

Análisis y predicción del comportamiento del precio del Bitcoin 2023-2024 mediante técnicas de machine learning

Carlos Jhovany Pedraza Prieto

Asesor

Jhoana Patricia Romero Leiton

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

Resumen

El Bitcoin ha presentado una volatilidad significativa a lo largo del tiempo, influenciado por factores económicos, regulatorios y de percepción del mercado. En este trabajo se estudia la predicción del precio de cierre del Bitcoin desde mayo de 2023 hasta mayo de 2024 utilizando tres enfoques principales: regresión lineal múltiple, modelos de series de tiempo ARIMA y redes neuronales recurrentes (RNN). Se realiza un análisis exploratorio de los datos con el propósito de encontrar patrones, tendencias históricas y relaciones importantes entre variables como el precio de apertura, máximo, mínimo y volumen, mostrando la alta volatilidad del mercado. Los resultados muestran que la regresión lineal múltiple capturó eficazmente las tendencias generales, mientras que ARIMA es adecuado para patrones temporales estables, y las RNN ofrecen la mayor precisión en la predicción al adaptarse mejor a los cambios dinámicos del mercado. El mejor modelo para predecir el precio del Bitcoin fue el de regresión lineal múltiple, debido a las métricas, lo cual indica una fuerte relación lineal entre las variables independientes (precio de apertura, máximo, mínimo y volumen) y el precio de cierre. En este estudio se muestra la importancia de seleccionar el modelo adecuado según el contexto y se busca explorar enfoques híbridos y el uso de variables externas en futuras investigaciones.

Palabras claves: Regresión lineal múltiple, series de tiempo ARIMA, RNN (redes neuronales recurrentes), predicción, criptomonedas, volatilidad.

Abstract

Bitcoin has exhibited significant volatility over time, influenced by economic, regulatory and market perception factors. This paper studies the prediction of the Bitcoin closing price from May 2023 to May 2024 using three main approaches: multiple linear regression, ARIMA time series models, and recurrent neural networks (RNN). An exploratory analysis of the data is performed to identify patterns, historical trends and important relationships between variables such as opening price, high, low and volume, showing the high volatility of the market. The results show that multiple linear regression effectively captured general trends, while ARIMA is suitable for stable temporal patterns, and RNNs offer the highest prediction accuracy by better adapting to dynamic market changes. The best model for predicting the Bitcoin price was the multiple linear regression model due to the metrics, which indicates a strong linear relationship between the independent variables (opening price, high, low and volume) and the closing price. This study shows the importance of selecting the appropriate model according to the context and seeks to explore hybrid approaches and the use of external variables in future research.

Keywords: Multiple linear regression, ARIMA time series, RNN (recurrent neural networks), forecasting, cryptocurrencies, volatility.

Tabla de Contenido

Introducción	9
Justificación	12
Planteamiento del Problema	15
Pregunta de Investigación	16
Objetivos	17
Objetivo General	17
Objetivos Específicos	17
Antecedentes Sobre el Análisis del Precio del Bitcoin	18
Desarrollo Histórico del Bitcoin	18
Evolución del Precio del Bitcoin	18
Estudios Previos (Análisis de Datos) Acerca de la Evolución del Precio del Bitcoin	19
Marco Referencial	23
Marco Teórico	23
Series de Tiempo ARIMA	24
Regresión Lineal Simple y Múltiple	25
Modelo de Redes Neuronales Recurrentes (RNN)	26
Metodología	28
Resultados	32
Recolección y Limpieza de Datos	32
Análisis Exploratorio de los Datos	33
Desarrollo de Modelos Predictivos	39
Modelo de Regresión Lineal Múltiple	39

Modelo ARIMA	41
Modelo de Redes Neuronales Recurrentes (RNN).....	45
Conclusiones.....	48
Recomendaciones	52
Referencias Bibliográficas	53
Apéndices.....	57

Lista de Tablas

Tabla 1 <i>Objetivo Específico 1</i>	28
Tabla 2 <i>Objetivo Específico 2</i>	29
Tabla 3 <i>Objetivo Específico 3</i>	30
Tabla 4 <i>Objetivo Específico 4</i>	31
Tabla 5 <i>Descripción de las Variables Involucradas en el Estudio</i>	32
Tabla 6 <i>Medidas de Tendencia Central, Dispersión y Localización</i>	33
Tabla 7 <i>Coefficientes de Correlación Dos a Dos</i>	34
Tabla 8 <i>Coefficientes del Modelo de Regresión Lineal Múltiple</i>	39
Tabla 9 <i>Comparación de Métricas de Rendimiento</i>	47

Lista de Figuras

Figura 1 <i>Gráfico de Caja para 'Volume'</i>	35
Figura 2 <i>Pie Chart para la Variable 'Close Category'</i>	36
Figura 3 <i>Diagrama de Violín para el Precio de Apertura</i>	37
Figura 4 <i>Histogramas de Estandarización</i>	38
Figura 5 <i>Valores Predichos vs. Valores Reales</i>	40
Figura 6 <i>Gráfica de la Serie Temporal para la Variable 'Close'</i>	42
Figura 7 <i>Gráficos para Determinar Valores Óptimos de los Parámetros</i>	43
Figura 8 <i>Validación de los Parámetros Seleccionados Para el Modelo</i>	44
Figura 9 <i>Predicción del Precio del Bitcoin Usando RNN</i>	46

Lista de Apéndices

Apéndice A *Código de Python Utilizado en la Implementación de Modelos Predictivos* 57

Apéndice B *Código de Python Utilizado en el Análisis Exploratorio de los Datos* 57

Introducción

El presente trabajo se fundamenta en la predicción del precio de cierre del Bitcoin, una criptomoneda de alta importancia a nivel mundial debido a su creciente adopción y notable volatilidad en los mercados financieros. Este problema es sumamente desafiante por las características dinámicas y los comportamientos cíclicos de este activo digital, que lo convierten en un caso de estudio ideal para explorar técnicas avanzadas de modelado y predicción.

Por otro lado, los estudios previos sobre la dinámica del precio del Bitcoin presentan la aplicación de diversas técnicas de análisis de datos y modelos predictivos. Hua, (2020) utilizó un modelo ARIMA para presentar una comparación clásica de la previsión de series temporales, con el fin de realizar una predicción eficaz limitada a un intervalo de tiempo corto; por otro lado, Krishna et al. (2023) utilizaron modelos de redes neuronales LSTM para predecir precios basados en patrones temporales, mientras que Gajula et al. (2023) emplearon regresiones Lasso y Ridge, logrando precisiones de hasta 93.05%. Andi (2021) abordó el sobreajuste mediante LSTM y regresión logística, logrando una gestión eficaz de grandes datos. Carrión (2020) aplicó Random Forest, obteniendo coeficientes de determinación entre 95% y 99%, mostrando un ajuste óptimo. Kristoufek (2015) encontró que factores económicos y búsquedas en internet influyen en los precios. Estos estudios fundamentan el análisis del precio del Bitcoin mediante modelos avanzados.

Para abordar este problema, se desarrolló un análisis exhaustivo que abarcó desde la recolección y limpieza de los datos hasta la implementación de modelos predictivos avanzados. Inicialmente, se realizó un análisis exploratorio de los datos para identificar tendencias, patrones y relaciones relevantes entre variables como el precio de apertura, máximo, mínimo y volumen. Posteriormente, se implementaron tres enfoques distintos para predecir el precio de cierre: un

modelo de regresión lineal múltiple, un modelo ARIMA y redes neuronales recurrentes (RNN). Cada modelo fue evaluado utilizando métricas como el error cuadrático medio (MSE), el error absoluto medio (MAE) y la raíz del error medio cuadrático para determinar su efectividad en un entorno caracterizado por alta volatilidad.

Los resultados mostraron que, aunque los tres modelos ofrecieron perspectivas valiosas, cada uno posee sus fortalezas y limitaciones. La regresión lineal múltiple se destacó por su capacidad de capturar tendencias generales gracias a su alto coeficiente de determinación ($R^2 = 0.99928$), pero presentó notables limitaciones en la predicción de variaciones específicas, debido a un alto valor del MSE. Por otro lado, el modelo ARIMA evidenció ser eficaz para patrones temporales estables, aunque su desempeño se vio afectado por la naturaleza abrupta de las variaciones del mercado. Finalmente, las RNN, gracias a su capacidad de capturar dependencias a largo plazo en las series temporales, ofrecieron predicciones más precisas en escenarios dinámicos, mostrándose como una herramienta prometedora para este tipo de análisis.

Este estudio no solo contribuye a la comprensión de la manera en la cual diferentes modelos enfrentan el desafío de predecir un mercado tan complejo, sino que también abre camino para investigaciones futuras que integren enfoques híbridos y variables externas, como eventos económicos o noticias, para mejorar la precisión en las predicciones.

El principal aporte de este estudio radica en la evaluación comparativa de tres enfoques predictivos desde un punto de vista integral, analizando no sólo la precisión de los modelos, sino también su capacidad de adaptación a las características volátiles del mercado del Bitcoin. Por otra parte, se brinda una propuesta metodológica accesible para futuros investigadores, destacando las fortalezas y debilidades de cada técnica en contextos reales. Este trabajo también incluye elementos novedosos, tales como la integración de múltiples modelos desde un enfoque

comparativo, el cual ayuda a identificar el modelo más adecuado según las condiciones específicas del mercado y las necesidades del usuario. En ese orden de ideas, se ofrecen nuevas perspectivas para la predicción en mercados volátiles, con una posible aplicación en otros contextos financieros dinámicos.

Justificación

La volatilidad en el precio de las criptomonedas, particularmente del Bitcoin ha sido una causa de preocupación constante para los inversores. Esta falta de estabilidad se traduce en notables desafíos para quienes necesitan tomar decisiones financieras basadas en un entorno que cambia constantemente (Manuel et al., 2020). Por tal razón se hace sumamente necesario utilizar herramientas analíticas avanzadas que ayuden a comprender el comportamiento de este mercado y a prever posibles tendencias futuras.

La importancia de estudiar la temática en este trabajo, radica en la creciente relevancia de las criptomonedas en las operaciones financieras a nivel mundial. En los últimos años las criptomonedas no solamente se han aceptado como activos de inversión, sino que también se están utilizando en transacciones comerciales y transferencias de valor (Gómez, 2016). En consecuencia, entender y predecir el comportamiento del precio del bitcoin es fundamental para inversores y empresas.

Por otro lado, en este trabajo se pretende brindar una gran utilidad social al ofrecer herramientas y conocimientos apropiados para la población, esto puede motivar a individuos o comunidades a tomar decisiones financieras más informadas y reducir los riesgos de inversión en un mercado tan volátil.

En síntesis, este trabajo no solamente estudia una problemática económica de suma importancia, sino que también busca generar un impacto social positivo al brindar herramientas apropiadas para la toma de decisiones en el ámbito financiero y mejorar la educación financiera en el contexto de las criptomonedas. Su relevancia se apoya en la notable importancia de las criptomonedas en la economía mundial y el constante avance de las tecnologías de análisis de datos.

Por otra parte, la novedad de este trabajo se fundamenta en la aplicación de técnicas avanzadas de machine learning para analizar y predecir el comportamiento del precio del Bitcoin, considerando enfoques innovadores en el análisis de datos con una temática de gran importancia mundial. Ciertamente existen diversos estudios relacionados con las criptomonedas, sin embargo, gran parte de ellos se centran en enfoques tradicionales de modelado estadístico, permitiendo un espacio significativo para la exploración de métodos más modernos y precisos como los modelos de aprendizaje automático (Díaz López et al., 2023).

En este sentido, lo que hace este trabajo especialmente relevante es su enfoque en el uso de algoritmos de machine learning, los cuales ofrecen una capacidad superior para encontrar patrones no lineales, analizar grandes volúmenes de datos, y adaptarse a la naturaleza altamente volátil y compleja de las criptomonedas. En contraste con estudios previos, este proyecto busca no solo predecir precios a corto plazo, sino también analizar qué aspectos tienen mayor peso en la variabilidad del precio, permitiendo a inversores e instituciones financieras entender mejor los riesgos asociados (Díaz López et al., 2023).

Por otra parte, este estudio aborda la falta de una evaluación comparativa integral al analizar múltiples modelos predictivos, como regresión lineal, ARIMA y RNN, en un entorno de alta volatilidad. También explora en detalle las variables financieras más relevantes para el precio del Bitcoin, mostrando claridad sobre su dinámica. También, desarrolla modelos avanzados como RNN para identificar dependencias a largo plazo, un enfoque poco explorado en trabajos anteriores. Finalmente, ofrece una perspectiva práctica con el fin de que inversores y empresas mejoren sus predicciones y reduzcan riesgos financieros.

Sin embargo, es importante incluir factores externos, tales como noticias y eventos macroeconómicos, para evaluar su influencia en el precio del Bitcoin. Por otro lado, sería

fundamental desarrollar modelos híbridos, por ejemplo, ARIMA-RNN, y analizar comportamientos en escalas de tiempo más específicas. También, es necesario investigar el impacto de regulaciones y adopciones tecnológicas en la dinámica de precios y extender la metodología a otras criptomonedas buscando validar la generalización de los hallazgos.

En conclusión, este trabajo no solamente contribuye a los estudios previos al implementar enfoques novedosos en el análisis del precio del Bitcoin, sino que también muestra la importancia de adoptar tecnologías actuales tales como el machine learning para estudiar problemas económicos contemporáneos. Su relevancia radica en ofrecer una perspectiva innovadora y práctica que permita a los inversores y tomadores de decisiones moverse con mayor confianza en un mercado financiero cada vez más digitalizado y complejo.

Planteamiento del Problema

El mercado de criptomonedas, y particularmente el Bitcoin, se caracteriza por una alta volatilidad influenciada por factores externos como noticias económicas, eventos globales y cambios en la percepción de los inversionistas. Esta volatilidad representa un desafío significativo para analistas, inversionistas e instituciones financieras, quienes requieren herramientas confiables para predecir las fluctuaciones de precio y tomar decisiones informadas (Carrión R., 2020). No obstante, la naturaleza compleja y dinámica del mercado dificulta el uso de modelos tradicionales, dado que estos no siempre logran capturar patrones que no son lineales ni responder adecuadamente a cambios abruptos en las tendencias.

En este sentido, existen diversos enfoques para la predicción del precio del Bitcoin, que van desde modelos estadísticos clásicos hasta técnicas avanzadas de machine learning; sin embargo, aún no existe un acuerdo sobre cuál de ellos ofrece la mejor precisión en contextos de alta volatilidad (Díaz López et al., 2023). Por otro lado, la calidad y disponibilidad de los datos utilizados en estos modelos, así como la selección de variables relevantes, pueden afectar significativamente los resultados y limitar la capacidad predictiva de las metodologías existentes.

Ante esta problemática, se hace necesario realizar un estudio riguroso que integre un análisis exploratorio de datos históricos del Bitcoin con la implementación y comparación de modelos predictivos basados en tres enfoques principales: regresión lineal múltiple, modelos de series temporales ARIMA y redes neuronales recurrentes (RNN). Este estudio permitirá evaluar la efectividad de cada modelo de acuerdo con la precisión y desempeño, brindando información valiosa para mejorar la predicción del precio del Bitcoin y fortalecer el desarrollo de herramientas analíticas en un entorno financiero altamente dinámico.

Pregunta de Investigación

¿Cuál de los tres modelos predictivos: regresión lineal múltiple, ARIMA o redes neuronales recurrentes proporciona la mayor precisión en la predicción del precio de cierre del Bitcoin durante el período de mayo de 2023 a mayo de 2024?

Objetivos

Objetivo General

Proponer diferentes modelos predictivos para analizar el comportamiento de la variación del precio del bitcoin durante mayo de 2023 hasta mayo de 2024 utilizando tres algoritmos de machine learning con el fin de determinar el mejor modelo en términos de precisión de la predicción.

Objetivos Específicos

Recopilar datos históricos acerca del precio del Bitcoin en fuentes confiables tales como investing.com, en un periodo de tiempo comprendido entre mayo de 2023 hasta mayo de 2024.

Implementar técnicas de limpieza de datos para identificar y corregir valores atípicos, inconsistencias y posibles errores en la base de datos histórica del precio del Bitcoin, garantizando la calidad y confiabilidad de los datos para su posterior análisis y modelado predictivo.

Identificar patrones y tendencias en los datos recopilados por medio de un análisis descriptivo y exploratorio de las variables involucradas en la base de datos recopilada.

Desarrollar los modelos de machine learning, particularmente: regresión lineal, series de tiempo y redes neuronales para la predicción de precios del Bitcoin.

Antecedentes Sobre el Análisis del Precio del Bitcoin

Desarrollo Histórico del Bitcoin

El Bitcoin se creó en 2009, en ese momento, Satoshi Nakamoto inventa la primera moneda virtual, independiente y descentralizada, la cual nombró como “Bitcoin”; la principal motivación para crear esta criptomoneda fue la necesidad de establecer un sistema financiero descentralizado y resistente a la censura, inspirado por los problemas observados en el sistema financiero tradicional, especialmente durante la crisis económica global de 2008 (Carrión R., 2020). Nakamoto identificó varias debilidades en los sistemas de pago existentes y en las instituciones financieras tradicionales, lo que llevó a la concepción de Bitcoin como una alternativa más transparente, segura y autónoma. A partir de esta época han transcurrido 17 años y actualmente la moneda ha mostrado un crecimiento permanente con respecto a su producción, y a la vez bastante volátil respecto a su valor debido a las especulaciones; por otro lado, el mercado de oferta de las criptomonedas se compone por alrededor de 700 clases de monedas digitales (Álvarez-Díaz, 2019).

Más adelante, en 2009 también se inauguró el software de código abierto que desarrolla el protocolo Bitcoin y se minó el primer bloque de Bitcoin, conocido como el "bloque génesis"(Regal et al., 2019). Desde ese momento, Bitcoin se ha transformado de una curiosidad tecnológica a un activo financiero de importancia global.

Evolución del Precio del Bitcoin

Desde el momento de su creación, el precio del Bitcoin ha sufrido una volatilidad bastante notable. En sus primeros años (entre 2009 y 2010), el Bitcoin solo valía unos centavos de dólar; en 2011 el precio empieza a tener un cambio significativo posicionándose en 30 USD por unidad; en septiembre de 2012 se dio nacimiento a una entidad que pueda estandarizar,

proteger, informar y promocionar el protocolo de código abierto que conforma el Bitcoin como mecanismo de aceleración del crecimiento global de la moneda (Álvarez-Díaz, 2019), razón por la cual el precio disminuyó a 13 USD por unidad; en 2013 debido a que se presenta el primer conflicto legal en relación a la moneda a causa de que la Administración de Control de Drogas de los Estados Unidos incautó 11,02 Bitcoin e informó del procedimiento al Departamento de Justicia de los Estados Unidos, el bitcoin alcanza una subida histórica a un precio de 1242 USD (Sánchez de Diego Martínez-Cabrera Directora & Jesús Giménez Abad Madrid, 2014); en contraste, para 2014 el Bitcoin empieza a ser aceptado como moneda de inversión, varias empresas de que operan en internet manifiestan la aceptación del Bitcoin como moneda de pago, en este momento su precio se posiciona en 450 USD; en 2017, alcanzó un máximo histórico de casi 19728 USD, seguido por una fuerte caída para el siguiente año a 6050 USD. En los años siguientes, el precio ha fluctuado notoriamente, alcanzando nuevos máximos históricos en 2020 y 2021.

Estudios Previos (Análisis de Datos) Acerca de la Evolución del Precio del Bitcoin

La dinámica del precio del Bitcoin ha sido un tema de gran interés por parte de numerosos investigadores enfocados particularmente en la analítica de datos. Por ejemplo, Krishna, Sapthami, Banka y Ravela (2023) crearon un modelo que busca predecir el precio de la criptomoneda utilizando LSTM (Long Short Term Memory) con el fin de generar evaluaciones basadas en consultas de acciones de bitcoin, mientras que Gajula, Pavankumar, J.Velmurugan (2023) crearon modelos de regresión de Lasso (método para simplificar modelos lineales) y Ridge que permitan predecir el precio de las criptomonedas en los últimos años, allí determinaron un valor de precisión del 93,05% para el modelo lineal Ridge, también encontraron que método de regresión de Lasso tiene una puntuación de 91,01% de corrección; por otro lado,

se destaca el trabajo de Andi (2021) en el cual se busca resolver el problema del sobreajuste mediante una gestión adecuada de grandes datos utilizando el modelo LSTM (aprendizaje de patrones en secuencias) y regresión logística, allí el problema se logró resolver exitosamente mediante una gestión adecuada de grandes datos; también cabe mencionar la investigación desarrollada por Carrión (2020) en la cual se busca implementar algunos procesos de Machine Learning, en particular el algoritmo de Random Forest, con el fin de predecir el precio de las criptomonedas de mayor popularidad, allí se concluyó que el coeficiente de determinación generalmente se incrementa cuando se recopilan más datos. Este valor se sitúa entre un 95 y 99 % para estos modelos implementados después de haber elaborado una búsqueda de la configuración óptima de hiperparámetros, esto muestra que los errores son bastante pequeños y que el ajuste ha sido satisfactorio; finalmente destacamos el trabajo de Kristoufek (2015) quien encontró que tanto factores económicos como de búsqueda en internet afectan su precio. Estas investigaciones brindan una base teórica para el análisis por medio de modelos de regresión del precio del Bitcoin.

Por otro lado, los estudios mencionados se centran en el uso de modelos predictivos avanzados, como LSTM, regresión regularizada, Random Forest, y análisis multivariable para entender y predecir el precio del Bitcoin. Su mención en este trabajo se justifica por su enfoque en técnicas de machine learning y análisis estadístico aplicado a datos históricos, los cuales coinciden con los objetivos y métodos de este estudio. Este tipo de investigaciones representa la base teórica y metodológica necesaria para desarrollar modelos de predicción más robustos y útiles para un mercado tan volátil como el del Bitcoin.

No obstante, existen otros estudios sobre el precio de Bitcoin que no se incluyeron aquí debido a diversas razones:

Foco en métodos tradicionales: Algunos estudios emplean métodos más convencionales, como análisis econométrico o modelos ARIMA básicos, los, aunque son útiles, no garantizan la misma capacidad predictiva ni el enfoque innovador de las técnicas de *Machine Learning*.

Limitaciones en el alcance: Otras investigaciones se centran solamente en el impacto de un factor aislado, como las políticas gubernamentales o la adopción tecnológica, sin elaborar un análisis integral del comportamiento del precio.

Enfoques teóricos sin aplicaciones prácticas: Algunos trabajos presentan análisis teóricos o simulaciones que no consideran datos reales ni ofrecen herramientas prácticas para predecir el precio.

Falta de actualización: En un mercado tan dinámico como el de las criptomonedas, las investigaciones más antiguas pueden no reflejar las condiciones actuales ni las tendencias más recientes del mercado.

En resumen, este trabajo se fundamenta en investigaciones que integran datos históricos, técnicas modernas de análisis y machine learning, puesto que ofrecen una aproximación más sólida y precisa para analizar la volatilidad del precio del Bitcoin y predecir tendencias futuras.

A diferencia de las investigaciones mencionadas, que han abordado la predicción del precio del Bitcoin utilizando exclusivamente modelos tradicionales o enfoques basados en series temporales, este proyecto elabora un análisis completo que combina técnicas de limpieza de datos, exploración visual y modelos avanzados de predicción como regresión lineal múltiple, series de tiempo ARIMA y redes neuronales recurrentes (RNN). Por otro lado, se hace énfasis en la comparación entre modelos, evaluando su desempeño bajo distintas métricas para identificar el enfoque más adecuado según las características del comportamiento histórico del Bitcoin. Este trabajo también presta atención particular en la transformación y preprocesamiento de datos para

garantizar su calidad, lo que permite reducir sesgos y mejorar la capacidad predictiva de los modelos, destacándose por un enfoque metodológico integral que abarca desde la recolección de datos hasta la visualización de los resultados.

Marco Referencial

El propósito de este trabajo es brindar una base teórica y conceptual que facilite comprender y sustentar la investigación acerca de la evolución del precio del Bitcoin. Por medio de la revisión de literatura y teorías adecuadas, se definirán las bases necesarias para la recolección, análisis y visualización de datos históricos sobre el precio del Bitcoin, también para la aplicación de modelos de predicción utilizando técnicas de machine learning tales como la regresión lineal, el modelo de series de tiempo ARIMA y las redes neuronales recurrentes aplicadas a estos datos recolectados.

Para desarrollar lo anterior, se recolectarán datos históricos del precio del Bitcoin en el periodo de mayo de 2023 a mayo de 2024 de la página web <https://es.investing.com/crypto/bitcoin/historical-data> , se realizará un análisis exploratorio o caracterización de estos datos, se aplicarán modelos predictivos y se visualizarán los resultados, todo esto será desarrollado en el lenguaje de programación Python.

Marco Teórico

En este apartado se presentan los fundamentos teóricos necesarios para el análisis de la evolución del precio del Bitcoin. Se introduce el concepto de series temporales y se describen los modelos que se implementarán: regresión lineal, ARIMA y redes neuronales recurrentes. Estos modelos han sido seleccionados por su capacidad para manipular series temporales y su efectividad demostrada en estudios recientes sobre predicción de precios en mercados financieros y criptomonedas.

En este sentido, el análisis de series temporales se consolida como una técnica clave en la modelización de datos secuenciales, utilizada ampliamente en el ámbito financiero para predecir tendencias futuras. Los modelos de predicción seleccionados han sido aplicados en

investigaciones recientes para evaluar la volatilidad del mercado de criptomonedas (Zhang et al., 2023; Liu & Chen, 2022).

Series de Tiempo ARIMA

El modelo ARIMA es ampliamente utilizado en el análisis y predicción de series temporales financieras (Box et al., 2016). Se compone de tres elementos fundamentales:

- AR (Autoregresivo): Expresa el valor actual de la serie como una función lineal de valores pasados.
- I (Integrado): La serie se diferencia una o más veces para hacerla estacionaria.
- MA (Media Móvil): Representa el valor actual de la serie como una combinación de errores pasados.

El modelo ARIMA se representa como ARIMA(p, d, q), donde:

- p es el orden de la parte autoregresiva (AR).
- d es el grado de diferenciación (I).
- q es el orden de la media móvil (MA).

La ecuación general del modelo ARIMA (p, d, q) se expresa como:

$$\Phi_p(B)(1 - B)^d Y_t = \Theta_q(B)\epsilon_t \quad (4)$$

Donde:

- $\Phi_p(B)$ es el polinomio autoregresivo de grado p :

$$\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

- $(1 - B)^d$ es el operador de diferenciación de orden d :

$$(1 - B)^d Y_t = \Delta^d Y_t$$

- $\Theta_q(B)$ es el polinomio de media móvil de grado q :

$$\Theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

- B es el operador de retraso (backshift operator), tal que $BY_t = Y_{t-1}$.
- Y_t es el valor de la serie temporal en el tiempo t .
- ϵ_t es el término de error en el tiempo t .

En aplicaciones recientes, el modelo ARIMA ha mostrado un buen desempeño en la predicción de precios de activos digitales, aunque con limitaciones en la captura de cambios abruptos en la volatilidad del mercado (Rahman et al., 2021).

Regresión Lineal Simple y Múltiple

Es un modelo simple que bajo ciertas condiciones asume una relación lineal de la forma $\tilde{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$ entre las variables independientes y la variable dependiente.

Para el caso de una sola variable independiente se pueden encontrar los parámetros β_0 y β_1 del modelo $\tilde{y} = \beta_0 + \beta_1 x$ estableciendo la métrica conocida como el error medio cuadrático, definido por:

$$E(\beta_0, \beta_1) = \frac{1}{2} \sum_{i=1}^n (\tilde{y} - y_i)^2 = \frac{1}{2} \sum_{i=1}^n (\beta_0 + \beta_1 x_i - y_i)^2 \quad (1)$$

Siendo (x_i, y_i) los pares de valores correspondientes a la base de datos de entrenamiento para las variables independiente en relación con la dependiente, los parámetros β_0 y β_1 del modelo son los valores que minimizan el error medio cuadrático, un cálculo con derivadas parciales demuestra que estos valores son:

$$\beta_0 = \frac{(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n x_i y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (2)$$

$$\beta_1 = \frac{n(\sum_{i=1}^n x_i y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (3)$$

Aunque la regresión lineal ha sido aplicada en estudios previos para predecir precios de criptomonedas (Sharma et al., 2022), su capacidad predictiva es limitada en mercados con alta volatilidad debido a la no linealidad de los datos.

Modelo de Redes Neuronales Recurrentes (RNN)

Las Redes Neuronales Recurrentes (RNN) consisten en una clase de red neuronal creada para procesar y analizar secuencias de datos. Las RNN tienen conexiones que forman ciclos, lo que les ayuda a mantener una memoria de estados anteriores y utilizar esta información para predecir la siguiente salida en la secuencia. (Matich, 2001)

De acuerdo con esta definición, las RNN tienen un estado oculto que depende de un tiempo t representado como h_t , el cual se calcula usando el estado h_{t-1} , con este modelo se pretende encontrar la entrada actual x_t y la salida actual y_t a partir del estado oculto.

Considerando las variables anteriores, las ecuaciones del modelo RNN son:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b_h) \quad (5)$$

Donde:

- h_t es el estado oculto en el tiempo t .
- σ es la función de activación (puede ser \tanh).
- W_h es la matriz de pesos que conecta el estado oculto anterior al estado oculto actual.
- W_x es la matriz de pesos que conecta la entrada actual al estado oculto.
- b_h es el sesgo del estado oculto.

$$y_t = \phi(W_y h_t + b_y) \quad (6)$$

Donde:

- y_t es la salida en el tiempo t .
- ϕ es la función de activación de la salida.
- W_y es la matriz de pesos que conecta el estado oculto a la salida.
- b_y es el sesgo de la salida.

Estudios recientes han demostrado que variantes como las Long Short-Term Memory (LSTM) y las Gated Recurrent Units (GRU) mejoran significativamente la capacidad de predicción en el mercado de criptomonedas (Goodfellow et al., 2016; Kim & Kim, 2023).

Metodología

A continuación, detallamos las actividades a realizar durante el desarrollo del presente proyecto, en relación con cada uno de los objetivos específicos propuestos:

Tabla 1

Objetivo Específico 1

Actividad	Justificación
Descargar la información de la base de datos investing.com	Garantizar la obtención de datos históricos confiables y representativos del precio del Bitcoin, los cuales son esenciales para el análisis exploratorio y la implementación de modelos predictivos.

Nota. Recopilar datos históricos acerca del precio del Bitcoin en fuentes confiables tales como investing.com, en un periodo de tiempo comprendido entre mayo de 2023 hasta mayo de 2024.

Tabla 2*Objetivo Específico 2*

Actividad	Justificación
Depuración y estructuración de la base de datos	Eliminar errores e inconsistencias que podrían afectar la calidad del análisis y la precisión de los modelos predictivos.
Tratamiento datos nulos	Garantizar la integridad de los datos al eliminar o imputar valores faltantes, reduciendo su impacto en los resultados.
Categorización de la variable de precio de cierre	Facilitar la identificación de patrones específicos mediante la segmentación de los precios en rangos, lo cual permite un análisis descriptivo más detallado.

Nota. Implementar técnicas de limpieza de datos para identificar y corregir valores atípicos, inconsistencias y posibles errores en la base de datos histórica del precio del Bitcoin, garantizando la calidad y confiabilidad de los datos para su posterior análisis y modelado predictivo.

Tabla 3*Objetivo Específico 3*

Actividad	Justificación
Análisis estadístico de los datos (moda, media, mediana, desviación estándar, entre otros)	Comprender las propiedades fundamentales de los datos, identificando tendencias generales, valores atípicos y variabilidad en los precios del Bitcoin.
Visualización de datos mediante gráficos de barras, diagramas de dispersión, gráficos de caja y gráficos circulares	Identificar relaciones entre variables, detectar anomalías y patrones complejos que no son evidentes en los análisis numéricos.
Generación de gráficos de tendencias del precio del Bitcoin	Evaluar la evolución temporal del precio, identificando ciclos, períodos de alta volatilidad y eventos relevantes en el mercado.

Nota. Identificar patrones y tendencias en los datos recopilados por medio de un análisis descriptivo y exploratorio de las variables involucradas en la base de datos recopilada.

Tabla 4*Objetivo Específico 4*

Actividad	Justificación
División de los datos en conjuntos de entrenamiento (80%) y prueba (20%)	Asegurar que los modelos sean entrenados y evaluados correctamente, equilibrando precisión y capacidad de generalización.
Implementación del modelo de regresión lineal múltiple	Establecer una referencia inicial para predecir el precio de cierre del Bitcoin utilizando variables como apertura, máximo, mínimo y volumen.
Implementación del modelo ARIMA	Modelar y predecir tendencias temporales y estacionales en el precio del Bitcoin, utilizando información histórica.

Nota. Desarrollar los modelos de machine learning, particularmente: regresión lineal, series de tiempo y redes neuronales para la predicción de precios del Bitcoin.

Resultados

Recolección y Limpieza de Datos

Para el desarrollo del presente proyecto, se descargó la base de datos a utilizar por medio de este [link](#), allí se registran precios diarios de apertura, cierre máximo, mínimo y volumen en un intervalo de tiempo comprendido entre mayo de 2023 y mayo de 2024, esta base de datos descargada en el archivo de Excel puede encontrarse [aquí](#).

En esta base de datos, se observa inicialmente que las columnas 'Close y 'Adj Close' presentan exactamente los mismos valores para todos los registros, procedemos a eliminar la columna 'Adj Close' ya que no presenta información relevante para el análisis. Las demás columnas son importantes para el análisis financiero, por lo cual no se eliminan, a continuación, detallamos en la siguiente tabla el significado de cada una:

Tabla 5

Descripción de las Variables Involucradas en el Estudio

Variable	Descripción
Date	Fecha de la observación
Open	Precio de apertura.
High	Precio más alto
Low	Precio más bajo
Close	Precio de cierre
Volume	Volumen de operaciones
Daily range	Diferencia entre el precio máximo y el precio mínimo

Nota. Estas variables involucradas en el estudio, son las obtenidas después de hacer la limpieza de datos de la base original.

También, se identifica una única fila en la cual todas las variables excepto la fecha presenta valores nulos, esta fila se elimina; por otro lado, con el fin de elaborar un análisis más detallado, se agrega la columna de rango diario definiéndola como la resta entre el precio máximo y el precio mínimo, además se realiza una categorización de la columna de precio de cierre con base en los quintiles. La nueva base de datos con la limpieza realizada puede encontrarse en la hoja ‘Datos posteriores-Trans o Prepr’ del archivo presente en este [link](#).

Análisis Exploratorio de los Datos

A continuación, desarrollaremos un Análisis Exploratorio de Datos (EDA) sobre la evolución del precio del Bitcoin en relación con la base de datos obtenida para la cual se describen las variables en la Tabla 1. Los códigos de Python utilizados para llevar a cabo este análisis, pueden encontrarse en el Apéndice B.

Tabla 6

Medidas de Tendencia Central, Dispersión y Localización

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
Open	366	39687.1002	14421.7697	25121.6738	27629.4736	34839.3984	44191.5205	73079.375
High	366	40427.7014	14858.6057	25735.3086	28045.3794	35271.0293	45557.6904	73750.0703
Low	366	38950.2463	13877.1214	24797.168	27151.7427	34286.1328	43619.0195	71334.0904
Close	366	39769.4846	14446.7142	25124.6758	27631.0107	34987.8066	44283.6475	73083.5
Volume	366	2.1942E+10	1.4142E+10	5331172801	1.2635E+10	1.7445E+10	2.6006E+10	1.028E+11
Daily Range	366	1477.45509	1436.35849	101.910156	574.178711	947.741211	1787.69727	9846.71484

Nota. Aquí se presentan las medidas de tendencia central, dispersión y localización de las variables cuantitativas incluidas en este estudio (ver Tabla 1) incluyendo media, desviación estándar, máximo, mínimo y cuartiles.

En esta tabla se observa que los precios de Bitcoin (Open, High, Low, Close) están estrechamente relacionados, con alta volatilidad. El volumen de transacciones muestra una gran dispersión, con picos significativos en ciertos días; el rango diario promedio es alto, reflejando

grandes fluctuaciones en los precios. En general, los datos indican alta volatilidad tanto en los precios como en el volumen y las fluctuaciones diarias.

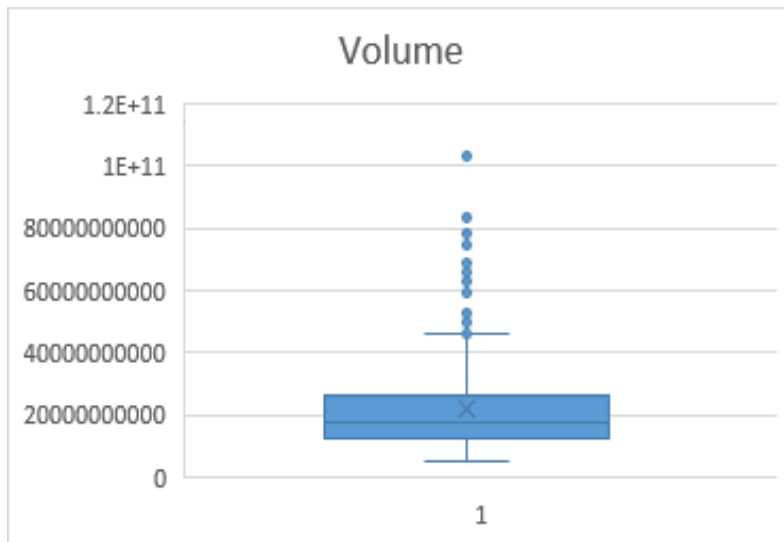
Tabla 7

Coefficientes de Correlación Dos a Dos

Variable	Open	High	Low	Close	Volume	Daily Range
Open	1	0.998359	0.997696	0.996546	0.724451	0.692766
High	0.998359	1	0.997361	0.998767	0.742687	0.712981
Low	0.997696	0.997361	1	0.998381	0.700447	0.660188
Close	0.996546	0.998767	0.998381	1	0.725293	0.690368
Volume	0.724451	0.742687	0.700447	0.725293	1	0.918512
Daily Range	0.692766	0.712981	0.660188	0.690368	0.918512	1

Nota. Aquí se presentan los coeficientes de correlación dos a dos para las diferentes variables cuantitativas involucradas en este estudio (ver Tabla 1).

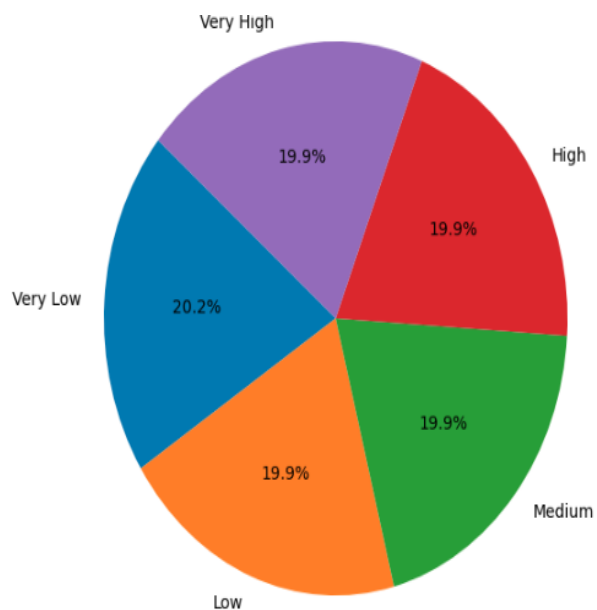
La tabla muestra que las variables Open, High, Low y Close están fuertemente correlacionadas entre sí, lo que indica una relación lineal casi perfecta entre los precios. El volumen tiene una correlación moderada con los precios, sugiriendo una relación positiva, pero no tan fuerte. El rango diario también muestra una correlación moderada con los precios, y la alta correlación entre el rango diario y el volumen (0.92) indica que los días con mayores volúmenes de transacciones suelen tener mayores fluctuaciones de precios, lo que es clave para entender la volatilidad en el mercado de Bitcoin.

Figura 1*Gráfico de Caja para 'Volume'*

El gráfico de caja muestra que la mayoría de los volúmenes de transacciones están en un rango bajo, pero existen valores atípicos significativamente altos. La mediana, ubicada cerca de 1.8×10^{10} , indica una distribución asimétrica con sesgo a la derecha, reflejando transacciones de volumen inusualmente alto en el conjunto de datos.

Figura 2

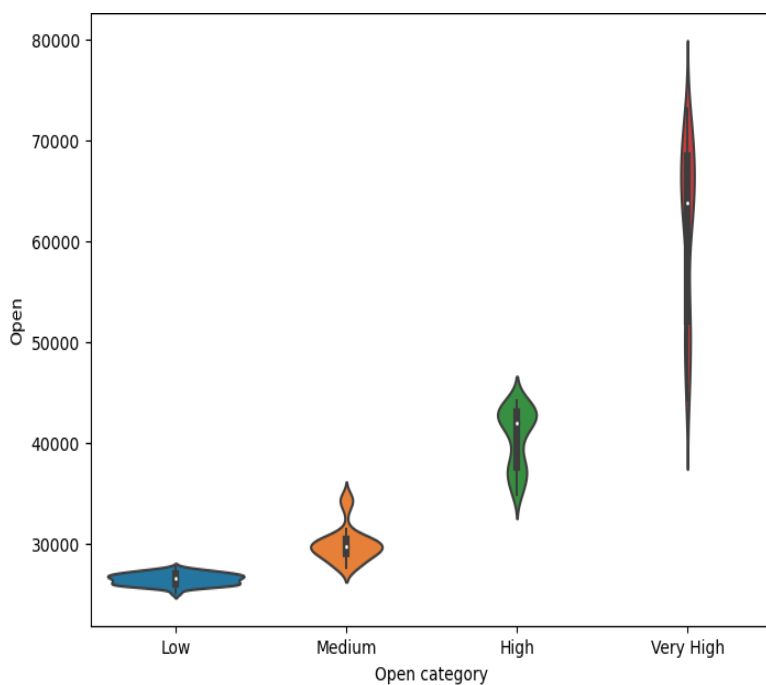
Pie Chart para la Variable 'Close Category'



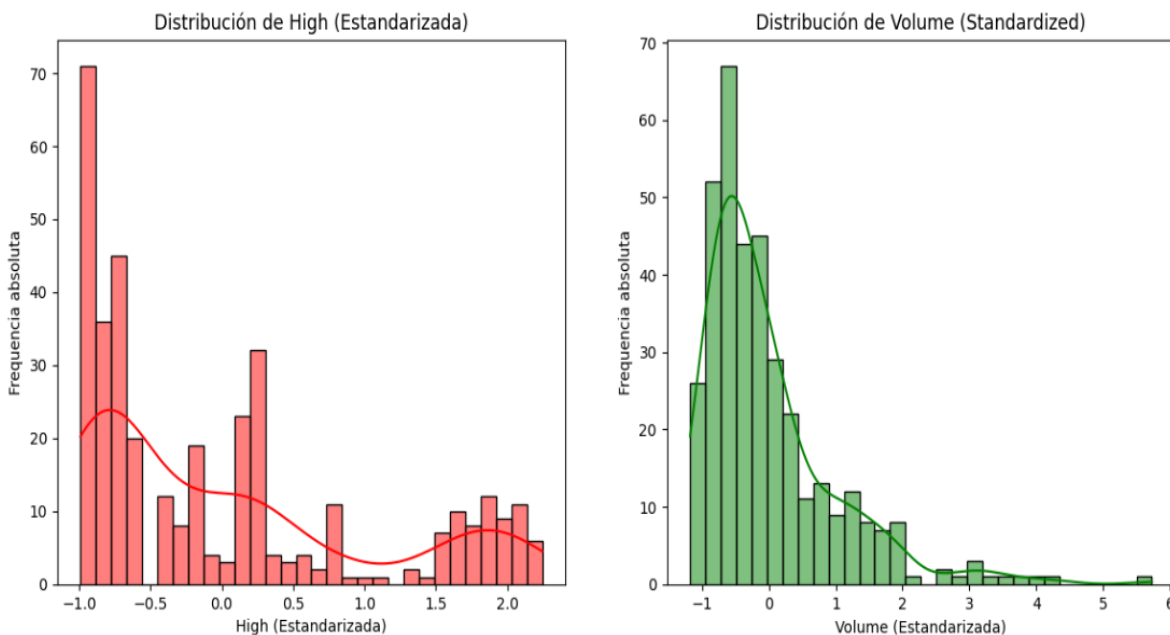
El gráfico muestra que la categoría "Very Low" predomina ligeramente, mientras que las demás categorías tienen una distribución uniforme (19.9% cada una). Esto indica que los precios de cierre del Bitcoin varían considerablemente, sin una tendencia clara hacia una categoría específica.

Figura 3

Diagrama de Violín para el Precio de Apertura



El gráfico muestra que las categorías "Low" y "Medium" tienen distribuciones más compactas y con menos variabilidad en los precios de apertura, mientras que "High" presenta más variabilidad y simetría. "Very High" muestra la mayor variabilidad y una distribución sesgada, con valores atípicos posibles, alcanzando hasta los 80,000 USD.

Figura 4*Histogramas de Estandarización*

Nota. Aquí se presenta la estandarización de las variables de precio máximo y volumen, también la elaboración de histogramas para estas dos variables.

El histograma del precio máximo estandarizado muestra una distribución asimétrica hacia la derecha, con la mayoría de los datos concentrados en valores negativos y varios picos, sugiriendo subgrupos en el mercado. En cuanto al volumen de transacciones de Bitcoin, la distribución también es asimétrica, con una cola larga hacia la derecha, lo que indica días con volúmenes altos. Comparando ambas variables, ambas presentan sesgo positivo, pero el volumen tiene mayor dispersión y una cola más larga, mientras que el precio máximo muestra varios picos.

Desarrollo de Modelos Predictivos

El desarrollo de modelos predictivos para el precio del Bitcoin se basa en técnicas como regresión lineal, ARIMA y redes neuronales recurrentes (RNN). Estos modelos se evalúan para identificar el más adecuado en cuanto a precisión y capacidad de interpretación, con el objetivo de comprender mejor la dinámica del mercado de criptomonedas. Los códigos de Python para la implementación de cada uno de estos modelos pueden encontrarse en el Apéndice A.

Modelo de Regresión Lineal Múltiple

Para la implementación de este modelo, consideramos como variable dependiente o de respuesta (target) el precio de cierre, y como variables independientes los precios de apertura, máximo, mínimo y volumen. La siguiente tabla muestra los resultados obtenidos para el modelo:

Tabla 8

Coefficientes del Modelo de Regresión Lineal Múltiple

β_0 (intercepto)	β_1	β_2	β_3	β_4
-45.823	$-5.196 \cdot 10^{-1}$	$-9.309 \cdot 10^{-1}$	$5.891 \cdot 10^{-1}$	$-6.309 \cdot 10^{-9}$

Nota. Aquí se muestran los coeficientes del modelo de regresión lineal múltiple para el precio de cierre en función de los precios de apertura, máximo y mínimo y volumen.

Estos resultados muestran que una posible ecuación que relaciona el precio de cierre en función de las demás variables mencionadas es:

$$y = -45.823 - 5.196 \cdot 10^{-1}x_1 - 9.309 \cdot 10^{-1}x_2 + 5.891 \cdot 10^{-1}x_3 - 6.309 \cdot 10^{-9}x_4$$

Siendo

y : precio de cierre.

x_1 : precio de apertura.

x_2 : precio máximo.

x_3 : precio mínimo.

x_4 : volumen de transacciones.

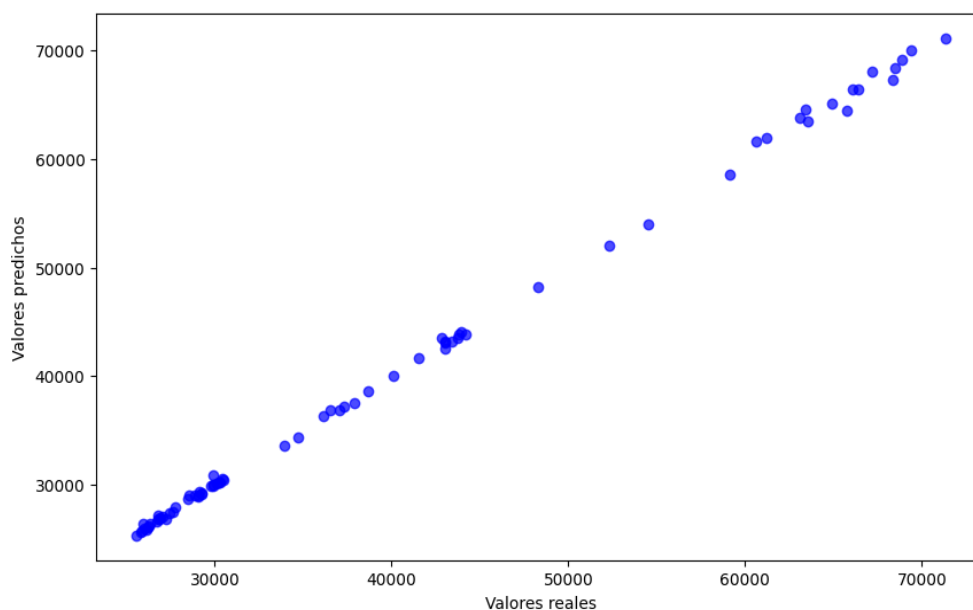
Este modelo podría utilizarse para realizar predicciones si se prevén ciertos valores para precio de apertura, precio máximo precio mínimo y volumen de transacciones.

Por otro lado, el modelo de regresión lineal muestra una fuerte correlación (0.99928) entre los valores predichos y reales, lo que indica que captura la tendencia general del precio del Bitcoin. Sin embargo, su error cuadrático medio (MSE) de 159817 sugiere que tiene limitaciones para predecir variaciones puntuales, especialmente en momentos de alta volatilidad. Aunque útil para análisis generales, el modelo podría necesitar técnicas más avanzadas para mejorar su precisión en mercados dinámicos como el de las criptomonedas.

En este sentido, presentamos a continuación un gráfico de dispersión que muestra la comparación entre los valores predichos por el modelo utilizando la base de prueba y los valores reales:

Figura 5

Valores Predichos vs. Valores Reales



El gráfico muestra una fuerte correlación entre los valores reales y predichos del modelo de regresión lineal múltiple para el precio de cierre del Bitcoin, ya que los puntos están cercanos a la diagonal $y = x$. Esto indica un buen desempeño del modelo al capturar la relación entre las variables independientes (precio de apertura, máximo, mínimo y volumen) y el precio de cierre, sin evidencias de sesgo sistemático.

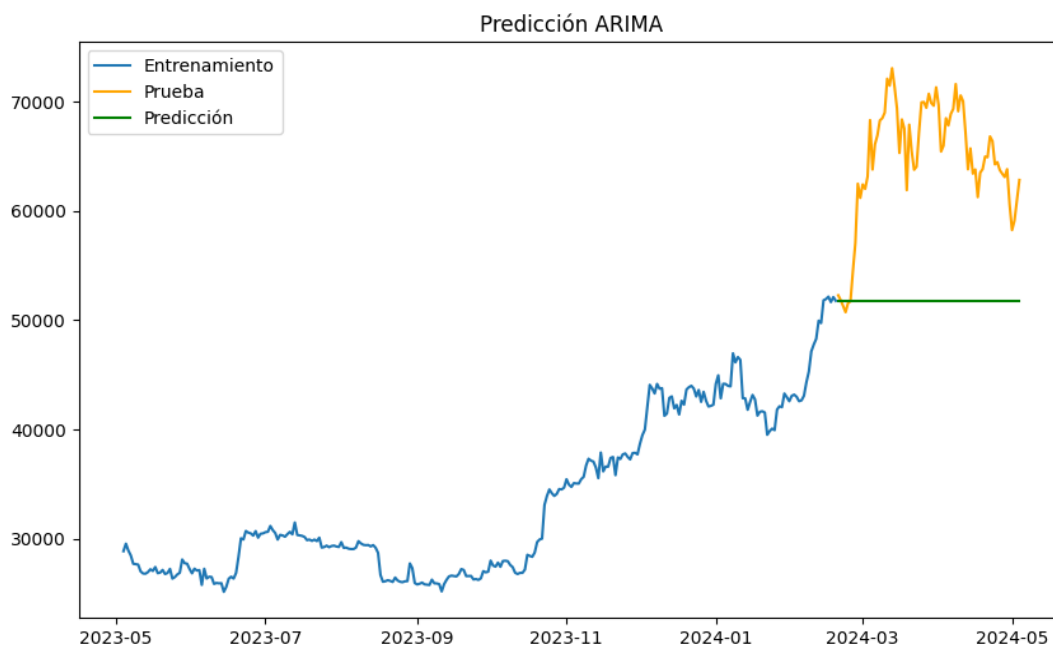
Modelo ARIMA

La implementación del modelo ARIMA (Autoregressive Integrated Moving Average) en este trabajo se enfoca en analizar y predecir la evolución del precio del Bitcoin, por medio de un enfoque basado en datos históricos. El modelo fue configurado eligiendo los valores óptimos de los parámetros (p,d,q) , seleccionados como punto de partida para capturar las tendencias, diferencias estacionarias y componentes de ruido presentes en la serie temporal. Los datos fueron preparados cuidadosamente, convirtiendo la columna de fechas en un índice temporal y seleccionando la columna de precios de cierre como la variable de interés.

En este orden de ideas, para seleccionar los valores óptimos de los parámetros, verificamos si la serie temporal generada por medio de la variable de precio de cierre es estacionaria por medio del comando `test_stationarity()`, el código indica que no lo es, por lo cual se hace necesario convertirla en estacionaria, tomando en principio el valor de diferenciación $d = 1$, por otra parte, esto se puede comprobar también observando el gráfico de la serie temporal correspondiente:

Figura 6

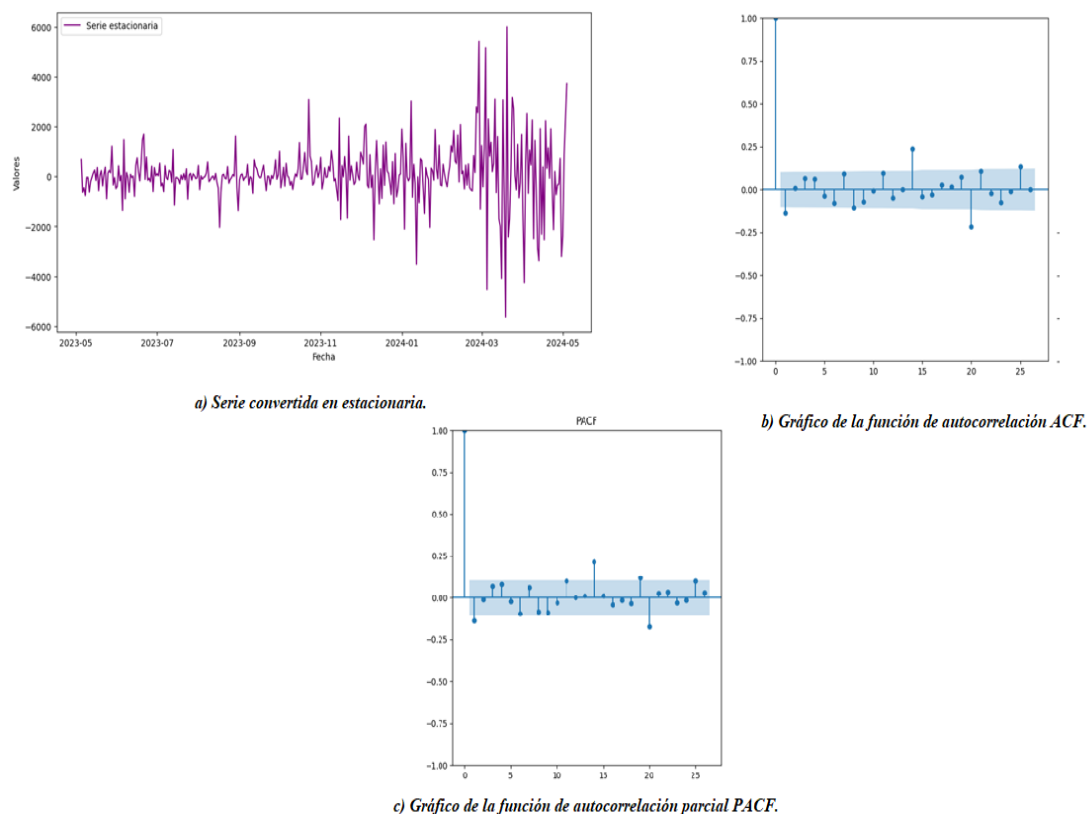
Gráfica de la Serie Temporal para la Variable 'Close'



En este gráfico se evidencia que los valores de la serie parecen volverse más dispersos a medida que avanza el tiempo, lo cual sugiere que la varianza no es constante, además los valores al principio están concentrados y más adelante tienen grandes oscilaciones, por lo cual la serie no es estacionaria. Seguidamente se procede a convertirla en estacionaria utilizando el comando `series.diff().dropna()` con el fin de calcular la primera diferencia $Y'_t = Y_t - Y_{t-1}$. En la figura 8 a), se observa que basta con esta primera diferenciación para convertir la serie en estacionaria.

Figura 7

Gráficos para Determinar Valores Óptimos de los Parámetros

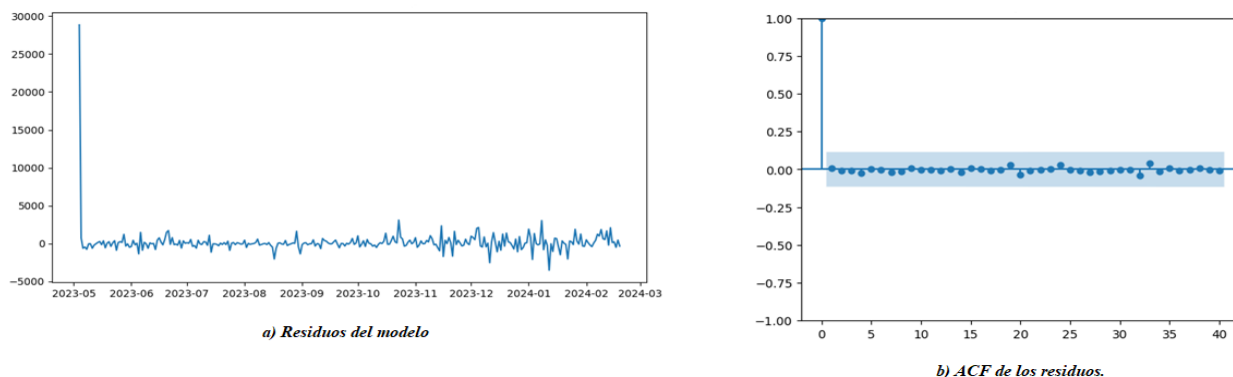


Una vez se ha convertido la serie temporal en estacionaria, se procede a dividir la base de datos en conjuntos de entrenamiento (80%) y prueba (20%) con el fin de evaluar el rendimiento del modelo. Las predicciones realizadas sobre el conjunto de prueba se analizaron mediante métricas tales como el error cuadrático medio (MSE), que proporciona una medida cuantitativa de la precisión del modelo. Además, se visualizó el comportamiento del modelo en comparación con los datos reales mediante gráficos que muestran el ajuste del modelo a la serie temporal histórica y su capacidad predictiva. Esta implementación demuestra cómo el modelo ARIMA puede ser aplicado para capturar patrones históricos y proyectar tendencias futuras en mercados financieros volátiles como el de las criptomonedas.

Posterior a la conversión de la serie en estacionaria, se procede a visualizar los gráficos de ACF y PACF (Figuras 7 b) y 7 c) con el fin de calcular los parámetros p y q . En el gráfico de ACF, se observa un pico significativo en el rezago 1 (dado que esta barra se encuentra por fuera de la franja azul) y que luego las autocorrelaciones se acercan rápidamente a cero. Esto sugiere que un valor de $q = 1$ podría ser suficiente para capturar la parte de promedios móviles (MA) de la serie. Por otro lado, en el gráfico de PACF se observa que no hay rezagos significativos después del rezago 0, ya que estos están dentro de la franja azul o bien no se alejan significativamente de ella. Esto sugiere que no se necesita incluir un componente autorregresivo, por cual podemos tomar $p = 0$. Una vez elegidos estos parámetros óptimos, procedemos a graficar los residuos y el ACF de los residuos para verificar si el modelo está ajustando correctamente los datos, con lo cual se pretende evaluar las métricas del modelo:

Figura 8

Validación de los Parámetros Seleccionados Para el Modelo



En la Figura 8 se observa que los residuos parecen oscilar alrededor de cero sin mostrar patrones evidentes de autocorrelación ni tendencia, lo que indica que el modelo está ajustando bien los datos. Por otra parte, las autocorrelaciones de los residuos caen dentro del intervalo de confianza (bandas azules) en casi todos los rezagos, esto confirma que no hay información

significativa en los residuos que el modelo no haya capturado, lo cual valida los valores seleccionados de $p = 0$, $d = 1$, $q = 1$.

Después de calcular y validar los valores anteriores para el modelo, se observa que el valor del error absoluto medio (MSE) obtenido fue de 13135.31, el cual es un valor aceptable teniendo en cuenta que la serie tiene valores grandes en la escala de decenas de miles; sin embargo, dado que el RMSE fue de 14150.05 (valor significativamente mayor al MSE), esto puede indicar que hay algunos errores grandes en las predicciones que afectan fuertemente el rendimiento del modelo (tal como se observa en la Figura 6). Este comportamiento puede presentarse debido a que el precio de las criptomonedas puede presentar movimientos bruscos y tendencias inesperadas, lo que puede dificultar la predicción utilizando modelos estadísticos tradicionales como ARIMA.

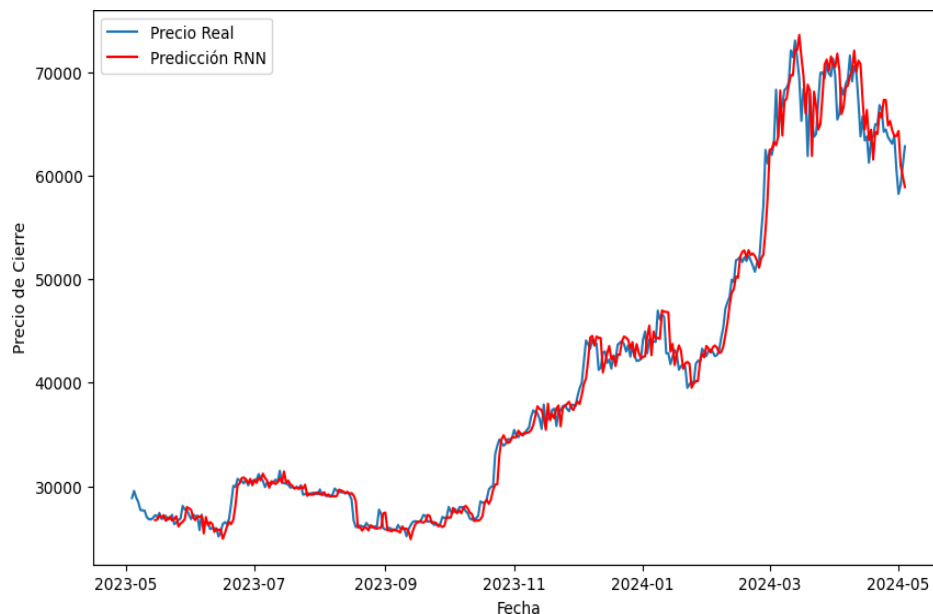
Modelo de Redes Neuronales Recurrentes (RNN)

Con el propósito de mejorar la capacidad predictiva del modelo en el análisis de los precios del Bitcoin, se implementó un modelo de Redes Neuronales Recurrentes (RNN). Las RNN son una clase de redes neuronales que son especialmente efectivas para trabajar con secuencias de datos temporales, como las series de tiempo. Como el precio del Bitcoin sigue una dinámica temporal, las RNN tienen la ventaja de recordar las dependencias pasadas a través de sus conexiones recurrentes, lo que les permite capturar mejor las tendencias y patrones en datos secuenciales. En este modelo, se utiliza una capa LSTM (Long Short-Term Memory), la cual es una variante de las RNN diseñada para mitigar el problema del desvanecimiento del gradiente y permitir que el modelo recuerde información importante a largo plazo. A continuación, se detalla el código implementado para entrenar y evaluar este modelo, con el objetivo de predecir los precios futuros del Bitcoin basándose en los datos históricos.

La siguiente gráfica muestra el resultado obtenido a partir de esta comparación entre los valores reales y las predicciones:

Figura 9

Predicción del Precio del Bitcoin Usando RNN



En la figura 9 se observa que el modelo RNN demuestra una alta precisión al seguir las tendencias generales y los picos más pronunciados en los precios del Bitcoin, reflejando de manera efectiva las fluctuaciones a largo plazo. Las líneas roja (predicciones) y azul (valores reales) son bastante similares en la mayor parte del gráfico, lo que sugiere que el modelo es competente para predecir el comportamiento general del mercado. Sin embargo, se observan ligeras desviaciones en algunas pequeñas fluctuaciones, lo cual es normal debido a la complejidad de las series temporales y la volatilidad inherente de los precios del Bitcoin. En general, la RNN es efectiva para capturar patrones a largo plazo, pero puede no ser tan precisa en las fluctuaciones a corto plazo, un desafío común en los modelos predictivos aplicados a series temporales financieras.

Una vez implementados los tres modelos predictivos, a continuación, resumimos en la siguiente tabla las métricas de cada uno de los 3 modelos implementados:

Tabla 9

Comparación de Métricas de Rendimiento

Modelo implementado	Error cuadrático medio (MSE)	Error absoluto medio (MAE)	Raíz del error medio cuadrático (RMSE)	Coficiente de determinación (R^2)
Regresión lineal múltiple	159817.00006	270.86	399.7711	0.9992835
Series de tiempo ARIMA	200223842.16	13135.31	14150.05	No aplica
Redes neuronales recurrentes	7329095.5199	2210.26100	2707.2302	No aplica

Nota. Aquí se muestra la comparación de métricas de rendimiento para los modelos de regresión

lineal múltiple, series de tiempo ARIMA y redes neuronales recurrentes.

Conclusiones

En este trabajo se desarrolló un estudio sobre la evolución del precio del Bitcoin, abarcando desde la recolección y limpieza de los datos hasta la implementación y evaluación de modelos predictivos avanzados. Inicialmente, se realizó un análisis exploratorio de datos para comprender las tendencias históricas, patrones y características clave del precio del Bitcoin, lo que permitió identificar su alta volatilidad y comportamientos cíclicos. Posteriormente, se implementaron tres enfoques distintos para la predicción del precio: regresión lineal, el modelo ARIMA y redes neuronales recurrentes (RNN). Cada modelo fue evaluado utilizando métricas como el error medio cuadrático (MSE) y el coeficiente de correlación, lo que permitió determinar su desempeño. Los resultados mostraron que mientras la regresión lineal ofreció una buena aproximación global, el modelo ARIMA destacó en su capacidad para capturar patrones temporales estables, y las RNN sobresalieron en la precisión de las predicciones al seguir de cerca las tendencias del mercado.

Por otra parte, este trabajo no solo comparó la eficacia de diferentes enfoques para modelar un activo financiero altamente volátil, sino que también destacó la importancia de la limpieza de datos, el análisis exploratorio y la selección adecuada de modelos en el desarrollo de sistemas de predicción robustos y confiables. Con respecto al análisis exploratorio de los datos del precio del Bitcoin se identificaron patrones clave que muestran su alta volatilidad y comportamiento en el mercado. Después de eliminar registros nulos y columnas redundantes, se categorizaron los precios de cierre en quintiles, observando que la mayoría pertenecían a la categoría "Muy Bajo". Se calculó el rango diario, evidenciando una significativa variabilidad de precios y dispersión en el volumen de transacciones. También, las correlaciones mostraron fuertes relaciones lineales entre precios diarios y una relación positiva entre el volumen y el

rango diario, indicando que altos volúmenes están asociados con mayor variabilidad. Por otro lado, algunos gráficos tales como el de violín y los histogramas confirmaron la asimetría en la distribución de precios y volúmenes, y sugirieron que, aunque existe cierta relación entre volumen y precios, factores externos como noticias y eventos económicos también son determinantes. Este análisis brinda información valiosa para entender los días más volátiles del mercado, útil para inversionistas y analistas.

En referencia a los resultados del modelo de regresión lineal múltiple, estos muestran que la relación entre el precio de cierre del Bitcoin y las variables de entrada (precio de apertura, precio máximo, precio mínimo y volumen de transacciones) está representada por una ecuación específica. Aunque el modelo presenta un coeficiente de determinación (R^2) de 0.99928, lo que indica una relación lineal fuerte entre las predicciones y los valores reales, el error cuadrático medio (MSE) de 159817.00008 evidencia una desviación significativa entre los valores predichos y los reales. Este contraste refleja que el modelo es adecuado para captar tendencias generales, pero presenta limitaciones para predecir variaciones específicas en un mercado volátil.

Por otro lado, el modelo ARIMA presenta un error absoluto medio (MAE) aceptable de 13135.31, lo cual es razonable dada la magnitud de los precios de Bitcoin. Sin embargo, el valor del RMSE de 14150.05 sugiere que existen errores significativos en las predicciones, lo que puede afectar negativamente el rendimiento del modelo. Este comportamiento refleja las dificultades de predecir movimientos bruscos e inesperados en los precios de las criptomonedas, lo que limita la efectividad de modelos estadísticos tradicionales como ARIMA en un mercado tan volátil.

Con respecto a la implementación del modelo de Redes Neuronales Recurrentes (RNN), se pudo mejorar la capacidad predictiva en escenarios dinámicos como el mercado de

criptomonedas. Las RNN representan una alternativa avanzada y prometedora para analizar la evolución del precio del Bitcoin y generar predicciones más precisas en comparación con los métodos estadísticos tradicionales.

De acuerdo con las observaciones anteriores, el mejor modelo para predecir el precio del Bitcoin fue el de regresión lineal múltiple, debido a los valores de las métricas que fueron significativamente bajos en comparación con los otros dos modelos (ver Tabla 5). A pesar de un error cuadrático medio (MSE) considerable, este modelo logró capturar con precisión la tendencia general de los datos, siendo efectivo para análisis iniciales y predicciones generales. En comparación, los modelos ARIMA y RNN presentaron mayores dificultades para manejar la alta volatilidad y los patrones complejos de la serie temporal, lo que disminuyó su desempeño predictivo.

A partir de los hallazgos obtenidos en este trabajo, es posible identificar diversas áreas para trabajos futuros, así como las limitaciones inherentes a los modelos implementados. Una recomendación clave para investigaciones futuras es explorar técnicas más avanzadas, como modelos híbridos que combinen la regresión lineal con redes neuronales profundas o enfoques basados en aprendizaje automático más sofisticados, como las redes neuronales convolucionales (CNN), los cuales podrían mejorar la capacidad de predicción en mercados tan volátiles como el de Bitcoin (Jiang, 2020). Del mismo modo, sería beneficioso incorporar variables externas, como noticias económicas o eventos de impacto global, para enriquecer el análisis y mejorar la precisión de las predicciones, dado que estos factores tienen una influencia significativa en los precios de las criptomonedas. Con respecto a las limitaciones, es importante señalar que los modelos ARIMA y RNN presentaron dificultades para manejar la alta volatilidad y las fluctuaciones inesperadas del mercado, lo que limita su efectividad en escenarios con cambios

abruptos. Por lo tanto, se sugiere continuar explorando modelos que integren la capacidad de adaptarse a cambios repentinos y que utilicen métodos de optimización más robustos.

Finalmente, la implementación de un sistema en tiempo real para monitorear y predecir los precios podría ser una herramienta valiosa para traders y analistas, contribuyendo a la toma de decisiones informadas en un entorno financiero tan dinámico.

Recomendaciones

Para estudiar el comportamiento del Bitcoin, es crucial comprender su alta volatilidad y sensibilidad a factores externos como noticias económicas y regulaciones, realizando un análisis exhaustivo de datos históricos y patrones del mercado. Se sugiere evaluar múltiples modelos predictivos, incluyendo enfoques estadísticos, aprendizaje automático y redes neuronales, involucrando variables externas como indicadores económicos y análisis de sentimientos para mejorar la precisión. Es fundamental considerar un enfoque multidisciplinar que combine conocimientos en finanzas, estadística y programación, además de utilizar herramientas de monitoreo en tiempo real para identificar cambios abruptos en los precios. Finalmente, se recomienda mantenerse actualizado en tendencias del mercado y tecnologías emergentes, y evaluar cuidadosamente los riesgos antes de tomar decisiones de inversión o realizar análisis predictivos.

Referencias Bibliográficas

- Álvarez-Díaz, L. J. (2019). Cryptocurrencies: Evolution, growth and perspectives of Bitcoin. *Población y Desarrollo*, 25(49), 130–142.
- Andi, H. K. (2021). An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model. *Journal of Soft Computing Paradigm*, 3(3), 205–217.
- Benjamin, J. J., Surendran, R., & Sampath, T. (2022). A professional strategy for Bitcoin and Ethereum using Machine Learning for Investors. *4th International Conference on Inventive Research in Computing Applications, ICIRCA 2022 - Proceedings, Icirca*, 798–804.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5^a ed.). Wiley
- Carrión R. (2020). *Predicción de precios de criptomonedas con random forest* [Trabajo de fin de grado, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid]. Repositorio Institucional de la UPM.
- Chávez Quisbert, N. (1997). Modelos ARIMA. In *Revista Ciencia y Cultura* (Issue 1). Universidad Católica Boliviana San Pablo.
- Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, 365, 112395.
- Díaz López, M. H., King-Domínguez, A., Améstica-Rivas, L., Díaz López, M. H., King-Domínguez, A., & Améstica-Rivas, L. (2023). Estimación de Precios de Bitcoin mediante Regresión Lineal Múltiple y Redes Neuronales. *PODIUM*, 44, 119–132.

- Duarte, K. (2016). Diseño de investigación para la comparación de algoritmos de machine learning aplicados a la predicción del valor del precio de criptomonedas, a través de pruebas estadísticas de contraste y post hoc, para seleccionar aquellos con el mejor desempeño. *Universidad de San Carlos de Guatemala*, 4, 1–75.
- Dutta, A., Kumar, S., & Basu, M. (2020). A Gated Recurrent Unit Approach to Bitcoin Price Prediction. *Journal of Risk and Financial Management*, 13(2).
- Frank, K. (2017). *Hands-On Data Science and Python Machine Learning*.
- Gómez, S. (2016). *Análisis de burbujas en el precio del Bitcoin* [Tesis de pregrado, Universidad de los Andes]. Repositorio Institucional de la Universidad de los Andes.
- González, P. (2007). *Análisis de series temporales económicas: modelos ARIMA*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Jay, P., Kalariya, V., Parmar, P., Tanwar, S., Kumar, N., & Alazab, M. (2020). Stochastic neural networks for cryptocurrency price prediction. *IEEE Access*, 8, 82804–82818.
- Jiang, X. (2020). Bitcoin Price Prediction Based on Deep Learning Methods. *Journal of Mathematical Finance*, 10(01), 132–139.
- Kim, J., & Kim, H. (2023). Predicting cryptocurrency prices using LSTM networks. *Journal of Financial Data Science*, 5(2), 45-62.
- Kim, J., Wimmer, H., Liu, H., & Kim, S. (2021). A Streaming Data Collection and Analysis for Cryptocurrency Price Prediction using LSTM. *Proceedings - 2021 IEEE/ACIS 6th International Conference on Big Data, Cloud Computing, and Data Science, BCD 2021*, 45–52.

- Krishna, B. M., Sapthami, I., Banka, V. R., & Ravela, C. (2023). Cryptocurrency Price Prediction using Long Short Term Memory Method in Machine Learning. *Proceedings of the 2023 2nd International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2023, Icaiss*, 468–471.
- Lemus-Delgado, D., & Navarro, R. P. (2020). Data science and global studies: Contributions and methodological challenges. *Colombia Internacional*, 102, 41–62.
- Liu, C., & Chen, Y. (2022). Time series forecasting for Bitcoin price prediction: A hybrid approach. *IEEE Transactions on Computational Intelligence*, 18(4), 230-245.
- Matich, D. (2001). Redes Neuronales: Conceptos Básicos y Aplicaciones. *Historia*, 55.
- Mauricio, J. A. (2022). *Análisis de Series Temporales*.
- Nuzulia, A. (1967). Enhancing accuracy of Bitcoin cost prediction using ridge linear classification over lasso regression Gajula. *Angewandte Chemie International Edition*, 6(11), 951–952.
- Pavankumar, G., & Velmurugan, J. (2023). Enhancing Accuracy of Bitcoin Cost Prediction using Ridge Linear Classification over Lasso Regression. *2023 Intelligent Computing and Control for Engineering and Business Systems, ICCEBS 2023*, 1–4.
- Rahman, M. S., Hasan, M. A., & Ahmed, F. (2021). Comparative analysis of ARIMA and machine learning models for Bitcoin price prediction. *Applied Economics Letters*, 28(5), 678-685.
- Ramos Hoogwout, D. (2022). *Predicción del precio del bitcoin mediante redes neuronales recurrentes*.
- Regal, A., Morzán, J., Fabbri, C., Herrera, G., Yaulli, G., Palomino, A., Gil, C., Regal, A., Morzán, J., Fabbri, C., Herrera, G., Yaulli, G., Palomino, A., & Gil, C. (2019).

- Proyección del precio de criptomonedas basado en Tweets empleando LSTM. *Ingeniare. Revista Chilena de Ingeniería*, 27(4), 696–706.
- Şaşmaz, E., & Tek, F. B. (2021). Tweet Sentiment Analysis for Cryptocurrencies. *Proceedings - 6th International Conference on Computer Science and Engineering, UBMK 2021*, 613–618.
- Sharma, R., Gupta, S., & Verma, P. (2022). Evaluating linear regression models for cryptocurrency price estimation. *International Journal of Data Science and Analytics*, 10(3), 215-228.
- Shridevi, S., Aiyyer, H. K., & Manthripragada, A. (2021). Cryptocurrency and Associated Bourse Analysis using Machine Learning and Knowledge Graphs. *3rd IEEE International Virtual Conference on Innovations in Power and Advanced Computing Technologies, i-PACT 2021*, 1–8.
- Yang, S. Y., & Kim, J. (2015). Bitcoin market return and volatility forecasting using transaction network flow properties. *Proceedings - 2015 IEEE Symposium Series on Computational Intelligence, SSCI 2015*, 1778–1785.
- Zhang, L., Wang, Y., & Zhao, X. (2023). Machine learning-based forecasting models for Bitcoin: A systematic review. *Journal of Cryptoeconomics*, 7(1), 1-19.

Apéndices

Apéndice A

Código de Python Utilizado en la Implementación de Modelos Predictivos

https://drive.google.com/file/d/1TLB4ILmu8muOAO4K_OIK9qVFpDn3BzF6/view?usp=sharing

g

Apéndice B

Código de Python Utilizado en el Análisis Exploratorio de los Datos

[https://colab.research.google.com/drive/16OR2SyEXz_m6wQpIpwnI7-](https://colab.research.google.com/drive/16OR2SyEXz_m6wQpIpwnI7-7a0RZ5kaMC?usp=sharing)

[7a0RZ5kaMC?usp=sharing](https://colab.research.google.com/drive/16OR2SyEXz_m6wQpIpwnI7-7a0RZ5kaMC?usp=sharing)