

# Predicción de precios de 3 acciones del mercado de valores colombiano para el segundo semestre de 2023

Luis Ángel Rodríguez Hernández <sup>1,\*</sup>

<sup>1</sup>Estudiante de Maestría en Ingeniería, Fundación Universitaria Los Libertadores

\*Autor de correspondencia: larodriguezh@libertadores.edu.co.



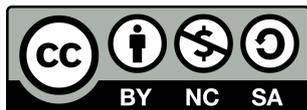
Facultad de Ingeniería y  
Ciencias Básicas



Recibido: 4 de julio de 2023

Aceptado: 4 de agosto de 2023

Publicado: 24 de noviembre de 2023



Copyright: ©2023 por los autores. Este artículo es de acceso abierto distribuido bajo los términos y condiciones de Creative Commons Licencia de atribución (CC BY NC SA) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

## Resumen

Los inversionistas del mercado de valores en Colombia y en cualquier parte del mundo requieren pronósticos del precio de las acciones. Por ello se apoya en econometría, estadística y aprendizaje automático. El pronóstico de mercados financieros generalmente se modela como una serie temporal donde los datos no son independientes entre sí y la secuencia junto con orden temporal son muy significantes.

Como resultado de la revisión de literatura y del preprocesamiento se obtuvo que las técnicas autoregresivas para series temporales son muy idóneas para resolver y modelar esta situación.

En este estudio se realizó predicción del precio de las acciones de Bancolombia, Davivienda y Ecopetrol para el segundo semestre de 2023. Los datos se obtuvieron de <https://es.investing.com/> desde el 1 de enero de 2018 hasta el 16 de junio de 2023. Se entrenaron 3 máquinas de aprendizaje (una por cada acción), las cuales fueron consumidas por un módulo web creado en Streamlit.

Para los tres activos, los resultados fueron de alta variabilidad para los 2 primeros meses del segundo semestre y estabilidad sin tendencia (ni alcista ni bajista) para los cuatro últimos.

**Palabras clave:** Predicción, Pronósticos, Mercado de valores, Series temporales, Modelos autoregresivos

## Como citar este artículo

Luis Ángel Rodríguez Hernández. "Predicción de precios de 3 acciones del mercado de valores colombiano para el segundo semestre de 2023", *Revista Apuntes de Ciencia e Ingeniería*, 2, 1, nov, pag 5-12. 2023. doi: [10.37511/apuntesci.v2n1a2](https://doi.org/10.37511/apuntesci.v2n1a2)

## 1. Introducción

El futuro ha sido una de las obsesiones más grandes del hombre a través de la historia, es decir, conocer lo que sucederá y anticiparse a los hechos. No obstante, la incertidumbre es una realidad que lo acompaña y con la cual ha aprendido a vivir para posteriormente desafiarla mediante la predicción; para ello, el hombre ha acudido a la parapsicología y clarividencia, como también a la ciencia y las matemáticas. Por fortuna para la humanidad, estas últimas se han visto favorecidas por el uso de la potencia, automatización y velocidad de cálculo que ofrecen las computadoras, como también a la vastísima cantidad de datos que se han producido hasta la actualidad; lo cual permite aplicar los recursos científicos y tecnológicos para lograr los pronósticos que requiere el hombre moderno.

Este estudio tiene como propósito predecir el precio de 3 acciones del mercado de valores colombiano para el segundo semestre de 2023 empleando técnicas de aprendizaje maquina y análisis de series temporales. Las acciones estudiadas fueron tres (Ecopetrol, Bancolombia y Davivienda) importantes activos del mercado colombiano y por lo tanto la unidad de medida del precio está dada en pesos de Colombia (COP). El conjunto de datos empleados tanto para el entrenamiento como para las pruebas de aprendizaje corresponde a una serie diaria de precios de cada activo que contiene datos desde enero 1ro de 2018 hasta junio 16 de 2023. Otras investigaciones similares se han aplicado a otros países, sin embargo, este estudio fue aplicado al mercado colombiano. Hablando de tiempo, el pronóstico fue aplicado al segundo semestre del año 2023. El tipo de análisis que se desea realizar es de tipo predictivo-descriptivo, es decir, realizar predicción y describir el pronóstico.

## 2. Marco teórico

En un comienzo se ha necesitado predecir determinados fenómenos naturales, tal es el caso del clima, terremotos, huracanes, eclipses, entre otros. No obstante además de los ya mencionados, se requiere realizar pronósticos de fenómenos modernos tales como el comportamiento del mercado de valores, y de esta manera tomar decisiones de tipo económico. Sin embargo, esta tarea no es fácil debido a su naturaleza compleja e inestable [1]. Es muy importante pronosticar actualmente para el proceso de toma de decisiones inteligentes [2]. Además, la ingente cantidad de datos (bigdata) y lo anteriormente mencionado ha contribuido al desarrollo de nuevos métodos científicos e ingenieriles para almacenar, procesar y analizar datos, contribuyendo a información y conocimiento en diferentes sectores [3], de los cuales el mercado accionario no escapa. No se puede dejar de lado los modelos econométricos tradicionales, que en convivencia con la tecnología de machine learning forman una alianza estratégica [4]. El aprendizaje maquina se extiende a la economía y las finanzas por los avances tecnológicos, lo cual implica un cambio de paradigma respecto a los tradicionales modelos de inferencia estadística y econometría clásica [4].

En cuanto al aprendizaje maquina, se aplican algunos métodos que de manera armoniosa combinan la estadística junto con la ciencia de datos con el firme propósito de lograr pronósticos de alta calidad de precisión, pues los avances recientes en estadística han permitido el desarrollo de técnicas automatizadas más potentes, principalmente de regresión y clasificación [5]. En el caso de la regresión, la cual intenta modelar la relación entre variables predictoras (independientes) y una variable de pronóstico (variable dependiente). De esta técnica existe un caso particular llamado regresión lineal simple, donde se intenta conocer la forma como aumenta o disminuye la variable dependiente respecto a la independiente [5], en la cual existe una variable predictor; mientras que en la regresión lineal múltiple serán varias [5]. Otro tipo de regresión es la polinómica, donde se modela la relación entre variables mediante expresiones con términos polinomiales [5]. En esta técnica se pueden presentar ondulaciones no deseables las cuales pueden ser resueltas mediante interpolaciones suaves entre puntos llamados Splines [5].

Como complemento a las anteriores está la correlación, que se apoya en 2 importantes conceptos: El coeficiente y la matriz de correlación. El primero, mide que tan fuerte es la relación entre 2 variables [6], las cuales deben ser de tipo numérico (no aplica para variables categóricas) y cuyo valor debe estar en el rango de -1 a 1, siendo los valores extremos indicadores de alta correlación, mientras que el cero y sus valores cercanos indican baja o nula correlación [5]. Por otro lado, la segunda es una matriz en la cual se disponen en filas y columnas formando celdas que contiene el coeficiente de correlación entre las



diferentes variables [5].

Otra técnica que es apta para enfrentar retos de regresión como de clasificación, son las máquinas de soporte vectorial. Estas construyen su solución basándose en subconjuntos de datos de entrenamiento llamados vectores de soporte que tienen como objetivo encontrar un límite para separar dos clases de objetos [3]. La separación (forma del límite) puede ser lineal o no serlo. El primer caso es aplicado cuando un conjunto de datos se puede separar linealmente por una línea recta que permite la clasificación (de tipo binario) [3]. Esta línea de separación también se llama calle y deber ser lo más estrecha posible. Cuando el conjunto de datos no es separable linealmente, se introducen algunas variables no negativas de tal manera que permite formulación mediante Multiplicadores de Lagrange y Transformación de Wolfe [3]. Por otro lado, los avances de la estadística moderna han permitido el desarrollo y aplicación de técnicas automatizadas de mayor potencia para hacer predicciones. Un ejemplo es el método de K-vecinos, que consiste en clasificar registros de acuerdo a la forma como han sido clasificados otros de similares características [5].

Las anteriores técnicas están clasificadas dentro de la categoría del aprendizaje automático supervisado, ya que se basan en un conjunto de datos de entrenamiento, y usan otro para el aprendizaje. En contraposición a esto, existen los métodos de aprendizaje automático no supervisado, por ejemplo: Agrupación K-means, que consiste en dividir los datos en grupos cuyos registros cercanos tienen muchas características similares [5]. Existe además una técnica de agrupación alternativa, la jerárquica, que permite visualizar el efecto de especificar diferentes números de agrupaciones facilitando la representación gráfica para la interpretación de los grupos [5].

Otra parte de la inteligencia artificial, es el aprendizaje profundo siendo representado principalmente por las redes neuronales artificiales. Aunque sabemos que las neuronas biológicas son células nerviosas en el cerebro que se conectan con otras transmitiendo señales químicas y eléctricas [7], también existen las neuronas artificiales, las cuales pueden ser representar una tarea de clasificación binaria mediante una función de decisión que se conecta con otra formando una estructura [7]. La estructura más usual (multicapa) consiste en capas bien diferenciadas: una capa de entrada, otra de salida, y una intermedia que normalmente está oculta. Las entradas junto con sus pesos asociados alimentan a la red, y estos valores son enviados a las capas intermedias, lugar donde los pesos son recalculados y la red se adapta. De aquí se obtiene un resultado que puede ser enviado nuevamente a la capa de entrada generando retroalimentación. Finalmente, el resultado definitivo es enviado a la capa de salida para lograr de esta manera el valor pronosticado [3].

El pronóstico de mercados financieros se modela como una serie temporal, donde los datos no son independientes entre sí y la secuencia junto con orden temporal son muy significantes [3]. El análisis de series temporales ha sido muy importante para el estudio de los mercados financieros durante muchos años [8]. Un caso particular de análisis es la predicción lineal o autoregresión, en la cual es común tomar X como variable dependiente que se traza como función del tiempo y la predicción tiene como objetivo una fórmula que pronostique entradas de la serie a partir de entradas anteriores [3]. También son conocidos las técnicas ARIMA (Auto Regressive Integrated Moving Average), estos sirven para modelar series estacionarias, y además permiten predecir una vez se han realizado ajustes de acuerdo a las propiedades estadísticas de la serie [3].

Como se mencionó en un principio, la incertidumbre es una realidad que acompaña al hombre con la cual ha aprendido a vivir y también a enfrentar. En la actualidad sus armas son la estadística y la tecnología. La estadística le ha enseñado a realizar juicios inteligentes y tomar decisiones informadas para enfrentar la incertidumbre y la variación [6]. La tecnología, destinada a la manipulación, gestión y análisis de grandes volúmenes de datos para extraer valor y beneficio a alta velocidad y a menor coste [3].

### 3. Metodología

Para comenzar con la fundamentación teórica, se realizó revisión sistemática de literatura relacionada a predicción de mercado accionario aplicando aprendizaje maquina, análisis de sentimientos y diferentes teorías económicas. La búsqueda original arrojó como resultado que las mejores técnicas de predicción de



mercados eran los pronosticadores autoregresivos. Por otro lado, el análisis de sentimientos era más apto para aplicarse a conjuntos de datos que contengan texto (opiniones, tweets, comentarios, entre otros), sin embargo, para el tipo de conjuntos de datos que fue empleado en el presente estudio no era viable debido a que las variables son plenamente numéricas.

Para la obtención de datos se procedió a descargar desde el sitio web <https://es.investing.com/> el histórico de precios para cada una de las siguientes acciones del mercado de valores: Ecopetrol, Bancolombia y Davivienda. El criterio de selección consistió en buscar los activos cuyo conjunto de datos estaba actualizado y que tuviera la menor cantidad de datos faltantes en la serie temporal. La descarga se hizo aplicando filtro de fechas desde el año 2018 (no hay datos de años anteriores a la fecha de este informe) hasta el 16 de junio de 2023. Otro filtro aplicado fue la periodicidad del plazo, la cual se seleccionó diaria con el fin de tener una temporalidad con el mayor detalle posible. El resultado de la obtención de datos es un archivo CSV (comma-separated values) por cada acción del mercado, cuya estructura está conformada por: fecha, cierre, apertura, máximo, mínimo, volumen, y finalmente, el porcentaje de variación.

También, se realizó entendimiento de los conjuntos de datos, es decir, conocer que significa cada una de las variables, el tipo de datos y evaluar la forma de procesamiento. Por lo tanto, se obtuvo y documentó los metadatos (datos sobre datos que amplían el conocimiento preliminar de un conjunto de datos [9]).

Los datos obtenidos requirieron una preparación para su posterior procesamiento. Para que el lenguaje de programación pueda interpretar correctamente las variables de importes (Último, Apertura, Máximo, Mínimo) se eliminó el separador de miles, se reemplazó el separador de fracción decimal de coma a punto, y se cambió el tipo de datos de texto a tipo numérico. Sin embargo, la única columna que se tuvo en cuenta fue la de Último y además fue renombrada por Precio. Se eliminó la columna de Volumen y Porcentaje de Variación ya que no se necesitan para este estudio. Por otro lado, se cambió el tipo de dato de la variable fecha, de texto a tipo fecha con el formato YYYY-MM-DD. Para finalizar la preparación, se verificó la serie temporal y se llenaron los datos vacíos con el precio inmediatamente anterior. Esto porque el mercado de divisas en Colombia no opera los fines de semana y por lo tanto los precios permanecen constantes, es decir, que los viernes serán fijados al sábado, domingo y lunes cuando este último es feriado.

Una vez los datos han sido preparados, se procede a realizar un pre-procesamiento con el objetivo de determinar cuál técnica de machine learning será empleada para procesar los datos. En primera instancia se obtiene la matriz de correlación entre variables para validar la viabilidad de aplicar algún modelo de regresión lineal. Posteriormente se probó con regresión lineal simple, múltiple y máquinas de soporte vectorial. De las anteriores, ninguna prometía obtener buenos resultados de predicción. Para finalizar, se probó con pronosticadores autoregresivos en series temporales, que arrojó los mejores resultados y por lo tanto fue elegida para la generación del modelo.

Para cada una de las 7 acciones del mercado de valores se realizó partición del conjunto de datos obtenido y preparado previamente así: 70 % para el entrenamiento y 30 % para la prueba del modelo. Con lo anterior se procedió a generación del modelo, para luego y persistirlo (guardarlo en disco) en archivo binario para su posterior ejecución. Se crearon 3 archivos (máquinas de aprendizaje) para cada acción.

Para visualizar la predicción en forma gráfica y tabular se creó una aplicación web con la librería Streamlit (un marco de trabajo de código abierto que permite crear aplicaciones web a los investigadores para compartir conjuntos de datos de sus estudios [10]). La aplicación desarrollada permite al usuario final seleccionar cuantos días a futuro desea pronosticar (cuyo valor puede estar entre 1 y 200 días a futuro) ejecutando las 3 máquinas de aprendizaje entrenadas anteriormente de acuerdo a selección en una lista desplegable (control de interfaz de usuario). El despliegue de la aplicación se hace en el protocolo de comunicaciones https por el puerto 8501.

En resumen, se obtienen, entienden, analizan y preparan los datos. Además se entrena un modelo que luego será persistido en un archivo Pickle con extensión PKL. Posteriormente, se hace la predicción en un módulo web que facilita la interacción del usuario con la máquina de aprendizaje de la siguiente manera: se parametriza el modelo restaurado desde el archivo PKL para luego ejecutarse y finalmente mostrar los



resultados de la predicción en el módulo web en un navegador de internet (Figura 1).

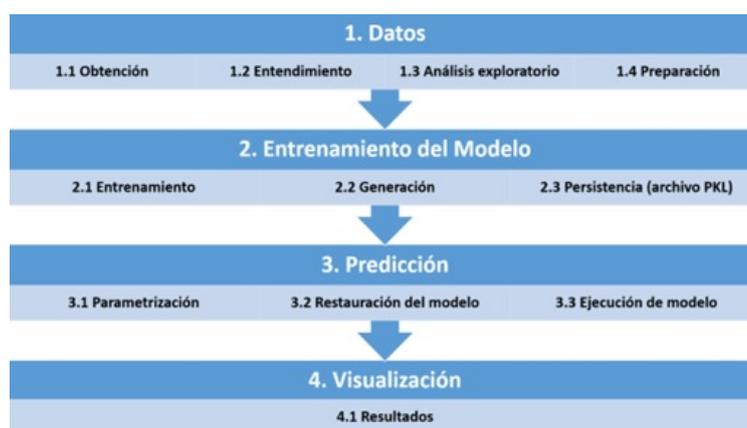


Figura 1: Resumen de la metodología.

## 4. Resultados

Se puede apreciar que los precios de las acciones bancarias (en este estudio Bancolombia y Davivienda) son muy similares, por otro lado, la acción de Ecopetrol se encuentra en una escala muy por debajo (Figura 2). También se puede apreciar el impacto de la pandemia por COVID 19 aproximadamente a mitad del primer semestre del 2020.



Figura 2: Comparativo histórico de acciones de Bancolombia, Davivienda y Ecopetrol desde el 1ro de enero de 2018 hasta el 16 de junio de 2023.

### 4.1. Predicción

Un resumen de las predicciones realizadas a los tres activos que fueron objeto de estudio se pueden ver en Tabla 1.

Activo	Promedio	Mínimo		Máximo	
		Fecha	Precio	Fecha	Máximo
Bancolombia	\$29,088.30	30-Jul	\$28,152.00	13-Jul	\$29,359.90
Davivienda	\$18,983.99	28-Jul	\$18,255.20	1-Jul	\$20,335.10
Ecopetrol	\$2,167.52	1-Jul	\$2,159.87	21-Nov	\$2,176.52

Tabla 1: Resumen de principales estadísticos de la predicción.



#### 4.1.1. Bancolombia

Luego de realizar el pronóstico para el segundo semestre del 2023 se puede apreciar que existe un máximo de la acción el 13 de julio y un mínimo el 30 de julio (Figura 3). Además el promedio general (aunque no sea una medida muy aplicada) es de 29,088.30 (Tabla 1).

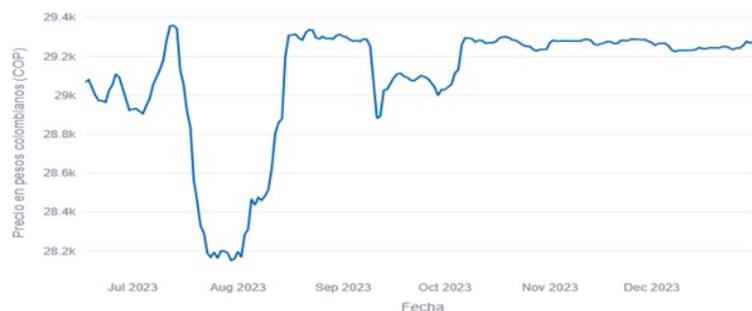


Figura 3: Predicción de Bancolombia para el segundo semestre del 2023.

#### 4.1.2. Davivienda

El pronóstico para el segundo semestre del 2023 indica que existe un máximo de la acción el 1 de julio y un mínimo el 30 de julio (Figura 4). Además el promedio general (aunque no sea una medida muy aplicada) es de 18,983.99.30 (Tabla 1).

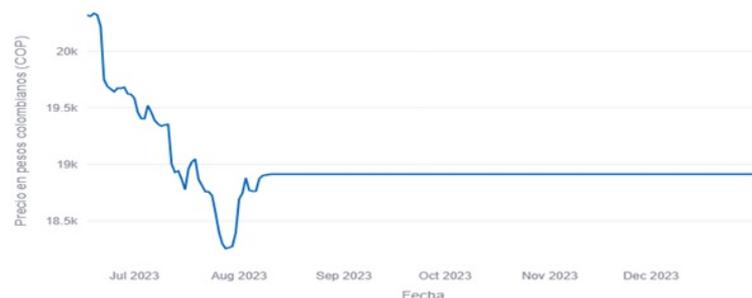


Figura 4: Predicción de Davivienda para el segundo semestre del 2023.

#### 4.1.3. Ecopetrol

El pronóstico para el segundo semestre del 2023 indica que existe un máximo de la acción el 29 de noviembre y un mínimo el 1ro de julio (Figura 5). Además el promedio general (aunque no sea una medida muy aplicada) es de 2,176.52 (Tabla 1).

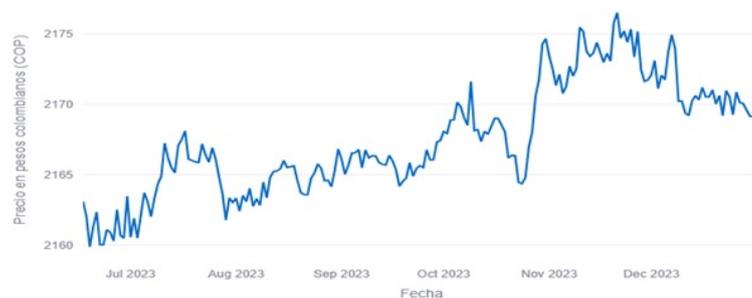


Figura 5: Predicción de Ecopetrol para el segundo semestre del 2023.



#### 4.2. Análisis

Respecto a Bancolombia se esperan 2 meses de variabilidad con tendencia a la baja en julio, aumento en agosto y con estabilidad el resto del semestre (Figura 6). La tendencia general es levemente alcista.

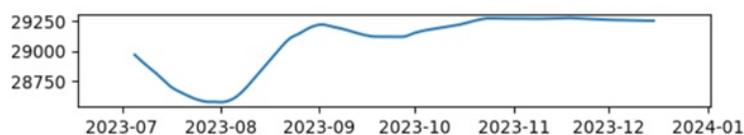


Figura 6: Predicción de Ecopetrol para el segundo semestre del 2023.

De Davivienda se puede inferir que al igual que Bancolombia, se esperan 2 meses de variabilidad con tendencia a la baja en julio, leve aumento en agosto y con estabilidad el resto del semestre (Figura 7). La tendencia general es levemente alcista.

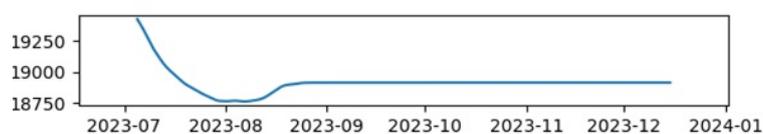


Figura 7: Tendencia Davivienda para segundo semestre de 2023.

De Ecopetrol se puede inferir que al igual que Bancolombia, se esperan 2 meses de variabilidad con tendencia alcista el resto del semestre (Figura 8).

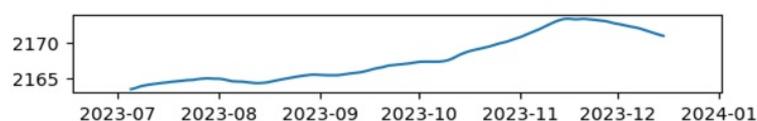


Figura 8: Tendencia Ecopetrol para segundo semestre de 2023.

También se puede apreciar que los precios de las acciones bancarias (en este estudio Bancolombia y Davivienda) son muy similares, por otro lado, la acción de Ecopetrol se encuentra en una escala muy por debajo (Figura 1). También se puede apreciar el impacto de la pandemia por COVID 19 aproximadamente a mitad del primer semestre del 2020.

## 5. Conclusiones

Se cumplió el objetivo de predecir el precio de 3 acciones del mercado de valores colombiano para el segundo semestre de 2023 empleando técnicas de aprendizaje maquina y análisis de series temporales realizando pronóstico del precio de una acción en una fecha futura.

Las máquinas de aprendizaje entrenadas en este estudio predicen variabilidad los 2 primeros meses, sin embargo, el resto del semestre predicen estabilidad. Por lo tanto, se puede concluir que no es confiable el método aplicado para realizar predicciones a mediano y largo plazo (más de 2 meses) teniendo en cuenta la volatilidad del mercado de valores.

El análisis realizado en este estudio fue de tipo univariado, siendo insuficiente para lograr predicciones más precisas ya que solamente se cuenta con el precio a lo largo del tiempo. Por lo tanto, se recomienda aplicar un enfoque multivariado incluyendo otras variables económicas, por ejemplo: S&P Index, DXY Index, entre otros.



Por otro lado, también se recomienda comparar aplicando otras técnicas de aprendizaje automático o aprendizaje profundo.

Para finalizar y como recomendación final, se sugiere adicionar análisis econométrico a los estudios de naturaleza económico independiente de la aplicación de IA.

## Referencias

- [1] A. E. Khedr, S.E.Salama y N. Yaseen, «Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis,» *International Journal of Intelligent Systems and Applications*, vol. 9, n.º 7, págs. 22-30, jul. de 2017, ISSN: 2074-9058. DOI: [10.5815/ijisa.2017.07.03](https://doi.org/10.5815/ijisa.2017.07.03). dirección: <http://dx.doi.org/10.5815/ijisa.2017.07.03>.
- [2] E. Moreno Sarmiento, «Predicción con series de tiempo y regresión,» *Panorama*, vol. 2, n.º 4, jun. de 2013, ISSN: 1909-7433. DOI: [10.15765/pnrm.v2i4.262](https://doi.org/10.15765/pnrm.v2i4.262). dirección: <http://dx.doi.org/10.15765/pnrm.v2i4.262>.
- [3] J. Herrero, *Ciencia de datos: técnicas analíticas y aprendizaje estadístico en un enfoque práctico*. Alfaomega, 2018, ISBN: 9786075382524. dirección: <https://books.google.com.co/books?id=0WUyxQEACAAJ>.
- [4] y. J. M. C. Alonso-Robisco, *Inteligencia artificial y finanzas: una alianza estratégica*, Online, 2022. dirección: <https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSerias/DocumentosOcasionales/22/Fich/do2222.pdf>.
- [5] P. Bruce, A. Bruce y P. Gedeck, *Estadística práctica para ciencia de datos con R y Python*. Marcombo, 2022, ISBN: 9788426734549. dirección: <https://books.google.com.co/books?id=IZxeEAAAQBAJ>.
- [6] J. Devore, *PROBABILIDAD Y ESTADISTICA PARA INGENIERIA Y CIENCIAS* (Ciencias e ingenierías). THOMSON LEARNING, 2001, ISBN: 9789706860675. dirección: <https://books.google.com.co/books?id=FJrxngEACAAJ>.
- [7] S. Raschka y V. Mirjalili, *Python machine learning: aprendizaje automático y aprendizaje profundo con Python, scikit-learn y TensorFlow*. Marcombo, 2019, ISBN: 9788426727206. dirección: <https://books.google.com.co/books?id=c6a2wgEACAAJ>.
- [8] P. Fernández y A. Monsalve-Cobis, *Introducción al Análisis de Series de Tiempo con Aplicaciones a la Econometría y Finanzas*. ago. de 2015, ISBN: 978-980-261-163-8.
- [9] R. Pearson, *Exploratory Data Analysis Using R* (Chapman & Hall/CRC data mining and knowledge discovery series). CRC Press LLC, 2018, ISBN: 9781498730235. dirección: <https://books.google.com.co/books?id=s-N7swEACAAJ>.
- [10] C. Nantasenamat, A. Biswas, J. Nápoles-Duarte, M. I. Parker y R. L. Dunbrack, «Building bioinformatics web applications with Streamlit,» en *Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development*. Elsevier, 2023, págs. 679-699, ISBN: 9780443186387. DOI: [10.1016/b978-0-443-18638-7.00001-3](https://doi.org/10.1016/b978-0-443-18638-7.00001-3). dirección: <http://dx.doi.org/10.1016/B978-0-443-18638-7.00001-3>.

