la teoría de portafolio desarrollada. Para orientar esta optimización, se han tomado bases contenidas en el libro de Agudelo (2014).

Las variables de entrada requeridas son los retornos esperados de cada acción y los riesgos y correlaciones. Para los riesgos y correlaciones se elaboró una matriz a partir de los datos históricos, mientras que los retornos esperados fueron calculados a partir del modelo CAPM, por recomendación de Agudelo (2014).

Las ecuaciones utilizadas (5, 6 y 7) se listan a continuación.

$$E(R_i) = R_f + \beta \left(E(R_{(m)}) - R_f \right) \quad (5)$$

En donde:

$E(R_i)$ = rentabilidad esperada de la acción i

R_f = rentabilidad esperada libre de riesgo TES 5 años

β = medida de sensibilidad (extraída de Damodaran)

$E(R_m)$ = rentabilidad esperada del mercado (COLCAP)

Para determinar la rentabilidad esperada del portafolio y el riesgo asociado a este, se partió de las ecuaciones 6 y 7:

$$E(R_p) = \sum_{i=1}^N W_i \cdot E(R_i) \quad (6)$$

En donde:

$E(R_p)$ = rentabilidad esperada del portafolio

W_i = porcentaje de inversión del portafolio destinada a la acción i

$E(R_i)$ = rentabilidad calculada con CAPM

$$\sigma_p^2 = \sum_{i=1}^N W_i^2 \cdot \sigma_i^2 + \sum_{i=1; i \neq j}^N \sum_{j=1}^N W_i \cdot W_j \cdot \rho_{ij} \cdot \sigma_i \cdot \sigma_j \quad (7)$$

σ_i = desviación estándar de la acción i

σ_i^2 = covarianza de la acción i

ρ_{ij} = covarianza entre acción i y j

El período evaluado corresponde a los mismos cinco años utilizados con las técnicas de *machine learning*. Se utilizan las acciones más líquidas para conformar el portafolio, y para los emisores de acciones ordinarias y preferenciales sólo se incluirán aquellas que no tienen derecho a voto. Se utiliza la herramienta computacional Solver de Excel, para llevar a cabo la optimización.

3.10. Retornos por activo

Se han considerado los retornos a partir de las diferencias logarítmicas en los precios de cada acción alcanzados durante los días de bolsa, en el intervalo de estudio 2018-2022, siguiendo la ecuación 8.

$$R_i = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (8)$$

En donde:

R_i = Rendimiento de la acción i

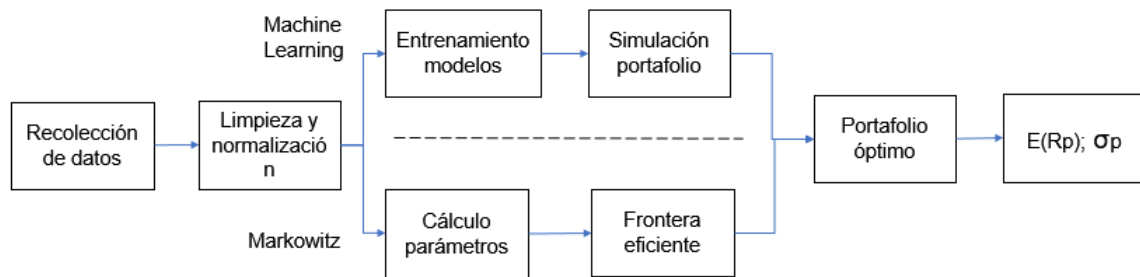
P_t = precio de la acción en el día t

P_{t-1} = precio de la acción en el día $t - 1$

3.11. Supuestos

Con el fin de diversificar el portafolio de acuerdo con las acciones más rentables, y disminuir el riesgo, se hace necesario asignar porcentajes hipotéticos de inversión con restricción de no negatividad y suma total del 100%. No se consideraron costos de transacciones, obteniendo una rentabilidad bruta en los portafolios.

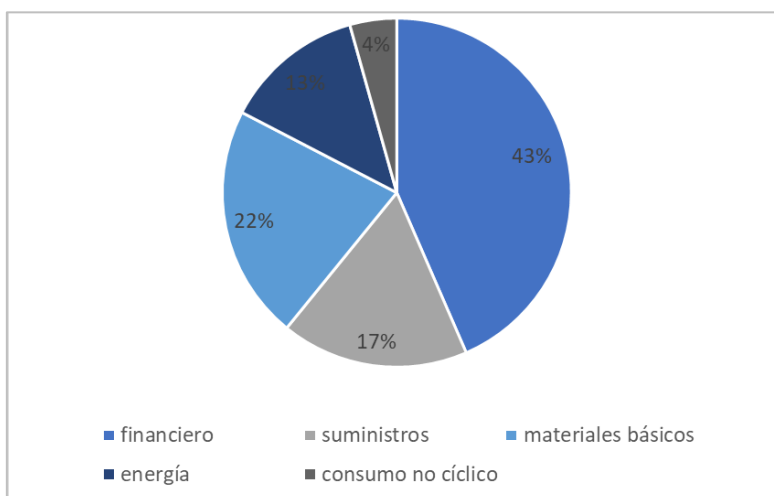
En la figura 7, se evidencia resumido el procedimiento para llevar a cabo.

Figura 7*Esquema metodológico propuesto*

4. Resultados

4.1. Análisis descriptivo

Se analizó el comportamiento en el rendimiento de las acciones seleccionadas para la optimización del portafolio. Cabe resaltar que el 43% de las acciones estudiadas corresponden al sector financiero, tal como se puede observar en la figura 8.

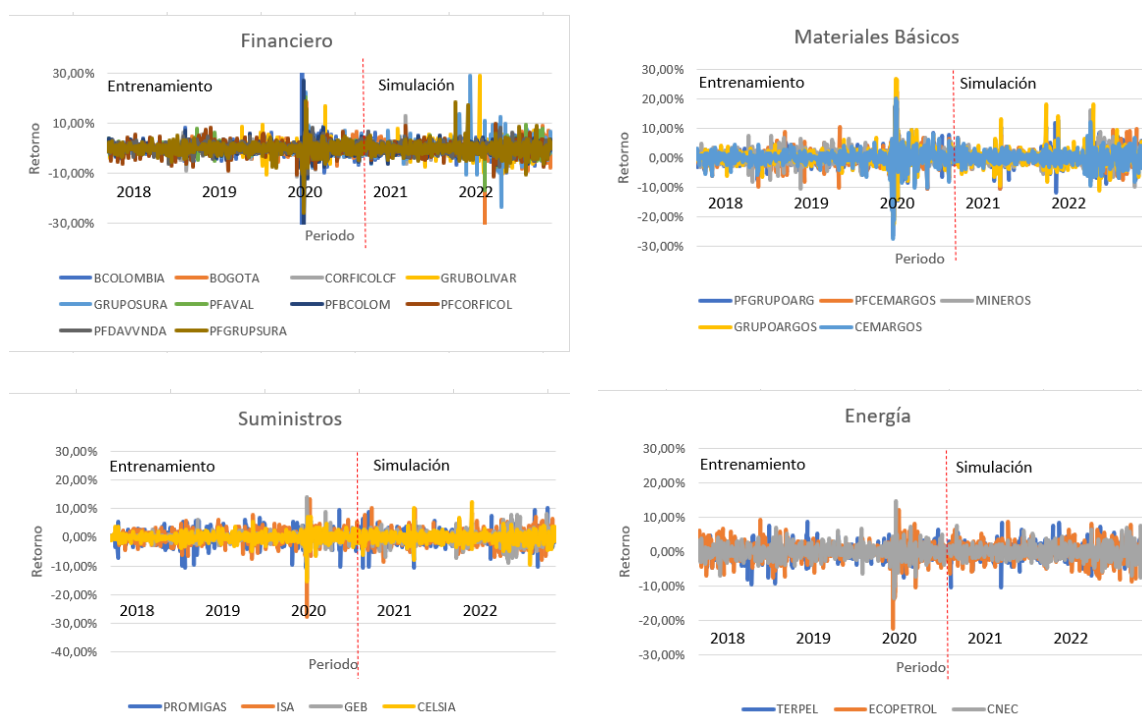
Figura 8*Composición por sectores económicos de las acciones estudiadas*

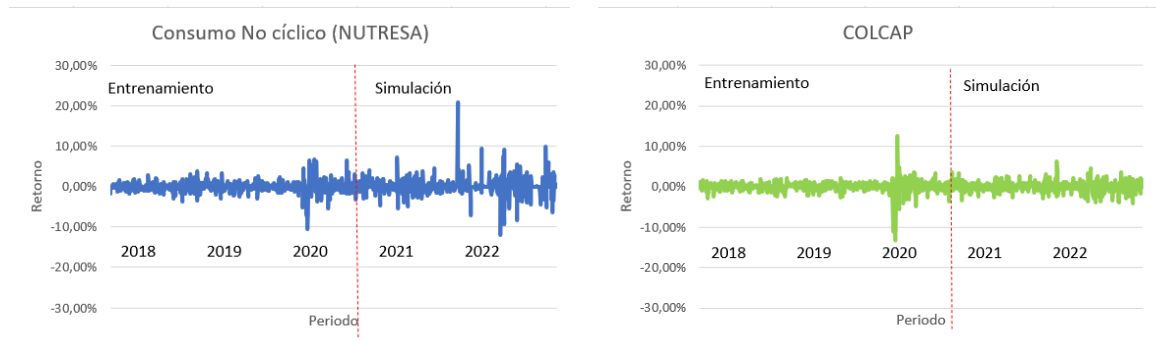
El comportamiento en estos rendimientos se ve similar a lo largo del período de estudio, y se destacan períodos de alta volatilidad en todos los sectores económicos durante 2020, producto de la pandemia, y algunas perturbaciones en los sectores de consumo no cíclico y materiales básicos en 2022, derivados de la alta inflación, los conflictos sociales y los efectos colaterales del covid-19 (El Tiempo, 2022).

En la figura 9 se resumen a continuación las gráficas de los retornos por sector.

Figura 9

Rendimientos de las acciones durante el período estudiado (2018-2022)





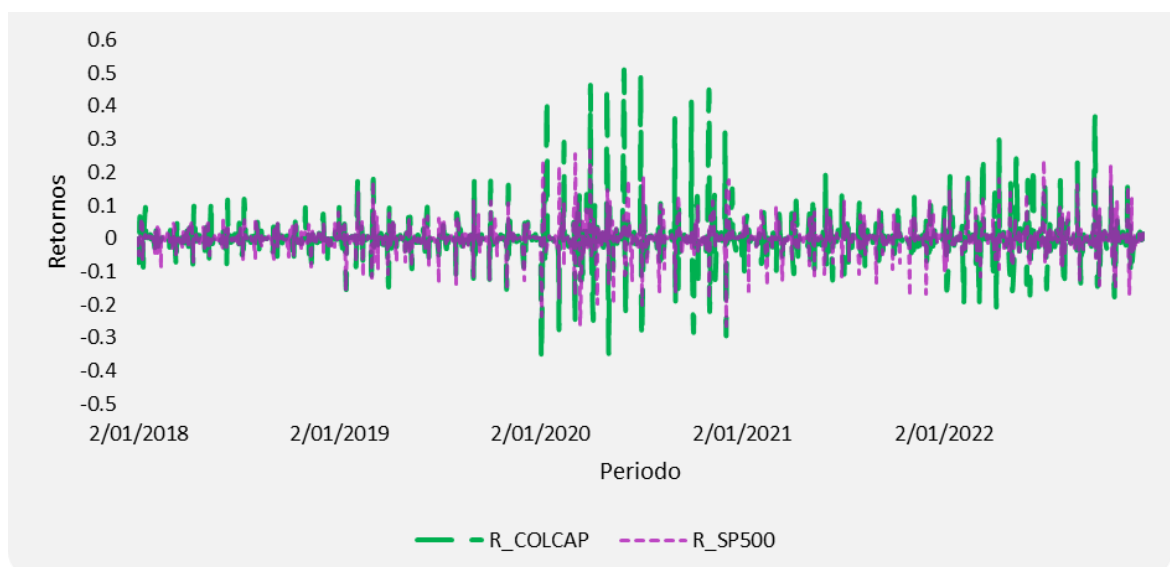
Nota. Elaborados a partir de la data descargada de Yahoo Finance.

En el caso de acciones ordinarias y preferenciales de un mismo emisor, el comportamiento a priori se visualiza similar, en este caso, al momento de aplicar la teoría de portafolio tradicional, se trabajará sólo con las ordinarias de cada emisor.

Adicionalmente, el comportamiento de retornos de los índices mencionados durante el período 2018-2022 se analiza en la figura 10.

Figura 10

Comportamiento histórico retornos de los índices Colcap y S&P500



Nota. Elaborado a partir de la data descargada de Investing.com.

A priori, se encuentra un nivel de volatilidad superior en el Colcap en comparación con el S&P500, debido a la pronunciada dispersión en los retornos, lo que obliga una mayor rentabilidad esperada para el inversionista.

Aunque se aprecian diferencias en términos de volatilidad, ambos índices se mostraron sensibles ante sucesos macroeconómicos tales como la pandemia por covid-19 en 2020, los choques inflacionarios y recesiones en Estados Unidos para el 2021, la crisis y la inestabilidad en el sector financiero por la iliquidez de bancos como el SVB a comienzos de 2022 y la escasez de suministros y el aumento de precios en exportaciones a finales de ese año, derivados de la coyuntura del 2020. Se destaca que, a nivel país, para el 2021 Colombia afrontó dificultades por el estallido social y perdió grado de inversión, al pasar de BBB- a BB+.

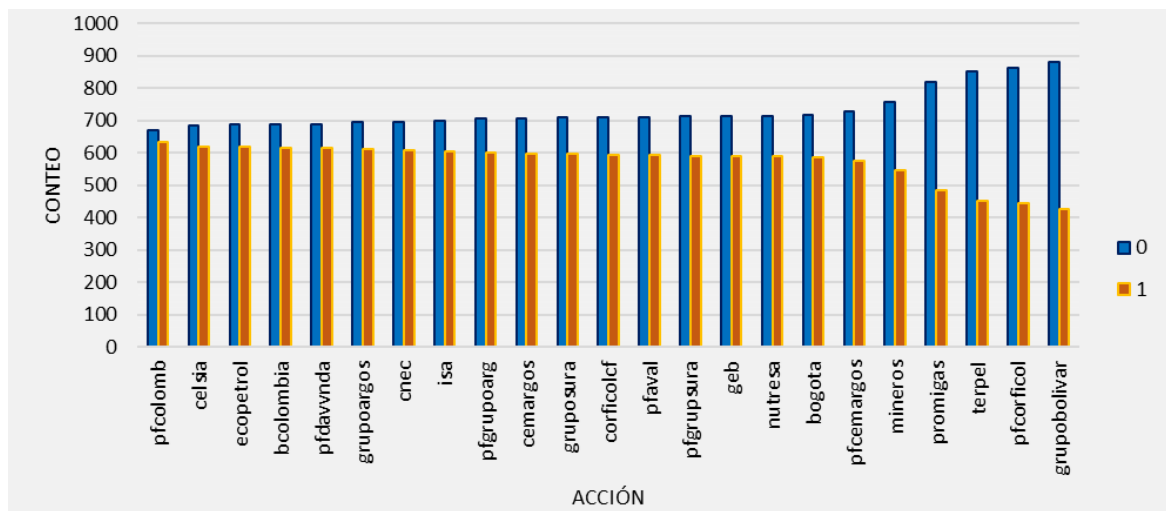
4.2. *Machine learning*

Con el fin de incorporar otras estrategias de *machine learning* en la conformación de portafolio, se propone evaluar las tres acciones cuyas predicciones en el movimiento del precio sea al alza, y su interacción con los índices Colcap, S&P500 y TRM, con el fin de observar la rentabilidad esperada y volatilidad obtenida.

Según se aprecia en la figura 11, se puede observar mayor cantidad de ceros en las acciones Bogotá, Pfcemargos, Mineros, Promigas, Terpel, Pfcorficol y Grupo Bolívar. Esto indica que el precio tenderá a la baja, y no se presentan atractivas para un portafolio *long-only*.

Figura 11

Comportamiento de la variable de respuesta sin predicción



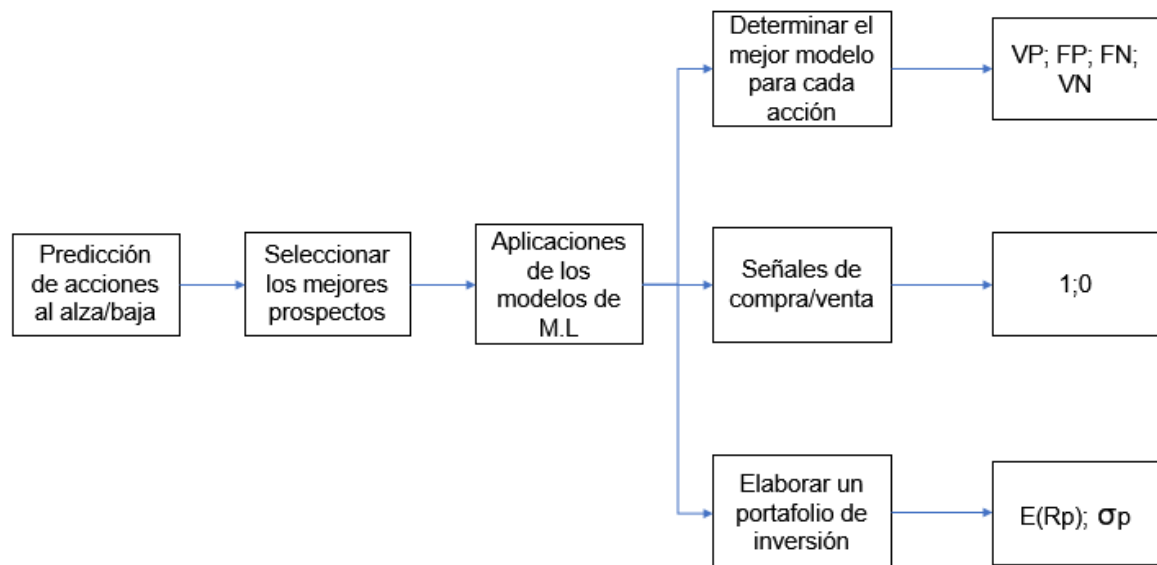
Nota. Elaborado a partir de los datos descargados de Yahoo Finance e Investing.com.

Caso contrario ocurre con las acciones Bancolombia, Celsia, Ecopetrol, Davivienda, Grupoargos, Cnec, ISA y Cemargos, las cuales tienden a presentar incremento en el precio al día siguiente, y serán consideradas para formular la estrategia de inversión.

Lo anterior arroja las variables de entrada que se les aplicaron en las metodologías de *machine learning* a las acciones que conformarán la estrategia de portafolio, juntando los resultados de los datos con recomendaciones de expertos, y se encontraron las variables más determinantes en la predicción y los modelos más precisos siguiendo la ruta diagramada en la figura 12.

Figura 12

Aplicación de inteligencia artificial a la teoría de portafolio



4.3. Estrategia

4.3.1. Elección de acciones

Con el fin de crear una estrategia para una cartera óptima, se eligieron tres de los activos que obtuvieron señales prometedoras de comportamiento al alza, y se incorporaron criterios de volatilidad, liquidez y defensiva presentados en la tabla 4. Las principales estadísticas descriptivas de las acciones que conforman la estrategia de inversión se pueden consultar en el anexo 1.

Tabla 4*Estrategia para portafolio*

Activo	Acción
Liquidez	Ecopetrol
Volátil	Celsia
Defensiva	Bancolombia

La acción defensiva es aquella que ante los posibles choques de las otras y del debilitamiento de la cartera funciona de soporte. Las acciones volátiles presentan una fluctuación pronunciada en su precio dentro de un intervalo de tiempo determinado, siendo consecuente con que, a mayor volatilidad, mayor riesgo recompensado con mayores beneficios potenciales.

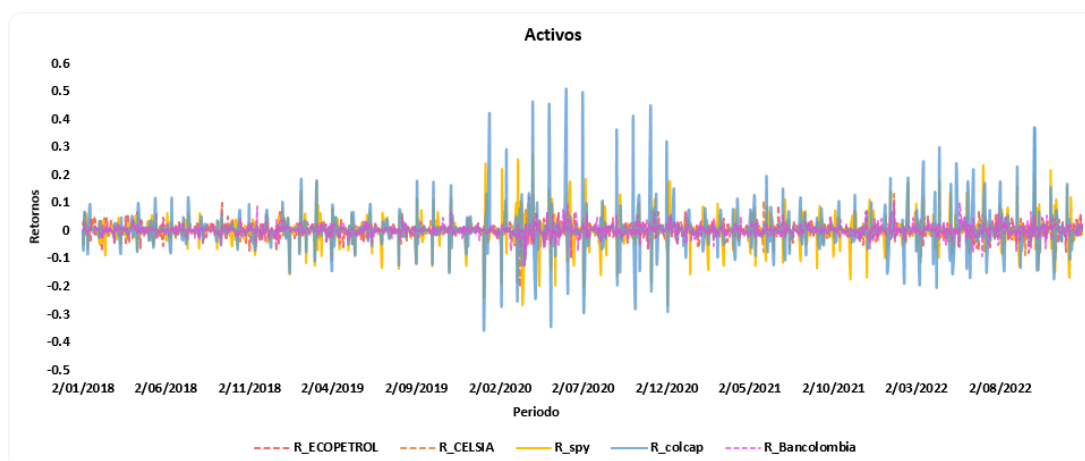
De acuerdo con su peso en el Colcap, la acción con más peso y con mayor valor de mercado es Ecopetrol, ya que, al ser de alta liquidez, inmediatamente se vuelve referencia, y Ecopetrol domina más de 50% del mercado (Gómez, 2022).

Para elegir una acción volátil, se consultó TradingView (s. f.), y se encontró que en el primer puesto aparecía Celsia, con una volatilidad del 6,51%, un volumen de 691.341.000 y un rendimiento por dividendo del 9,87% con corte a noviembre de 2023. Adicionalmente, en la elección de una acción defensiva se escogió Bancolombia, la cual, al estar dentro del segmento de servicios financieros, puede ser usada como la acción que conforme el portafolio en posición defensiva, ya que el sector financiero se mantiene un poco más resistente a las fluctuaciones económicas.

En la figura 13 se evidencia el comportamiento en los retornos de los activos mencionados.

Figura 13

Retorno por acción de la estrategia machine learning en el período 2018-2022



Nota. Elaborado a partir de la data descargada de Investing.com.

La partición óptima utilizada en la data para realizar las predicciones en el precio de las acciones a partir de las metodologías de *machine learning* fue 70% entrenamiento y 30% predicción. Esta predicción del movimiento del precio le permitirá al inversionista tomar decisiones de inversión, en las que se entrará en posiciones largas cuando el precio de la acción suba al día siguiente, y se tomarán posiciones cortas cuando la predicción del movimiento del precio indique que no subirá.

A continuación, se analizan las acciones seleccionadas para la estrategia y sus resultados en las predicciones.

4.3.2. Acción de Ecopetrol

Para el caso de Ecopetrol, la mejor predicción en el movimiento del precio se obtuvo del modelo *random forest*, con una precisión del 72%, y las variables más significativas en la predicción fueron retorno de cierre de la acción, retorno de la TRM y retorno de cierre ajustado del Colcap.

En la tabla 5 se resume la matriz de confusión de los modelos utilizados, en la que se puede apreciar que la mayor precisión fue obtenida por el modelo *random forest*. Adicionalmente, el indicador Kappa demuestra una proximidad moderada en sus predicciones.

El comportamiento del precio de esta acción por trimestre se puede consultar en el anexo 2.

Tabla 5*Desempeño de los modelos de machine learning con la acción de Ecopetrol*

Modelo	Predicción	Verdaderos positivos	Falsos positivos	Verdaderos negativos	Falsos negativos	Precisión	Kappa
Random Forest (RF)	Alza (1)	136	54	147	55	0,72	0,44
	Baja (0)	147	55	136	54		
Decition Tree (DT)	Alza (1)	122	73	136	61	0,66	0,32
	Baja (0)	136	61	122	73		
Regresión Log (R-LOG)	Alza (1)	87	26	164	115	0,64	0,29
	Baja (0)	164	115	87	26		

Como se mencionó anteriormente, el 30% de los datos fueron entrenados para predecir, y, de acuerdo con los resultados, la matriz de confusión junto con los demás indicadores muestra buen desempeño en cada modelo propuesto en la interacción retorno de precio de la acción junto con la TRM, la cual explica la volatilidad en las acciones estudiadas.

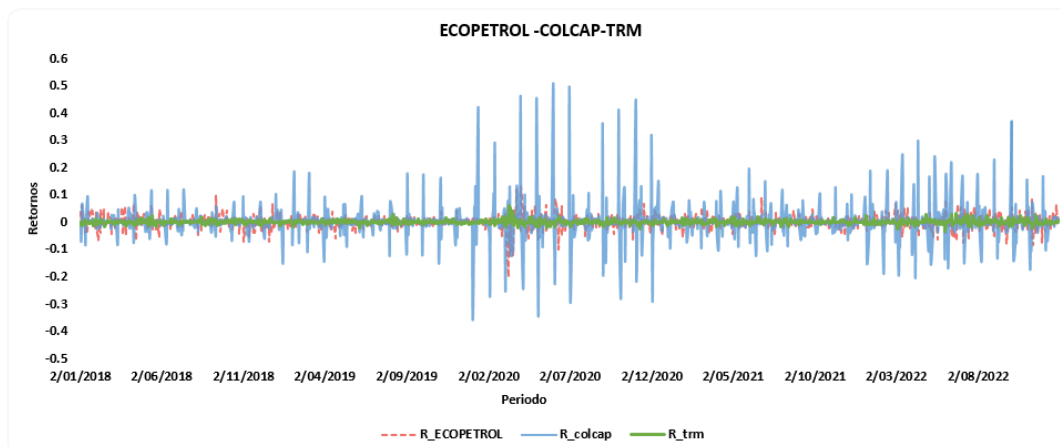
De los tres modelos evaluados, el *random forest* se destaca por una precisión del 72%; es decir, cuando el precio tiende al alza, efectivamente el modelo predice su tendencia correctamente. Por tanto, muestra un buen indicio para la estrategia de inversión del activo evaluado; asimismo, se demuestra con los verdaderos positivos 136 valores predichos correctamente.

Por otro lado, el indicador de Kappa como medida de concordancia refleja el 0,44 de probabilidad, indicando que el 44% de las veces predice la verdadera tendencia a la alza o baja, siendo una predicción moderada para tomar posiciones en corto.

En la figura 14 se evidencia el comportamiento de las variables que mejor describen las tendencias del precio de la acción y que corresponden a retorno de cierre de la acción, retorno de la TRM y retorno de cierre ajustado del Colcap.

Figura 14

Retornos de Ecopetrol, Colcap y TRM en el período 2018-2022



Nota. Elaborado a partir de la data descargada de Investing.com.

Ecopetrol, que pertenece al sector energético, es la empresa más grande de Colombia, con un valor de mercado de \$901.669.360.400 a noviembre 2023. A pesar de las coyunturas económicas por las que atraviese, ya sea por decisiones políticas o situaciones adversas tales como el derrame de crudo en el 2020, además de los conflictos internos que ha tenido el país desde la pandemia por covid-19, la empresa vendió \$38,9 billones para 2021, obteniendo un crecimiento del 20%. Así, Ecopetrol se reafirma en sus objetivos de gestión conservando una buena liquidez a nivel operativo y de producción (Ecopetrol, 2020).

4.3.3. Acción de Bancolombia

En un panorama general, la predicción del comportamiento de la acción Bancolombia, para la que el modelo *random forest* presenta con mejor precisión para predecir el precio, con un 71% y una afinidad moderada en la predicción tal y como se muestra en la tabla 6. El comportamiento del precio de esta acción por trimestre se puede consultar en el anexo 3.

Tabla 6*Desempeño de los modelos de machine learning con la acción de Bancolombia*

Modelo	Predicción	Verdaderos positivos	Falsos positivos	Verdaderos negativos	Falsos negativos	Precisión	Kappa
Random Forest (RF)	Alza (1)	119	56	156	61	0,71	0,4
	Baja (0)	156	61	119	56		
Decition Tree (DT)	Alza (1)	133	66	137	56	0,69	0,38
	Baja (0)	137	56	133	66		
Regresión Log (R-LOG)	Alza (1)	97	26	177	92	0,7	0,39
	Baja (0)	177	92	97	26		

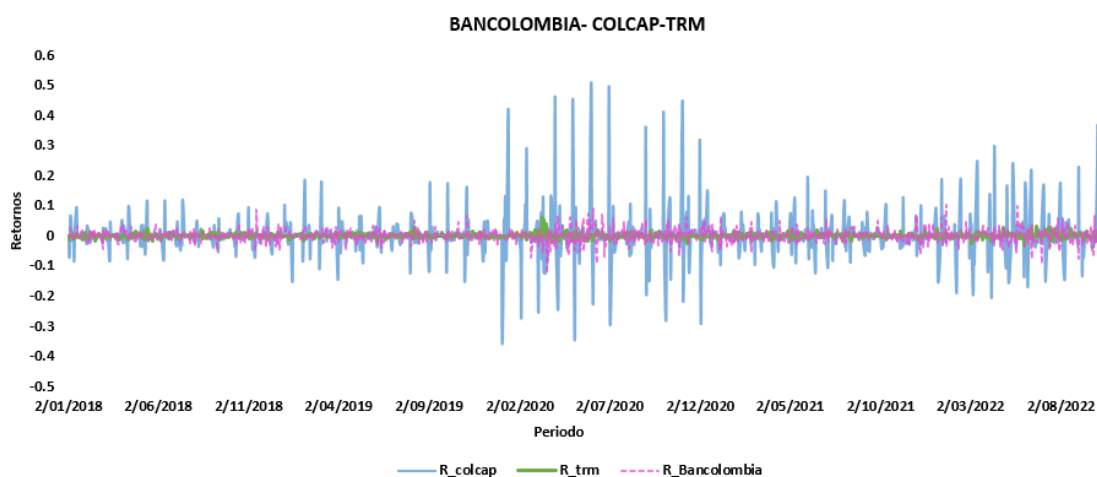
La tabla 6, muestra los modelos evaluados y calibrados para Bancolombia, se continuó con el ajuste de cada modelo a partir de los datos de entrenamiento, en la selección del mejor algoritmo de clasificación en conjunto es *random forest*.

El desempeño en conjunto para clasificar correctamente da una precisión del 71%, reflejando así que es adecuado según sus valores positivos correctamente clasificados (119).

Por otra parte, la métrica Kappa registrara un desempeño del 40%, sin caer en sobreajuste según los verdaderos positivos y verdaderos negativos. La figura 15 muestra los retornos de la acción Bancolombia para el periodo de estudio junto con los retornos de los indicadores Colcap y TRM.

Figura 15

Retornos de Bancolombia, Colcap y TRM en el período 2018-2022



Nota. Elaborado a partir de la data descargada de Investing.com.

A lo largo de varios años, Bancolombia ha consolidado su posición como un actor clave en el mercado de capitales colombiano (Bancolombia, 2022). Su relevancia para el país, así como el tamaño y la liquidez de sus acciones, confirman su papel privilegiado en las operaciones bursátiles diarias. También se destaca la sólida posición de Bancolombia en el mercado, evidenciada por su participación activa en la BVC y en el peso significativo de sus acciones en el índice Colcap (Mora, 2022). Asimismo, a pesar de los críticos desafíos del 2020, Bancolombia ha demostrado resiliencia, reflejada en atractivas rentabilidades y en la distribución de generosos dividendos a sus accionistas en 2021, lo que respalda la confianza en la entidad financiera.

4.3.4. Acción de Celsia

Evaluando el desempeño de los modelos de *machine learning* con la acción de Celsia, en la tabla 7 se aprecia una precisión entre el modelo *random forest* del 71%, y la regresión logística del 72%; sin embargo, el indicador Kappa presenta un resultado moderado, con una tasa de aciertos más precisa que la de los otros modelos planteados. El comportamiento del precio de esta acción por trimestre se puede consultar en el anexo 4.

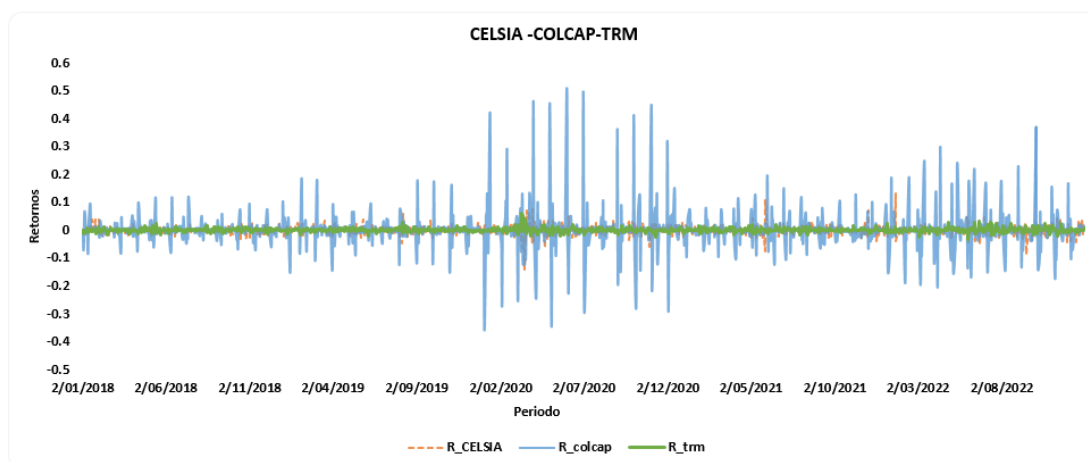
Tabla 7*Desempeño de los modelos de machine learning con la acción de Celsia*

Modelo	Predicción	Verdaderos positivos	Falsos positivos	Verdaderos negativos	Falsos negativos	Precisión	Kappa
Random Forest (RF)	Alza (1)	123	50	161	58	0,72	0,44
	Baja (0)	161	58	123	50		
Decition Tree (DT)	Alza (1)	134	77	125	56	0,66	0,32
	Baja (0)	125	56	134	77		
Regresión Log (R-LOG)	Alza (1)	108	43	167	74	0,7	0,39
	Baja (0)	167	74	43	108		

En cuanto al crecimiento de la demanda de energía eléctrica en Colombia, entre el 2000 y el 2020 este fue del 5% anual en promedio (Celsia, s. f.). En el 2019, hubo una caída significativa en la demanda eléctrica de Colombia (Castañeda, 2021). En el 2019, hubo una caída significativa en la demanda de energía eléctrica en Colombia. Para el 2021, la demanda de energía eléctrica en Colombia fue de 74.117 GWh-año, superior en 5,24% a la demanda en el 2020, cuando estuvo por el orden de 70.422 GWh, recuperándose tras la reducción de la demanda como consecuencia de la pandemia y estableciéndose como un sector prometedor (UPME, 2022). En la figura 16 se puede evidenciar a continuación el comportamiento histórico en los retornos y su relación con las variables predictoras.

Figura 16

Retornos de Celsia, Colcap y TRM en el período 2018-2022



Nota. Elaborado a partir de la data descargada de Investing.com.

En la tabla 8 se resume el porcentaje de precisión balanceada de los modelos propuestos, en donde 1 representa las predicciones al alza mientras que 0 las predicciones a la baja. Se destaca que el modelo *random forest* es aquel que mejor se ajusta a las acciones del Colcap utilizadas en la estrategia (Echeverri, 2021).

Tabla 8

Porcentaje de precisión balanceada de los modelos de machine learning

	Ecopetrol			Bancolombia			Celsia		
	RF	DT	R-log	RF	DT	R-log	RF	DT	R-log
1	72%	63%	77%	68%	67%	79%	71%	64%	70%
0	73%	69%	59%	72%	71%	66%	74%	69%	69%

4.4. Señales de compra y venta

Una vez se tienen los resultados del entrenamiento del modelo con su respectiva predicción, la *data set* se organizó creando un campo que permita conocer si se puede comprar o vender, o tomar una posición neutral. La decisión queda determinada de la siguiente manera:

- Si la predicción del movimiento del precio subió (es decir es igual a 1), entonces la decisión es comprar.
- Si la predicción del movimiento del precio bajó (es decir es igual a 0), entonces la decisión es vender o tomar una posición neutral.

La tabla 9 describe el comportamiento del portafolio de acuerdo con la variable respuesta definida, la predicción y la decisión de tomar una posición en corto o largo. Ecopetrol muestra un 51,79%, en términos de operar ya sea en comprar o vender (suma 25,77% y 26,02%); Bancolombia, 45,66%, y Celsia, 55,33%; sin embargo, se aprecia que la acción de Celsia representa una posición hacia la venta mayor que la de las demás las acciones del portafolio, y Ecopetrol una posición más acertada a comprar. En cuanto a la decisión de no operar, también está basado en el comportamiento del Colcap al momento de la operación.

Tabla 9*Predicción y decisión de posición para las acciones analizadas*

Ecopetrol			Bancolombia			Celsia		
Predicción	Decisión	Proporción	Predicción	Decisión	Proporción	Predicción	Decisión	Proporción
0	comprar	25,77%	0	comprar	22,70%	0	comprar	27,55%
1	vender	26,02%	1	vender	22,96%	1	vender	28,07%
0	No operar	25,00%	0	No operar	8,42%	0	No operar	27,55%
1	operar	23,21%	1	operar	45,92%	1	operar	16,84%
Total general		100%	Total general		100%	Total general		100%

4.5. Teoría de portafolio

A partir de los datos históricos de los rendimientos de las acciones del Colcap se obtuvieron las volatilidades de cada activo, y las covarianzas descontando la tasa IBR como libre de riesgo, tal y como sugiere Agudelo (2014).

Esto ha permitido obtener una volatilidad anual que se ha promediado por sector, resumido en la tabla 10, en donde se destaca que son el sector suministros, con una volatilidad media inferior, y el de consumo no cíclico y materiales básicos los de mayor volatilidad, corroborando lo afirmado por el diario *El Tiempo* (2022).

Tabla 10

Volatilidad media por sector Colcap

Sector	Volatilidad
Suministros	21,2%
Energía	26,7%
Financiero	27%
Materiales básicos	30,4%
Consumo no cíclico	36,9%

Con dichas volatilidades, e incorporando la rentabilidad RF de los bonos TES a 10 años y el rendimiento esperado de las acciones colombianas según Damodaran (2023), se obtuvieron los rendimientos esperados de cada activo y se usaron en combinaciones ponderadas para optimizar portafolios en función del riesgo admitido por el inversionista. De aquí se extrajeron los portafolios más dicientes en términos de rentabilidad y riesgo recopilados en la tabla 11

Tabla 11*Portafolios modelados*

Portafolio	σ_{Rp}	E[Rp]	R. Sharpe	Beta
Máxima rentabilidad	35%	21%	0,30	1,33
Máxima razón de Sharpe	16%	17%	0,41	0,86
Mínimo riesgo	12%	15%	0,34	0,50
<i>Machine learning</i>	20,60%	18,96%	0,40	1,06

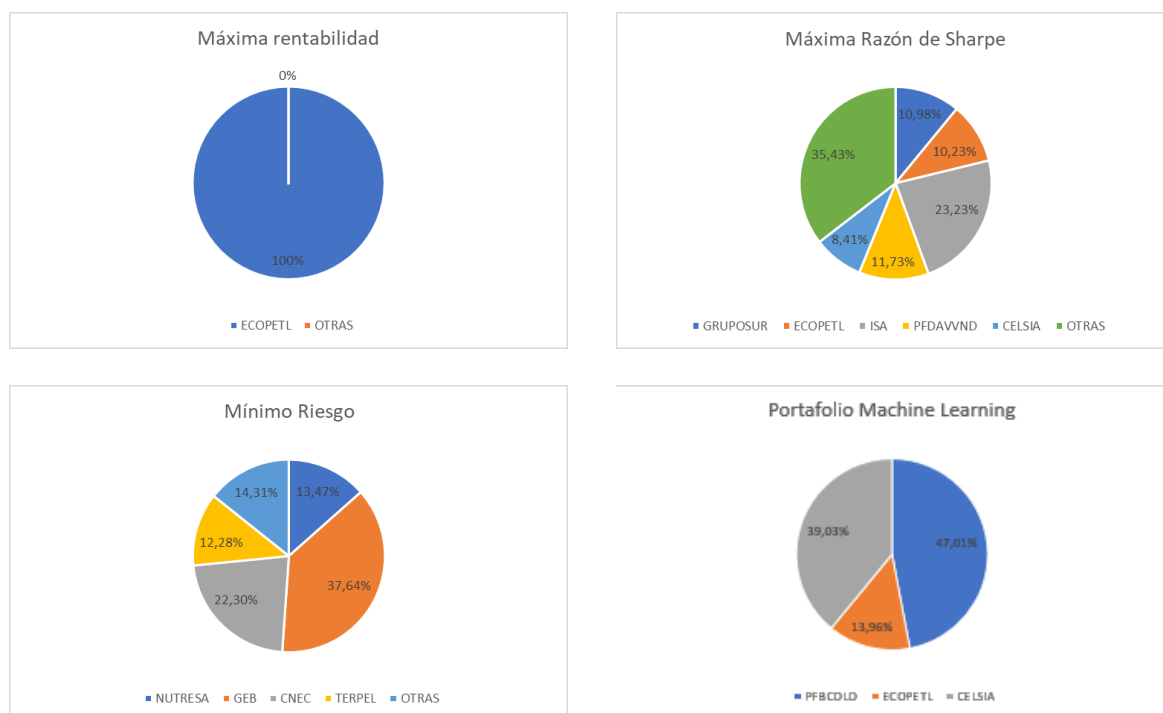
De los resultados observados en la tabla anterior se extrae que la máxima rentabilidad en la conformación de un portafolio se obtiene de invertir la totalidad del capital en el sector energía, particularmente en Ecopetrol.

Por otro lado, la mayor rentabilidad por unidad de riesgo, razón de Sharpe, se obtiene al diversificar el 65% de la inversión en los sectores suministros, financiero y energía. Por último, el mínimo riesgo proviene de invertir un 72% en los sectores energía y suministros.

En la figura 17 se puede observar la conformación de cada uno de los portafolios mencionados optimizados Colcap, a partir de la información analizada en el período 2018-2022.

Figura 17

Conformación portafolios optimizados Colcap en el período 2018-2022

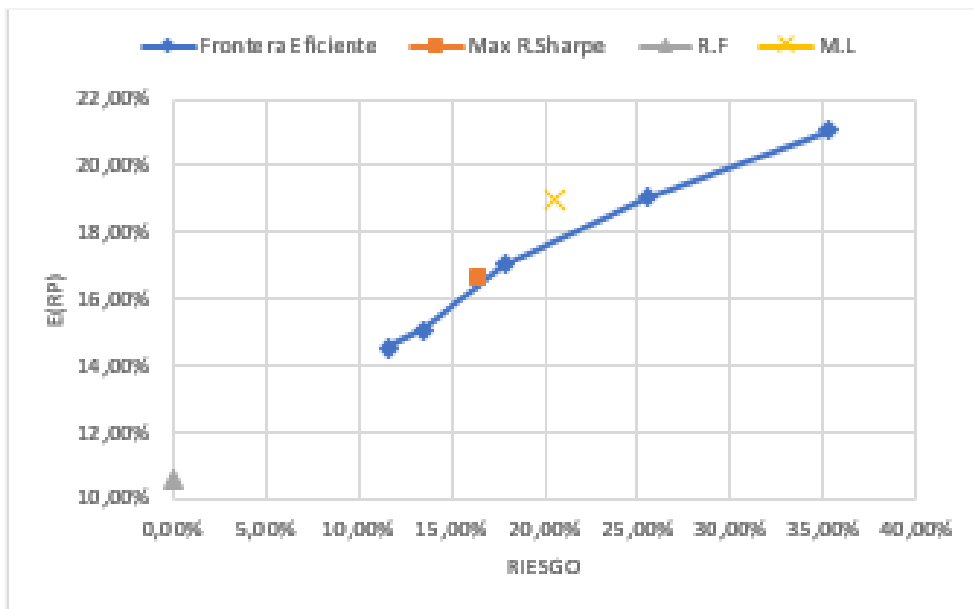


Las recomendaciones de los algoritmos de *machine learning*, por su parte, proponen la conformación de un portafolio de tres acciones, que son: Ecopetrol, Bancolombia y Celsia, basadas en los sectores energético, financiero y suministros.

En la figura 18 se presentan los diagramas de los portafolios obtenidos y la frontera eficiente de una de las iteraciones estocásticas en la conformación de portafolios.

Figura 18

Diagrama de los portafolios obtenidos y frontera eficiente de una de las iteraciones estocásticas



La gráfica anterior muestra cómo, a partir de las recomendaciones de la inteligencia artificial, se puede optimizar un portafolio que genere mayor rentabilidad por unidad de riesgo al tener en cuenta la tendencia en el precio futuro de la acción a partir de los datos históricos.

5. Conclusiones

Para responder al objetivo central de la presente investigación, el modelo que le permite al inversionista la selección del portafolio óptimo para activos financieros de renta variable en Colombia es *random forest*. Dicho modelo arrojó los mejores resultados en materia de precisión. Asimismo, los activos del Colcap que mejor se adecúan a un portafolio de inversión diversificado, según la inteligencia artificial, son Ecopetrol, Bancolombia y Celsia.

5.1. Inteligencia artificial, modelación, simulación

A mayor número de acciones incorporadas en el modelamiento con técnicas de inteligencia artificial, se encontró una mayor imprecisión en la predicción de los diferentes modelos estudiados, lo que fue solucionado mediante la aplicación de la estrategia de inversión. Esto se debe a la naturaleza volátil diferenciada de cada acción y de cada sector, lo cual influye en la decisión final del mejor modelo de *machine learning*.

Para la elección de las variables que mejor describen las tendencias futuras en el precio para cada acción, los procesos de *forward* y *backward* son fundamentales, en simultánea con la evaluación de los modelos por medio de la matriz de confusión y de otros indicadores asociados, como lo son la precisión y Kappa.

El modelo con mejor rendimiento, y el que arrojó mejores resultados en términos de predicción en los cambios de los precios para las acciones Colcap, ha sido

random forest, que en la matriz de confusión obtuvo una precisión entre el 71% y 74% en las acciones estudiadas. Esto puede atribuirse a la ventaja de interactuar mucho mejor entre variables cuando se ha estimado el modelo y a su capacidad de realizar numerosas iteraciones.

Para lograr un portafolio más eficiente, las predicciones en los precios de las acciones Colcap obtenidas por la inteligencia artificial y la simulación fueron complementados con estadísticas, criterios de volatilidad, defensivo, liquidez y juicio de expertos. En este orden de ideas, a partir de los resultados mencionados se toman las predicciones de cada acción y se complementa el análisis con un criterio que permita decidir cuándo comprar o cuando vender.

Al comprar una acción, contra el índice representativo Colcap y por defecto S&P500, siempre que esté por debajo, este precio bajo indica que puede ser aprovechado para comprar, ya que, de acuerdo con cifras del Gobierno nacional, hay expectativas de crecimiento a partir del 2023 sobre todo para el sector energético, lo que se traduce en confianza en la inversión.

5.2. Rentabilidad

La rentabilidad obtenida en esta investigación se ha planteado bruta y, por ende, no considera gastos de transacciones ni retenciones a las que haya lugar por las ganancias obtenidas. Al incluir los costos transaccionales en el análisis será más provechoso seguir la estrategia que realice menos transacciones, en donde la

estrategia obtenida por el análisis de datos genera mayor valor y esté respaldado en la historia reciente del comportamiento de las acciones colombianas.

De igual manera, vale la pena resaltar que la conformación de portafolios ha sido *long-only*, lo que indica que, a partir de las señales de crecimiento en ciertos activos obtenidos por la inteligencia artificial, se toma una posición larga en estos prospectos más prometedores.

Rentabilidades más altas pueden ser obtenidas a través de operaciones *intra-day*, lo que implica estar realizando constantemente operaciones de compra y venta en función de las predicciones en los precios de las acciones, dando lugar a modelaciones en *software* más complejas y a diferentes requerimientos computacionales.

Las rentabilidades obtenidas en acciones Colcap, sin apalancamiento y sin combinaciones con la tasa libre de riesgo, oscilan entre un 14% y un 21%. Las recomendaciones de inversión a partir de la inteligencia artificial se encontraron sobre la frontera eficiente, y guardan consistencia con los resultados de la metodología tradicional. Para ambas metodologías se encontró que los sectores de energía y suministros son los prometen mayores rentabilidades futuras, respaldado con el juicio de expertos y con un análisis fundamental.

6. Recomendaciones

Para futuras investigaciones en los campos tanto de análisis de datos como de finanzas se podrían abordar los siguientes aspectos:

- Utilizar las técnicas de *machine learning* para programar la teoría tradicional de portafolio y determinar su eficiencia en la optimización de portafolios en comparación con el Solver de Excel.
- Incluir los perfiles de riesgo del inversionista en la optimización con *machine learning*, para comparar rentabilidades esperadas y máximas pérdidas esperadas con los métodos tradicionales.
- Restringir las transacciones de compra y venta de acciones a partir de la adición de costos transaccionales por operación e incluir combinaciones con la tasa libre de riesgo, en aras de obtener rentabilidades netas ajustadas a la capacidad de las manos débiles del mercado.
- Aplicar las metodologías de *machine learning* en otros mercados; por ejemplo, MILA, para determinar su comportamiento ante la inclusión del riesgo por tasa de cambio, y ayudar a generar valor a una plataforma que no ha tenido el impacto en las operaciones y volúmenes de negociación esperados (Montes, 2021).
- Dentro de las alternativas para alternar el riesgo de volatilidad, liquidez y una acción estable, de acuerdo con los análisis vistos, las acciones de Argos, Promigas

o Bancolombia son buenas opciones para alternar frente a los choques económicos del país.

- Al incluir otra acción para la combinación del portafolio, deben considerarse otros factores importantes que cambian en función del contexto macroeconómico por el que atraviese el país. De esta manera, se optimiza el modelo de IA, para minimizar el riesgo inherente y maximizar rentabilidad, ya que ayuda a mejorar el modelo existente, o explorar otros que también sean robustos para predecir el movimiento del precio.

- Botía, J. A. (4 de octubre, 2007). *Árboles de decisión en aprendizaje automático y minería de datos*. UNAM.
http://fcaenlinea1.unam.mx/anexos/1566/1566_u6_act1a.pdf
- Caparrini López, A. (2021). *Optimización de cartera de activos financieros aplicando aprendizaje automático* [trabajo fin de Máster, Universitat Oberta de Catalunya]. Repositorio Institucional.
<https://openaccess.uoc.edu/bitstream/10609/147326/7/acaparriniTFM0123presentacion.pdf>
- Castañeda, M. P. (18 de marzo, 2021). Colombia | Sector eléctrico colombiano: retos y oportunidades. *BBVA Research*.
<https://www.bbva.com/publicaciones/colombia-sector-electrico-colombiano-retos-y-oportunidades/>
- Castro Buitrago, O. I., y Guerrero Calderón, D. G. (2021). *Modelo predictivo de propensión de ahorro e inversión en productos de banca patrimonial* [tesis de Maestría, Universidad de los Andes]. Repositorio Institucional.
<https://repositorio.uniandes.edu.co/entities/publication/836a54e7-1b82-44dd-9342-e70065e7f72f>
- Castro Mejía, A. F. (10 de junio, 2020). *Machine learning investment strategies and the liquidity premium for the Colombian yield curve*. EAFIT.
<https://repository.eafit.edu.co/server/api/core/bitstreams/867953a3-d654-485a-8491-bc8e637987b8/content>
- Celsia (s. f.). *Qué hacemos*. <https://www.celsia.com/es/quienes-somos/que-hacemos/>
- Chen, M., Zhang, Z., Shen, J., Deng, Z., He, J., & Huang, S. (2020). A Quantitative Investment Model Based on Random Forest and Sentiment Analysis. *Journal*

of *Physics: Conference Series*, 1575. <https://doi.org/10.1088/1742-6596/1575/1/012083>

Choudhry, R., & Garg, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting. *World Academy of Science, Engineering and Technology*, 39, 315-318.

Coronado Vaca, M. (2022). *Aplicación del machine learning al factor investing en renta fija corporativa* [tesis de Maestría, Comillas Universidad Pontificia]. Repositorio Institucional. <https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/56412/TFG%20-%20Periel%20Lopez%2c%20Jose.pdf?sequence=2>

Damodaran (octubre, 2023). *Country Default Spreads and Risk Premiums*. https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/ctryprem.html

De Lara Haro, A. (2005). *Medición y control de riesgos financieros* (3.^a ed.). Limusa.

Echeverri Sánchez, L. C. (2021). *Conformación automática de portafolios de inversión usando analítica financiera* [tesis de Maestría, Universidad Nacional de Colombia]. Repositorio Institucional. <https://repositorio.unal.edu.co/handle/unal/81198>

Ecopetrol (2020). *Informe de gestión*. <https://www.ecopetrol.com.co/wps/portal/Home/es/ResponsabilidadEtiqueta/InformesGestionSostenibilidad/Informesdegestion>

El Tiempo (18 de diciembre, 2022). *Los hechos que fueron noticia en Colombia durante este 2022*. <https://www.eltiempo.com/colombia/los-hechos-que-fueron-noticia-en-colombia-durante-este-2022-726836>

- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3-56.
[https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
- Fama, E. F., & French, K. R. (2005). A five-factor asset pricing model. *Journal of Financial Economics*, 116, 1-22.
<https://doi.org/10.1016/j.jfineco.2014.10.010>
- Gómez, L. M. (12 de septiembre, 2022). Acciones de Ecopetrol y Bancolombia continúan como las de mayor liquidez a agosto. *La República*.
<https://www.larepublica.co/finanzas/acciones-de-ecopetrol-y-bancolombia-continuan-como-las-de-mayor-liquidez-a-agosto-3445362>
- IBM (s. f.). ¿Qué es el etiquetado de datos? <https://www.ibm.com/es-es/topics/data-labeling#:~:text=Las%20etiquetas%20identifican%20los%20vectores,a%20realizar%20las%20mejores%20predicciones>
- Koehrsen, W. (27 de diciembre, 2017). Random Forest Simple explanation. *Medium*. <https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d>
- Leung, M. T., Daouk, H., & Chen, A.-S. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*, 30(6), 901-923.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=277774
- Madeeh, O., & Abdullah, H. (2021). *An Efficient Prediction Model based on Machine Learning Techniques for Prediction of the Stock Market*. *Journal of Physics: Conference Series*, 1804. <https://doi.org/10.1088/1742-6596/1804/1/012008>

Mangram, M. E. (2013). A simplified perspective of the Markowitz portfolio theory. *Global Journal of Business Research*, 7(1), 59-70.
<https://ssrn.com/abstract=2147880>

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

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press.

Mendizábal Zubeldía, A., Miera Zabalza, L. M., y Zubia Zubieaurre, M. (2002). El modelo de Markowitz en la gestión de carteras. *Cuadernos de Gestión*, 2(1), 33-46.
https://www.academia.edu/27207556/El_modelo_de_Markowitz_en_la_gest_i%C3%B3n_de_carteras

Ministerio de Hacienda (2005). *Parte 11 autoridades: Libro 2 superintendencia financiera*.
https://www.urf.gov.co/webcenter/ShowProperty?nodeId=/ConexionContent/WCC_CLUSTER-107279

Miskolczi, P. (2017). Note on simple and logarithmic return. *Applied Studies in Agribusiness and Commerce – Abstract*, 11(1-2), 127-136.
<https://doi.org/10.19041/APSTRACT/2017/1-2/16>

Montes, S. (19 de julio, 2021). Mila: ¿Qué tan integrado está diez años después? *Forbes Colombia*. <https://forbes.co/2021/07/19/negocios/mila-que-tan-integrado-esta-diez-anos-despues>

Mora Aguilar, K. V. (21 de noviembre, 2022). Bancolombia es la acción más atractiva según encuesta de Opinión Financiera (EOF). *La República*.
<https://www.larepublica.co/finanzas/bancolombia-es-la-accion-mas-atractiva-segun-encuesta-de-opinion-financiera-eof-3493059>

Nickolas, S. (8 de julio, 2023). SPY ETF: The SPDR S&P 500 ETF Trust and What It Holds. *Investopedia*.

<https://www.investopedia.com/articles/investing/122215/spy-spdr-sp-500-trust-etf.asp>

Rubiales, E. (2018). Kappa de Cohen. *Samiuc*.

<https://www.samiuc.es/estadisticas-variables-binarias/medidas-de-concordancia/kappa-de-cohen/>

Saetia, K., & Yokrattanasak, J. (2022). Stock Movement Prediction Using Machine Learning Based on Technical Indicators and Google Trend Searches in Thailand. *International Journal of Financial Studies*, 11(1), 5.

<https://doi.org/10.3390/ijfs11010005>

Sharpe, W. F. (1964), capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425-442.

<https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>

Surayagari, H. K. S., Anderson, C., Ben-Hur, A., & Stein, C. (2021). *Stock market predictions using machine learning* [tesis de Maestría, Colorado State University]. Repositorio Institucional.

<https://mountainscholar.org/items/1306b5ad-f62d-40f4-8fa8-7c36c5e84b48>

TradingView (s. f.). Acciones colombianas con mayor volatilidad. *Acciones colombianas*. Consultado el 1 de noviembre de 2023.

<https://es.tradingview.com/markets/stocks-colombia/market-movers-most-volatile/>

Unir (7 de mayo, 2021). Árboles de decisión: en qué consisten y aplicación en Big Data. *Ingeniería y Tecnología*.

<https://www.unir.net/ingenieria/revista/arboles-de-decision/>

Unidad de Planeación Minero Energética – UPME (2022). *Proyecciones de demanda*. <https://www1.upme.gov.co/DemandayEficiencia>

Wei, X., Chen, W., & Li, X. (2021). Exploring the financial indicators to improve the pattern recognition of economic data based on machine learning. *Neural Computing & Applications*, 33(2), 723-737. <https://doi.org/10.1007/s00521-020-05094-0>

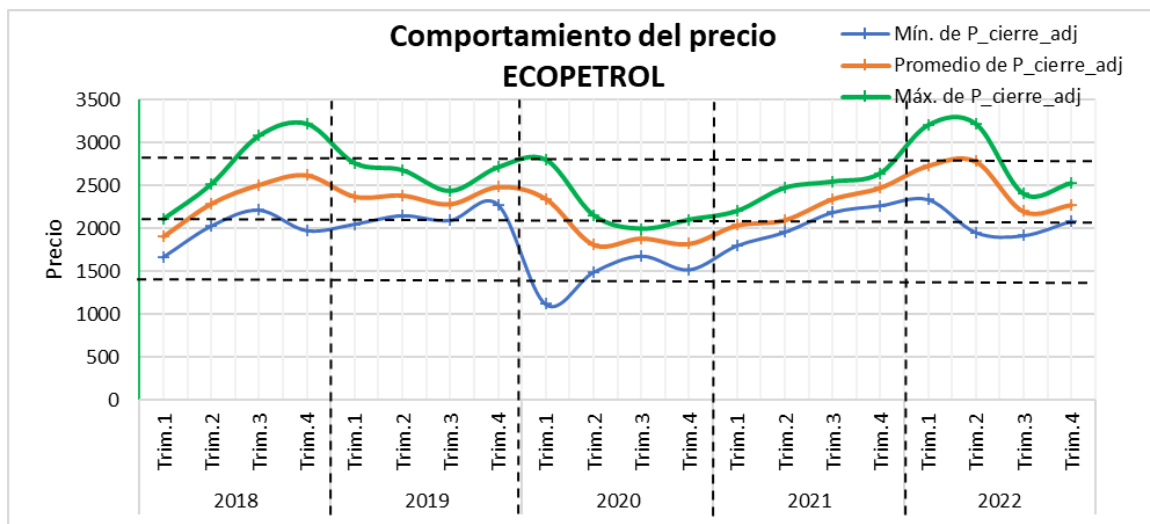
Anexos

Anexo 1 Estadísticas después de la simulación de los tres modelos *machine learning* (2018-2022)

ECOPETROL	Min	Max	Mean	Std. deviation	Variance	Skewness	Kurtosis	Overall sum
volumen	-	308,627,836.0000	9,738,503,1043	11,572,476,8500	133,922,220,442,674.0000	14,8423	356,1659	12,699,008,047,9999
R_min	0.9480	0.1068	0.0004	0.0357	0.0013	14.7089	381.2361	0.4645
R_max	0.9480	0.1395	0.0004	0.0350	0.0012	15.2183	413.5101	0.5080
R_cierre_ADJ	0.9575	0.1310	0.0002	0.0359	0.0013	14.8901	388.2038	0.2244
R_P-adj_spy	0.9175	0.2758	0.0006	0.0534	0.0029	3.8851	70.8806	0.8462
R_P-adj_colcap	0.5883	0.5091	0.0013	0.0657	0.0043	1.2407	22.3323	1.7267
R_trm	0.0347	0.0611	0.0004	0.0072	0.0001	1.0887	8.1825	0.4738
BANCOLOMBIA	Min	Max	Mean	Std. deviation	Variance	Skewness	Kurtosis	Overall sum
volumen	-	992,865,662.0000	986,889,9295	27,480,169,9810	755,159,742,182,586.0000	36,1165	1,304,6007	1,287,865,258,0000
R_min	0.1415	0.3654	0.0000	0.0183	0.0003	5.9408	128.2052	0.0528
R_max	0.0888	0.3868	0.0001	0.0179	0.0003	8.0884	171.5870	0.0655
R_cierre_ADJ	0.2320	0.3735	0.0000	0.0198	0.0004	4.0880	121.2785	0.0134
R_P-adj_spy	0.9175	0.2758	0.0008	0.0534	0.0029	3.8851	70.8806	0.8462
R_P-adj_colcap	0.5883	0.5091	0.0013	0.0657	0.0043	1.2407	22.3323	1.7267
R_trm	0.0347	0.0611	0.0004	0.0072	0.0001	1.0887	8.1825	0.4738
CELSIA	Min	Max	Mean	Std. deviation	Variance	Skewness	Kurtosis	Overall sum
volumen	-	992,865,662.0000	986,889,9295	27,480,169,9810	755,159,742,182,586.0000	36,1165	1,304,6007	1,287,865,258,0000
R_min	0.1415	0.3654	0.0000	0.0183	0.0003	5.9408	128.2052	0.0528
R_max	0.0888	0.3868	0.0001	0.0179	0.0003	8.0884	171.5870	0.0655
R_cierre_ADJ	0.2320	0.3735	0.0000	0.0198	0.0004	4.0880	121.2785	0.0134
R_P-adj_spy	0.9175	0.2758	0.0006	0.0534	0.0029	3.8851	70.8806	0.8462
R_P-adj_colcap	0.5883	0.5091	0.0013	0.0657	0.0043	1.2407	22.3323	1.7267
R_trm	0.0347	0.0611	0.0004	0.0072	0.0001	1.0887	8.1825	0.4738

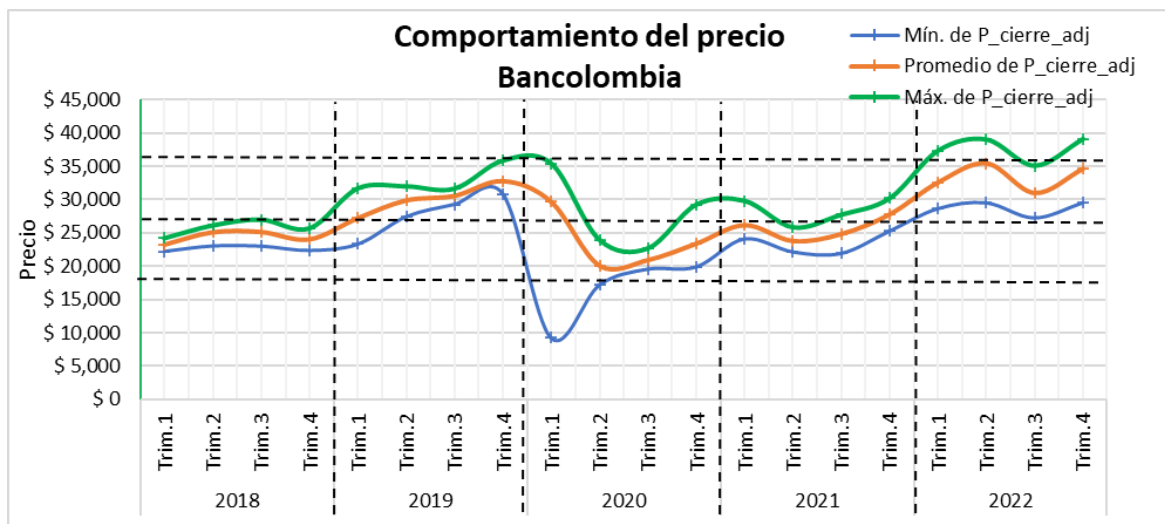
Nota. Elaboración propia a partir de los datos recopilados de Yahoo Finance en el periodo (2018-2022).

Anexo 2 Comportamiento del precio de la acción de Ecopetrol en el período 2018-2022



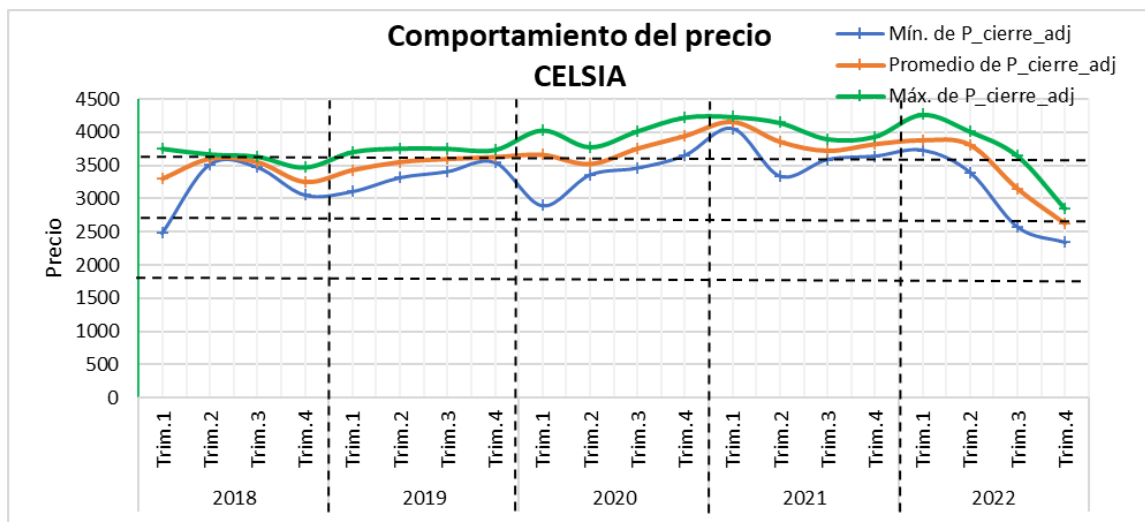
Nota. Elaboración propia a partir de los datos recopilados de Yahoo Finance en el periodo (2018-2022).

Anexo 3 Comportamiento del precio de la acción de Bancolombia en el período 2018-2022



Nota. Elaboración propia a partir de los datos recopilados de Yahoo Finance en el periodo (2018-2022), unidades en miles de pesos colombianos

Anexo 4 Comportamiento del precio de la acción de Celsia en el período 2018-2022



Nota. Elaboración propia a partir de los datos recopilados de Yahoo Finance en el periodo (2018-2022)