

**Implementación del modelo ARIMA para la predicción del stock financiero de Bancolombia S.A**  
**Implementation of the ARIMA model for the prediction of the financial stock of Bancolombia S.A.**  
**Implementação do modelo ARIMA para previsão do estoque financeiro do Bancolombia S.A.**

**Kevin Camilo BARRETO VARON<sup>1</sup>**

*Artículo de investigación científica y tecnológica<sup>2</sup>*

### **Resumen**

En esta investigación se analizaron datos históricos de la acción de Bancolombia S.A. en la bolsa de Nueva York, con el objetivo de desarrollar un modelo de predicción a corto plazo que considerara su volatilidad y tendencias temporales. Se aplicaron técnicas de ingeniería de datos para centralizar la información y asegurar la incorporación periódica de nuevas fuentes de datos. El modelo de predicción fue construido utilizando ARIMA, ajustando sus parámetros mediante ciclos iterativos para lograr una calibración precisa mediante los parámetros de diferenciación, promedio móvil y autoregresivo. Finalmente, el modelo fue expuesto y gestionado en producción a través de una orquestación con ML Flow, lo que permitió su implementación eficiente para su uso práctico.

**Palabras Claves:** *ARIMA, predicción, modelo de datos, inteligencia artificial.*

### **Abstract**

*In this research, historical data of Bancolombia S.A. shares on the New York Stock Exchange were analyzed, with the aim of developing a short-term prediction model that considered its volatility and time trends. Data engineering techniques were applied to centralize the information and ensure the periodic incorporation of new data sources. The prediction model was built using ARIMA, adjusting its parameters through iterative cycles to achieve precise calibration using the differentiation, moving average and autoregressive parameters. Finally, the model was exposed and managed in production through an orchestration with ML Flow, which allowed its efficient implementation for practical use.*

**Keywords:** *ARIMA, prediction, data model, artificial intelligence.*

### **Resumo**

*Nesta pesquisa foram analisados dados históricos sobre as ações do Bancolombia S.A. na Bolsa de Valores de Nova York, com o objetivo de desenvolver um modelo de previsão de curto prazo que considerasse sua volatilidade e tendências temporais. Foram aplicadas técnicas de engenharia de dados para centralizar a informação e garantir a incorporação regular de novas fontes de dados. O modelo de predição foi construído utilizando ARIMA, ajustando seus parâmetros através de ciclos iterativos para obter calibração precisa utilizando os parâmetros de*

---

<sup>1</sup> Estudiante de ingeniería en sistemas y telecomunicaciones. Analista de datos, Universidad de Manizales(Manizales, Caldas, Colombia). Correo electrónico: kbarreto@umanizales.edu.co

<sup>2</sup>Artículo proveniente del proyecto *Implementación del modelo ARIMA para la predicción del stock financiero de Bancolombia S.A.*, ejecutado en el periodo fecha 2024-03 –2024-10, e inscrito en el grupo de investigación *Semillero de inteligencia artificial* de la *Universidad de Manizales*. [Si es una opción de grado, se especifica el título obtenido y el director o presidente del trabajo].

*diferenciação, média móvel e autorregressivos. Por fim, o modelo foi exposto e gerenciado em produção através de uma orquestração com ML Flow, o que permitiu sua implementação eficiente para uso prático.*

**Palavras-chave:** *ARIMA, previsão, modelo de dados, inteligência artificial.*

## Introducción

El análisis financiero desde sus inicios ha estado lleno de probabilidades, sesgos, metodologías y técnicas que se han creado con el objetivo de dar luz a un pronóstico, a un ambiente futuro que dé razón a hacia dónde se dirige la empresa, el país o en este caso, un activo financiero, tanto así que los interesados en este campo *«se han basado en una serie de herramientas econométricas y metodologías tales como ARIMA, GARCH, EGARCH, TGARCH»*(Camargo, 2013, 201) además de otros enfoques que en la actualidad están al alcance de los académicos y estudiosos de las finanzas con ayuda de la inteligencia artificial como servicios de análisis y predicción en línea.

Hoy se aborda la predicción y forecasting en distintas industrias que van desde el mercadeo, optimización logística y la gestión de riesgos hasta llegar a la previsión financiera, por ejemplo *«Los pronósticos se utilizan en diversas áreas de las empresas y del gobierno con el fin de tener una guía en la toma de algunas decisiones. En actividades agropecuarias, como la productiva, se utilizan para decidir qué producir (un producto específico o una combinación de productos), cuándo producir y dónde producir. También se han utilizado en pronósticos de precios y en la disponibilidad de insumos futuros»*(Delgadillo-Ruiz, 2016, 1) lo que hace que su uso e implementación sea totalmente transversal y no esté inclinada hacia una industria o área de conocimiento específica.

La aplicación de estos métodos para la previsión de valores o precios futuros en los mercados financieros o también llamados Stock Market Exchange exigen una alta estabilidad predictiva así como también una extensa fase de evaluación y corrección por parte de los conocedores del campo, por lo que varios estudios en el pasado han llevado a concluir que la dificultad radica en el *«hecho de que los precios bursátiles y las tasas de retornos no tienen auto-correlación a través del tiempo pero se caracterizan por una alta volatilidad y tiempos apacibles en el rango de la serie de tiempo»*(Camargo, 2013, 201) lo que plantea un desafío pero a su vez esta volatilidad ha generado un *«gran interés por parte de los académicos, inversionistas, y autoridades reguladoras del mercado»*(Camargo, 2013, 201) en desarrollar una metodología mediante los procesos de diferenciación de datos que impulse y simplifique estos pronósticos en el tiempo mediante la orquestación de procesos de ingeniería y analítica moderna.

Definida la volatilidad como *«la desviación estándar (o raíz cuadrada de la varianza) de los rendimientos de un activo o un portafolio»*(Camargo, 2013, 202) es uno de los panoramas que se han buscado estudiar para lograr desarrollar algoritmos, metodologías propias o adaptadas a una variable de análisis como lo puede ser el crecimiento de una empresa, análisis de ventas futuras, entre otros casos de uso que a medida avanzan los estudios se tiende a complejizar su configuración debido a parámetros de correlación parcial o simple de acuerdo a

los datos de entrada que se estén usando. Una de las herramientas más usadas en la actualidad para el análisis financiero, identificación de riesgos y que además permite un correcto análisis de la volatilidad son las series tiempo a las que se «le denomina a cualquier variable que conste de datos reunidos, registrados u observados sobre incrementos sucesivos de tiempo. Por lo tanto, se concluye que es una secuencia ordenada de observaciones sobre una variable en particular.»(Murillo & Olaya, 2003, 1).

Dado este contexto, desde el punto de vista académico e investigativo se identifican varios puntos de mejora referente a la puesta en marcha de modelos de predicción fiables apoyados en modelos estadísticos y de inteligencia artificial, para ello las nuevas propuestas deben seguir el ciclo de ingeniería de datos modernos para centralizar y medir los datos cuantitativos a lo largo del tiempo en la evolución de los proyectos para al final de este proceso tener como resultado unos datos limpios y fiables para sobre ellos realizar entrenamientos y evaluaciones «de un buen modelo que represente la serie» (Murillo & Olaya, 2003, 6) de tiempo que se requiere analizar.

## **1. Fundamento Teórico**

Desde la masificación de la inteligencia artificial en 2021 se ha creado una competencia en cuanto a la implementación de soluciones empresariales que tengan en su núcleo Machine Learning y procesamiento de datos, para empresas pequeñas y de mediano tamaño esto ha supuesto una carrera por recopilar datos de sus respectivas actividades económicas e iniciar estrategias que permitan generar valor e impacto futuro, tanto en su crecimiento como en su cliente final. Desde entonces los términos como análisis predictivo, inteligencia artificial generativa o análisis prescriptivo han venido tomando fuerza, y esto se ha visto desde el mundo académico donde mes a mes cientos de iniciativas se han publicado con el fin de aclarar todos los puntos oscuros que esta nueva revolución supone y mostrar todo tipo de innovaciones. A continuación se muestran algunos de los antecedentes más relevantes donde se emplean metodologías y modelos relacionados con esta investigación.

Según Hernandez & Torres (2022, 20) el campo del análisis de pronósticos y generación de modelos estadísticos que apoyen la idea de obtener un horizonte de datos y variables se deben considerar los modelos de estadística clásica como las series de tiempo o los modelos de Machine Learning y Deep Learning, los cuales con la correcta configuración están en capacidad de entregar predicciones que sean insumo para actividades estratégicas en la toma de decisiones en los diferentes sectores económicos que conforman la economía global. En su estudio mencionan que la analítica sobre históricos de una variable permiten analizar el comportamiento pasado e identificar patrones desde la etapa de visualización.

Así mismo en otra investigación se menciona que *«la previsión de la demanda como estimaciones sobre la cantidad que va a ser demandada de un producto o productos en un periodo de tiempo determinado utilizando datos históricos y otras informaciones»*(Barrera, 2020, 41), será de vital importancia en estos tipos de proyecto debido a que *«las proyecciones de la demanda se pueden englobar en dos grandes grupos: proyecciones cualitativas y proyecciones cuantitativas»*(Barrera, 2020, 43), para el caso específico de esta investigación el tipo de proyección que más despierta interés es la proyección cuantitativa, debido a que se basa en datos numéricos y aplicación de modelos matemáticos y estadísticos para generar el análisis de series de tiempo temporales, además su impacto puede verse multiplicado cuando se reúnen las siguientes condiciones:

- Relación clara entre las variables que se analizan(en el caso de previsión financiera, tiempo y valor de mercado).
- Datos históricos gestionados eficientemente mediante técnicas de ingeniería de datos.
- Trazabilidad e identificación de eventos o patrones en la serie de tiempo.

Con esto en mente se evidencia que para que un análisis de proyecciones futuras se debe crear un ambiente de co-creación entre diferentes roles o equipos que gestionen la calidad del dato y luego su correcto análisis.

Hernandez & Torres (2022, 20) concluye en su investigación que el uso de modelos predictivos como ARIMA y SARIMA(Modelo Autorregresivo Integrado de Media Móvil Estacional) contribuyen al objetivo de encontrar patrones y predecir valores futuros en base a observaciones y comportamientos pasados de alguna característica o variable modificada a través del tiempo que puede ser datos relacionados a la evolución de valor de una acción bursátil, se resalta que este tipo de modelos permiten establecer una serie de parámetros como (P,D,Q)s donde *«P es el parámetro autorregresivo del proceso estacional, Q el parámetro de media móvil, D es el parámetro de diferenciación la parte estacional y s representa la periodicidad de la serie»*(Gomez et al., 2022). Una vez definidos estos modelos se deben seguir algunos pasos por convención para lograr replicar y obtener datos precisos y posteriormente evaluar la utilidad, precisión y aplicación de estos horizontes de datos futuros que se han obtenido.

Conforme lo menciona Morales (2022, 20) luego de la fase de limpieza y transformación de datos históricos, se debe realizar un análisis exploratorio y calibración de los modelos seleccionados que mejor se ajusten al caso de uso que se esté desarrollando, para posteriormente *«teniendo en cuenta los hallazgos en el proceso exploratorio, con el objetivo de optimizar su rendimiento y obtener las mejores métricas posibles»*(Morales , 2022, 6) y de esta manera lograr una iteración exitosa en el desarrollo del modelo de predicción. Para obtener un buen

rendimiento, serán necesarios varios procesos correctivos con el fin de identificar los parámetros correctos.

En conclusión, la evidencia científica dicta que estos nuevos modelos y metodologías de desarrollo de productos de inteligencia artificial han logrado establecer un hito, tanto así que *«el análisis en series de tiempo ha sido provisto de modelos estadísticos tales como ARIMA, SARIMAX y modelos de volatilidad GARCH, que a lo largo de los años han provisto a bancos, gobiernos y grandes organizaciones realizar el análisis de este tipo de información conformada por secuencias, en los últimos años el uso de modelos de deep learning como los Long short Term Memory networks ha resaltado por encima de la estadística en el análisis en series de tiempo»*(Suárez & Ríos, 2019, 3), esto visto desde un punto de vista de revisión científica sienta bases sólidas para futuros análisis usando estos modelos de regresión constituidos para conjuntos de datos estacionarios. De esta forma los profesionales y organizaciones que mayor interés cobran son las que abren esta puerta a la innovación.

## **2. Metodología**

Esta investigación no se desarrolló bajo ninguna metodología científica establecida, sino que por el contrario se tomaron algunos puntos de la metodología de observación y recolección de datos debido a que *«otorga al investigador una mejor comprensión de lo que está ocurriendo en la cultura, y otorga credibilidad a las interpretaciones que da a la observación»*(Kawulich, 2005, 6), en este caso específicamente en la fase de exploración de datos. A su vez, en la fase de calibración del modelo recurrimos a la experimentación, lo cual es muy común cuando usamos modelos como ARIMA, ya que esto implica un proceso iterativo en el que se prueban diferentes valores de los parámetros  $p$ ,  $d$ ,  $q$ , y posiblemente  $P$ ,  $D$ ,  $Q$ ,  $m$  si se considera un modelo SARIMA.

A continuación se expone paso a paso el proceso definido para lograr el desarrollo del modelo de previsión de datos financieros.

### **1. Definición del Problema**

El primer paso consistió en identificar el problema de fondo: la necesidad de predecir los precios de cierre de una acción bursátil con precisión(BIC para Bancolombia S.A), utilizando datos históricos de la serie temporal. La volatilidad del mercado y la ausencia de datos en días festivos y fines de semana fueron desafíos clave a abordar. El objetivo fue desarrollar un modelo predictivo capaz de anticipar el comportamiento de los precios a corto plazo.

## 2. Recolección y Preparación de los Datos

Se recopilaron los datos históricos de precios de cierre de una acción bursátil mediante el uso de una API pública(Polygon). Estos datos fueron organizados posteriormente en un modelo de base de datos pertinente que soporte una o varios activos financieros y sus datos históricos, de los datos completos que la API retorna, para el análisis son de interés dos columnas: "fecha\_precio" (fecha de los registros) y "c" (precio de cierre en dólares). El periodo histórico abarcaba varios años, pero contenía lagunas en los días no hábiles, como fines de semana y festivos.

Subpasos:

1. **Diseño modelo de datos:** Inicialmente se desarrolló un modelo de datos que soportará los requerimientos de la investigación incluyendo persistencia de datos históricos, gestión de varios stock financiero, así como también la persistencia de notificaciones para cliente final y manejo de usuarios.
2. **Consumo y gestión del dato:** Para la carga de los datos históricos(1995 - 2015) se realizó por medio de una carga en bloque a una base de datos en Postgresql, las fechas posteriores se alimentaron por medio del consumo de una API. La segunda parte de la gestión del dato se realiza mediante una lógica preestablecida para no duplicar registros en la base de datos, ya que este escenario generaría confusión para el modelo.
3. **Cargar los datos:** Se cargaron los datos de la serie temporal en un formato adecuado, utilizando pandas.
4. **Conversión de fechas:** Se aseguraron que las fechas estuvieran en un formato de tiempo adecuado, convirtiéndolas con `pd.to_datetime()`.
5. **Establecer el índice temporal:** La columna de fechas se estableció como el índice del dataframe para facilitar el análisis y la modelización.
6. **Frecuencia diaria:** Se rellenaron los días faltantes (días festivos y fines de semana) con valores NaN utilizando la función `asfreq('D')` para asegurarse de que la serie temporal tuviera una periodicidad diaria, esto debido a que el modelo ARIMA exige un flujo de datos regular, es decir, que no falten periodos de tiempo.
7. **Interpolación de datos faltantes:** Los valores faltantes se completaron mediante técnicas de interpolación temporal para mantener la consistencia en los datos antes de ajustar el modelo.

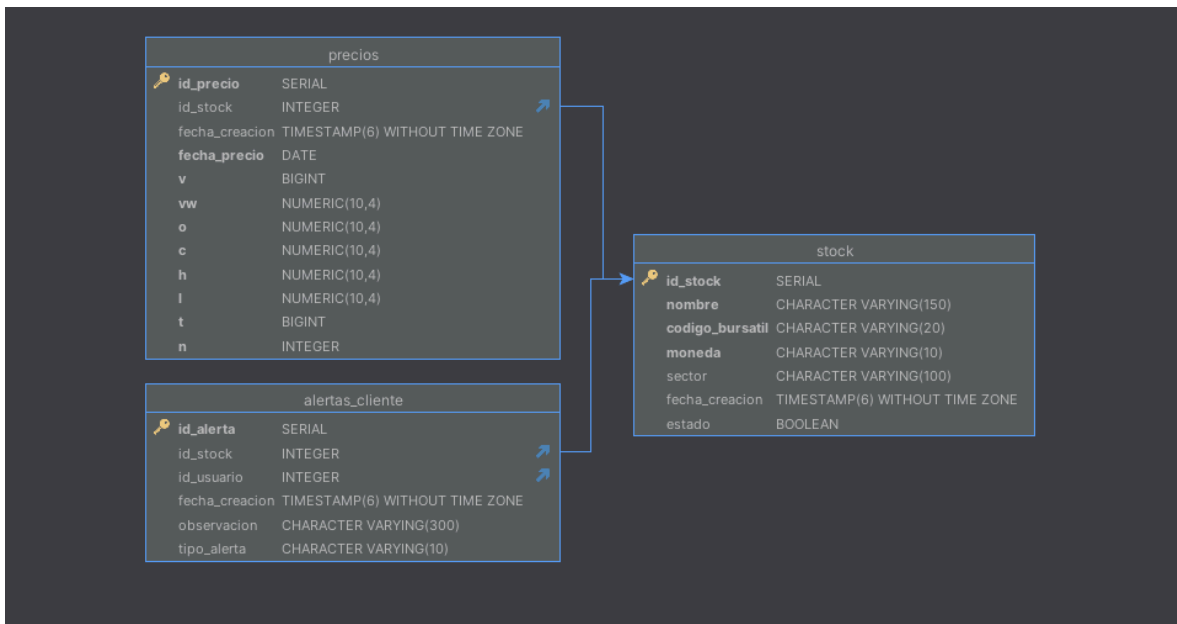


Imagen 1. Modelo de datos para gestionar históricos de varios activos financieros.

### 3. Análisis de Estacionariedad

El siguiente paso fue verificar si la serie temporal era estacionaria, lo cual es un requisito para ajustar correctamente el modelo ARIMA. Para ello, se aplicó la Prueba de Dickey-Fuller Aumentada (ADF).

Subpasos:

1. **Prueba de ADF:** Se implementó la prueba de Dickey-Fuller utilizando la función `adfuller()` para evaluar si la serie era estacionaria o no. Los resultados de la prueba (valor p y estadístico de la prueba) determinaron si era necesario diferenciar los datos.
2. **Diferenciación:** Dado que la serie no era estacionaria, se aplicó la primera diferenciación a la serie con `.diff()` para eliminar las tendencias y convertirla en estacionaria.

Los datos financieros debido a su volatilidad son conjuntos de datos en su mayoría no estacionarios, es decir que no tienen un patrón definido a lo largo del tiempo y sus fluctuaciones pueden llegar a ser impredecibles, para estos casos existe el proceso de la diferenciación cuya aplicación «*elimina la tendencia, mientras que el Logaritmo Natural se encarga de la heterocedasticidad, por lo que las series objetos de estudios se encuentran lista para ser analizadas*»(Camargo, 2013, 216), posteriormente, una vez nuestros datos sean estacionarios podemos

proceder con el siguiente paso. A continuación se muestra la serie de tiempo original frente a la serie diferenciada junto con su respectiva gráfica de correlación.

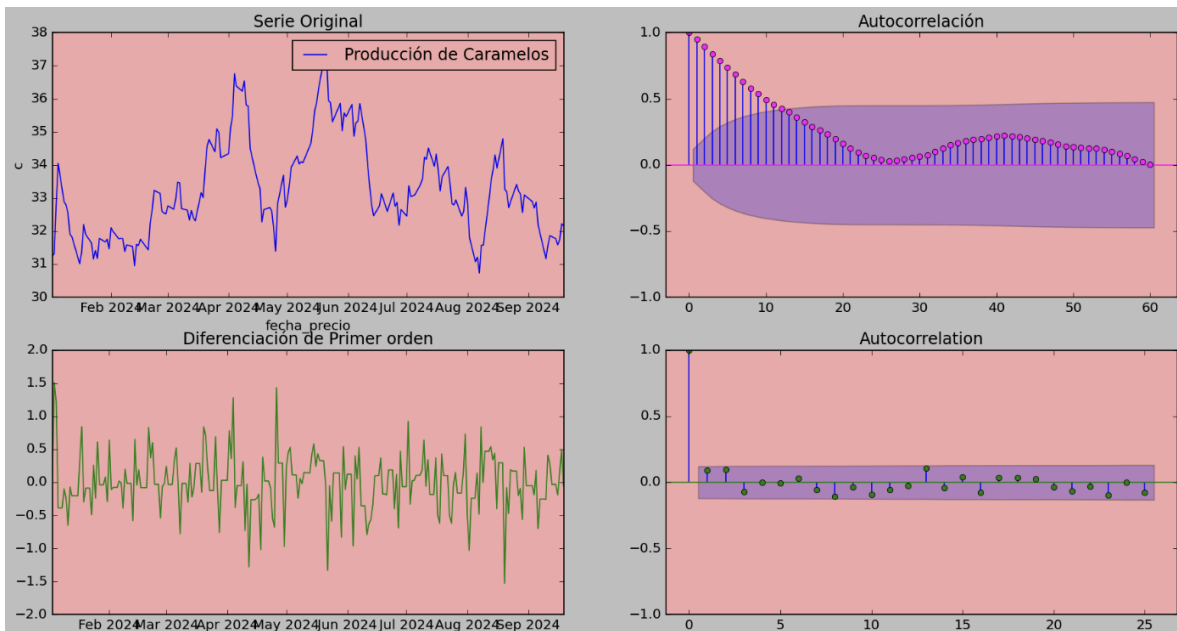


Imagen 2. Aplicación de proceso de diferenciación de los datos.

#### 4. Selección de Parámetros $p$ , $d$ , y del Modelo ARIMA

Se utilizaron gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF) para determinar los parámetros del modelo ARIMA. Estos gráficos ayudan a identificar el orden de los términos autorregresivos  $p$ , el grado de diferenciación  $d$ , y los términos de media móvil  $q$ .

Subpasos:

1. **Gráfico ACF:** Este gráfico indicó el número de rezagos a considerar para el término de media móvil  $q$ .
2. **Gráfico PACF:** Este gráfico ayudó a identificar el número de rezagos a utilizar para el componente autorregresivo  $p$ .
3. **Selección de  $d$ :** El valor  $d$  se seleccionó basado en el análisis de la serie diferenciada para asegurarse de que la serie fuera estacionaria.

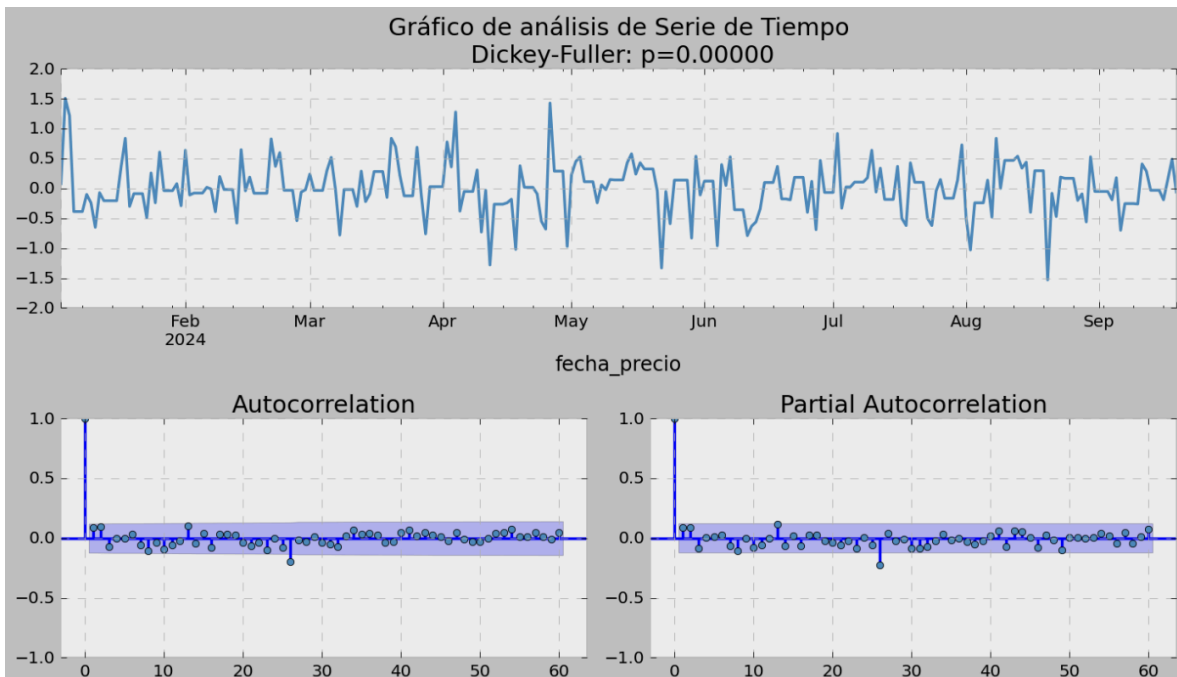


Imagen 3. Gráficos de autocorrelación y autocorrelación parcial sobre la serie de tiempo diferenciada.

## 5. Ajuste del Modelo ARIMA

Con los parámetros  $p$ ,  $d$  y  $q$  identificados, se ajustó el modelo ARIMA utilizando la librería statsmodels. El modelo se entrenó con los datos históricos para poder realizar predicciones sobre el futuro comportamiento de los precios.

Subpasos:

1. **Construcción del modelo:** Se creó el modelo ARIMA con los parámetros seleccionados usando `ARIMA(df['candy'], order=(p, d, q))`.
2. **Ajuste del modelo:** Se entrenó el modelo utilizando el método `.fit()` para ajustarlo a los datos históricos de precios de cierre.
3. **Evaluación del modelo:** Se examinó el Akaike Information Criterion (AIC) y otros indicadores estadísticos para evaluar la calidad del ajuste del modelo y asegurarse de que los residuales fueran aleatorios.

## 6. Predicciones con el Modelo ARIMA

Una vez ajustado el modelo, se realizaron predicciones de los precios futuros para un horizonte de 30 días. Además, se probó una segunda versión del modelo

utilizando técnicas de validación cruzada y optimización de parámetros para mejorar la precisión de las predicciones.

Subpasos:

1. **Predicción de valores futuros:** Se utilizó el modelo ajustado para predecir los valores de precios de cierre para los próximos días con el método `forecast(steps=30)`.
2. **Intervalos de confianza:** Se generaron intervalos de confianza para las predicciones utilizando `conf_int()`, lo que permitió visualizar el rango de posible variación de los precios futuros.
3. **Validación de resultados:** Las predicciones se validaron comparándolas con un subconjunto de datos de prueba, evaluando la precisión del modelo con métricas como el Mean Squared Error (MSE).

Durante esta etapa de la investigación fue de crucial importancia tener en cuenta los intervalos de confianza debido que a través de ellos se representa la incertidumbre asociada a las predicciones, esto se gráfica mediante variaciones mínimas y máximas que puede tomar el valor futuro, también se debe tener en cuenta que cuando «*una serie de tiempo en estudio tiene intervalos de observación menores a un año, entonces es frecuente que estas tengan variaciones ó patrones sistemáticos cada cierto periodo, estas variaciones sistemáticas inferiores a un año por ejemplo semestral, mensual, diario, etc. Deben ser captadas en los llamados Factores Estacionales, dentro de la estructura del modelo a construirse.*» (Villavicencio, 2010, 24). En el caso de esta investigación se definieron los intervalos de confianza al 95%, se eligió así por convención de la metodología.

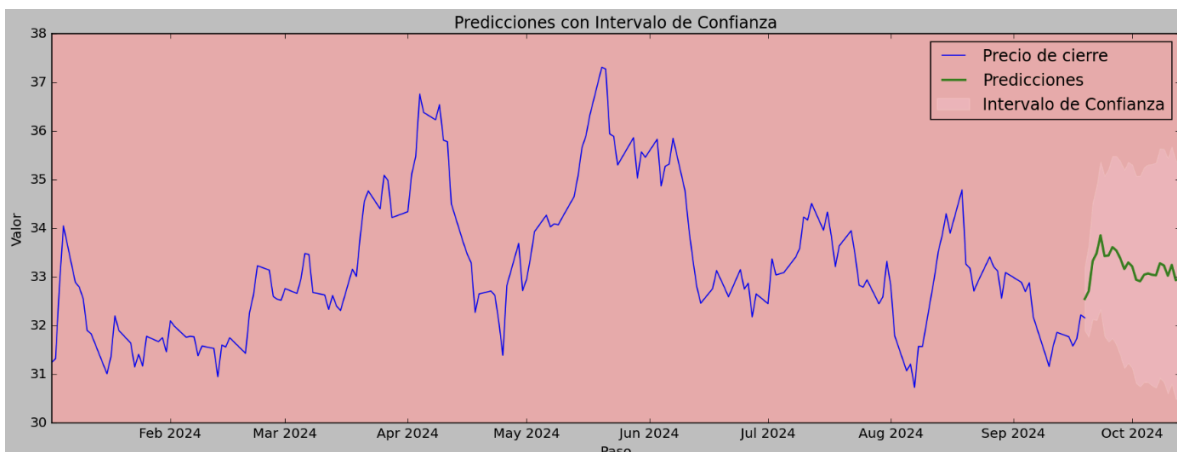


Imagen 4. Graficación de predicción horizonte próximo de 30 días, incluye intervalo de confianza para el mismo periodo de tiempo.

## 7. Implementación de la API para Consultas del Modelo

Finalmente, el modelo ARIMA entrenado fue implementado como un servicio web mediante una API REST utilizando Flask. Esto permitió que otros sistemas o usuarios pudieran consultar las predicciones del modelo de manera programática.

Subpasos:

1. **Guardar el modelo:** El modelo entrenado fue guardado utilizando joblib para su posterior uso.
2. **Desarrollo de la API:** Se creó una API en Flask con un endpoint /predict que recibe solicitudes POST con el número de días para predecir y devuelve las predicciones en formato JSON.
3. **Despliegue del modelo:** La API fue configurada para funcionar con Gunicorn y Nginx, permitiendo su acceso desde entornos externos.

## 8. Visualización de los Resultados

Se generaron gráficos que permitieron visualizar tanto los datos históricos como las predicciones generadas por el modelo ARIMA.

Subpasos:

1. **Gráfico de la serie temporal:** Se creó un gráfico de la serie temporal con los precios de cierre históricos y las predicciones, permitiendo observar visualmente el comportamiento pasado y futuro de la acción.
2. **Intervalos de confianza:** Se visualizaron los intervalos de confianza de las predicciones para mostrar la posible variabilidad de los precios futuros.

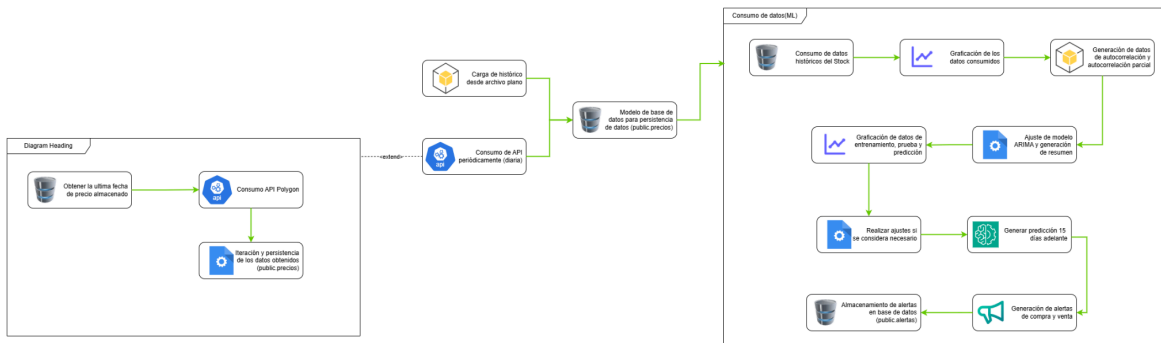


Imagen 5. Flujo completo de gestión de datos y creación del modelo de predicción financiera.

Esta metodología proporcionó una estructura clara y sistemática para desarrollar un modelo ARIMA robusto, capaz de predecir los precios futuros de las acciones

bursátiles(BIC). Al seguir estos pasos detallados, la investigación abordó el problema de la volatilidad del mercado y la falta de datos en ciertos días, ofreciendo una solución práctica y escalable a través de una API accesible para los usuarios.

### **3. Resultados y discusión**

#### **3.1 Descripción de resultados**

Esta investigación a pesar de estar en dos fases generales, una orientada a la ingeniería de datos y orquestación de procesos; la otra es la que más cobra relevancia al momento de exponer los resultados debido a que esta segunda fase tiene que ver con la estabilización y puesta en marcha del modelo de predicción usando el modelo ARIMA para la predicción univariado de valores asociados a series de tiempo, a continuación se resaltan los hallazgos y resultados mas relevantes:

#### **Resultado 1: Serie de tiempo diferenciada**

El conjunto de datos inicial de referencia para el histórico era un conjunto de datos que no cumplía con los parámetros para poder aplicar el modelo ARIMA, por ello, se hizo uso del proceso de diferenciación con el objetivo de hacer la serie de tiempo estacionaria, esta es una técnica común en el análisis de series de tiempo, y su objetivo principal es eliminar las tendencias o patrones no estacionarios que puedan estar presentes en los datos.

Al aplicar la diferenciación, transformamos la serie en una serie estacionaria, lo que mejora el rendimiento del modelo ARIMA, ya que ahora puede trabajar con una serie cuyas propiedades no cambian con el tiempo. Esto facilita que el modelo capture las relaciones entre los valores pasados y los futuros.

En esta investigación, se utilizó la primera diferenciación ( $d = 1$ ) en los datos de la serie temporal, restando el valor anterior de cada dato del valor actual, lo que permitió eliminar las tendencias lineales. Esto hizo que el modelo ARIMA pudiera capturar la dinámica subyacente de la serie de tiempo sin ser influenciado por fluctuaciones no estacionarias.

#### **Resultado 2: Predicción e intervalos de confianza**

Los resultados del modelo ARIMA ofrecieron predicciones del precio de la acción para un número determinado de días(el alcance de la investigación es la predicción a corto plazo), así como sus correspondientes intervalos de confianza.

- **Predicción de precios:** El modelo predijo el precio de cierre de la acción para los próximos días (los próximos 30 días). Los valores predichos estuvieron influenciados por los patrones históricos de la serie temporal, como la volatilidad y las tendencias a corto plazo.
- **Visualización de los intervalos de confianza:** Se graficaron tanto los valores predichos como los intervalos de confianza. La línea principal del gráfico muestra la predicción puntual, mientras que el área sombreada alrededor de la predicción muestra el rango de los intervalos de confianza (por ejemplo, entre el 5% y el 95% de probabilidad). A medida que se predicen más días hacia el futuro, los intervalos se expanden, lo que indica una mayor incertidumbre sobre los valores futuros.

### **Resultado 3: Modelo estabilizado**

Para esta investigación el modelo entrenado permite generar predicciones de precios de cierre para periodos futuros, como los próximos días o semanas. En este proyecto, se realizaron predicciones a corto plazo (por ejemplo, para los próximos 30 días), lo cual es útil para la toma de decisiones en un contexto de inversión o análisis financiero.

Ejemplo: Si hoy el precio de cierre de la acción es \$35 USD, el modelo puede predecir cómo se comportará ese precio durante los próximos días, ajustándose a las tendencias observadas en el pasado.

### **Resultado 4: Exposición de predicciones**

El modelo entrenado se implementó a través de una API usando Flask, lo que permite que terceros (usuarios finales o sistemas cliente) puedan consultar las predicciones automáticamente, facilitando la integración de este modelo en aplicaciones o sistemas financieros. Esta implementación reduce el trabajo manual y facilita el acceso a las predicciones del modelo en tiempo real.

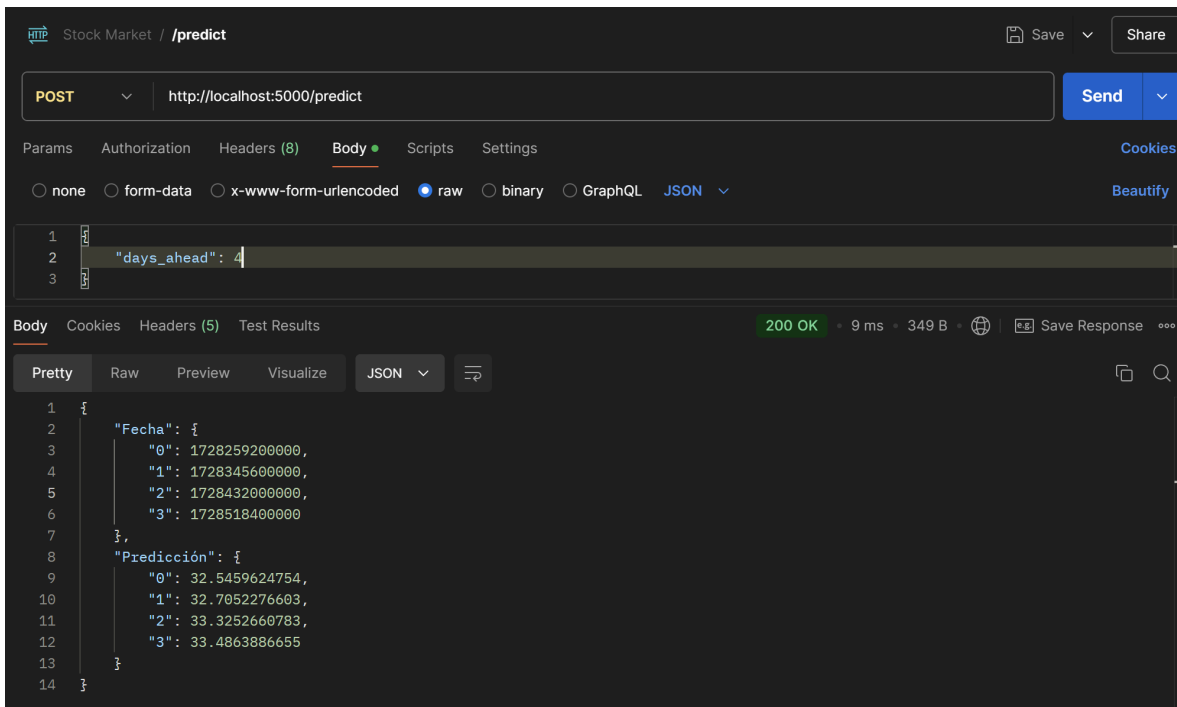


Imagen 6. Prueba de consumo de API desarrollada para exponer la predicción de los precios de cierre según la cantidad de días que se soliciten.

### 3.2 Discusión de resultados

Los resultados obtenidos en este proyecto muestran que el modelo ARIMA es una herramienta poderosa para predecir el precio de acciones a corto plazo, con una buena precisión y la capacidad de generar intervalos de confianza que ofrecen un marco para evaluar la incertidumbre. Sin embargo, dado que el modelo se basa en datos históricos, su capacidad para predecir en mercados altamente volátiles y en horizontes temporales largos es limitada, lo que requiere un uso cuidadoso de los resultados. La integración con MLFlow y su adaptabilidad a nuevos datos fortalecen el impacto práctico del modelo, haciéndolo escalable y útil en aplicaciones reales de análisis financiero.

## 4. Conclusiones

- El modelo ARIMA ha sido efectivo en la predicción de valores futuros para en análisis de stock de los activos analizados en este trabajo con una

efectividad del 90% luego de realizar las comparaciones y revisiones de los datos proveídos.

- La integración de MLOps para procesos de Machine Learning e inteligencia artificial es indispensable para el ecosistema tecnológico actual debido a que proporciona un versionamiento y control de nivel superior afectando de esta manera el despliegue y salida a producción de los sistemas o algoritmos de Machine Learning

Como trabajos futuros es interesante evaluar un panorama de integración del sistema desarrollado a un sistema superior de análisis de datos para mercados financieros debido a que fue desarrollado siguiendo los principios de ingeniería de Software funcional y empresarial de bajo acoplamiento y alta cohesión, de esta manera permitiendo el análisis de cualquier stock financiero desde el modelo de datos diseñado.

## 5. Agradecimientos

Quiero expresar mi más sincero agradecimiento a la Universidad de Manizales y a los docentes de la Facultad de Ciencias e Ingeniería, quienes, con su disposición y colaboración incondicional, han sido pilares fundamentales en la culminación de este proyecto. Su dedicación, conocimientos y constante apoyo han sido clave en el desarrollo de esta investigación. Gracias por guiarme a lo largo de este proceso y por fomentar un ambiente académico que impulsa la curiosidad científica y el crecimiento profesional.

## Referencias bibliográficas

- ALBA SUÁREZ, M. A., & PINEDA RÍOS, W. D. Natural lenguaje processing para la predicción de series de tiempo en el mercado de commodities energético con el uso de modelos en inteligencia artificial.[enlínea].<<https://repository.usta.edu.co/handle/11634/28677>> [consulta: 07/10/2024]
- BARRERA LÓPEZ, M. (2020). Desarrollo de una aplicación para el estudio de los modelos de previsión de la demanda.[enlínea].<[https://biblus.us.es/bibing/proyectos/abreproy/92782/descargar\\_fichero/TFG-2782+BARRERA-L%C3%93PEZ%2C+MANUEL.pdf](https://biblus.us.es/bibing/proyectos/abreproy/92782/descargar_fichero/TFG-2782+BARRERA-L%C3%93PEZ%2C+MANUEL.pdf)> [consulta: 07/10/2024]
- CAMARGO, E. A. P. (2013). Volatilidad y pronóstico del Índice General de la Bolsa de Valores de Colombia (IGBC) utilizando la metodología Box-Jenkins. *La administración frente a los desafíos de la economía global*, 50(4), 201.[enlínea].<[https://www.academia.edu/download/36420332/Memorias\\_Encuentro\\_Investigacion\\_2013.pdf#page=214](https://www.academia.edu/download/36420332/Memorias_Encuentro_Investigacion_2013.pdf#page=214)> [consulta: 07/10/2024]

- DELGADILLO-RUIZ, O., RAMÍREZ-MORENO, P. P., LEOS-RODRÍGUEZ, J. A., SALAS GONZÁLEZ, J. M., & VALDEZ-CEPEDA, R. D. (2016). Pronósticos y series de tiempo de rendimientos de granos básicos en México. *Acta universitaria*, 26(3), 23-32.[enlínea].<[https://www.scielo.org.mx/scielo.php?pid=S0188-62662016000300023&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S0188-62662016000300023&script=sci_arttext)> [consulta: 07/10/2024]
- HERNÁNDEZ RODRÍGUEZ, I. Y., & TORRES APONTE, L. M. (2022). Modelo de pronóstico de demanda para productos del sector eléctrico..[enlínea].<<https://repositorio.uniandes.edu.co/entities/publication/186deae8-fedf-4e7f-bbc2-1b05988df62a>> [consulta: 07/10/2024]
- KAWULICH, B. B. (2005). La observación participante como método de recolección de datos.[enlínea].<<http://biblioteca.udgvirtual.udg.mx/jspui/handle/123456789/2715>> [consulta: 07/10/2024]
- MURILLO, J., TREJOS, A., & OLAYA, P. C. (2003). Estudio del pronóstico de la demanda de energía eléctrica, utilizando modelos de series de tiempo. *Scientia et technica*, 3(23).[enlínea].<> [consulta: 07/10/2024]
- VILLAO GÓMEZ, J. I., BARRE ROMERO, J. W., & BAUZ, S. (2022). Análisis predictivos de ventas de categorías de productos en una empresa de retail con un sistema de recomendaciones de compra para clientes, empleando técnicas de visualización de datos.[enlínea].<<https://www.dspace.espol.edu.ec/handle/123456789/54269>> [consulta: 07/10/2024]
- VILLAVICENCIO, J. (2010). *Introducción a series de tiempo*. Puerto Rico.[enlínea].<[https://d1wqtxts1xzle7.cloudfront.net/38458362/manual\\_intro\\_series\\_tiempo-libre.pdf](https://d1wqtxts1xzle7.cloudfront.net/38458362/manual_intro_series_tiempo-libre.pdf)> [consulta: 07/10/2024]