



PROYECTO DE GRADO

Presentado ante la ilustre UNIVERDIDAD DE LOS ANDES como requisito final para
obtener el Título de INGENIERO DE SISTEMAS

DISEÑO DE ESTRATEGIAS DE TRADING CON CRIPTOMONEDAS HACIENDO USO DE SERIES TEMPORALES Y MINERIA DE DATOS

www.bdigital.ula.ve
POR

Br. ANGEL JESUS VARELA RODRIGUEZ

Tutor: Prof. Francisco L. Barillas Q.

Febrero 2021

@2021 Universidad de los Andes Mérida, Venezuela

C.C. Reconocimiento

DISEÑO DE ESTRATEGIAS DE TRADING CON CRIPTOMONEDAS HACIENDO USO DE SERIES TEMPORALES Y DE MINERIA DE DATOS

Br. Varela Rodríguez Ángel Jesús

Proyecto de Grado – Investigación de Operaciones, páginas

Escuela de Ingeniería de Sistemas, Universidad de los Andes, 2021

Resumen: En la actualidad las monedas virtuales se han convertido en un fenómeno de la economía mundial, si bien no es entendido todavía por muchas personas, entidades bancarias, gobiernos y algunas compañías son conscientes de su importancia. Por tanto, surge una necesidad de conocer si las monedas virtuales son seguras.

La principal razón por la que ha evolucionado el mundo en las monedas virtuales es porque se ubican en un sistema dispersado, es decir, que no está regulado por ningún país ni gobierno vigente, esa es la principal diferencia que existe entre las criptomonedas y el dinero tradicional, pues este último está a cargo de un banco.

Con la crisis mundial que comenzó hace una década, las monedas virtuales han mostrado una evolución de progreso constante y siguen experimentándolo, de hecho, ya encontraron un lugar en la sociedad de hoy en día, por tanto, ya están muy integradas porque muchas personas la están utilizando como método de pago de diferentes productos o servicios, no solo se pueden utilizar como forma de pago, sino que también se puede invertir en ellas, para aumentar el capital propio, así como para mantenerlo. Son una buena opción a largo plazo en caso de que la economía mundial vuelva a empeorar, y así se podrían evitar futuras dificultades que pueden afectar la propia economía.

Las criptomonedas están presentes en el mundo hace una década, en principio la sociedad no confiaba en su uso, esa perspectiva ha ido cambiando un poco con los años, todo el mundo tenía como costumbre utilizar el dinero tradicional y físico para pagar, se produjo un verdadero rechazo hacia estas monedas virtuales.

Las criptomonedas son el futuro ya que mejora el mundo bancarizado y también está transformando la vida de las personas desbancarizadas que afecta a un tercio de la población mundial dotándoles de una herramienta financiera por lo que sería recomendable actualizarse, comprándolas e invirtiendo en ellas.

Los factores económicos que inciden en su comercialización y cuál será su precio a lo largo del tiempo llevó a este trabajo de grado a obtener modelos matemáticos estadísticos que predicen el precio de las monedas virtuales, haciendo uso de técnicas como la minería de datos y series temporales (modelos de Arima) los cuales se usan para proponer o diseñar las estrategias de trading.

Palabras claves: Monedas virtuales, predicción, minería de datos, series temporales, análisis de multivariantes.

www.bdigital.ula.ve

Índice General

Introducción.....	9
Capítulo1: Identificación del proyecto.....	10
1.1.Antecedentes	10
1.2.Planteamiento del problema	13
1.3.Objetivos.....	14
1.4.Justificación.....	14
1.5.Alcance	15
Capítulo 2: Marco teórico.....	16
2.1. Criptomonedas.....	16
2.1.1. ¿Qué son las criptomonedas?.....	17
2.1.2. ¿Cómo funcionan las criptomonedas?.....	17
2.1.3. Características de una criptomoneda.....	18
2.1.4. Ventajas y desventajas de las criptomonedas.....	18
2.2. Criptomonedas en estudio.....	19
2.2.1. Bitcoin BTC.....	19
2.2.2. Ethereum ETH.....	20
2.2.3. Dash DSH.....	21
2.3. Blockchain.....	22
2.3.1. ¿Qué es el Blockchain?.....	22
2.3.2. ¿Cómo funciona el Blockchain?.....	22
2.3.3. Características principales del Blockchain.....	23
2.3.4. Beneficios y casos de uso del Blockchain.....	25
2.4. Series Temporales.....	26
2.4.1. Series de tiempo.....	26
2.4.2. Modelos estadísticos para series de tiempo.....	27
2.4.3. Medidas de dependencia: Autocorrelacion y Correlación cruzada.....	28
2.4.4. Series de tiempo estacionarias.....	29

2.4.5. Estimadores de correlación.....	31
2.4.6. Modelos ARMA.....	32
2.4.7. Modelos ARIMA.....	34
2.4.8. Funciones de Autocorrelacion (ACF) y Autocorrelacion (PACF).....	37
2.4.9. Pronósticos.....	39
2.5. Machine Learning (Aprendizaje Automático).....	40
2.5.1. Conceptos Básicos.....	41
2.5.2. Tipos de aprendizaje automático.....	42
2.5.2.1. Métodos de aprendizaje automático.....	43
2.5.3. Reglas de asociación.....	44
2.5.4. Algoritmo a priori.....	46
Capítulo 3: Aspectos Metodológicos.....	48
3.1. Estadística descriptiva.....	48
3.2. Metodología a utilizar para series de temporales.....	49
Capítulo 4: Análisis y resultados.....	56
4.1 Predicciones para el Bitcoin (BTC).....	58
4.2 Predicciones para el Dash (DHS).....	67
4.3 Predicciones para el Ethereum (ETH).....	77
4.4 Análisis de las reglas de asociación.....	89
4.5. Diseño de estrategias de trading.....	94
4.5.1. Primera simulación de la estrategia de trading backtesting.....	97
4.5.2. Segunda simulación de la estrategia de trading backtesting.....	99
4.5.3. Tercera simulación de la estrategia de trading backtesting.....	101
4.5.4. Observación de las simulaciones.....	103
Capítulo 5: Conclusiones y Recomendaciones.....	104
Bibliografía.....	108

Índice de tablas

Tabla 3.1. Casos para la definición del modelo adecuado.

Índice de figuras

Figura 4.1. Cuadro estadísticos descriptivos del BTC

Figura 4.2. Comportamiento del BTC

Figura 4.3. Correlograma simple del BTC

Figura 4.4. Correlograma parcial del BTC

Figura 4.5. Prueba de Dickey-Fuller del BTC

Figura 4.6. Correlograma simple serie diferenciada del BTC

Figura 4.7. Correlograma parcial serie diferenciada del BTC

Figura 4.8. Prueba de Dickey-Fuller serie diferenciada del BTC

Figura 4.9. Serie BTC diferenciada

Figura 4.10. Coeficientes propuestos modelo ARIMA del BTC

Figura 4.11. Diagramas ACF y PACF del BTC

Figura 4.12. Tabla de predicciones del BTC (25 de enero de 2020 – 23 de febrero 2020)

Figura 4.13. Gráfico pronóstico precios del Bitcoin

Figura 4.14. Cuadro estadísticos descriptivos del DSH

Figura 4.15. Comportamiento del DSH

Figura 4.16. Correlograma simple del DSH

Figura 4.17. Correlograma parcial del DSH

Figura 4.18. Prueba de Dickey-Fuller del DSH

Figura 4.19. Correlograma simple serie diferenciada del DSH

Figura 4.20. Correlograma parcial serie diferenciada del DSH

Figura 4.21. Prueba de Dickey-Fuller serie diferenciada del DSH

Figura 4.22. Serie DSH diferenciada

Figura 4.23. Coeficientes propuestos modelo ARIMA del DSH

Figura 4.24. Diagramas ACF y PACF del DSH

Figura 4.25. Tabla de predicciones del DSH (25 de enero de 2020 – 23 de febrero 2020)

Figura 4.26. Gráfico pronóstico precios del Dash

Figura 4.27. Cuadro estadísticos descriptivos del ETH

Figura 4.28. Comportamiento del ETH

Figura 4.29. Correlograma simple del ETH

Figura 4.30. Correlograma parcial del ETH

Figura 4.31. Prueba de Dickey-Fuller del ETH

Figura 4.32. Correlograma simple serie diferenciada del ETH

Figura 4.33. Correlograma parcial serie diferenciada del ETH

Figura 4.34. Segunda prueba Dickey-Fuller del ETH

Figura 4.35. Serie ETH diferenciada

Figura 4.36. Segundo correlograma simple del ETH

Figura 4.37. Segundo correlograma parcial del ETH

Figura 4.38. Gráfico de la segunda diferenciación del ETH

Figura 4.39. Coeficientes propuestos modelo ARIMA del ETH

Figura 4.40. Diagramas ACF y PACF del ETH

Figura 4.41. Tabla de predicciones del ETH (25 de enero de 2020 – 23 de febrero 2020)

Figura 4.42. Gráfico pronostico precios del Ethereum

Figura 4.43. Rendimiento de las diferentes criptomonedas

Figura 4.44. Matriz de transacciones

Figura 4.45. Reglas de asociación arrojadas por A priori

Figura 4.46. Gráfico de las reglas de asociación arrojadas A priori

Figura 4.47. Coeficientes de variaciones de las criptomonedas

Figura 4.48. Plataforma MetaTrader 4

Figura 4.49. Software Forex Simulator

Figura 4.50. Valores de predicción para la primera simulación

Figura 4.51. Resultados de la primera simulación

Figura 4.52. Valores usados para la primera simulación

Figura 4.53. Valores de predicción para la segunda simulación

Figura 4.54. Resultados de la segunda simulación

Figura 4.55. Valores usados para la segunda simulación

Figura 4.56. Valores de predicción para la tercera simulación

Figura 4.57. Resultados de la tercera simulación

Figura 4.58. Valores usados para la tercera simulación

www.bdigital.ula.ve

Introducción

Las monedas virtuales lentamente han logrado tener un lugar en el panorama de las transferencias económicas mundiales, se han convertido en medios digitales de intercambio, además tienen costos casi imperceptibles de transacción, depende de los movimientos internos de la plataforma de oferta y demanda, entonces visto de esta forma podrían cambiar la forma actual de hacer negocios en el mundo, la más conocida de ellas es el Bitcoin, pero existen otras como el Ethereum y Dash, en el cual son con las que se trabajara en el presente trabajo.

Las criptomonedas están todavía en desarrollo, sin embargo, son muchas las empresas e inversores de todo el mundo que ya están utilizándolas, se trata de una innovación que permite a sus usuarios una mayor velocidad en sus pagos, la ventaja es que significa la libertad regulatoria que gozan respecto a gobiernos o entidades puede a su vez resultar una desventaja dado que las mismas sufren de una elevada volatilidad y no existen autoridades que se responsabilicen ante eventuales fraudes.

Esta investigación consiste en experimentar para poder predecir el precio de las monedas virtuales, así como también encontrar factores o variables económicas que incidan en su comercialización, haciendo uso de datos públicos que se encuentran en la web. De esta manera poder recolectar toda la información necesaria para así aplicar técnicas estadísticas, y a su vez predecir el precio y encontrar distintas estrategias de inversión entre ellas, con la finalidad de mostrar si las monedas virtuales pueden reemplazarse o ser usadas como las monedas reales.

Capítulo 1

Identificación del proyecto

1.1 Antecedentes

En base a la búsqueda de información en entorno al tema se han encontrado diferentes investigaciones las cuales se listan a continuación:

María Andrea Cajiao Hoyos y Daniela Fonseca Medellín (2016), realizaron un trabajo de investigación donde analizan si la moneda virtual Bitcoin se puede relacionar al entorno económico y social en Colombia. Para esto realizaron un análisis de variables económicas y financieras de Colombia, junto con la opinión de algunos expertos en finanzas y en el mercado internacional.

Castillo Fonseca Diego y Garcia Marcano Roberto (2017), realizaron un trabajo donde proponen un sistema que a través de las criptomonedas cualquier residente en el exterior pueda mandar remesas instantáneamente a sus familiares en Venezuela, al mismo tiempo que los ciudadanos del país puedan tener la oportunidad de resguardar el valor de su dinero en criptomonedas. Además, los autores propusieron un modelo que estudiaron a través de un estudio de factibilidad.

Bibiana Moreno R., Francely Soto, Nancy Valencia y Adelina Sánchez (2018), estudiaron las Criptomonedas como alternativa de inversión en Colombia, analizaron la operatividad, riesgos, su estado actual de regulación teniendo en cuenta la posición del Banco de la República y la Superintendencia Financiera, y su posibilidad de usarse o hacerse efectiva en Colombia. Analizaron el comportamiento y avances de estas monedas, plataformas de operatividad, postura de entidades internacionales como NYSE (New York Stock Exchange, en español, Bolsa de Nueva York), CFTC (Commodity Futures Trading Commission, en español, Comisión de Negociación de Futuros de Productos Básicos de Estados Unidos) frente a licencias y regulación de brókeres, estadísticas y análisis de varianza de volatilidad, y otros análisis a partir de

investigaciones que al respecto se han realizado, a fin de fundamentar al lector frente a las Criptomonedas como modelo de inversión.

Ignacio Martín Gallardo Urbini (2018) aplicó arquitectura de certificados digitales o infraestructura de clave pública lo que allá en el 2009 logró exitosamente Satoshi Nakamoto (creador del Bitcoin), un rediseño, reingeniería y cambio de paradigma en el comercio electrónico, de una arquitectura centralizada y jerárquica a una completamente descentralizada por medio de su innovación tecnológica llamada blockchain o cadena de bloques.

López Álvarez René David (2018) analizó la factibilidad en la implementación del método de criptomonedas en el sistema financiero en el mundo; con sus fundamentos teóricos en Javier, (2011), Ozcáriz, (2014) y Cullerés, (2014), entre otros. El tipo de investigación fue documental, descriptivo no experimental. La técnica documental empleada se centra en la observación y análisis de diversas fuentes documentales, análisis e interpretación de documentales bibliográficas, información de los razonamientos hermenéuticos. Obtuvo como resultado, que existe una diversidad de opiniones sobre la influencia de las criptomonedas o Bitcoin en el sistema financiero, lo cierto es que la tecnología ha dado un gran empuje a la economía mundial a través del método de moneda virtual.

Gerard Maynés Moreno (2018) realizó un trabajo sobre la tecnología blockchain, sus conceptos relacionados y un análisis de distintas criptomonedas. También realizó un estado del arte de la situación económico-financiera de las distintas monedas, así como una justificación del análisis de estas. Continuó con una serie de conceptos básicos para entender cómo funciona la tecnología blockchain. Hizo un análisis breve de las tres criptomonedas que decidieron estudiar (Bitcoin, Dash y Litecoin), como una serie de características únicas en cada una de ellas. Usaron un pequeño estudio del minado de cada moneda, así como de su eficiencia y del impacto medioambiental que generan.

Max Puvoguel Rojas (2018) intentó realizar una aproximación jurídica a las tecnologías blockchain y monedas virtuales. Específicamente, el autor buscó demostrar que dichas tecnologías no encajan con la taxonomía jurídica actual, encontrándose en un

vacío legal. Para desarrollar dicha hipótesis, el autor agrupó dichas tecnologías en una misma entidad, realizando una introducción técnica y económica de la materia para luego abordar el análisis jurídico desde diversas materias.

Miguel Andrés Porro Vargas (2018) desarrolló un modelo de análisis de criptomonedas basado en aprendizaje automático y series temporales para realizar pronósticos de precios de las mismas.

Gilary Dayan Perez Silva (2018) analizó la dinámica que ha tenido la Criptomoneda Bitcoin en el mundo para uso comercial, y la proyección en las transacciones globales futuras, teniendo en cuenta su evolución entre 2010 y 2018. Esto lo hizo teniendo en cuenta factores como la confianza por parte de los usuarios en las criptomonedas, la forma en realizar los pagos con Bitcoin, las principales preocupaciones económicas que acarrea el uso de las monedas virtuales para transacciones globales, la seguridad que ofrecen las diferentes plataformas, como se encuentra el panorama legal para las criptomonedas. También analizó los casos de Colombia y Venezuela respecto al uso de las criptomonedas para el cual se enfatizó en la información encontrada sobre las estadísticas del uso de criptomonedas en Venezuela y Colombia, crecimiento de las criptomonedas y el nacimiento del Petro en Venezuela, ya que estos dan un panorama de la evolución de estas, especialmente de Bitcoin para el futuro.

1.2 Planteamiento del problema

En la actualidad, la economía mundial ha ido evolucionando para adaptarse al nuevo panorama, ya que deben acoplarse al proceso tan abrumador como es la globalización. Los antiguos modelos son una de las causas que la actividad económica también debe superar, es en la actualidad cuando el uso de las criptomonedas se hace más visible, dado a la aceptación que estas han tenido a nivel mundial por su forma fácil y en algunos casos rápida forma de generar ganancias, una de las ventajas más importantes que tienen este tipo de moneda es que brinda a las personas una forma segura de realizar sus transacciones sin tener que depender de un banco u otro ente que les supervise. Desde la creación de las criptomonedas las personas que están pensando en invertir o han invertido en ellas, no tienen certeza de lo que pueda pasar con su dinero en el futuro, normalmente solo tienen las palabras de algunos conocidos que garantizan que ganarán dinero a futuro sin poder asegurarles el mejor momento para invertir en éstas, hasta este momento hay un notable desconocimiento sobre los modelos que existen para predecir la variación de la criptomonedas.

La oportunidad que nos brinda hacer trading con criptomonedas es provechosa, pues se puede realizar las 24 horas del día y todos los días de la semana, el problema es que son inversiones con riesgos altos, sino se realizan las estrategias adecuadas pueden incurrir en pérdida de incluso todo su capital.

En los modelos estadísticos matemáticos contruidos en algunos trabajos no se realizan como tal estrategias trading, solo se limitan a pronosticar valores futuros de la moneda, pero no se genera como tal una estrategia de cuando es recomendable comprar, vender, o quedarse con las criptomonedas.

1.3 Objetivos

Objetivo general

Desarrollar modelos basados en series de tiempo, que permitan estudiar y eventualmente predecir el comportamiento de los valores del precio de un conjunto a seleccionar de criptomonedas con la finalidad de diseñar estrategias de trading.

Objetivos específicos

1. Recolectar las variables asociadas a las criptomonedas y entender el fundamento teórico de las mismas.
2. Recoger información asociada a las series de tiempo y modelos ARIMA.
3. Ajustar modelos de series temporales a los precios de las monedas.
4. Discutir y analizar las estrategias de trading ajustadas a los modelos predictivos obtenidos.

www.bdigital.ula.ve

1.4 Justificación

La idea de este proyecto nace al observar el notable incremento en el uso de las criptomonedas, no solo en Venezuela sino a nivel mundial. Un aspecto muy importante es que a pesar del crecimiento tanto en popularidad como en cantidad de este tipo de moneda, y que un número cada vez mayor de personas están invirtiendo o por lo menos están pensando invertir en éstas, no tienen el conocimiento básico sobre que es este tipo de moneda y su funcionamiento. No es un secreto que actualmente hay un alto grado de desconocimiento por parte de la población acerca de estas monedas virtuales y esto conlleva a que se hagan inversiones en momentos en los cuales no es recomendable hacerlo, esta inversión se hace con argumentos pocos convincentes como recomendaciones de personas que se dejan llevar por sentimientos y no realizan estudios previos para lograr mejores resultados.

En el siguiente estudio se estarán desarrollando modelos basados en técnicas de series temporales y de minería de datos con el fin de predecir el valor en el precio que están tomando las criptomonedas y así brindar una orientación para conocer el momento idóneo para invertir. Con estos modelos se busca disminuir el nivel de incertidumbre a la hora de hacer una inversión, además, brindar una herramienta de mayor confianza y no depender solo de recomendaciones sin estudio previo.

1.5 Alcances

Las criptomonedas son un nuevo tipo de monedas que han alcanzado un alto nivel de popularidad a nivel mundial. El estudio que se presenta está enfocado a cualquier persona que tenga el deseo de invertir en las criptomonedas y que no sepa cuál es el momento indicado para hacerlo, de forma general este proyecto tendrá como alcance al público latino americano, esto no quiere decir, que las estrategias de trading aplicándole algunos ajustes de los valores de las monedas locales de cada país puedan ser usadas.

Capítulo 2

Marco teórico

El marco teórico, ofrece a lector una serie de conceptos que permiten abordar el problema dentro de un ámbito donde este cobre sentido para lograr el objetivo deseado.

2.1. Criptomonedas

A través de la historia, ha habido múltiples medios de intercambio de bienes y servicios. Todos ellos han ido sufriendo evoluciones y cambios o, directamente, han desaparecido. Es el caso del trueque o la utilización de materiales preciosos, los cuales dejaron paso al sistema que hasta la actualidad ha funcionado, los billetes y monedas.

No obstante, en una época donde la tecnología cada día gana más y más terreno, se hace necesario un medio de intercambio más ajustado a los tiempos: la criptomoneda. A raíz del movimiento Cypherpunk en los años 80, surgió la Criptomoneda. Este arte defendía el uso extendido de escribir con claves secretas que tan sólo podría entender quién supiese descifrarlas.

Una década después, David Chaun crea Digicash para proporcionar un sistema centralizado de dinero electrónico que permitía transacciones más seguras y anónimas. En el mismo período, Adam Black propone Hashcash, un sistema basado en prueba de trabajo para controlar el spam y los ataques de denegación de servicio.

Aunque no sería hasta el 2009, cuando surgiese el Bitcoin, la primera criptomoneda completamente descentralizada. [‘economipedia.com’ (2019)]

2.1.1 ¿Qué son las criptomonedas?

En un artículo sobre economía en la página web 'economipedia.com' (2019), define a la criptomoneda o criptodivisa como un tipo de moneda digital que utiliza la criptografía para proporcionar un sistema de pagos seguro. Estas técnicas de cifrado sirven para regular la generación de unidades monetarias y verificar la transferencia de fondos. No necesitan de un banco central u otra institución que las controle.

Las criptomonedas son un tipo de moneda digital, son aquellas que no existen de forma física pero que sirven como moneda de intercambio, permitiendo transacciones instantáneas a través de internet y sin importar las fronteras. Otros tipos de moneda digital son las monedas virtuales (normalmente controlada por los desarrolladores), dinero electrónico y cupones de internet.

Estas monedas digitales se crean a través de un proceso llamado “minería”, donde los mineros van añadiendo los registros de las operaciones que se realizan al libro público mayor de determinada criptomoneda. Todas las transacciones ocurren de inmediato en cada una de las plataformas y son agregadas al libro público, donde pueden ser visualizadas por cualquier persona de la red. ['economipedia.com' (2019)]

2.1.2. ¿Cómo funcionan las criptomonedas?

Cada una de las transacciones que se realizan con las criptomonedas son enviadas entre pares desde las billeteras de las personas que se encuentran involucradas en la transacción, justo en el momento en que coinciden los códigos públicos. Estos códigos públicos se encuentran relacionados con cada una de las contraseñas privadas o claves criptográficas creadas por los usuarios y se encargan de que cada transacción llegue al monedero digital de la otra persona. Luego, las transacciones que se realizaron se registran en el libro mayor público, mejor conocido como “Blockchain” o en español: cadena de bloques. Cualquier usuario que forme parte de la misma criptomoneda podrá acceder al libro mayor, siempre y cuando haya descargado una billetera de nodo

completo en lugar de un monedero virtual. Una vez hayan accedido podrán visualizar los montos de cada una de las operaciones, más no el número de cuenta (ya que está encriptado).

El motivo por el cual se le considera una “cadena de bloques” es porque, cada vez que un usuario recibe o envía una criptomoneda, bien sea a través de contraseñas públicas o privadas, la transacción queda en lista de espera esperando ser agregada al libro mayor. [‘economipedia.com’ (2019)]

2.1.3. Características de una criptomoneda

En artículos de ‘economipedia.com’ (2019) se pueden encontrar las principales características de las criptomonedas:

- **Criptografía:** Utilizan técnicas de cifrado para realizar cobros y pagos seguros.
- **Descentralización:** No necesitan ser controladas por ninguna institución.
- **No hay posibilidad de falsificación o duplicación:** Gracias al sistema criptográfico que protege a los usuarios.
- **No hay intermediarios:** Contacto directo persona a persona.
- **Las transacciones son irreversibles:** Una vez se efectúe el pago, no hay posibilidad de cancelación.
- **Se pueden intercambiar por otras divisas.**
- **Privacidad de uso:** No es necesario revelar tu identidad al hacer negocios.

2.1.4. Ventajas y desventajas de las criptomonedas

Las ventajas principales que caracterizan a la criptomoneda son:

- **Bajos costes de transacción:** Como consecuencia de la no presencia de intermediarios.
- **Seguro:** Cada moneda pertenece solo a su dueño.
- **Transparente:** Las transacciones se incorporan en un registro de acceso libre.

- Se acumula en un espacio ínfimo como es un USB.

Por su parte, las desventajas son:

- Volatilidad de sus precios
 - Falta de aceptación actual por algunas empresas
 - Al no necesitar un regulador como el Gobierno o el Banco Central y proporcionar privacidad pueden ser utilizadas para transacciones ilegales.
- [‘economipedia.com’ (2019)]

2.2. Criptomonedas de interés en este estudio

2.2.1. Bitcoin BTC

El Bitcoin es una moneda digital o criptomoneda que utiliza criptografía para asegurar las transacciones dentro de su infraestructura, lo que representa una base de datos online distribuida, también llamada blockchain. La abreviatura de Bitcoin es BTC y con ella se aplica el mismo principio que con el USD (dólar estadounidense) y el EUR (Euro), es decir, se puede emparejar con otras monedas. Sirve para intercambiar bienes y servicios. Sin embargo, a diferencia de otras monedas, Bitcoin es una divisa electrónica que presenta novedosas características y destaca por su eficiencia, seguridad y facilidad de intercambio.

Su mayor diferencia frente al resto de monedas, se trata de una moneda descentralizada, por lo que nadie la controla. Bitcoin no tiene un emisor central como los dólares o los euros, la criptomoneda es producida por las personas y empresas de alrededor del mundo dedicando gran cantidad de recursos a la minería.

Sin lugar a dudas lo que hace distinto a Bitcoin es la descentralización. O lo que es lo mismo, Bitcoin está fuera del control de cualquier gobierno, institución o entidad financiera, ya sea de tipo estatal o privado, como pueden ser el euro, controlado por el Banco Central Europeo o el Dólar por la Reserva Federal de los EEUU.

En Bitcoin el control lo realizan, de forma indirecta mediante sus transacciones, los propios usuarios a través de los intercambios P2P (Peer to Peer o Punto a Punto). Esta estructura P2P y la falta de control imposibilita que cualquier autoridad manipule su valor o provoque inflación produciendo más cantidad.

De hecho, su producción y valor se basa en la ley de la oferta y la demanda. Otro detalle interesante es que Bitcoin tiene un límite fijado de 21 millones de monedas, que se alcanzará en 2030. [‘economipedia.com’ (2019)]

2.2.2. Ethereum ETH

Según Acosta J. (2018) Ethereum es una red Blockchain o Tecnología de Contabilidad Distribuida (DTL) con un lenguaje de programación Turing completo integrado, una computadora blockchain, que permite que cualquiera pueda escribir contratos inteligentes y aplicaciones descentralizadas simplemente escribiendo la lógica en unas pocas líneas de código.

El protocolo Ethereum fue concebido originalmente como una versión mejorada de la criptomoneda Bitcoin, para superar las limitaciones de su lenguaje de programación, proporcionando características avanzadas tales como custodia sobre la blockchain, límites de retiro, contratos financieros, mercado de juegos de azar y similares a través de un lenguaje de programación muy generalizado.

En Ethereum, una versión básica de Namecoin, la base de datos de registro de nombres descentralizada, se puede escribir en dos líneas de código, y otros protocolos como monedas y sistemas de reputación se pueden construir en menos de veinte líneas. Contratos inteligentes, «cajas» criptográficas que contienen valor y sólo se desbloquean si se cumplen ciertas condiciones, también pueden ser incorporados encima de la plataforma, con mucho más poder que la ofrecida por el script de Bitcoin debido al poder agregado del Turing completo.

El protocolo Ethereum no «soporta» cualquiera de las aplicaciones directamente, pero la existencia de un lenguaje de programación Turing completo permite que contratos arbitrarios teóricamente se puedan crear para cualquier tipo de transacción o

aplicación. Lo que es más interesante de Ethereum, sin embargo, es que el protocolo Ethereum se mueve mucho más allá de la moneda. Protocolos alrededor de almacenamiento de archivos descentralizados, computación descentralizada y mercados de predicción descentralizados, entre docenas de otros conceptos similares, tienen el potencial para aumentar sustancialmente la eficiencia de la industria computacional, y proporcionar un gran impulso a otros protocolos peer-to-peer adicionando por primera vez una capa económica. Por último, también hay una variedad considerable de aplicaciones que no tienen nada que ver con el dinero en absoluto.

Ether es la criptomoneda de Ethereum, el combustible que impulsa esta plataforma de aplicaciones distribuidas. Es una criptomoneda utilizada por los clientes de la plataforma Ethereum para realizar pagos a otras personas o a máquinas que ejecutan operaciones solicitadas. Es decir, Ether es el incentivo que asegura que los desarrolladores escriban aplicaciones de calidad (la codificación innecesaria cuesta más) y que la red permanezca saludable (la gente es recompensada por los recursos aportados).

2.2.3. Dash DSH

Conocida como Digital Cash (del inglés, efectivo digital), Dash se presenta en su libro blanco como una criptomoneda basada en Bitcoin fuertemente anónima. Se trata de una red que no solo incentiva a los mineros por validar transacciones y proteger la blockchain, como en la mayoría de las plataformas entre pares (p2p), sino que también lo hace con los nodos maestros o masternodos para su sistema de gobernanza.

Dash cuenta con funciones como PrivateSend para incrementar la privacidad de los usuarios (ocultando el origen de los fondos) y mantener así la fungibilidad de las criptomonedas al eliminar su historial.

Otra de las características de Dash es InstaSend que apunta a que los usuarios puedan pagar, en cualquier establecimiento comercial, con sus dispositivos móviles de forma instantánea sin esperar minutos u horas por las confirmaciones de rigor.

Ambos aspectos de la red se sustentan en los quórum de los nodos maestros, un consenso entre un número determinado de este tipo de nodos, de mayor jerarquía que los nodos tradicionales, y que permiten agilizar las operaciones. Su equivalente en Bitcoin serían los nodos completos. [Caurin J. (2018)]

2.3. Blockchain

2.3.1 ¿Qué es el Blockchain?

Un artículo de economía publicado por el diario ‘Gestion.pe’ (2019), el cual se titula “*¿Qué es el Blockchain y cómo funciona?*”, lo define como una tecnología que permite la transferencia de datos de una manera completamente segura gracias a una codificación muy sofisticada. Se suele comparar con un libro de contabilidad de una empresa donde están registradas todas las entradas y salidas de dinero. Claro que, en este caso, se trataría de un libro de acontecimientos digitales. Esta transferencia no necesita de un intermediario que compruebe y apruebe la información, sino que está distribuida en diversos nodos independientes (usuarios) entre sí que la registran y validan. Así, una vez que la información es introducida no podrá ser borrada, solo se podrán añadir nuevos registros. Además, no será legitimada a menos que la mayoría de ellos se pongan de acuerdo para hacerlo. Básicamente, el Blockchain, elimina a los intermediarios, descentralizando toda la gestión. El control del proceso es de los usuarios y son ellos los que se convierten parte de un enorme banco con miles, millones de nodos, cada uno de los cuales se convierte en partícipe y gestor de los libros de cuenta del banco.

2.3.2. ¿Cómo funciona el Blockchain?

Como lo describe el artículo citado anteriormente por ‘Gestion.pe’ (2019), hasta hace algún tiempo era necesario un tercero en el que las dos partes confiaran para garantizar la autenticidad de las transacciones (banco, auditor, notario o Paypal) y que tuviese un registro o sello de veracidad. Ahora, con el blockchain, este problema y

doble gasto desaparece al combinar la tecnología P2P de intercambio entre pares con la moneda virtual y así crear una nueva forma de comunicación digital. Asimismo, las transacciones incluidas en los bloques son creadas por los integrantes del sistema. Estas son registradas y transmitidas a todos los nodos de la red. Así, todos los integrantes tienen la información de las transacciones actualizadas en tiempo real.

Las transacciones se realizan desde los llamados “wallets” o monederos electrónicos. Los wallets son archivos encriptados que funcionan de forma similar a una cuenta bancaria, estos poseen dos claves: pública y privada. La clave pública es una cadena alfanumérica entre 26 y 35 caracteres, es la dirección de Bitcoin y hace las veces de número de cuenta. De esta manera, para que alguien le envíe bitcoins, previamente debe darle la clave pública para luego recibir las monedas virtuales. En tanto, la clave privada sirve para autorizar operaciones desde su wallet. Este proceso es el que se conoce como criptografía asimétrica. [‘Gestion.pe’ (2019)]

Como de manera constante están fluyendo nuevas transacciones por la red, estas serán añadidas a un “pool” (similar a una cooperativa) de transacciones sin verificar. Aquí es donde aparece la figura de los llamados “mineros”, que son los encargados de escoger las transacciones de estos pools para crear un nuevo bloque de transacciones confirmadas.

Los mineros se dedican a verificar las transacciones que están ocurriendo en estos momentos y cumplen dos importantes funciones: Crear nuevos bitcoins por cada bloque que se mina y asegurar que las transacciones son reales y legítimas. Los grupos de minería son conjuntos de mineros que trabajan en conjunto para resolver un bloque y se dividen las recompensas otorgadas. [‘Gestion.pe’ (2019)]

Gracias al uso del blockchain que se sincroniza entre los nodos se logra la irreversibilidad de las transacciones, lo que permite que nadie haga fraudes para beneficiarse, modificando el libro de cuentas para desviar los bitcoins de un lado a otro sin que otros se enteren. Ese libro de cuentas es seguro pues los bloques enlazados

cuentan con un puntero “hash” (codificado) que enlaza al bloque anterior; además, toda la información es pública. Lo que significa que aunque protege la privacidad de sus usuarios, sí permite controlar la trazabilidad de esas transacciones. [‘Gestion.pe’ (2019)]

Cabe resaltar que cada bloque es una parte de la cadena con los siguientes elementos:

- Un código alfanumérico que enlaza con el bloque anterior.
- Un paquete de transacciones.
- Otro código alfanumérico que enlazará con el bloque posterior.

2.3.3. Características principales de Blockchain

Porro, A. (2018), menciona en su tesis de grado algunas de las características más relevantes del sistema Blockchain:

1. Es un sistema seguro dado que su tecnología se basa en la criptografía de datos.
2. Las transacciones se concentran en bloques, en dichos bloques la información se almacena cronológicamente.
3. Una vez aceptada, la información no se puede borrar ni modificar, y puede ser consultada en cualquier momento.
4. Una Blockchain puede ser pública o privada, inclusive puede permitir ciertas consultas con un permiso especial.
5. La confianza, dada por la descentralización y no disponer de un nodo central donde se aglomere toda la información, hace que las partes intervinientes confíen plenamente en esta tecnología.
6. Transacciones en tiempo real: Su inmediatez minimiza el riesgo de contra-parte producido en otro tipo de transacciones, donde el compromiso de pago tarda varios días

en ejecutarse (evitando riesgos de impago, concursos de acreedores y fraudes durante el proceso de pago).

2.3.4. Beneficios y casos de uso del Blockchain

Blockchain es conocida principalmente en el mundo como la tecnología sobre la que se basan los intercambios de Bitcoin, una de las criptomonedas a estudiar en este trabajo, pero sus usos prácticos en la actualidad van más allá de las transacciones de criptomonedas. Por ejemplo, en finanzas las redes Blockchain permiten que los intercambios de valores se liquiden en minutos en lugar de días. Los contratos inteligentes (smart contracts) son una de las aplicaciones más prometedoras, dado que son capaces de ejecutar transacciones comerciales y acuerdos automáticamente, haciendo cumplir las obligaciones de todas las partes en un contrato, sin los costos adicionales de intermediarios. Entre otras de las diversas aplicaciones interesantes están el manejo de identidad, tramitación de documentos, protección de la identidad intelectual, registros médicos, hospitales e incluso los registros civiles se podrán basar en esta tecnología. [Acosta, J. (2018)]

Acosta, J. (2018), también menciona que uno de los países más proactivos en la materia en este momento es Estonia, donde su gobierno trabaja de la mano con Bitnation, para permitirle a sus residentes con identidad digital, registrar sus uniones matrimoniales y partidas de nacimiento en la Blockchain Horizon.

Blockchain es capaz de resolver incluso el problema de la votación por internet, brindando el anonimato del voto, la garantía de que una persona vote una sola vez y lo más importante de todo, establecer la confianza, dado que, al ser una red descentralizada, ninguna autoridad central que gestione la votación podría manipularla. [Porro, A. (2018)]

2.4. Series Temporales

2.4.1. Series de Tiempo

El análisis de series de tiempo desempeña un papel importante en el análisis requerido para el pronóstico de eventos futuros. Existen varias formas o métodos de calcular cual va a ser la tendencia del comportamiento del proceso en estudio. [Acosta, J. (2018)].

- **Definición**

Una serie de tiempo es un conjunto de observaciones X_t , cada una registrada a un tiempo específico t .

El análisis clásico de las series de tiempo, se basa en la suposición de que los valores que toma la variable de observación es la consecuencia de tres componentes, cuya actuación conjunta da como resultado los valores medidos, estas componentes son:

1. **Componente tendencia:** Se puede definir como un cambio a largo plazo que se produce en la relación al nivel medio, o el cambio a largo plazo de la media. La tendencia se identifica con un movimiento suave de la serie a largo plazo.
2. **Componente estacional:** Muchas series de tiempo presentan cierta periodicidad o dicho de otro modo, variación de cierto período (semestral, mensual, etc.). Por ejemplo las ventas al delta en Puerto Rico aumentan por los meses de noviembre y diciembre por las festividades navideñas. Estos efectos son fáciles de entender y se pueden medir explícitamente o incluso se pueden eliminar de la serie de datos, a este proceso se le llama desestacionalización de la serie.
3. **Componente aleatoria:** Esta componente no responde a ningún patrón de comportamiento, sino que es el resultado de factores fortuitos o aleatorios que inciden de forma aislada en una serie de tiempo.

De las tres componentes anteriores las dos primeras son componentes determinísticos, mientras que la última es aleatoria. Así, se puede denotar la serie de tiempo como:

$$X_t = T_t + E_t + \xi_t$$

Donde T_t es la tendencia, E_t es la componente estacional y ξ_t es la componente aleatoria.

2.4.2. Modelos Estadísticos para Series de Tiempo

Como principal objetivo en el análisis de series de tiempo, se tiene el desarrollo de modelos matemáticos que arrojen una descripción apropiada de los datos muestrales. En principio utilizaremos primeramente la definición de series de tiempo.

- **Definición**

Un **proceso estocástico** es una familia de variables aleatorias indexadas $x(\omega, t)$ ó $x_t(\omega)$ donde t pertenece a un conjunto de índices T y ω pertenece a un espacio muestral Ω . Si $t = t^*$ fijo, $(\omega, *)$ es una variable aleatoria. Si $\omega = \omega^*$ fijo, (ω^*, t) es una función de t , y se llama una realización del proceso. Una **serie de tiempo** es la realización de un proceso estocástico.

A continuación se muestran algunos ejemplos:

1) *Ejemplo: Ruido Blanco*

Una manera sencilla de generar series de tiempo, puede ser considerando una sucesión de variables aleatorias no correlacionadas, ω_t con media 0 y varianza σ^2 . Las series de tiempo generadas de esta manera son usadas como modelos para ruido en aplicaciones de ingeniería, donde ellas son llamadas *ruido blanco*, denotaremos este proceso como $\omega_t \sim \omega_n(0, \sigma_\omega^2)$. La designación blanca se origina de la analogía con luz

blanca e indica que todos los posibles períodos de oscilación están presentes con igual intensidad.

2) Ejemplo: Promedio móvil

Un promedio móvil se construye sustituyendo cada valor de una serie por la media obtenida con esa observación y algunos de los valores inmediatamente anteriores y posteriores.

$$v_t = \frac{1}{3}(\omega_{t-1} + \omega_t + \omega_{t+1}), \quad (2.1)$$

2.4.3. Medidas de Dependencia: Autocorrelación y Correlación Cruzada

- **Definición**

La **función de media** está definida como:

$$\mu_{xt} = \mathbb{E}(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx, \quad (2.2)$$

En caso de que exista, $\mathbb{E}$ denota el operador usual de esperanza. Ahora escribiremos μ_{xt} como μ_t . Es relevante entender que μ_t consiste en que es una medida teórica para la serie de tiempo en un punto particular, en donde la media se calcula sobre todos los posibles eventos que podrían haber producido X_t .

- **Definición**

La **función de autocovarianza** es definida como producto del segundo momento:

$$\gamma_x(s, t) = \mathbb{E}[(x_s - \mu_s)(x_t - \mu_t)], \quad (2.3)$$

Para todo t y s . Cuando no haya confusión en la existencia sobre a qué serie nos referimos, escribiremos $\gamma_x(s, t) = \gamma(s, t)$.

- **Definición**

La función de **autocorrelación (ACF)** se define como:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}. \quad (2.4)$$

La ACF mide la predictibilidad lineal de una serie de tiempo en tiempo t , digamos x_t usando solo el valor x_s , si podemos predecir x_t exactamente de x_s a través de la relación lineal $x_t = \beta_0 + \beta_1 x_s$ entonces la correlación será 1 cuando $\beta_1 > 0$ y -1 cuando $\beta_1 < 0$.

2.4.4. Series de Tiempo Estacionarias

Las series temporales se pueden clasificar en:

1. **Estacionarias:** Una serie es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo.
2. **No Estacionarias:** Son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

- **Definición**

Una serie de tiempo **estrictamente estacionaria** es una serie para la cual el comportamiento probabilístico de cada sucesión de valores:

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$$

Es idéntico a la serie trasladada en el tiempo:

$$\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$$

Esto es,

$$P\{x_{t_1} \leq c_1, \dots, x_{t_k} \leq c_k\} = P\{x_{t_1+h} \leq c_1, \dots, x_{t_k+h} \leq c_k\} \quad (2.5)$$

Para todo $k = 1, 2, \dots$, todos puntos de tiempo $t_1, t_2, \dots, t_k$ y números reales $c_1, c_2, \dots, c_k$ y todo salto $h = 0, \pm 1, \pm 2, \dots$.

Esta versión de estacionaridad es muy fuerte para la mayoría de aplicaciones. Más aún, es difícil conseguir estricta estacionaridad en un conjunto sencillo de datos.

- **Definición**

Una serie de tiempo **débilmente estacionaria** X_t , es un proceso de varianza finita tal que:

1. La función media μ_t , definida en (2.2) es constante y no depende del tiempo t , y.
2. La función de covarianza, (s, t) , definida en (2.3) depende solo de las diferencias de s y t , $|s-t|$.

Por consiguiente, usaremos el término **estacionaridad** para referirnos a estacionalidad débil; si un proceso es estacionario en el sentido estricto usaremos el término estrictamente estacionario.

Es claro de la definición de estrictamente estacionario que una serie de tiempo estrictamente estacionaria con varianza finita, también es una serie estacionaria.

- **Definición**

La **función de autocovarianza de una serie de tiempo estacionaria** se escribirá como:

$$\gamma(h) = E[(x_{t+h} - \mu)(x_t - \mu)]. \quad (2.6)$$

- **Definición**

La **función de autocorrelación (ACF) de una serie de tiempo estacionaria** será escrita, usando (2.4) como:

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)}. \quad (2.7)$$

Por la desigualdad de Cauchy-Schwartz tenemos que $-1 \leq \rho(h) \leq 1$ para todo h .

2.4.5. Estimadores de Correlación

- **Definición**

Sea $x_1, x_2, \dots, x_n$ una muestra de una serie de tiempo. La **media muestral** de $x_1, x_2, \dots, x_n$ es:

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t, \quad (2.8)$$

- **Definición**

La **función de autocovarianza muestral** se define como:

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}), \quad (2.9)$$

Con $\hat{\gamma}(-h) = \hat{\gamma}(h)$ para $h = 0, 1, 2, \dots, n-1$.

- **Definición**

La **función de autocorrelación muestral** se define como:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad (2.10)$$

2.4.6. Modelos ARMA

Los modelos ARMA deben su nombre a sus siglas en inglés, Auto Regressive Moving Average (Modelos Autorregresivos de Promedio Móvil).

- **Los Modelos Autorregresivos AR**

Los modelos autorregresivos están basados en la idea de que el valor actual de la serie x_t se puede explicar cómo una función de p valores pasados $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ donde p determina el número de pasos necesarios para predecir el valor actual.

1) Definición

Un **modelo autoregresivo de orden p** , abreviado **AR (p)**, es de la forma:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \omega_t \quad (2.11)$$

Donde x_t es estacionario, $\phi_1, \phi_2, \dots, \phi_p$ son constantes ($\phi_p \neq 0$). A menos que se declare lo contrario, se asume que ω_t es un ruido blanco gaussiano de media cero y varianza $\sigma^2 \omega$. La media de x_t en (2.11) es cero. Si la media μ de x_t no es cero, reemplazamos x_t por $x_t - \mu$ en (2.11), es decir:

$$x_t - \mu = \phi_1 (x_{t-1} - \mu) + \phi_2 (x_{t-2} - \mu) + \dots + \phi_p (x_{t-p} - \mu) + \omega_t,$$

O escribimos:

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \omega_t \quad (2.12)$$

Donde $\alpha = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$.

2) Definición

El operador autoregresivo se define como:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.13)$$

- **Los Modelos de Promedio Móvil MA**

Como una alternativa a la representación autoregresiva en la cual x_t del lado izquierdo de la ecuación se asume como una combinación lineal, en los modelos de promedio móvil de orden q abreviados MA (q) asumimos el ruido blanco ω_t del lado derecho de la ecuación que los define como una combinación lineal de los datos observados.

1) Definición

El **modelo de promedio móvil de orden q** o modelo **MA (q)**, se define como:

$$x_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q}, \quad (2.18)$$

Donde hay q pasos en el promedio móvil y $\theta_1, \theta_2, \dots, \theta_q (q \neq 0)$ son parámetros.

El ruido ω_t se asume como un ruido blanco gaussiano.

Podemos también escribir el proceso MA (q) en la forma equivalente:

$$x_t = \theta(B)\omega_t, \quad (2.19)$$

Usando la siguiente definición.

2) Definición

El **operador de promedio móvil** es:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.20)$$

En contraste con el proceso autoregresivo, el proceso de promedio móvil es estacionario para cada valor de los parámetros $\theta_1, \theta_2, \dots, \theta_q$.

- **Modelos Autorregresivos de Promedio Móvil (ARMA)**

Dada una serie temporal de datos X_t , el modelo ARMA es una herramienta para entender y, aún más, para predecir futuros valores de la serie. El modelo está formado

por dos partes, una parte autorregresiva (AR) y otra de media móvil (MA). El modelo se conoce con el nombre de modelo ARMA (p, q), donde p es el orden de la parte autorregresiva y q es el orden de la parte de media móvil.

2.4.7. Modelos ARIMA

Los modelos econométricos contemplan de forma explícita la información que aportan las variables causales del fenómeno de interés, de acuerdo con una determinada teoría económica. Una ventaja de este modelo consiste en que los resultados que se generan son más eficientes y poseen mayor poder explicativo que los modelos univariados. Sin embargo, en estos modelos, cuando se desea realizar predicciones, el desconocimiento de los valores de las variables explicativas en el futuro, determina la necesidad de utilizar predicciones para éstas, lo cual incrementa el nivel de incertidumbre con que se realiza la predicción econométrica. Por otra parte, cuando el futuro puede suponer una alteración de tendencias de comportamiento respecto al pasado reciente, es recomendable utilizar estos modelos para predecir a mediano plazo (1 a 5 años), por lo cual nos resultarán de gran ayuda en nuestro estudio.

Ya hemos discutido la importancia de los modelos ARMA para representar series estacionarias. Una generalización de estos modelos, que incorporan un amplio rango de series no estacionarias, es proporcionada por los procesos **ARIMA**, es decir, procesos que se reducen a procesos ARMA cuando se los diferencia infinitamente muchas veces.

Terminológicamente hablando: decimos que Y_t está integrado de orden 1, I (1), porque tiene que ser diferenciado una vez para obtener una serie temporal estacionaria. En general, una serie puede ser I (d), si debe ser diferenciada “d” veces para obtener una serie estacionaria.

- **Definición**

Si d es un entero no negativo, entonces X_t es un proceso ARIMA (p,d,q) si:

$$Y_t = (1 - B)^d X_t, \quad (2.20)$$

En un proceso ARMA casual (p,q).

Esta definición significa que X_t satisface una ecuación en diferencia de la forma:

$$\varphi * (B)X_t = \varphi(B)(1 - B)^d X_t = \theta(B)Z_t, \quad (2.21)$$

Donde $Z_t \sim \omega_n(0, \sigma^2 \omega)$, $\varphi(z)$ y $\theta(z)$ son polinomios de grado p y q , respectivamente, y $\varphi(z) \neq 0$ para $|z| \leq 1$. El polinomio $\varphi * (z)$ tiene un cero de orden d en $z = 1$. El proceso X_t es estacionario si y solo si $d=0$, en cuyo caso se reduce a un proceso ARMA (p, q). Hay que tener en cuenta que si $d \geq 1$, podemos agregar una tendencia polinómica arbitraria de grado $(d - 1)$ a X_t sin violar la ecuación de diferencia (2.24). Los modelos ARIMA son, por lo tanto, útiles para representar datos con tendencia.

Sin embargo, los procesos ARIMA también pueden ser apropiados para modelar series sin tendencia alguna.

NOTA: Para trabajar con los modelos ARMA y ARIMA nuestra serie de tiempo necesariamente debe ser estacionaria.

- **Definición**

Tipos de suavizado de la serie:

- Suavizado Logarítmico: Consta de tomarle el Logaritmo en base 10 o logaritmo neperiano a los datos para alisar o suavizar la serie de tiempo.
- Suavizado exponencial: Este método contiene un mecanismo de autocorrección que ajusta los pronósticos en dirección opuesta a los errores pasados. Es un caso particular de promedios móviles ponderados de los valores actuales y anteriores en el cual las ponderaciones disminuyen exponencialmente. Se emplea tanto para suavizar como para realizar pronósticos. [Suárez, M. (2014)]

Las desviaciones de la estacionaridad pueden ser sugeridas por el gráfico de la serie misma o por la función de autocorrelación de muestra o ambos.

1) Prueba de Dickey-Fuller

La Prueba de Dickey-Fuller busca determinar la existencia o no de raíces unitarias en una serie de tiempo. La hipótesis nula H_0 de esta prueba es que existe una raíz unitaria en la serie.

Esta prueba contrasta la hipótesis nula H_0 de existencia de una raíz unitaria contra la alternativa H_1 de que no existen raíces unitarias. Para verificar la existencia de una raíz unitaria en un proceso AR (p), se lleva a cabo la siguiente prueba:

$$H_0 : \rho = 1 \quad vs \quad H_1 : \rho < 1 \quad (2.22)$$

Aplicando la regresión:

$$X_t = c_t + \rho X_t + \sum_{i=1}^{p-1} \varphi_i \Delta X_{t-i} + \xi_t, \quad (2.23)$$

Donde c_t es una función determinista del tiempo t y $\Delta X_i = X_i - X_{i-1}$ es la serie de diferencias de X_t usada para aproximar la estructura ARMA de los errores, y el valor de ρ se fija de modo que el error ξ_t sea correlacionado serialmente. En la práctica, c_t puede ser cero, una constante o bien $c_t = \omega_0 + \omega_1 t$. El ADF se basa en las estimaciones de mínimos cuadrados y está dado por:

$$ADF - Test = \frac{\hat{\rho} - 1}{std(\hat{\rho})} \quad (2.24)$$

Donde $\hat{\rho}$ denota el valor estimado por mínimos cuadrados de ρ y $std(\hat{\rho})$ su desviación estándar.

El fundamento econométrico de la prueba es un simple modelo autoregresivo de orden (1):

$$X_t = \rho X_{t-1} + \xi_t \quad (2.25)$$

Donde X_t es la variable de interés, t es el índice de tiempo, ρ es un coeficiente, y ξ_t es el término de error. La raíz unitaria está presente si $\rho = 1$. En este caso, el modelo no sería estacionario.

Mientras el valor de ρ sea más cercano a cero se considerará más estacionaria la serie.

2.4.8. Funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF)

Villagarcía, T. (2017), menciona que cuando se quiere analizar la serie es necesario identificar la estructura que la genera, es decir cómo influyen las observaciones del pasado en las observaciones del futuro. Para identificar esta dependencia utilizamos dos fuentes de información, la Función de Autocorrelación Simple (ACF) y la Función de Autocorrelación Parcial (PACF).

- **Función de Autocorrelación simple (ACF)**

La función de autocorrelación simple de una serie proporciona la estructura de dependencia lineal de la misma.

Si denominamos X_t a la serie temporal, los valores que se observan van a ser:

$$X_1, X_2, \dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, \dots \quad (2.26)$$

Donde X_1 representa el primer valor de la serie, X_2 el segundo, y X_t será el valor actual de la serie. De este modo, X_{t+1} representa el valor de la serie para el próximo período, es decir, un valor futuro.

Si denominamos que X_1 influye sobre X_2 como $X_1 \rightarrow X_2$, la función de autocorrelación simple tiene por objetivo estudiar cómo influyen entre ellas las diversas observaciones:

$$X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_{t-1} \rightarrow X_t \rightarrow X_{t+1} \rightarrow \dots \quad (2.27)$$

La idea de la función de autocorrelación es proporcionar el coeficiente de correlación entre las observaciones separadas un número determinado de períodos. Así la ACF, será una sucesión de números:

$$\rho_1, \rho_2, \dots, \rho_k \quad (2.28)$$

Que representan cómo influye una observación sobre la siguiente (ρ_1), sobre la segunda posterior (ρ_2) o sobre la k períodos posteriores (ρ_k).¹⁶

$\rho_1 \rightarrow$ como influye una observación sobre la siguiente. $X_i \rightarrow X_{i+1}$

$\rho_2 \rightarrow$ como influye una observación sobre la 2 períodos posteriores. $X_i \rightarrow X_{i+2}$

$\rho_3 \rightarrow$ como influye una observación sobre la 3 períodos posterior es. $X_i \rightarrow X_{i+3}$

Los coeficientes de la ACF, $\rho_1, \rho_2, \dots, \rho_k, \dots$ están acotados entre $[-1, +1]$. Cuando un ρ_i vale cero, quiere decir que no existe efecto entre una observación y la i posiciones posterior.

Si ρ_i es próximo a 1 indica que hay mucha relación entre una observación y la i posiciones posterior, y que esa relación es positiva.

Si ρ_i es próximo a -1 indica que hay mucha relación entre una observación y la i posiciones posterior, y que esa relación es negativa.

La ACF proporciona cómo una observación influye sobre las postreras. Aunque la misma presenta un problema, y es que si por ejemplo ρ_1 es distinto de cero, entonces:

$$X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_{t-1} \rightarrow X_t \rightarrow X_{t+1} \rightarrow \dots$$

Es decir, existe una cadena de influencia separada por un retardo. Pero si $X_1 \rightarrow X_2$ y $X_2 \rightarrow X_3$, entonces $X_1 \rightarrow X_3$. Por lo tanto, la ACF en general, si ρ_1 es distinto de cero, encontrará que $\rho_2, \rho_3, \dots$, serán distintos de cero.

Sin embargo es necesario distinguir entre varias cadenas de influencia posibles:

- La cadena de influencia general ρ_1 .
- Las cadenas de influencia directa. Es decir, como influye X_1 sobre X_3 directamente, es decir, sin pasar a través de X_2 .

Para resolver este problema se construye la Función de Autocorrelación Parcial (PACF).

- **Función de Autocorrelación Parcial (PACF)**

La función de autocorrelación parcial (PACF) proporciona la relación directa que existe entre observaciones separadas por k retardos. Esta es una información muy valiosa sobre la estructura de la serie, ya que elimina el problema que presentaba la función de autocorrelación simple de que si $X_1 \rightarrow X_2$ y $X_2 \rightarrow X_3$, entonces $X_1 \rightarrow X_3$. En la función de autocorrelación simple, la primera barra (relación entre X_1 y X_2 ó X_2 y X_3) será significativa. Y el segundo también, ya que si $X_1 \rightarrow X_2$ y $X_2 \rightarrow X_3$, entonces $X_1 \rightarrow X_3$.

www.bdigital.ula.ve

2.4.9. Pronósticos

Mendoza, M. (2011), define un pronóstico como una afirmación sobre un evento cuya ocurrencia no es segura. Típicamente, los pronósticos se producen sobre eventos que pueden ocurrir en el futuro. Los fenómenos que dan origen a los eventos que son objeto de pronósticos se conocen con el nombre de fenómenos inciertos (sus resultados no se pueden anticipar con certeza). En virtud de que los eventos para los cuales se producen pronósticos no se pueden anticipar con certeza, una característica intrínseca de todo pronóstico es que puede fallar. La incertidumbre asociada a los fenómenos inciertos puede provenir de, al menos, dos fuentes distintas: la falta de conocimiento y la falta de control. La falta de control se manifiesta a través de la variabilidad de los resultados observados.

La producción de pronósticos está inevitablemente asociada con descripción de la incertidumbre. Producir pronósticos es relativamente fácil, pero producir pronósticos que acierten con frecuencia es difícil. Por otra parte, producir pronósticos que acierten siempre es imposible. Medir el grado de confiabilidad de un pronóstico, después de la eventual ocurrencia del evento relevante, es fácil. En cambio, medir el grado de confiabilidad de un pronóstico, antes de la eventual ocurrencia del evento relevante, es difícil. [Mendoza, M. 2011)]

Un pronóstico ideal debe incluir: una medida de su precisión, una medida de su confiabilidad (previa); además, debe obtenerse con un mecanismo reproducible. Las técnicas estadísticas de pronóstico no solamente reúnen estas tres características sino que constituyen una valiosa herramienta para la producción de pronósticos a partir del análisis de información previa. [Mendoza, M. (2011)]

2.5. Machine Learning (Aprendizaje Automático)

En la actualidad desarrollar o adquirir sistemas de información transaccionales no debe ser el único fin de los departamentos de tecnología. Por el contrario, uno de los fines principales debe ser aprovechar todos los datos que se han acumulado durante años y utilizarlos para aportar valor real al negocio y generar una ventaja competitiva en el mercado. El Aprendizaje Automático o mejor conocido como **Machine Learning** es un término que se escucha cada vez más con mayor frecuencia, dado que proporciona valor al negocio y arroja un nuevo uso y sentido a los datos con que cuentan las organizaciones. [Porro A. (2018)]

A pesar de ser un término recientemente muy sonado y en ocasiones relacionado con el concepto de inteligencia artificial, lo cual de hecho, es bastante cierto, pues está catalogado como un subconjunto del mismo, sus inicios tienen raíces en los años 50 en el MIT. Hoy en día el aprendizaje automático está presente en muchas de las actividades que nos rodean, los asistentes de voz de Siri y Alexa, el reconocimiento facial de Facebook y Microsoft, las recomendaciones de Amazon y Netflix y la tecnología que

evita que los vehículos autónomos se estrellen, son el resultado de los avances en el aprendizaje automático.

El término Machine Learning (Aprendizaje Automático) fue acuñado por **Arthur Samuel** en 1959, un pionero estadounidense en el campo de los juegos informáticos y la inteligencia artificial, afirmó que da a las computadoras la capacidad de aprender sin ser programadas explícitamente.

En 1997, **Tom Mitchell** dio una definición matemática y relacional a la que consideró como bien planteada, “se dice que un programa de computadora aprende de la experiencia E con respecto a alguna tarea T y alguna medida de rendimiento P, si su rendimiento en T, medido por P, mejora con la experiencia E”.

2.5.1 Conceptos Básicos

La Minería de Datos (Data Mining) es el estudio de la recolección, limpieza, procesamiento, análisis y obtención de información útil. Existe una amplia variación en términos de los dominios problemáticos, aplicaciones, formulaciones y representaciones de datos que se encuentran en aplicaciones reales. Por lo tanto, **Minería de datos** es un término general muy amplio que se utiliza para describir estos diferentes aspectos de procesamiento de datos. [Porro A. (2018)]

Aprendizaje Automático es como ya hablamos, una rama de la Inteligencia Artificial cuyo objetivo principal es desarrollar técnicas que permitan a las computadoras aprender. Más concretamente, se trata de crear algoritmos capaces de generalizar comportamientos y reconocer patrones a partir de una información suministrada. Entonces podemos decir que es un proceso de inducción del conocimiento, en otras palabras, un método que permite obtener por generalización un enunciado general, a partir de enunciados que describen casos particulares. Una vez observados todos los casos particulares, la inducción se considera completa, por lo que la generalización a la que da lugar se considera válida.

Sin embargo, en la mayoría de los casos es imposible obtener una inducción completa, por cual, el enunciado a que da lugar queda sometido a cierto grado de incertidumbre y por lo tanto, no se puede considerar como un esquema de inferencia totalmente válido ni se puede justificar empíricamente. En muchas ocasiones el campo de actuación del aprendizaje automático se solapa con el de minería de datos, ya que las dos disciplinas están enfocadas en el análisis de datos, sin embargo el Aprendizaje Automático se centra más en el estudio de la complejidad computacional de los problemas con la intención de hacerlos factibles desde el punto de vista práctico, no únicamente teórico. [Porro A. (2018)]

Básicamente, podríamos decir que una de las tareas del aprendizaje automático, es intentar extraer conocimiento sobre algunas propiedades no observadas de un objeto, basándonos en las propiedades que sí han sido observadas de ese mismo objeto, inclusive basándose en propiedades observadas en objetos similares. Esto significa, que podemos predecir un comportamiento futuro a partir de lo que ha ocurrido en el pasado.

Un ejemplo de mucha actualidad sería, el de predecir si un determinado producto le va a gustar a un cliente, tomando en cuenta las valoraciones que ese mismo cliente ha hecho sobre otros productos que sí ha probado.

2.5.2 Tipos de Aprendizaje Automático

Entre los tipos de Aprendizaje Automático tenemos los siguientes:

1. Supervisado: se genera una función que establece una correspondencia entre las entradas y las salidas deseadas del sistema, donde la base de conocimientos del sistema está formada por ejemplos etiquetados a priori (es decir, ejemplos de los que sabemos su clasificación correcta). Este consta de dos fases, una de entrenamiento y una de prueba:

- a.) En la fase de entrenamiento se cuenta con un conjunto de datos que por lo general está entre el 60% o 70% del total de los datos disponibles, que son con los que se entrena al algoritmo para encontrar los patrones y relaciones en el conjunto de datos.
- b.) Posteriormente en la fase de prueba, se cuenta con un conjunto de datos de prueba entre el 20% o 30% del total de los datos disponibles, el cual sirve para validar el rendimiento del algoritmo. [Porro A. (2018)]

2. No supervisado: es donde el proceso de modelado se lleva a cabo sobre un conjunto de ejemplos formados únicamente por entradas al sistema, sin conocer su clasificación correcta. En este caso, lo que se busca es que el sistema sea capaz de reconocer patrones para poder etiquetar las nuevas entradas.

3. Semi-supervisado: es una combinación de los dos algoritmos anteriores, teniendo en cuenta ejemplos clasificados y no clasificados.

2.5.2.1 Métodos de Aprendizaje Automático

Existen diversos métodos de Aprendizaje Automático, A.A, los cuales se diferencian entre sí por el tipo de objeto que intentan predecir. A continuación se listan tres de los más usados para abordar una amplia gama de problemas.

- 1. Regresión:** Intentan predecir un valor real. Por ejemplo, predecir el valor de la bolsa mañana a partir del comportamiento de la bolsa que está almacenado (datos pasados). Consiste en encontrar la mejor ecuación que atraviese de forma óptima un conjunto de puntos (n dimensiones). Se utiliza cuando la precisión no es crítica y el número de variables es pequeño. Este método está considerado dentro del tipo de aprendizaje supervisado.

2. **Clasificación:** Intentan predecir la clasificación de objetos sobre un conjunto de clases prefijadas. Por ejemplo, clasificar si una determinada noticia es de deportes, entretenimiento, política, etc. Si sólo se permiten 2 posibles clases, entonces se llama clasificación binaria; si se permiten más de 2 clases, estamos hablando de clasificación multiclase. Ej. Identificar si un tumor es maligno o benigno, dado su tamaño y edad del paciente. Este método también está considerado como aprendizaje supervisado. [Porro A. (2018)]
3. **Agrupación:** Este método se utiliza cuando se necesitan clasificar las instancias de datos pero no se conocen previamente las categorías. Esta agrupación permite construir grupos (cluster) coherentes de instancias, teniendo en cuenta las variables del conjunto de datos.

En palabras sencillas, permite encontrar qué se tiene en los datos. Por ejemplo se utiliza para la clasificación de personas dados sus genomas, clasificación de documentos, entre otros. Está considerado como un tipo de aprendizaje no supervisado.

www.bdigital.ula.ve

2.5.3 Reglas de Asociación

Propuesto por Agrawal et al. En 1993, a pesar de no ser un algoritmo diseñado para estudiar datos numéricos, nos será de gran utilidad para encontrar y determinar patrones existentes entre las diferentes criptomonedas a estudiar. Su uso inicial fue para el análisis del carro de supermercado (Market Basket Analysis), a fin de encontrar relaciones entre los distintos artículos en las compras. El objetivo principal de este algoritmo, es determinar qué operaciones se están realizando de manera conjunta, asociar de manera eficiente cuáles criptomonedas que poseen comportamientos similares o identificar criterios de dependencia. [Porro A. (2018)]

- **Definición**

El soporte de un elemento I se define como la fracción de transacciones en la base de datos $T = T_1, \dots, T_n$ que contienen I como un subconjunto. El soporte de un elemento I se denota por $\text{su}(I)$.

Claramente, los artículos que están correlacionados frecuentemente ocurren juntos en las transacciones. Tales conjuntos de elementos tendrán un gran soporte. Por lo tanto, el problema frecuente de minería de patrones es el de determinar conjuntos de elementos que tienen el requisito de nivel de soporte mínimo.

- **Definición**

Minería de conjuntos de elementos frecuentes Dado un conjunto de transacciones $T = T_1, \dots, T_n$, donde cada transacción T_i es un subconjunto de elementos de U , determina todos los conjuntos de elementos I que se producen como un subconjunto de al menos una fracción predefinida de mínimo soporte de las transacciones en T .

Propiedad de cierre descendente: cada subconjunto de un conjunto de elementos frecuentes, es también frecuente.

La propiedad de cierre descendente de los patrones frecuentes, es algorítmicamente muy conveniente porque proporciona una restricción importante en la estructura inherente de los patrones frecuentes. [Porro A. (2018)]

- **Definición**

Confianza: Sean X e Y dos conjuntos de elementos. La confianza $conf(X \cup Y)$ de la regla $X \cup Y$ es la probabilidad condicional de $X \cup Y$ que ocurre en una transacción, dado que la transacción contiene X . Por lo tanto, la confianza $conf(X \Rightarrow Y)$ se define de la siguiente manera:

$$conf(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)} \quad (2.29)$$

Se dice que los conjuntos de elementos X e Y son el antecedente y el consecuente de la regla, respectivamente.

- **Definición**

Reglas de Asociación: sean X e Y dos conjuntos de elementos. Entonces, la regla $X \Rightarrow Y$ se dice que es una regla de asociación con un mínimo de soporte *minsup* y mínimo de confianza *minconf*, si satisface los dos criterios siguientes:

1. El soporte del conjunto de elementos $X \cup Y$ es al menos *minsup*.
2. La confianza de la regla $X \Rightarrow Y$ es al menos *minconf*. [Porro A. (2018)]

El primer criterio garantiza que un número suficiente de transacciones sea relevante para la regla; por lo tanto, tiene la masa crítica requerida para que se considere relevante para la aplicación a mano. El segundo criterio asegura que la regla tiene fuerza suficiente en términos de probabilidad condicional. Por lo tanto, las dos medidas cuantifican diferentes aspectos de las reglas de asociación.

2.5.4. Algoritmo Apriori

El algoritmo Apriori comienza contando los soportes de los elementos individuales para generar los frecuentes 1-conjuntos de elementos.

Los 1-conjuntos de elementos se combinan para crear candidatos 2-conjuntos de elementos, cuyo soporte es contado. Los 2-conjuntos de elementos frecuentes se conservan. En general, los conjuntos de elementos frecuentes de longitud k , se usan para generar los candidatos de longitud $(k + 1)$ para valores crecientes de k . Los algoritmos que cuentan el soporte de los candidatos con mayor longitud son referidos como algoritmos de nivel. Sea F_k el conjunto de k -conjuntos de elementos frecuentes, y C_k denota el conjunto de candidatos k -conjuntos de elementos. El núcleo del enfoque es generar iterativamente los $(k + 1)$ - candidatos $C_k + 1$ de k -conjuntos de elementos frecuentes en F_k ya encontrados por el algoritmo. Las frecuencias de estos $(k + 1)$ candidatos se cuentan con respecto a la base de datos de transacciones. Al generar los candidatos $(k + 1)$, el espacio de búsqueda puede ser reducido chequeando

si todos los k -subconjuntos de $C_k + 1$ están incluidos en F_k . Entonces, ¿cómo se genera un candidato relevante $(k + 1)$ en $C_k + 1$ de patrones frecuentes de k en F_k ? [Porro A. (2018)]

Si un par de conjuntos de elementos X e Y en F_k tienen en común $(k - 1)$ elementos, entonces una unión entre ellos usando los elementos comunes $(k - 1)$ crearán un conjunto de elementos candidato de tamaño $(k + 1)$. Por ejemplo, los dos 3-conjuntos de elementos $\{abc\}$ y $\{abd\}$ o abc y abd para abreviar, cuando se unen en los dos elementos comunes a y b , arrojarán el candidato $abcd$. Por supuesto, es posible unirse a otros patrones frecuentes para crear el mismo candidato. Uno también podría unir $\{abc\}$ y $\{bcd\}$ para lograr el mismo resultado. Supongamos que cuatro de los 3 subconjuntos de $abcd$ están presentes en el conjunto de 3 conjuntos de elementos frecuentes. Uno puede crear el candidato $abcd$ en de $\binom{4}{2} = 6$ formas diferentes.

Para evitar la redundancia en la generación de candidatos, la convención es imponer un orden lexicográfico en los artículos y usar los primeros $(k - 1)$ elementos del conjunto de elementos para la unión. Por lo tanto, en este caso, la única forma de generar $abcd$ sería unirse usando los primeros dos elementos a y b . Por lo tanto, los conjuntos de elementos abc y abd tendrían que unirse para crear a B C D. Tenga en cuenta que, si cualquiera de abc y abd no son frecuentes, entonces $abcd$ no se generará como un candidato que usa este enfoque de unión. Además, en tal caso, se garantiza que $abcd$ no será frecuente debido a la propiedad de cierre descendente de conjuntos de elementos frecuentes. Por lo tanto, la propiedad de cierre descendente asegura que en el conjunto de candidatos generados utilizando este enfoque no se pierde ningún elemento que sea verdaderamente frecuente. [Porro A. (2018)]

Capítulo 3

Aspectos Metodológicos

En el presente capítulo se describe el procedimiento o secuencia de pasos a seguir a fin de resolver el problema planteado y dar solución a los objetivos propuestos en la investigación.

3.1. Estadística descriptiva

En este trabajo se utilizan bases de datos históricas que comprenden datos entre diciembre del 2013 a enero del 2020, dichos datos fueron extraídos de la página web ‘*coinmarketcap.com*’ para cada criptomoneda respectivamente. Al tratarse de datos cuantitativos los que representan a las variables de cada criptodivisa en estudio, en primera instancia, mediante la aplicación de estadística descriptiva se realiza el cálculo de indicadores estadísticos básicos (media, mediana, varianza, desviación estándar) aunado a la elaboración de gráficos de correlación, obtención de valores atípicos “*outliers*”, exploración de datos faltantes (NA’s), entre otros. Esto, con el propósito de analizar e interpretar el comportamiento de los mismos a partir de las relaciones existentes entre las variables, cómo se distribuyen sus observaciones y otros resultados que permitan clasificar, organizar y representar dicha información haciéndola comprensible para el estudio posterior.

Debido a que en el mundo de las finanzas existen ciertos patrones o tendencias anómalas y aún más haciendo referencia al mercado de las criptodivisas, las fluctuaciones en el rendimiento de las mismas son inciertas, por esa razón, en este proyecto se decidió no omitir los valores atípicos de rendimiento ya que disminuiría en gran medida el número de observaciones obtenidas en principio.

3.2. Metodología a utilizar para Series Temporales

Luego de hacer el estudio previo a los datos, se edita la base de datos, dejando las variables de interés (fecha de la observación y rendimiento) para modelar la serie de

tiempo de cada criptomoneda, a fin de observar el rendimiento real que ha tenido la misma a través del tiempo.

Generalmente, cuando se quiere generar un modelo predictivo es necesario dividir la data en dos subconjuntos, uno denominado **conjunto de entrenamiento** y otro llamado conjunto o **fase de prueba**. El propósito de hacer esto es que, a partir del set de entrenamiento el modelo sea capaz de predecir una salida que al compararla con el set de prueba se ajuste lo mayor posible a éste, determinando así su efectividad como buena o mala dependiendo de la aproximación que tenga el pronóstico a la realidad.

En este estudio, se modelará la serie temporal de cada criptomoneda utilizando el conjunto de datos de entrenamiento para crear el modelo ARIMA que mejor se ajuste para luego comparar los datos pronosticados con el conjunto de prueba y ver qué tan bueno es el modelo a partir del porcentaje de error medio que se obtenga.

Fernández, S. (2008), expone en su artículo “*Series temporales: Modelo ARIMA*”, a Box, G. and Jenkins M. como desarrolladores de modelos estadísticos para series temporales que tienen en cuenta la dependencia existente entre los datos, esto es, cada observación en un momento dado es modelada en función de los valores anteriores. Los análisis se basan en un modelo explícito, y dichos modelos se conocen con el nombre genérico de **ARIMA** (*Auto Regresive Integrated Moving Average*), que deriva de sus tres componentes **AR** (Auto Regresivo), **I** (Integrado) y **MA** (Medias Móviles). El modelo ARIMA permite describir un valor como una función lineal de datos anteriores y errores debidos al azar. Además, puede incluir un componente cíclico o estacional, es decir, debe contener todos los elementos necesarios para describir el fenómeno.

Partiendo de la metodología de Box & Jenkins, se siguen una serie de etapas o fases para la elaboración de un modelo ARIMA con fines predictivos, las cuales se desglosan a continuación:

➤ **Etapa 1: Identificación**

La primera fase consiste en identificar el posible modelo ARIMA que sigue la serie, detectando el tipo de proceso estocástico que han generado los datos. Es decir, se encuentran los valores (parámetros) adecuados de p , d y q del modelo ARIMA. Para la identificación de los parámetros p y q que hacen referencia al orden de la estructura autorregresiva y de media móvil del proceso, se hace uso de los gráficos arrojados por las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF).

Es importante mencionar una condición inicial que debe cumplir una serie temporal antes de utilizar los criterios de identificación ACF (q) y PACF (p) para ajustar el modelo ARIMA, la cual es cumplir con el supuesto de estacionaridad, es decir, que presente media y varianza constante a lo largo del tiempo. En caso de no ser estacionaria, se aplican diversas transformaciones dependiendo del conjunto de datos para suavizarla y así lograr que la misma sea estacionaria.

Como primer paso del estudio de series temporales, se realiza una representación gráfica de la serie para ver el comportamiento de los datos y a partir de este punto decidir sobre la estacionaridad, aplicando algún tipo de función para el suavizamiento a la serie, que puede ser: logarítmica, exponencial, cuadrática, trigonométrica, entre otras; dependiendo de los datos que contenga dicha serie a fin de lograr transformar los valores para que representen una media y varianza constante a través del tiempo.

Luego de aplicar el suavizado a la serie, para comprobar la estacionaridad de la misma, se procedió a realizar el test de *Dickey-Fuller*, también conocido como la prueba de la raíz unitaria. Un artículo publicado por Moffatt, M. (2019), menciona que esta prueba se utiliza para comprobar si la serie es estacionaria o no, partiendo de una hipótesis nula H_0 que alega la existencia de que una raíz unitaria se encuentra presente en la serie temporal. Dicha prueba consta de 2 parámetros fundamentales que determinan el rechazo o aceptación de H_0 , los cuales son: el valor del estadístico de prueba *Dickey-Fuller*, el cual debería ser negativo para cumplir con la hipótesis alternativa H_1 , junto con el valor de p -value, que debería ser menor a 0.05. (p -value <

0.05). Cuanto más negativo es el valor del estadístico de prueba ADF , más fuerte es el rechazo de la hipótesis de que hay una raíz unitaria.

En caso de no ser estacionaria, se procede a observar la gráfica de la serie temporal a ver si presenta algún tipo comportamiento donde se logre comprobar la existencia de algún componente estacional, de tendencia, o en su defecto ambos. En caso de ser así, es necesario eliminar dicha tendencia lineal o periodicidad mediante el uso de una diferenciación. Según expresa Mompeán, O. (2016), existen dos tipos de diferenciación: diferenciación regular y diferenciación estacional. La diferenciación regular se utiliza para eliminar la tendencia de una serie aleatoria, dicha diferenciación consiste en que a cada dato se le resta el anterior, suponiendo que la tendencia evoluciona lentamente, consiguiendo así una nueva serie con un dato menos que la original ($n-1$ observaciones) en la que se ha eliminado la tendencia. Si la tendencia no se ha eliminado por completo al realizar una primera diferenciación regular, se realiza una segunda vez.

$$\Delta Y_t = Y_t - Y_{t-1} \quad (3.1)$$

Por otra parte, para realizar una diferenciación estacional se le realiza a la serie una transformación similar a la anterior para la eliminación de la tendencia, solo que en este caso la diferencia que se toma es la frecuencia anual de los datos.

$$\Delta_a Y_t = Y_t - Y_{t-a} \quad (3.2)$$

Fernández, S. (2008), dice que una forma de saber cuántas veces es necesario diferenciar es a partir del cálculo de la varianza de la serie original y de la serie sometida a diferentes diferenciaciones, tomando como diferenciación adecuada aquella para la que la varianza es mínima. El método es tanto más adecuado cuanto mayor sea la diferencia entre las varianzas anteriores.

Una vez confirmada la estacionaridad de la serie al someterla nuevamente (en caso de ser necesario) al test ADF y superarlo, además de eliminar el componente estacional o de tendencia mediante la diferenciación de la serie, proceso mediante el cual se identifica el orden de integración $I(d)$ del modelo ARIMA, con la finalidad de volver aún más estacionaria la serie en caso de ya serlo y sino, diferenciarla d veces hasta

convertirla en estacionaria, se procede a la identificación de los parámetros p y q mediante los gráficos ACF y $PACF$ respectivamente. Para determinar los valores de dichos parámetros, se presenta a continuación un cuadro con los patrones teóricos de ACF y $PACF$ según el tipo de modelo, los cuales serán de utilidad para la identificación de p y q del modelo ARIMA:

Tabla 3.1. Casos para la definición del modelo adecuado.

Fuente: Rosales, R. (2008)

Tipo de Modelo	Patrón Típico de ACF	Patrón típico de PACF
$AR(p)$	Decrecimiento rápido de tipo geométrico puro, geométrico con alteración de signos, sinusoidal o mezcla de ambos tipos	Picos grandes en los p rezagos o corta abruptamente a partir del rezago p , es decir la FAP se anula luego del rezago p .
$MA(q)$	Picos grandes en los q rezagos o corta abruptamente a partir del rezago q , es decir la FAP se anula luego del rezago q .	Decrecimiento rápido de tipo exponencial y /o sinusoidal
$ARMA(p,q)$	Decrecimiento exponencial	Decrecimiento exponencial

Cabe anotar, que en el procedimiento de identificación de p y q , se consideran aquellos rezagos estadísticamente significativos, es decir, se consideran los p retardos de las relaciones directas que existen entre las observaciones, donde se determina el orden p del modelo ARIMA a partir del primer retardo hasta que la función se anula o no logra superar la franja o límites de confianza inferior o superior, dependiendo del tipo de correlación que exista entre los retardos, la cual puede ser positiva o negativa, donde si los ρ_i son próximos a 1 indica que hay mucha relación entre una observación y la i posiciones posterior, y que dicha relación es positiva. En caso contrario, si los ρ_i son próximos a -1 sucede lo mismo pero la relación resulta negativa. Por lo tanto, no se incluyen rezagos intermedios hasta p o q si éstos no son significativos.

➤ **Etapa 2: Estimación**

Seleccionado provisionalmente un modelo para la serie estacionaria, en esta etapa se estiman los coeficientes de los términos autorregresivos y de media móvil incluidos en el modelo, cuyo número de rezagos p y q ya han sido identificados en la etapa anterior. Algunas veces los parámetros AR y MA del modelo se estiman utilizando el método de mínimos cuadrados lineales, pero otras veces se recurre a la estimación por el método de máxima verosimilitud y a partir de allí se obtienen sus errores estándar y los residuos del modelo. Los paquetes estadísticos efectúan este procedimiento a través de rutinas de computador en las que se tienen definidos los parámetros iniciales, así como los criterios de convergencia. En el caso de estudio, la herramienta a utilizar, será RStudio, el cual es un software que contiene los paquetes necesarios para realizar dicho estudio de manera automática.

Debido a que en la práctica es difícil identificar con exactitud el orden p y q del modelo ARMA, se suelen plantear dos o más modelos plausibles, que luego de ser estimados son útiles para la elección del más apropiado.

➤ **Etapa 3: Verificación de Diagnóstico**

La tercera fase es el diagnóstico, luego de tener predeterminado el modelo ARIMA (p, d, q) a utilizar para modelar la serie de tiempo en cuestión, se genera dicho modelo mediante la aplicación del método de máxima verosimilitud que trae por defecto la librería de series temporales para el software R. Una vez obtenido el modelo se realizan diversas pruebas ajustando los parámetros del modelo p, d y q a fin de encontrar uno que mejor se ajuste o modele la serie. Para esto, se hace el uso del criterio de información de Akaike (AIC) el cual es uno de los parámetros de salida del modelo, y según Wikipedia (2019) representa una medida de calidad relativa de un modelo estadístico, para un conjunto dado de datos, es decir, proporciona un medio para la selección del modelo. Este criterio se basa en la utilización de la suma de los cuadrados de los errores, buscando minimizarla a partir de diversas combinaciones de p y q . A esta

prueba también se le conoce como la prueba de Parsimonia. La expresión del *AIC* también llamado criterio de Información Bayesiana (*BIC*) es la siguiente:

$$AIC = \ln\left(\frac{\widehat{\varepsilon\varepsilon}}{n}\right) \frac{2k}{n} \quad (3.3)$$

Donde k es el número agregado de parámetros incluido el intercepto si se ha incorporado en la estimación, es decir, $k = p + q + 1$ y n es el número utilizable de observaciones.

Rosales, R. (2008), menciona sobre este criterio que, al efectuar el diagnóstico se desea que el *AIC* dé lo menor posible al comparar modelos con diversas combinaciones de p y q . En este proceso es importante considerar que los rezagos adicionales deben ser significativos, puesto que si éstos no lo son aumenta k sin que la suma de cuadrados de los errores se reduzca.

En esta etapa, se comprueba además que los residuos no tengan una estructura de dependencia y sigan un proceso de ruido blanco, es decir, que todos los posibles períodos de oscilación se presentan con igual intensidad (con media igual a 0 y varianza σ^2) y no tienen correlación serial. Si los residuos muestran estructura de dependencia se modifica el modelo para incorporarla y se repiten las etapas anteriores hasta obtener un modelo adecuado. Para esto, se utiliza como soporte el test de *Ljung – Box*, mediante una prueba efectuada para determinar la significancia de la hipótesis nula H_0 que hace referencia a la NO autocorrelación de los residuos, donde el valor de significancia α ó $p\text{-value} > 0.05$ para aceptar H_0 , es decir, que los residuos son independientes. De lo contrario se aceptaría la hipótesis alternativa H_1 en la que se alega que si existe una dependencia entre ellos.

➤ **Etapa 4. Pronóstico**

La cuarta y última fase es la predicción o pronóstico, donde una vez obtenido un modelo adecuado, es decir, aquel que es el “el mejor” resultante de las etapas anteriores, se realizan las predicciones del mismo.

En esta última etapa, luego de obtener los valores pronosticados por el modelo, se realiza una tabla que contiene las observaciones del conjunto de datos originales resultantes, los cuales se dejaron como fase o período de prueba con el propósito de compararlos con dichos valores resultantes arrojados por el pronóstico de ARIMA.

La evaluación o comprobación del pronóstico se realiza calculando el porcentaje del error absoluto medio (*MAPE*) entre las observaciones reales y pronosticadas en ese período de tiempo. A continuación se expresa matemáticamente el cálculo de dicho error:

$$MAPE = \frac{\frac{\sum_{t=1}^n |A_t - F_t|}{|A_t|}}{n} \quad (3.4)$$

Donde A_t representa el valor real y F_t el valor pronosticado por el modelo. Por otra parte n es el total de observaciones que conforman dichos conjuntos. Es necesario mencionar que, el método de pronóstico que se utiliza para este estudio es relativamente bueno para pronósticos a corto plazo, es decir, su efectividad de predicción es buena sólo para un set de muestras pequeño ($n \leq 30$ observaciones).

Para concluir el proceso de predicción del rendimiento de las criptomonedas, se realiza gráfico que contiene la serie temporal original más la traza de la serie pronosticada por el modelo resultante del proceso. Así se puede observar a mejor detalle el comportamiento gráficamente de ambas series y determinar que tan bien modela el pronóstico la realidad. Antes de aplicar éste último paso, es recomendable realizar la prueba de los residuos del modelo para ver si siguen una distribución normal mediante el test de *Shapiro Wilk*, donde H_0 representa la distribución normal de los residuos y se acepta si el valor de *p-value* es mayor a 0.05, para los valores de α más pequeños se rechaza la hipótesis nula de normalidad de los datos al aceptar la alternativa H_1 .

Este proyecto, tiene como enfoque u objetivo la predicción en el cambio de las criptodivisas a partir de su rendimiento y no del precio de las mismas, planteándose la posibilidad de encontrar un modelo que se ajuste de la mejor manera posible a la serie temporal al evitar los fuertes cambios producidos por los continuos y violentos puntos

de inflexión en el precio de la moneda, disipando en buena medida las fluctuaciones abruptas, a fin de obtener una gráfica que presente el patrón de comportamiento de ruido blanco, y de no ser así, aplicar las técnicas mencionadas con antelación para lograrlo.

Capítulo 4

Resultados y Análisis

En este capítulo se explican los resultados obtenidos mediante la implementación metodológica descrita anteriormente, haciendo un análisis de cómo o de qué manera se desarrolla el proceso, logrando así alcanzar los objetivos propuestos en primera instancia.

Como primer punto en la realización de esta etapa, resulta indispensable la obtención de los datos históricos referentes a las distintas criptomonedas a estudiar, los cuales nos permitirán hacer los análisis correspondientes. Dicho esto, a continuación se presenta una breve explicación de cómo se realiza el proceso de obtención, selección de las criptomonedas, extracción y limpieza de los datos:

1. **La fuente de datos:** dichos datos son provenientes de la página ‘*www.coinmarketcap.com*’ que es de donde nos hemos guiado. A su vez, se ha utilizado este sitio web para monitorear y hacerle seguimiento a las diversas criptomonedas del mercado.
2. **La elección de las criptomonedas:** se han escogido las 3 criptomonedas más cotizadas y con mayor capitalización en el mercado durante los últimos años; siendo éstas: Bitcoin (*BTC*), Ethereum (*ETH*) y el DASH (*DSH*).
3. **La extracción de los datos:** para el proceso de extracción de los datos contenidos en la página web, se utiliza como apoyo una de las herramientas del desarrollador del explorador de Google Chrome para el Scraping web, llamado “*Data Miner*” el cual es una aplicación que representa la versión mejorada de la

opción web scraper del navegador, donde se configura la misma para obtener la tabla (archivo.xlsx) de los registros de datos históricos de cada una de las criptomonedas seleccionadas, una por una.

4. Limpieza de los datos: haciendo un estudio de los datos y dado que el interés para este estudio se decanta por los rendimientos de cierre diario (*Close*), decidimos eliminar las columnas que no aportaban información relevante al modelo que se plantea en principio, quedando así solamente nuestra base de datos con las columnas Fecha (*Date*) y precio de cierre (*Close*) en primera instancia. Las columnas del conjunto de datos están definidas de la siguiente manera:

- *Date*: la fecha en que ocurrió cada observación.
- *Open*: nos indica el precio inicial de la criptomoneda en el día de la observación.
- *Close*: nos indica el precio de cierre registrado por la criptomoneda el día de la observación. Ésta es la variable escogida para realizar el análisis.
- *Volume*: ésta columna nos indica el número de transacciones en \$ realizadas en las últimas 24 horas.
- *Market Cap*: es una forma de clasificar el tamaño relativo de una criptomoneda. Se calcula multiplicando el precio de la moneda por el suministro circulante.

Antes de profundizar en los estudios a realizar, se aplica estadística descriptiva para que cumpla la tarea de entrar en contacto con la información numérica, a fin de (en base al conjunto de datos representado) poder clasificar, organizar y representar la información suministrada, con el propósito de hacerla comprensible; es decir, de resumirla, traducirla o mejor dicho, hacer mucho más entendibles los datos obtenidos y así poder sacar relaciones y detectar comportamientos de los mismos. Lo cual resulta de gran ayuda para los estudios posteriores.

En principio, se carga la base de datos de cada criptomoneda por separado donde se reflejan todas las variables que comprenden o describen a cada una: Date, close Price,

Volume y Market Cap (debemos recalcar que finalmente nos quedamos con el precio de cierre) y se explica cómo se ha planteado el modelo para el análisis de series temporales.

Nuestro conjunto de datos presentan observaciones diarias respecto al precio con el que cierra a diario cada una de las criptomonedas a estudiar, por esa razón se puede decir que se está en presencia de una serie de tiempo econométrica, donde los modelos que mejor se adaptan a este tipo de series son los modelos ARMA y ARIMA, los cuales fueron estudiados a profundidad en el capítulo 2.

Es importante recordar que con estos modelos que se presentan a continuación, se busca predecir el comportamiento futuro y el precio de cada una de las criptomonedas en estudio, que a nuestro criterio son las que tienen mayor influencia actualmente en el mercado, con el propósito de proporcionar una herramienta que ayude a mejorar la toma de decisiones de inversión.

Como se menciona en el capítulo anterior, se utiliza el software estadístico R en su versión RStudio, en el cual se hace uso, o en caso de no tenerlas, se instalan las librerías/paquetes necesarias para la realización de este estudio, las cuales son: “*readr*”, “*tidyverse*”, “*tseries*”, “*forecast*”, “*ggplot2*”, “*dplyr*”, “*plotly*”, “*astsa*”, “*stats*”, “*lubridate*” y “*lmtest*”.

4.1 Estadísticos descriptivos para el Bitcoin (BTC)

Price			Volume			MarketCap		
Min.	:	178.1	Min.	:	2.858e+06	Min.	:	2.444e+09
1st Qu.:		433.5	1st Qu.:		3.570e+07	1st Qu.:		6.341e+09
Median :		915.8	Median :		1.546e+08	Median :		1.462e+10
Mean :		3435.0	Mean :		4.459e+09	Mean :		5.874e+10
3rd Qu.:		6529.7	3rd Qu.:		5.500e+09	3rd Qu.:		1.125e+11
Max.	:	19497.4	Max.	:	4.511e+10	Max.	:	3.265e+11

Figura 4.1: Cuadro estadísticos descriptivos del BTC.

En la Figura 4.1 podemos observar los estadísticos descriptivos de cada una de las 3 variables referentes al bitcoin en la base de datos, la primera llamada “Price” se refiere al precio diario de la moneda bitcoin desde el 27 de Diciembre del año 2013 hasta el 24 de enero del año 2020, el precio mínimo mostrado en la tabla es de 178.1 \$ el cual tuvo lugar el 14 de enero del año 2015, la mediana del precio es de 915.8\$ y el precio promedio es de 3435\$, el valor máximo del precio del Bitcoin es de 19497.4\$ y se alcanzó el 16 de Diciembre del año 2017.



Figura 4.2: Comportamiento del BTC.

La figura 4.2 nos muestra el comportamiento del precio del Bitcoin en el periodo en estudio, en el cual podemos notar una gran volatilidad, es decir, que el precio cambia rápidamente en el tiempo, observando el pico alcanzado con el precio máximo en Diciembre del año 2017.

Para verificar si la serie estacionaria o no construimos los correlogramas.

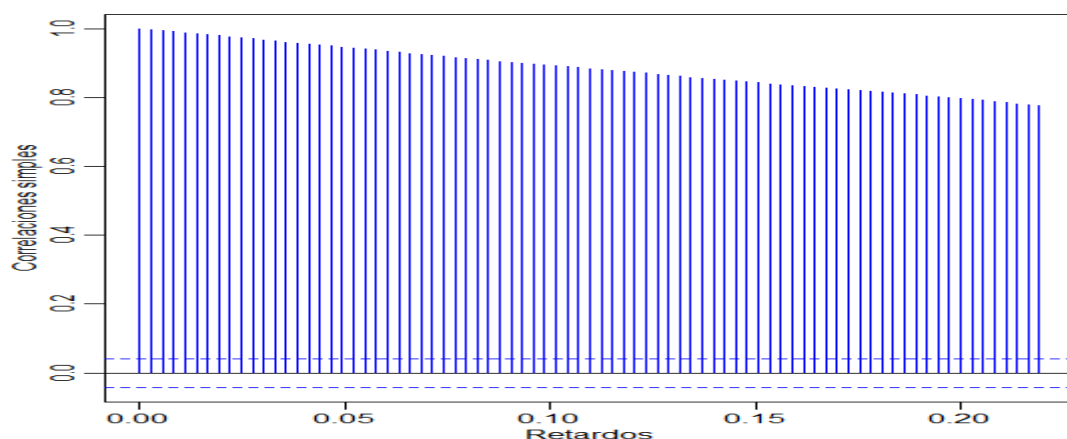


Figura 4.3: Correlograma simple del BTC.

A través de la figura 4.3 apreciamos que el valor (r_k) de la función de autocorrelación, las cuales todas son significativas y decaen lentamente, esto indica la presencia de una serie sin tendencia, no estacionaria, pero cuyas observaciones sucesivas están correlacionadas positivamente, es decir, una serie en la que a una observación por encima de la media, le suele seguir otra o más observaciones por encima de la media. El correlograma es de la siguiente forma: un valor relativamente alto de r_1 , seguido de algunos coeficientes r_k distintos de cero, pero cada vez más pequeños y valores de r_k aproximadamente cero para k grande.

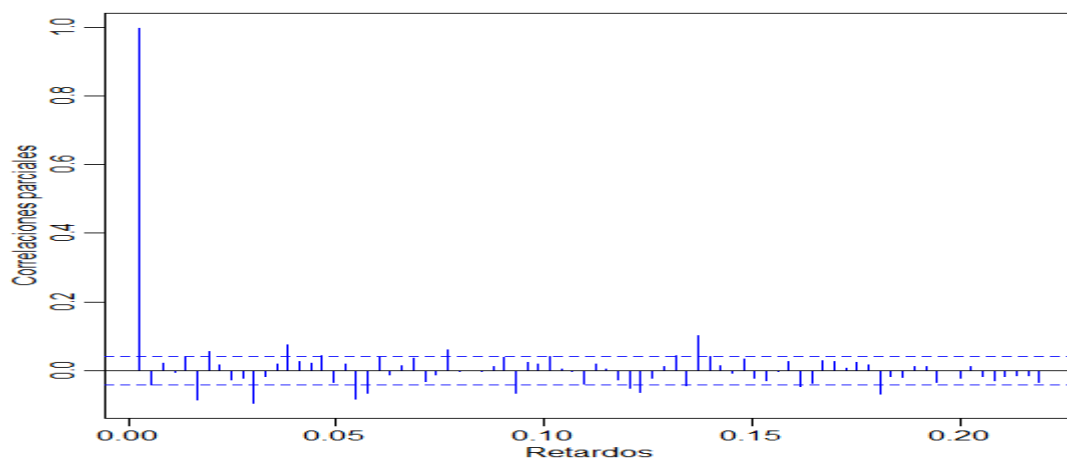


Figura 4.4: Correlograma parcial del BTC.

Las PACF de la figura 4.4 proporcionan la relación directa existente entre observaciones separadas por k retardos, también observamos algunas correlaciones parciales significativas.

El test de DickeyFuller aumentado nos proporciona una prueba formal para verificar la estacionariedad de la serie, a pesar de que a través de los correlogramas pudimos determinar de manera visual que la serie no es estacionaria, las hipótesis a probar son las siguientes:

H0: La serie es no estacionaria: tiene raíz unitaria.

H1: La serie es estacionaria: no tiene raíz unitaria.

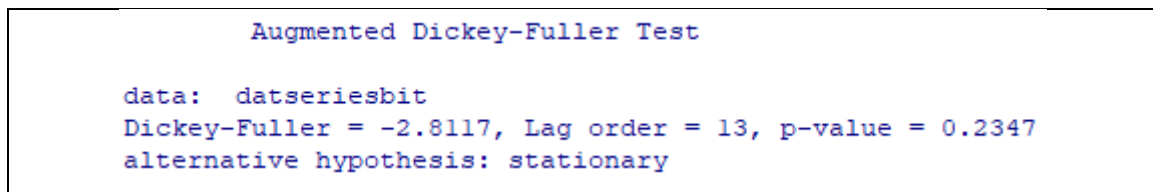


Figura 4.5: Prueba de Dickey-Fuller del BTC.

En la figura 4.5, observamos que el p-value es igual a 0.2347 y es mayor que el nivel de significancia de 0.05, por lo tanto, no rechazamos la hipótesis nula, que nos indica que efectivamente la serie es no estacionaria.

Diferenciamos una vez la serie para tratar de lograr que sea estacionaria.

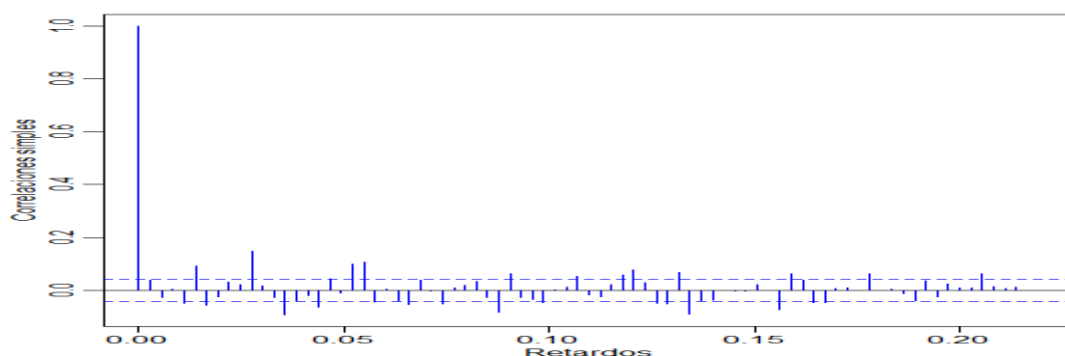


Figura 4.6: Correlograma simple serie diferenciada del BTC.

En la figura 4.6 se pueden observar algunas correlaciones simples significativas, lo cual nos puede dar indicio de que modelo ajustar si finalmente se ha logrado la estacionariedad de la serie, pero antes de esto, observamos el correlograma parcial en la siguiente figura:

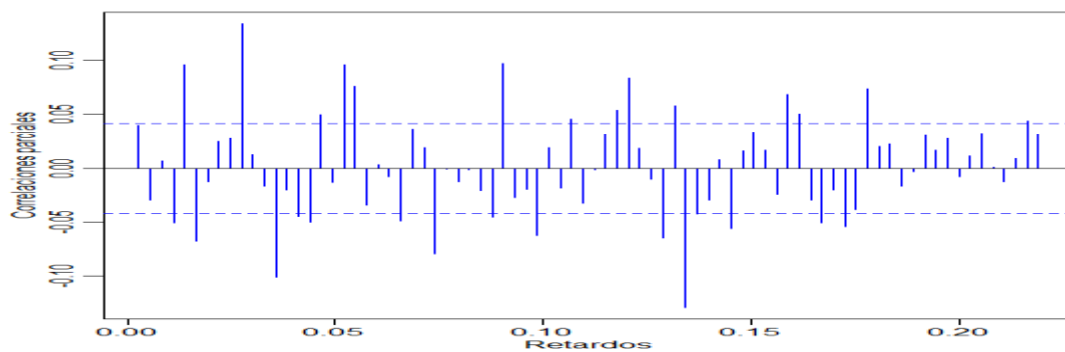


Figura 4.7: Correlograma parcial serie diferenciada del BTC.

En la figura 4.7 observamos algunas correlaciones parciales también significativas.

Aplicamos la prueba de DickeyFuller a la serie diferenciada

```
Augmented Dickey-Fuller Test

data: datseriesbitd
Dickey-Fuller = -12.507, Lag order = 13, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(datseriesbitd, alternative = "stationary") :
  p-value smaller than printed p-value
```

Figura 4.8: Prueba de Dickey-Fuller serie diferenciada del BTC.

En la figura 4.8, el p valor es igual a 0.01 el cual es menor que un nivel de significación del 5% por lo que rechazamos H_0 y concluimos que la serie es estacionaria.

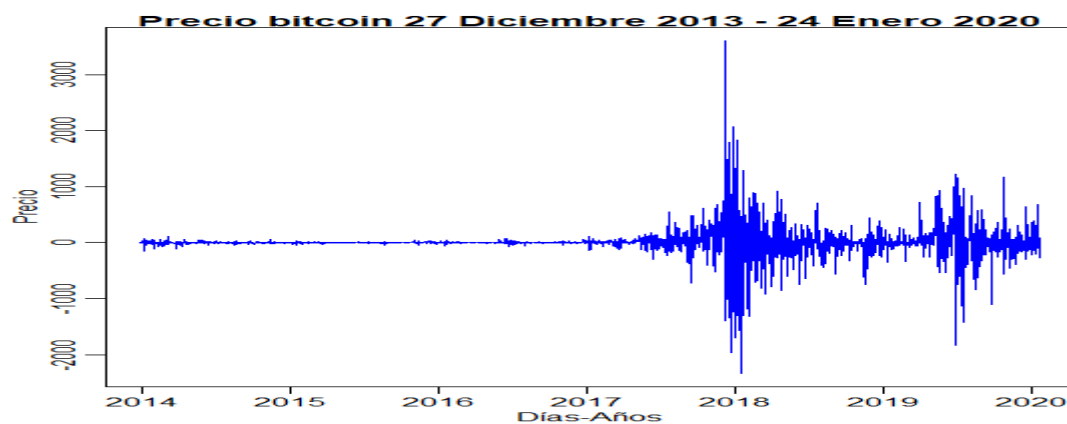


Figura 4.9: Serie BTC diferenciada.

En la figura 4.9, la serie diferenciada presenta mayores cambios del año 2017 en adelante pero se mantiene relativamente constante alrededor del cero.

Ajustamos el modelo: La volatilidad o cambio brusco en los precios del bitcoin, hace la tarea un poco complicada a la hora de identificar el modelo a través de los correlogramas, ya que siempre habrán correlaciones significativas.

El comando `auto.arima` del software R recomienda usar un modelo ARIMA (1, 1, 1), es decir 1 para el coeficiente autorregresivo (AR), 1 para el coeficiente de diferenciación (I) y 1 para el coeficiente de media móvil (MA).

```
Series: datseriesbit
ARIMA(1,1,1)
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
  Q), period = S), xreg = constant, transform.pars = trans, fixed = fixed,
  optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
      ar1      ma1  constant
    -0.7196  0.7749    3.5303
s.e.    0.0774  0.0697    5.7104

sigma^2 estimated as 67886:  log likelihood = -15485.49,  aic = 30978.99

$degrees_of_freedom
[1] 2215
```

```

$tttable
      Estimate      SE t.value p.value
arl    -0.7196 0.0774  -9.3004  0.0000
mal     0.7749 0.0697  11.1204  0.0000
constant 3.5303 5.7104   0.6182  0.5365

$AIC
[1] 13.96708

$AICc
[1] 13.96709

$BIC
[1] 13.97737

```

Figura 4.10: Coeficientes propuestos modelo ARIMA del B

Observando la figura 4.10 el modelo propuesto es $\widehat{Y}_{dt} = -0.7196Y_{t-1} + 0.7749\mathcal{E}_{t-1} + \mathcal{E}$

Notamos que los parámetros del modelo propuesto son significativos, ya que los p valores (última columna [p.value](#)) son menores que el nivel de significación alpha del 5%.

Si los parámetros de orden del modelo y la estructura se especifican correctamente, no esperaríamos que existieran autorrelaciones significativas.

Hemos ajustado a un modelo que puede producir un pronóstico, pero comenzamos examinando los gráficos ACF y PACF para los residuos del modelo. Si los parámetros de orden del modelo y la estructura se especifican correctamente, no esperaríamos que existieran autocorrelaciones significativas.

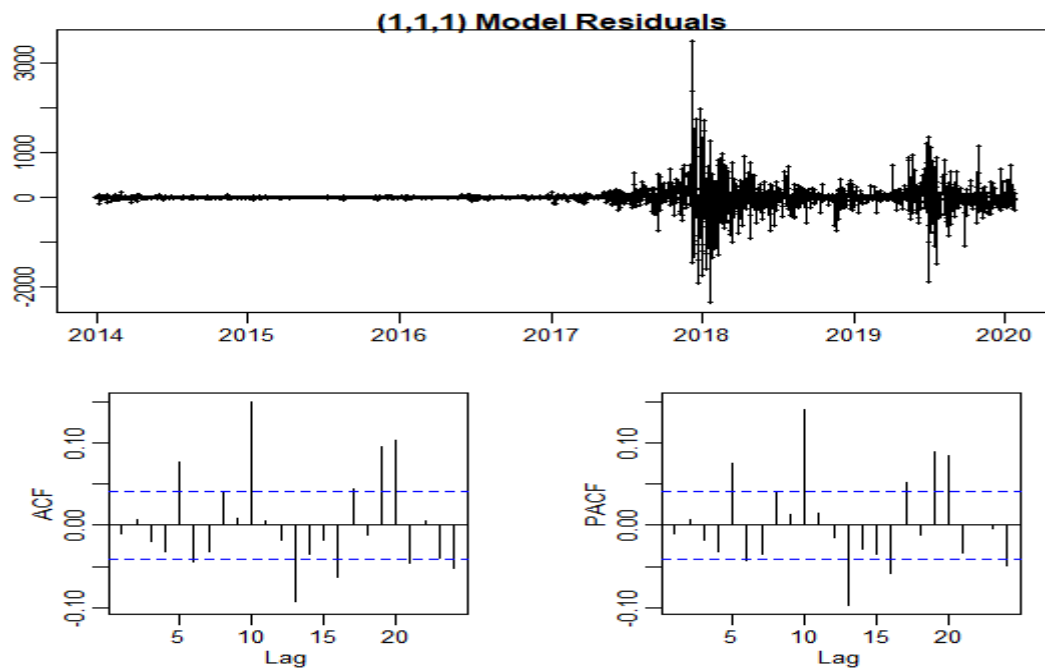


Figura 4.11: Diagramas ACF y PACF del BTC.

Hay un claro patrón presente en ACF que se repite en el retardo 5. Esto sugiere que nuestro modelo puede ser mejor con una especificación diferente como $p=5$ o $q=5$, más sin embargo esta aplicación no mejoró nuestro modelo, por lo que nos quedamos con el planteado inicialmente.

Podemos especificar el horizonte de pronóstico h periodos por delante para que se realicen las predicciones, y usar el modelo ajustado para generar dichas predicciones:

```

$pred
Time Series:
Start = c(2020, 25)
End = c(2020, 54)
Frequency = 365
 [1] 8394.472 8409.212 8404.676 8414.010 8413.365 8419.900 8421.268 8426.354
 [9] 8428.765 8433.101 8436.051 8439.999 8443.229 8446.975 8450.350 8453.992
[17] 8457.442 8461.030 8464.519 8468.079 8471.588 8475.134 8478.653 8482.191
[25] 8485.716 8489.250 8492.778 8496.310 8499.839 8503.370

$se
Time Series:
Start = c(2020, 25)
End = c(2020, 54)
Frequency = 365
 [1] 260.5502 378.8096 462.0685 536.2140 598.8728 657.1523 709.6139
 [8] 759.1645 805.1922 849.0557 890.5366 930.3246 968.3721 1005.0546
[15] 1040.3929 1074.6057 1107.7372 1139.9237 1171.2137 1201.6979 1231.4216
[22] 1260.4488 1288.8195 1316.5810 1343.7677 1370.4161 1396.5554 1422.2149
[29] 1447.4192 1472.1924

```

Figura 4.12: Tabla de predicciones del BTC (25 de enero de 2020-23 de febrero 2020)

En la figura 4.12 podemos observar los pronósticos un mes adelante por nuestro modelo, para el 25 de enero pronostica el precio del bitcoin igual 8394.472\$, es importante recalcar que los modelos ARIMA se usan para pronósticos a corto plazo, ya que a lo largo del tiempo tienden a estabilizarse en un valor constante por lo que se vuelve contraproducente utilizarlos a largo plazo.



Figura 4.13: Gráfico pronóstico precios del Bitcoin.

En la figura 4.13 observamos los pronósticos en color azul para el precio del bitcoin, la banda en color gris indican la amplitud en las predicciones.

4.2. Estadísticos descriptivos para el Dash (DSH)

Price		Volume		MarketCap	
Min.	: 0.3149	Min.	:9.604e+03	Min.	:1.047e+06
1st Qu.:	3.3800	1st Qu.:	1.560e+05	1st Qu.:	1.867e+07
Median :	16.5800	Median :	3.716e+06	Median :	1.176e+08
Mean :	117.9744	Mean :	8.940e+07	Mean :	9.392e+08
3rd Qu.:	153.4300	3rd Qu.:	1.541e+08	3rd Qu.:	1.280e+09
Max.	:1550.8500	Max.	:2.737e+09	Max.	:1.204e+10

Figura 4.14: Cuadro estadísticos descriptivos DSH.

En la figura 4.14 podemos observar los estadísticos descriptivos por cada una de las 3 variables de la moneda Dash , la primera llamada “Price” se refiere al precio diario de la moneda dash desde el 14 de Febrero del año 2014 hasta el 24 de enero del año 2020, el precio mínimo mostrado en la tabla es de 0.3149 \$ alcanzado el 15 de febrero del año 2014, con una mediana de 16.58\$ y un precio promedio de 117.98\$, el valor máximo del precio del dash es de 1550.85\$ y se alcanzó el 20 de Diciembre del año 2017.

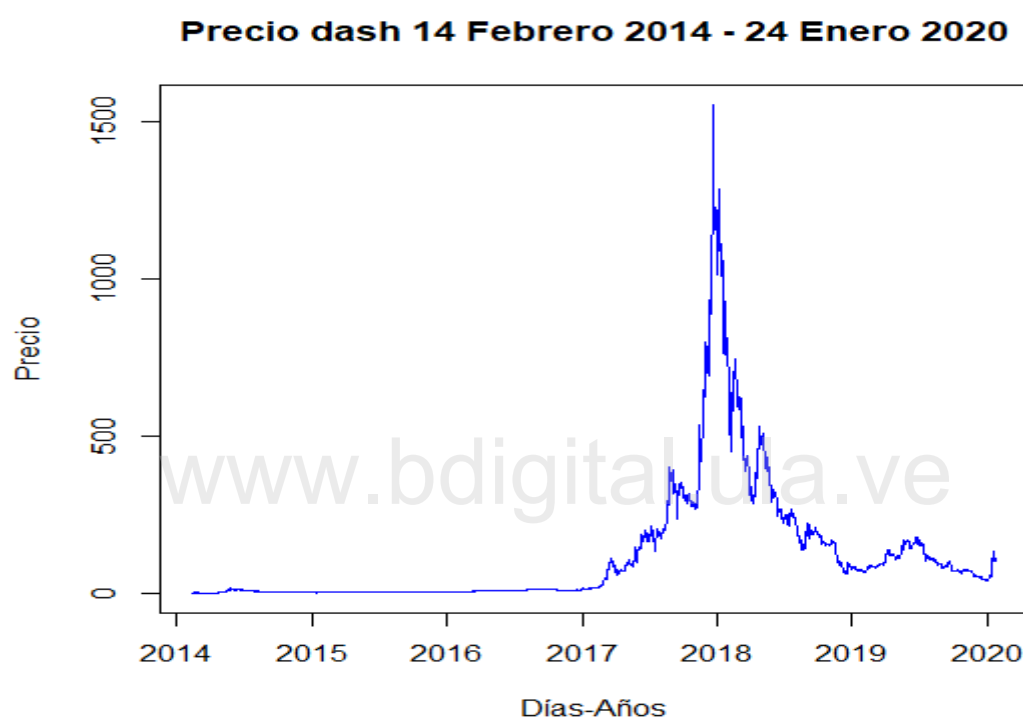


Figura 4.15: Comportamiento del DSH.

La figura 4.15 nos muestra el comportamiento del precio del dash en el periodo en estudio, del cual podemos notar una gran volatilidad, es decir, que el precio cambia rápidamente en el tiempo, con un pico claramente pronunciado en Diciembre del año 2017 (precio máximo alcanzado).

Para verificar si la serie estacionaria o no construimos los correlogramas.

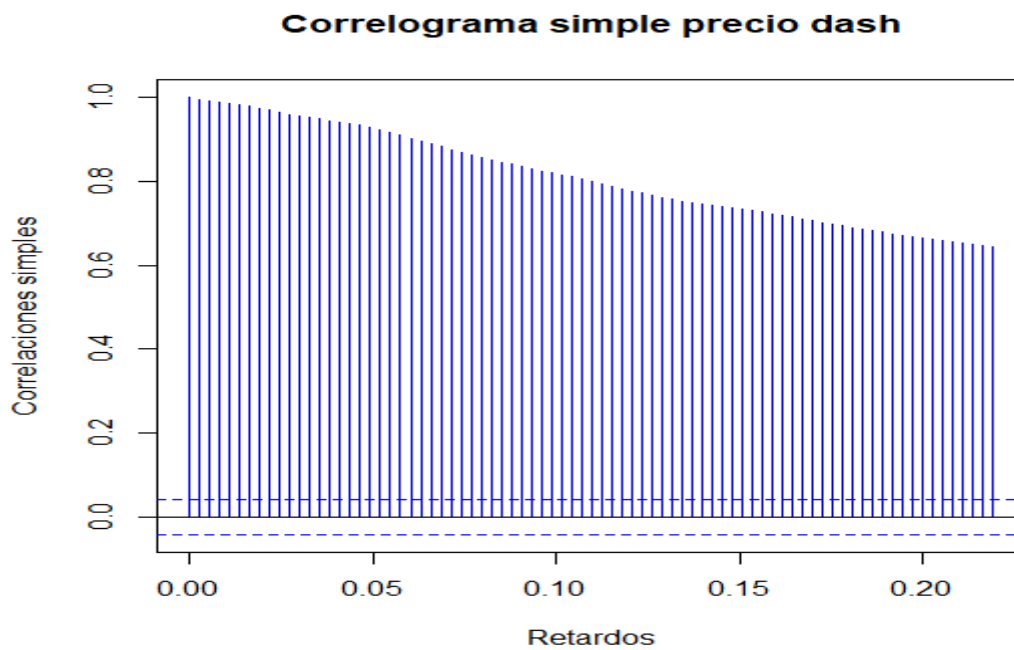


Figura 4.16: Correlograma simple del DSH.

A través de la figura 4.16 del correlograma simple apreciamos que el valor de la función de autocorrelación decae lentamente, y todas estas correlaciones son significativas, lo que muestra la no estacionariedad de la serie.

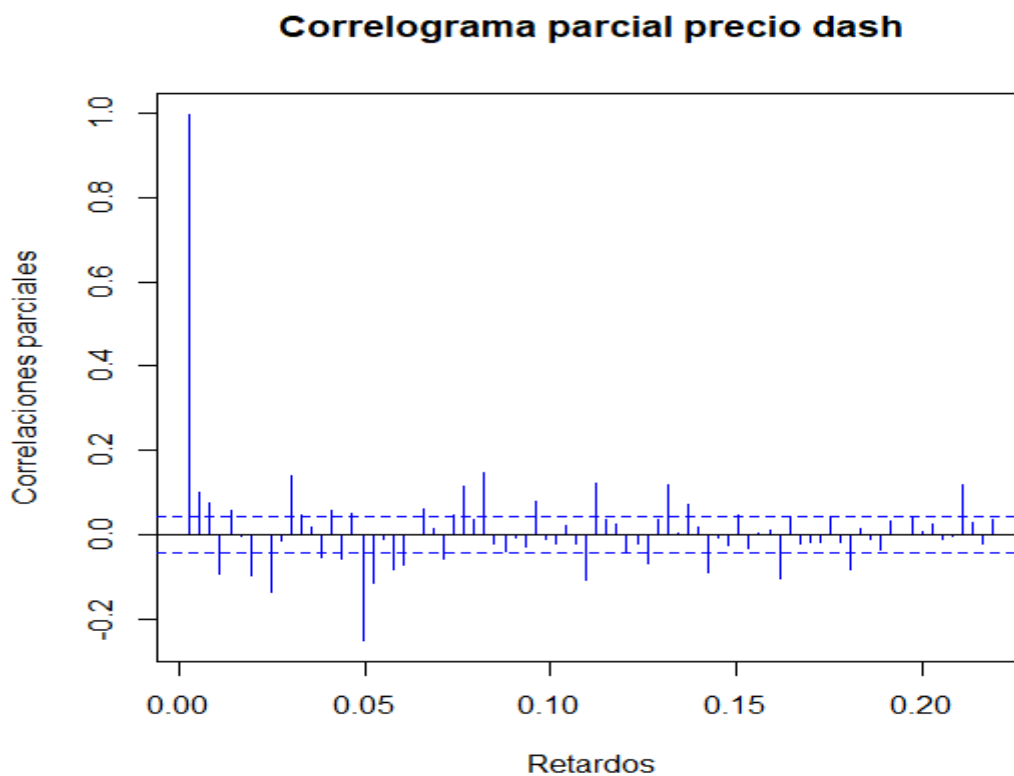


Figura 4.17: Correlograma parcial del DSH.

Las PACF de la figura 4.17 proporcionan la relación directa existente entre observaciones separadas por k retardos, observamos claramente correlaciones parciales significativas. Aplicando la prueba de Dickey Fuller a continuación corroboramos lo observado en los gráficos.

El test de DickeyFuller

H0: La serie es no estacionaria: tiene raíz unitaria.

H1: La serie es estacionaria: no tiene raíz unitaria.

```

Augmented Dickey-Fuller Test

data:  datseriesdash
Dickey-Fuller = -2.0278, Lag order = 12, p-value = 0.5666
alternative hypothesis: stationary

```

Figura 4.18: Prueba de Dickey-Fuller del DSH.

Observando la figura 4.18 con un nivel de significación del 5%, como el p-value es igual a 0.5666 y es mayor que el nivel de significancia de 0.05, no rechazamos la hipótesis nula, que nos indica que efectivamente la serie es no estacionaria.

Diferenciamos una vez la serie para tratar de lograr que sea estacionaria.

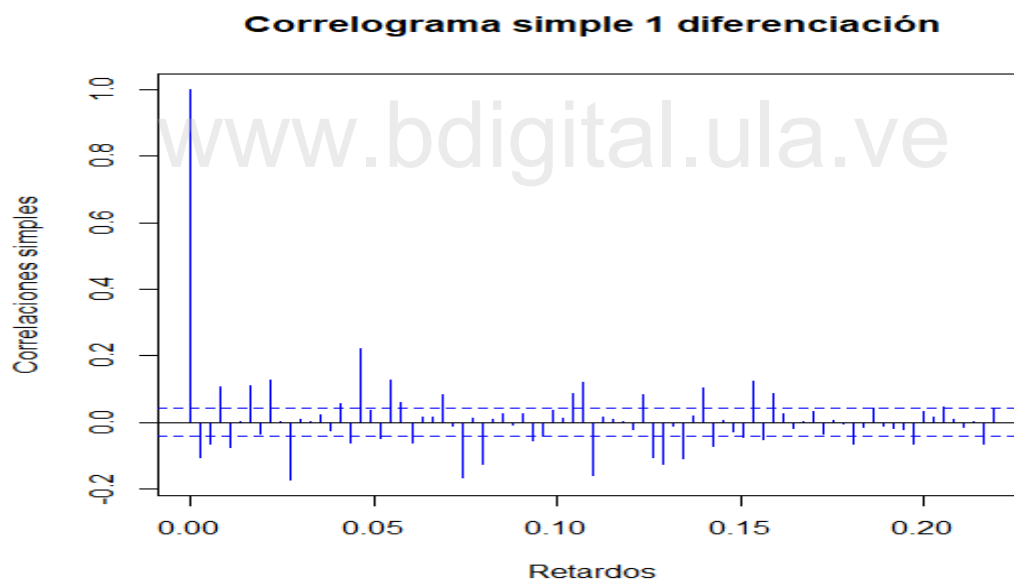


Figura 4.19: Correlograma simple serie diferenciada del DSH.

En el correlograma simple de la figura 4.19, observamos aún algunas correlaciones significativas, veamos que ocurre con el correlograma parcial a continuación:

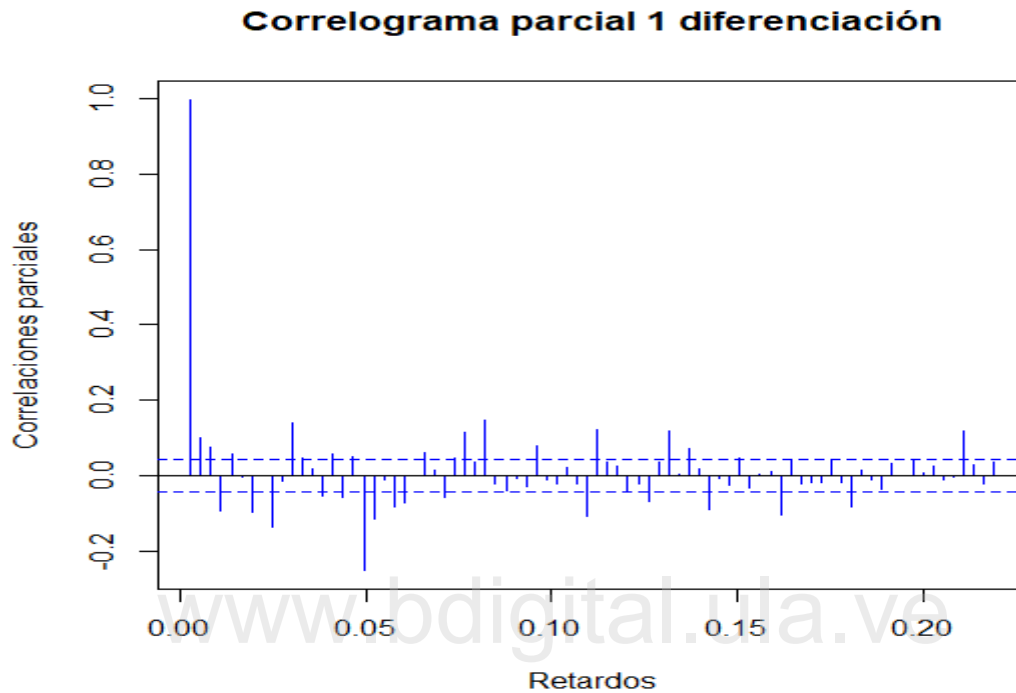


Figura 4.20: Correlograma parcial serie diferenciada del DHS.

En el correlograma parcial de la figura 4.20, observamos aún algunas correlaciones parciales significativas que nos pudieran ayudar a identificar el modelo adecuado a ajustar, pero primero debemos verificar si la serie diferenciada es estacionaria.

Aplicamos la prueba de DickeyFuller a la serie diferenciada


```
Augmented Dickey-Fuller Test

data:  datseriesdashd
Dickey-Fuller = -12.849, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(datseriesdashd, alternative = "stationary") :
  p-value smaller than printed p-value
```

Figura 4.21: Prueba de Dickey-Fuller serie diferenciada del DSH.

En la figura 4.21 vemos que el p valor es igual a 0.01, el cual es menor que un nivel de significación del 5% por lo que rechazamos H_0 y concluimos que la serie es estacionaria.

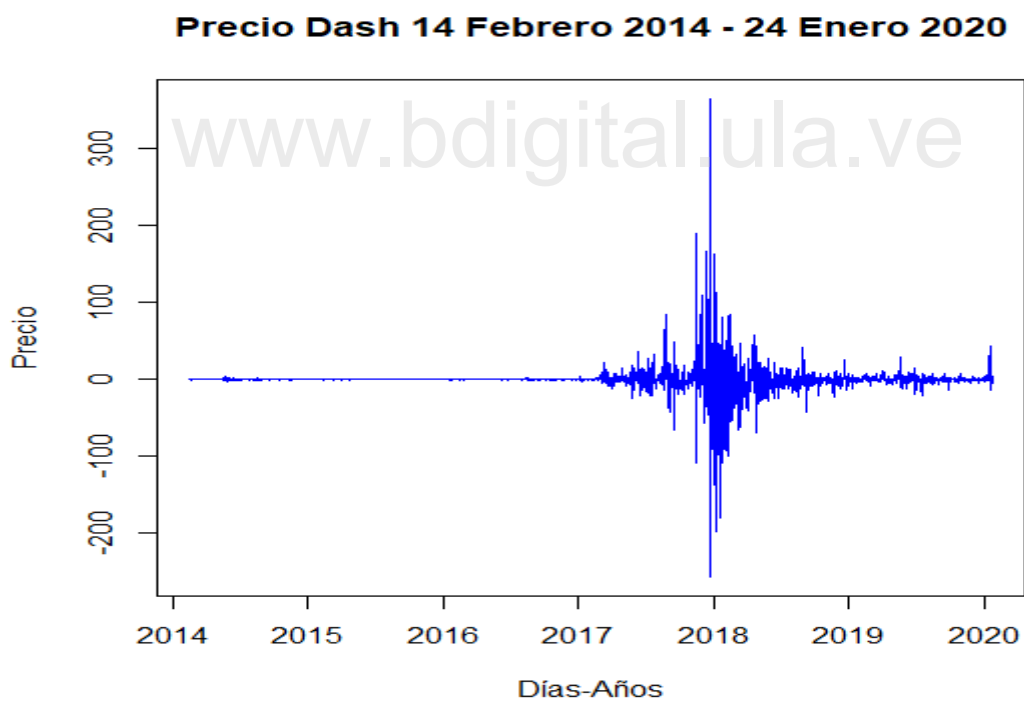


Figura 4.22: Serie DSH diferenciada.

Se muestra en la figura 4.22 que la serie diferenciada presenta mayores cambios del año 2017 en adelante pero se mantiene relativamente constante alrededor del cero. A continuación ajustamos el modelo:

El comando `auto.arima` del software R recomienda usar un modelo ARIMA (5, 1, 2), es decir 5 para el coeficiente autorregresivo (AR), 1 para el coeficiente de diferenciación (I) y 2 para el coeficiente de media móvil (MA).

```
Series: datseriesdash
ARIMA(5,1,2)

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1      ma2
      1.3338  -0.7062  0.1077  -0.2211  0.1646  -1.4584  0.8046
s.e.    0.0904   0.0458  0.0386   0.0378  0.0280   0.0933  0.0404

sigma^2 estimated as 343.4:  log likelihood=-9406.71
AIC=18829.42   AICc=18829.49   BIC=18874.88

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.03911418 18.49768 5.74856 -0.02919207 4.603689 0.03353266
              ACF1
Training set 0.003119854
```

```
$degrees_of_freedom
[1] 2161

$table
      Estimate      SE  t.value p.value
ar1         1.3330 0.0878  15.1841  0.0000
ar2        -0.7060 0.0456 -15.4778  0.0000
ar3         0.1076 0.0386   2.7891  0.0053
ar4        -0.2212 0.0376  -5.8759  0.0000
ar5         0.1647 0.0276   5.9768  0.0000
ma1        -1.4575 0.0905 -16.1129  0.0000
ma2         0.8042 0.0397  20.2796  0.0000
constant    0.0420 0.4278   0.0981  0.9218

$AIC
[1] 8.68207

$AICc
[1] 8.682101

$BIC
[1] 8.705647
```

Figura 4.23: Coeficientes propuestos modelo ARIMA del DSH

En la figura 4.23 observamos el valor de los coeficientes y su significación estadística.

El modelo propuesto es $\widehat{Y}_{dt} = 1.333Y_{t-1} - 0.7060Y_{t-2} + 0.1076Y_{t-3} - 0.2212Y_{t-4} + 0.1647Y_{t-5} - 1.4575\mathcal{E}_{t-1} + 0.8042\mathcal{E}_{t-2} + \mathcal{E}$

Hemos ajustado a un modelo que puede producir un pronóstico, pero comenzamos examinando los gráficos ACF y PACF para los residuos del modelo. Si los parámetros de orden del modelo y la estructura se especifican correctamente, no esperaríamos que existieran autocorrelaciones significativas.

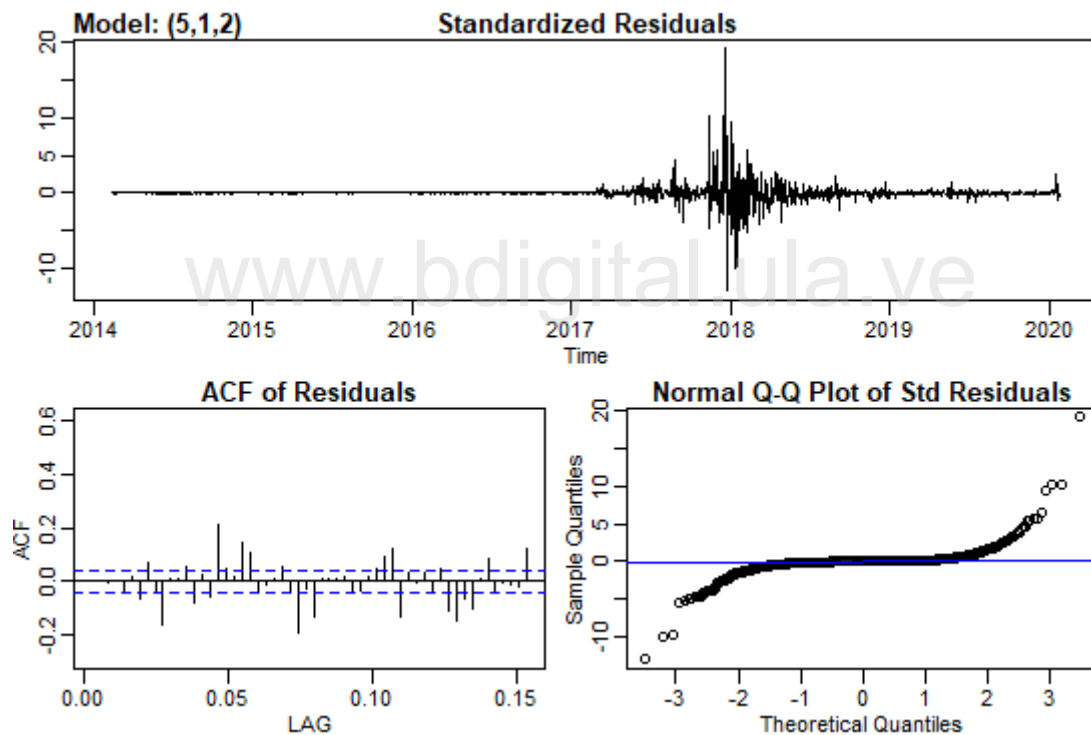


Figura 4.24: Diagramas ACF y PACF del DSH.

Finalmente en la Figura 4.24 podemos especificar el horizonte de pronóstico h periodos por delante para que se realicen las predicciones, y usar el modelo ajustado para generar dichas predicciones:

```
$pred
Time Series:
Start = c(2020, 25)
End = c(2020, 54)
Frequency = 365
[1] 100.02333 95.74126 92.09036 90.63378 90.48700 91.49418 92.89919
[8] 93.77940 93.87521 93.29929 92.42737 91.73232 91.49675 91.73627
[15] 92.25854 92.78386 93.09227 93.11046 92.91095 92.64873 92.47384
[22] 92.46468 92.60834 92.82613 93.02303 93.13392 93.14639 93.09494
[29] 93.03532 93.01494

$se
Time Series:
Start = c(2020, 25)
End = c(2020, 54)
Frequency = 365
[1] 18.50191 24.59071 28.77598 33.36790 37.01989 40.70943 44.86223
[8] 49.13431 53.38590 57.37577 60.91963 64.05254 66.89187 69.57680
[15] 72.22605 74.89334 77.56765 80.19896 82.73254 85.13886 87.42281
[22] 89.61399 91.74938 93.85756 95.95012 98.02217 100.05971 102.04929
[29] 103.98499 105.87011
```

Figura 4.25: Tabla de predicciones del DSH (25 de enero de 2020-23 de febrero 2020)

En la figura 4.25 podemos observar los pronósticos un mes adelante por nuestro modelo, para el 25 de enero pronostica el precio del Dash es igual 100.02333\$.



Figura 4.26: Gráfico pronóstico precios del Dash.

En la figura 4.26 observamos los pronósticos en color azul con sus respectivas bandas de amplitud en la predicción en color gris.

4.3. Predicciones para el Ethereum (ETH)

www.bdigital.ula.ve

Price		Volume		MarketCap	
Min.	: 0.4348	Min.	:1.021e+05	Min.	:3.221e+07
1st Qu.:	11.3500	1st Qu.:	1.580e+07	1st Qu.:	9.483e+08
Median :	148.8700	Median :	1.186e+09	Median :	1.574e+10
Mean :	203.3840	Mean :	2.393e+09	Mean :	2.024e+10
3rd Qu.:	286.6125	3rd Qu.:	3.149e+09	3rd Qu.:	2.814e+10
Max.	:1396.4200	Max.	:1.866e+10	Max.	:1.354e+11

Figura 4.27: Cuadro estadísticos descriptivos del ETH.

En la figura 4.27 podemos observar los estadísticos descriptivos por cada una de las 3 variables referentes a la moneda Ethereum, la primera llamada “Price” se refiere al precio diario de la moneda Ethereum desde el 7 de Agosto del año 2015 hasta el 24 de enero del año 2020, el precio mínimo mostrado en la tabla es de 0.4348 \$ alcanzado el 20 de octubre del año 2015, con una mediana de 148.87\$ y un precio promedio de 203.3840\$, el valor máximo del precio del ethereum es de 1396.42\$ y se alcanzó el 13 de Enero del año 2018.



Figura 4.28: Comportamiento del ETH.

Este grafico nos muestra el comportamiento del precio del Ethereum en el periodo en estudio, del cual podemos notar una gran volatilidad, es decir, que el precio cambia rápidamente en el tiempo, y un pico pronunciado alcanzado en Enero del año 2018.

Para verificar si la serie estacionaria o no construimos los correlogramas a continuación:

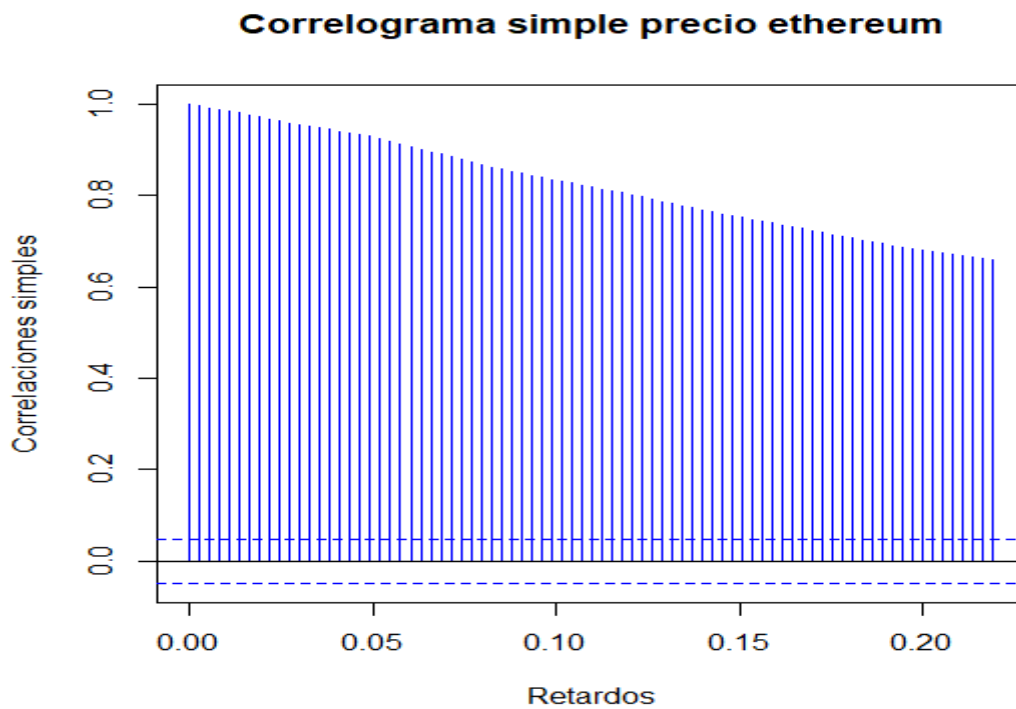


Figura 4.29: Correlograma simple del ETH.

En la figura 4.29 observamos que las correlaciones simples son todas no significativas y decrecientes en el tiempo, lo que nos indica la no estacionariedad de la serie.

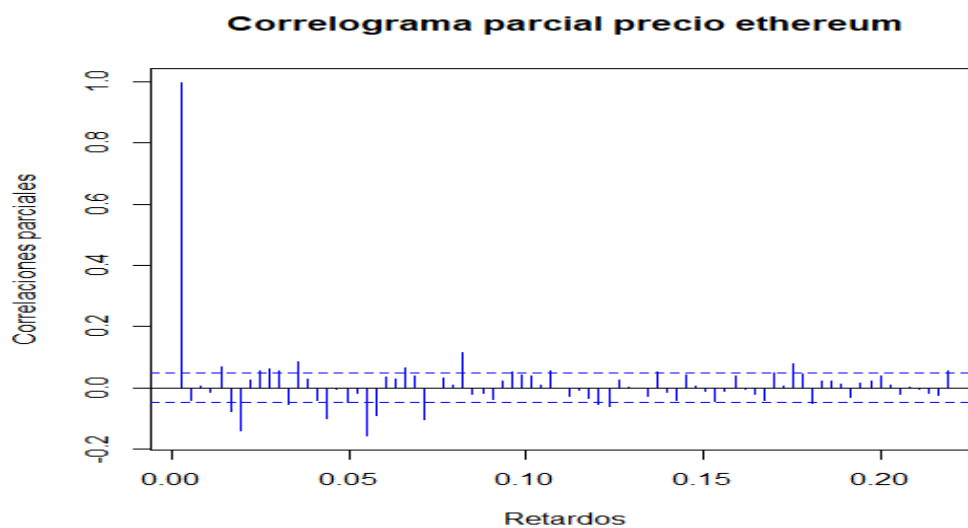


Figura 4.30: Correlograma parcial del ETH.

Las PACF proporcionan la relación directa existente entre observaciones separadas por k retardos, observamos algunas correlaciones parciales significativas. Aplicamos la prueba de Dickey Fuller para verificar lo mostrado en los gráficos de correlaciones anteriores.

El test de DickeyFuller

H0: La serie es no estacionaria: tiene raíz unitaria.

H1: La serie es estacionaria: no tiene raíz unitaria.

```
Augmented Dickey-Fuller Test  
  
data: datseriesethereum  
Dickey-Fuller = -1.7769, Lag order = 11, p-value = 0.6728  
alternative hypothesis: stationary
```

Figura 4.31: Prueba de Dickey-Fuller del ETH.

Observando la figura 4.31 con un nivel de significación del 5%, como el p-value es igual a 0.6728 y es mayor que el nivel de significancia de 0.05, no rechazamos la hipótesis nula, que nos indica que efectivamente la serie es no estacionaria.

Diferenciamos una vez la serie para tratar de lograr que sea estacionaria.

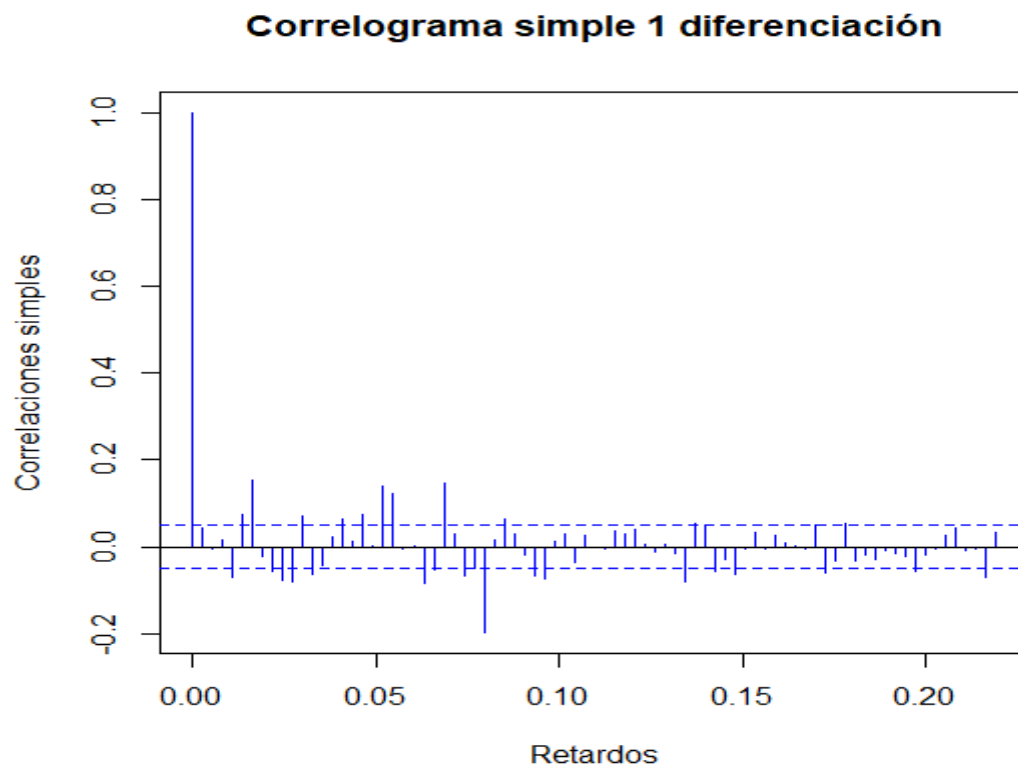


Figura 4.32: Correlograma simple serie diferenciada del ETH.

En la figura 4.32 se puede observar que entre retardos consecutivos, hay muchas correlaciones que se salen de las bandas lo que puede ser indicio de falta de estacionaridad.

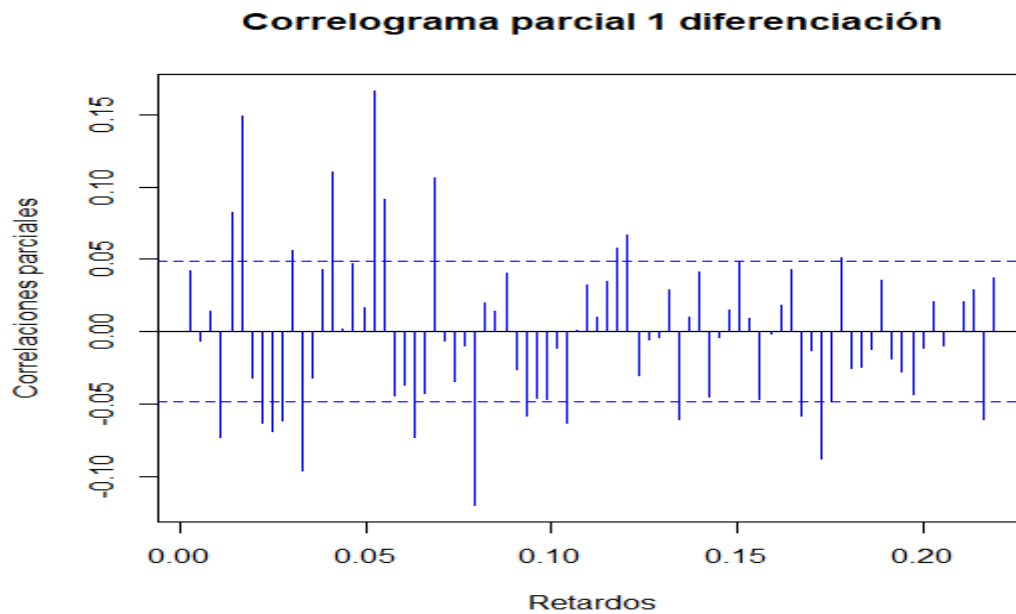


Figura 4.33: Correlograma parcial serie diferenciada del ETH.

Se observa en la figura 4.33 que entre k retardos consecutivos, hay muchas correlaciones que se salen de las bandas lo que puede ser indicio de falta de estacionaridad.

Aplicamos la prueba de DickeyFuller a la serie diferenciada.

```

Augmented Dickey-Fuller Test

data:  datseriesethereumd
Dickey-Fuller = -12.943, Lag order = 11, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(datseriesethereumd, alternative = "stationary") :
  p-value smaller than printed p-value

```

Figura 4.34: Segunda prueba de Dickey-Fuller del ETH.

En la figura 4.34 el p valor es igual a 0.01, el cual es menor que un nivel de significación del 5% por lo que rechazamos H_0 y concluimos que la serie es estacionaria.

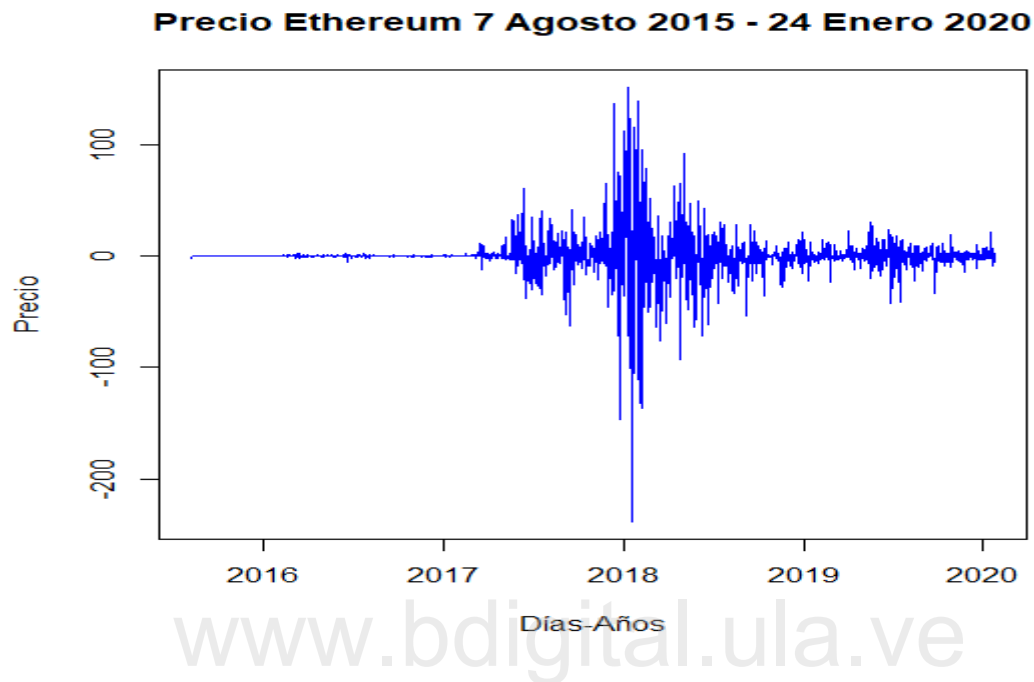


Figura 4.35: Serie ETH diferenciada.

En la figura 4.35, la serie diferenciada presenta mayores cambios entre el 2017 y el 2019.

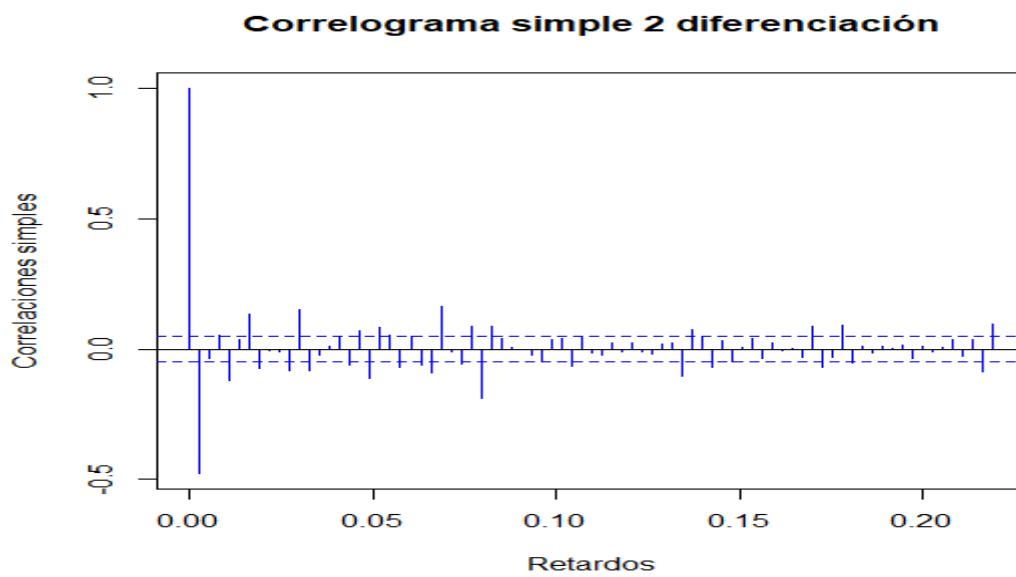


Figura 4.36: Segundo correlograma simple del ETH.

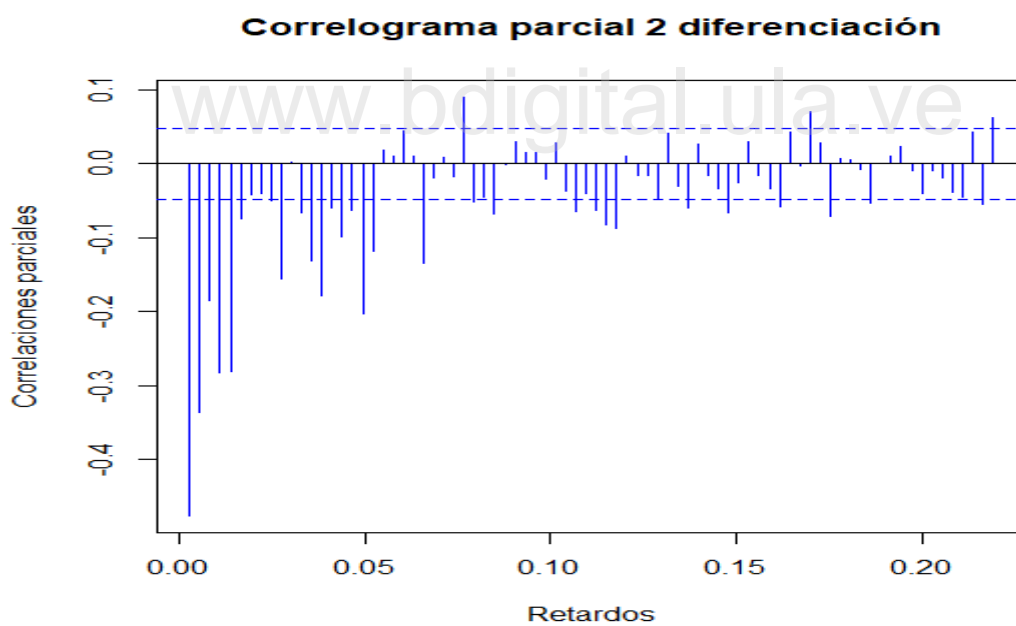


Figura 4.37: Segundo correlograma parcial del ETH.

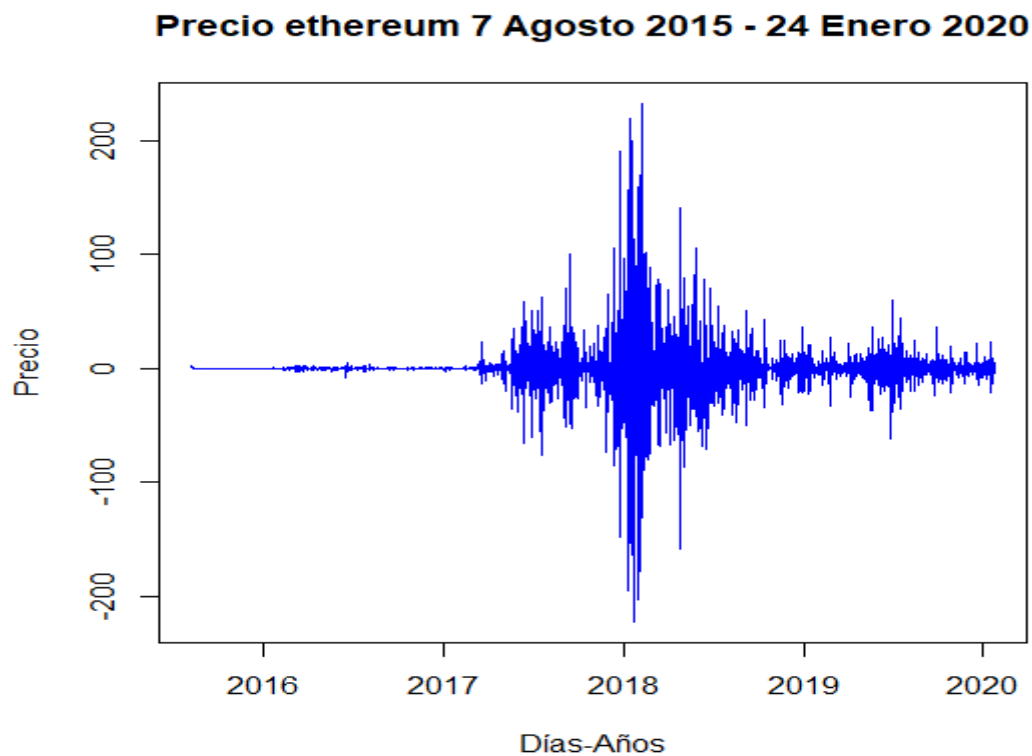


Figura 4.38: Gráfico de la serie segunda diferenciación del ETH.

El comando `auto.arima` del software R recomienda usar un modelo ARIMA (0, 2, 1), es decir 0 para el coeficiente autorregresivo (AR), no presencia de coeficiente AR, 1 para el coeficiente de diferenciación (I) y 1 para el coeficiente de media móvil (MA). A continuación ajustamos el modelo:

```

$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
      mal
      -1.0000
s.e.    0.0027

sigma^2 estimated as 393.6:  log likelihood = -7182.11,  aic = 14368.22

$degrees_of_freedom
[1] 1628

$tttable
      Estimate      SE    t.value p.value
mal         -1 0.0027 -372.6204      0

$AIC
[1] 8.820271

$AICc
[1] 8.820273

$BIC
[1] 8.826896

```

Figura 4.39: Coeficientes propuestos modelo ARIMA del ETH.

Observando la figura 4.39 el modelo propuesto es $\hat{Y}_{dt} = -\mathcal{E}_{t-1} + \mathcal{E}$ ya que sus parámetros son significativos.

Hemos ajustado un modelo que puede producir un pronóstico, pero comenzamos examinando los gráficos ACF y PACF para los residuos del modelo. Si los parámetros de orden del modelo y la estructura se especifican correctamente, no esperaríamos que existieran autocorrelaciones significativas.

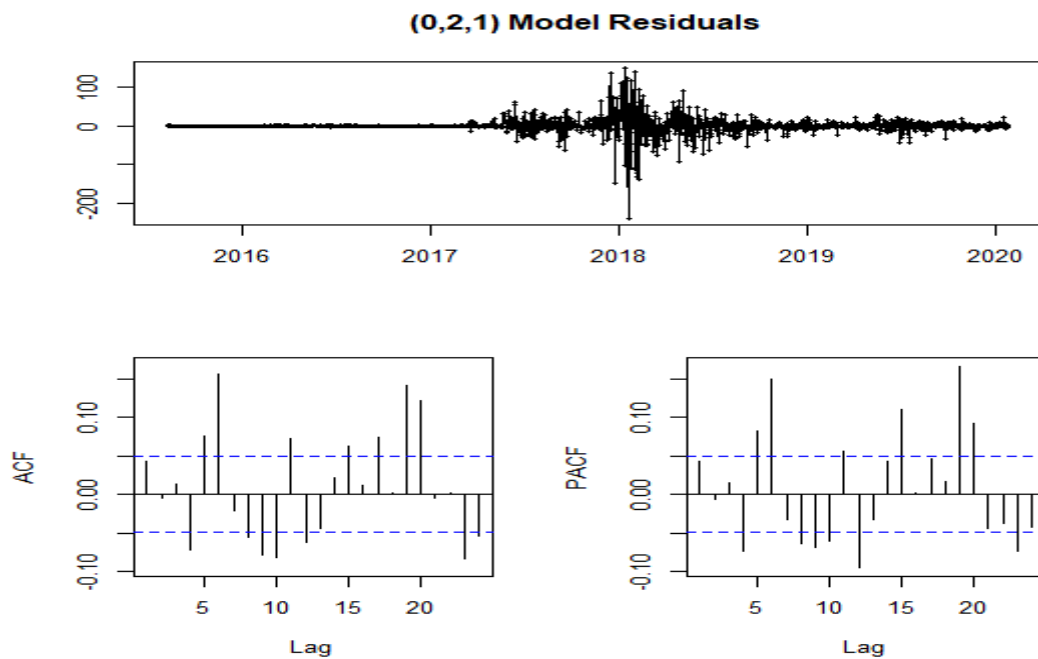


Figura 4.40: Diagramas ACF y PACF del ETH.

www.bdigital.ula.ve

Podemos especificar el horizonte de pronóstico h periodos por delante para que se realicen las predicciones, y usar el modelo ajustado para generar dichas predicciones:

```

$pred
Time Series:
Start = c(2020, 25)
End = c(2020, 54)
Frequency = 365
[1] 163.0283 163.1265 163.2248 163.3230 163.4213 163.5195 163.6178 163.7161
[9] 163.8143 163.9126 164.0108 164.1091 164.2073 164.3056 164.4039 164.5021
[17] 164.6004 164.6986 164.7969 164.8952 164.9934 165.0917 165.1899 165.2882
[25] 165.3864 165.4847 165.5830 165.6812 165.7795 165.8777

$se
Time Series:
Start = c(2020, 25)
End = c(2020, 54)
Frequency = 365
[1] 19.84607 28.07518 34.39547 39.72862 44.43154 48.68720 52.60425
[8] 56.25349 59.68404 62.93169 66.02344 68.98024 71.81877 74.55254
[15] 77.19267 79.74848 82.22782 84.63741 86.98305 89.26978 91.50202
[22] 93.68366 95.81816 97.90861 99.95776 101.96810 103.94190 105.88118
[29] 107.78782 109.66352

```

Figura 4.41: Tabla de predicciones del ETH (25 de enero de 2020-23 de febrero 2020)

En la figura 4.41 podemos observar los pronósticos un mes adelante por nuestro modelo, para el 25 de enero pronostica el precio del Ethereum igual a 163.0283\$.

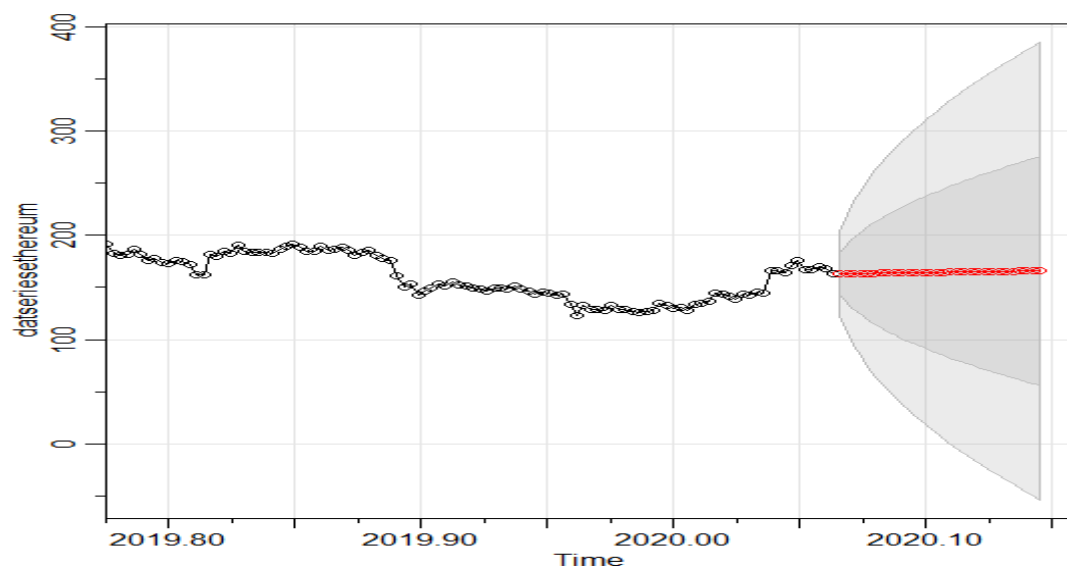


Figura 4.42: Gráfico pronóstico precios del Ethereum.

En la figura 4.42 observamos los pronósticos en color rojo con sus respectivas bandas de amplitud en la predicción en color gris.

4.4. Análisis de Reglas de Asociación

Para darle sentido a este análisis es necesario realizar algunas modificaciones a los datos con los que hemos venido trabajando, dado que los algoritmos de las reglas de asociación no están diseñados en principio para estudiar datos numéricos. Sin embargo, se crea una función determinista para poder asociar de forma correcta el comportamiento de las diferentes criptomonedas.

Se genera una nueva base de datos con 4 columnas, las cuales contienen los rendimientos diarios de cada una de las 3 criptomonedas en cuestión, además de la fecha de cada observación. La data resultante es de igual longitud a la criptomoneda con menor número de observaciones hasta la fecha que se obtuvieron los datos, la cual es el Ethereum(*ETH*), que cuenta con 1631 observaciones correspondientes, comenzando desde el 7 de Agosto del año 2015 hasta el 24 de enero del año 2020. Por lo que, utilizando esta columna como pivote, se aúnan las demás con su set de datos, siguiendo el mismo rango de fecha a la que se hace referencia en las líneas superiores. Nuestro nuevo conjunto de datos es muy similar al del principio de este capítulo, sólo que con las fechas y los valores de rendimientos de las diferentes criptodivisas en estudio. Por último, luego de tener ordenada la nueva matriz de datos, hacemos referencia a la columna (fecha), la cual se excluye del estudio ya que no es relevante para el desarrollo del mismo.

	BTC_Perform	ETH_Perform	DSH_Perform
1	0.358963	0.000000000	-7.3863636
2	-6.645683	-72.804151620	-5.5214724
3	1.563218	-6.826801181	0.9740260
4	-0.230119	0.933327824	-0.9646302
5	2.238439	51.034373730	2.5974026
6	-1.483043	14.018691590	-3.1645570
7	-0.863428	50.000000000	0.0000000
8	0.605877	0.000000000	-0.9803922
9	-1.554502	-7.650273224	-1.6501650
10	-1.162302	-7.100591716	-1.3422819
11	-0.205021	-23.566878980	-1.0204082
12	-18.179704	-9.166666667	-8.2474227
13	7.390563	15.596330280	-1.1235955
14	3.824775	15.873015870	5.3030303
15	-1.181219	-4.109589041	-0.7194245
16	-0.937352	-1.428571429	-2.8985507
17	-0.963583	-2.173913043	-1.4925373
18	-7.748608	-8.888888889	-9.8484848
19	5.282911	-7.317073171	6.3025210
20	1.904246	1.754385965	3.1620553
21	-0.469380	-0.862068966	-1.9157088
22	2.949682	3.478260870	1.5625000

Figura 4.43: Rendimiento de las diferentes criptomonedas.

Fuente: Elaboración Propia.

El siguiente paso para la realización de este análisis, se basa en eliminar aquellos valores donde el rendimiento de la criptodivisa es igual a cero, debido a que en este estudio un rendimiento nulo no es de importancia ya que no proporciona una respuesta de rendimiento positivo o negativo del activo. Por lo tanto, se procede a sustituir los valores iguales a 0 por NA's y posteriormente eliminar los mismos de la data.

Al tener limpia la base de datos (sólo valores positivos y negativos) se reduce la dimensionalidad de la misma a 1448 observaciones. Seguidamente, el próximo paso consta en transformar el conjunto de datos de la figura 4.43 en una matriz de “ceros” y “unos”, a través de una función indicadora, la cual recorre dicha matriz fila i por columna j sustituyendo por 0 o 1 el valor dependiendo de su condición, de la siguiente manera:

$$F(x) = \begin{cases} BTC_{Perform(i)} = 1 & \text{if } BTC_{Perform(i)} > 0 \\ BTC_{Perform(i)} = 0 & \text{if } BTC_{Perform(i)} < 0 \end{cases} \quad (4.1)$$

Aplicando la función 4.1 a la matriz de la figura 4.43, obtenemos la matriz binaria de transacciones que se muestra a continuación:

	BTC_Perform	ETH_Perform	DSH_Perform
1	0	0	0
2	1	0	1
3	0	1	0
4	1	1	1
5	0	1	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	1	1	0
11	1	1	1
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	1	0	1
17	1	1	1
18	0	0	0
19	1	1	1
20	0	1	0
21	1	1	1
22	0	0	0

Figura 4.44: Matriz de Transacciones.

Fuente: Elaboración Propia.

La matriz de transacciones de la figura 4.44 consta de 1448 observaciones, dado que el Ethereum tiene esa cantidad, representando la moneda con la menor cantidad de observaciones entre todas. A pesar de esto, la cantidad de observaciones para el estudio es considerablemente buena, las cuales describen bastante bien lo ocurrido en las fechas

de mayor volatilidad del mercado de criptoactivos, registrado desde mediados del año 2017 hasta principios de abril del 2018. La intención de cambiar los datos numéricos a factores del tipo “ceros” y “unos” es para lograr identificar de forma sencilla cuando el rendimiento de una criptomoneda baja o sube, de manera que si el rendimiento del Bitcoin, por ejemplo, baja, entonces colocamos un 0 en esa casilla, por el contrario, si el rendimiento sube en esa observación se sustituye dicho valor por un 1. Y así sucesivamente con todos los rendimientos de cada una de las distintas criptomonedas a estudiar, de tal manera que la data se encuentre apta para que el algoritmo A priori pueda establecer las diversas relaciones existentes entre ellas, teniendo como entrada las observaciones diarias y los valores que toma cada variable por observación, comparándolos entre sí y sacando a partir de acá las diversas reglas de asociación entre las variables.

En este punto, ya se puede aplicar el algoritmo A priori a la matriz de transacciones, para así extraer las reglas de asociación de las criptodivisas en cuestión.

La aplicación de este algoritmo, con 1448 observaciones, arroja un total de 9 reglas, seguidamente se muestran en la figura 4.45 las 5 más relevantes con su respectivo ítem:

```
> inspect(APRI)
  lhs                rhs      support confidence  lift count
[1] {BTC_Perform=0} => {ETH_Perform=0} 0.250000000 0.8080357 1.462545 362
[2] {ETH_Perform=1} => {BTC_Perform=1} 0.272790055 0.8586957 1.655648 395
[3] {BTC_Perform=0,
    ETH_Perform=0} => {DSH_Perform=0} 0.280386740 0.8731183 1.683456 406
[4] {BTC_Perform=0,
    DSH_Perform=0} => {ETH_Perform=0} 0.254143646 0.8000000 1.571777 368
[5] {ETH_Perform=1,
    DSH_Perform=1} => {BTC_Perform=1} 0.244475138 0.8009050 1.631098 354
```

Figura 4.45: Reglas de Asociación arrojadas por A priori.

Fuente: Elaboración Propia.

Relaciones 1 a 1 más relevantes:

2. $ETH_{Perform} = 1 \Rightarrow BTC_{Perform} = 1$, con un 86% de confianza y un 27% de soporte, es decir, que si Ethereum sube, entonces el Bitcoin también aumenta su rendimiento.

1. $BTC_{perform} = 0 \Rightarrow ETH = 0$, con un 81% de confianza y un 25% de soporte, es decir, que si Bitcoin baja, entonces Ethereum también baja.

Relaciones 2 a 1 más relevantes:

3. $BTC_{perform} = 0, ETH_{perform} = 0 \Rightarrow DSH_{perform} = 0$, con un 87% de confianza y un 28% de soporte, esto significa, que si el Bitcoin y Ethereum bajan, también lo hace el Dash.

4. $BTC_{perform} = 0, DSH_{perform} = 0 \Rightarrow ETH_{perform} = 0$, con un 80% de confianza y un 25% de soporte, lo que quiere decir que, cuando el Bitcoin y Dash bajan, el rendimiento del Ethereum tiende a bajar.

5. $ETH_{perform} = 1, DSH_{perform} = 1 \Rightarrow BTC_{perform} = 1$, con un 80% de confianza y un 24% de soporte, es decir, que si el Ethereum y Dash aumentan su rendimiento, el del Bitcoin también lo hace.

Estas son las reglas de asociación con mayor confianza y soporte entre las reglas arrojadas por el algoritmo Apriori. Por último, se muestran gráficamente en la figura 4.46 las reglas de asociación con el coeficiente de confianza y soporte respectivamente:

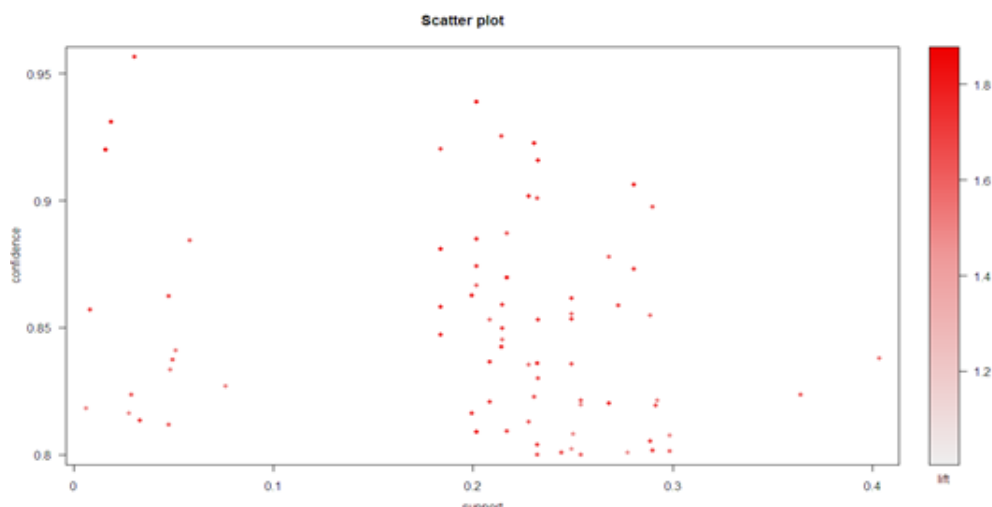


Figura 4.46: Gráfico de las reglas de asociación arrojadas A priori.

Finalmente, luego de haber desarrollado los modelos de pronósticos que mejor se ajustan a cada una de las criptomonedas en estudio, se procede a calcular el coeficiente de variación de las mismas, para determinar cuál de éstas presenta una mejor estabilidad en su rendimiento. Mediante la siguiente fórmula se calcula dicho parámetro:

$$coef_{var} = \left(\frac{SD_{var}}{\bar{x}_{var}} \right) * 100 \quad (4.2)$$

A partir de la ecuación 4.2, da como resultado lo que se muestra a continuación:

```
> coef_VAR_BTC
[1] 2033.897
> coef_VAR_ETH
[1] 1314.386
> coef_VAR_DSH
[1] 1626.527
```

Figura 4.47: Coeficientes de variaciones de las criptomonedas.

Fuente: Elaboración Propia.

Se puede apreciar en la figura 4.47 que el Ethereum resulta ser la criptomoneda con el menor índice de variación entre todas, encontrándose cerca el Dash. Por lo tanto, podemos decir que ésta es la criptodivisa que tiene el mejor comportamiento (el más estable) en cuanto a rendimiento se trata a lo largo de la serie.

4.5. Diseño de estrategias de trading

Existen diferentes plataformas que permiten operar en el mercado Fórex como: Prorealtime, Visual Chart, Sirix, MetaTrader, entre otras. Algunas gratuitas y otras pagas, la plataforma que sirvió de apoyo para la simulación por ser gratuita y su fácil forma de operar fue MetaTrader 4, además MetaTrader 4 ofrece la posibilidad de operar con una cuenta demo, así tener la posibilidad de hacer muchas pruebas antes de operar con dinero real, MetaTrader 4 es una plataforma gratuita destinada al comercio de mercados Fórex. En la siguiente imagen se puede observar la pantalla de inicio de la plataforma usada.

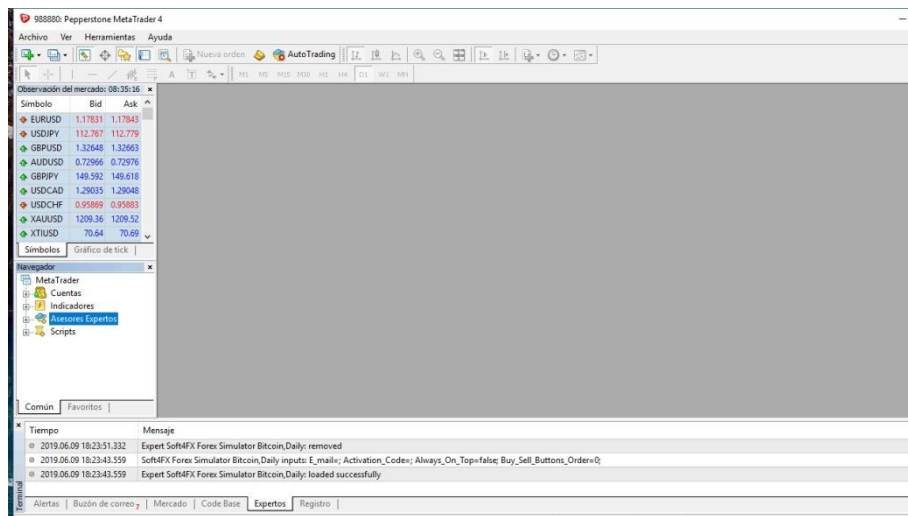


Figura4.48: Plataforma MetaTrader 4.

Fuente: MetaTrader 4 (Pepperstone).

MetaTrader 4 cuenta con la disponibilidad de agregar diferentes softwares de simulación de Fórex, para la simulación de las estrategias de trading del proyecto se contará con el software Fórex Simulator, el objetivo principal de este software es que permite hacer simulaciones en el mercado con una cuenta Demo, que puede llegar a ser de hasta 10.000\$, las restricciones que se encontraron en el uso de este software es que opera solo con 2 criptomonedas: Bitcoin y Ethereum, otra restricción es que solo permite tener 5 operaciones abiertas y permite un número limitado de 10 simulaciones por sesión. Es una herramienta útil para el aprendizaje y entrenamiento del trading, además es una herramienta sencilla en su forma de operar y con un alto parecido a la forma de operar con dinero real.

Antes de aplicar cualquier estrategia de trading es recomendable probarla en datos pasados, para ello existe una metodología llamada “backtesting” según (Rankia, 2017)

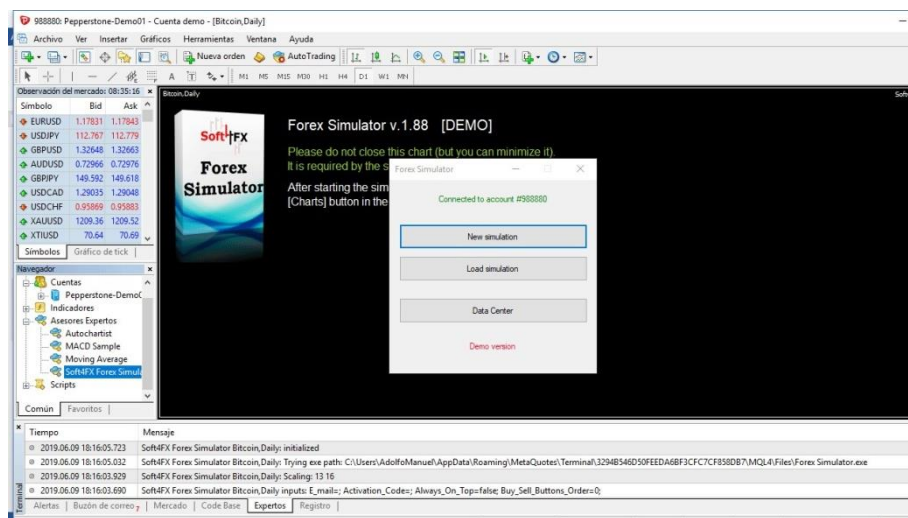


Figura 4.49: Software Forex Simulator.

Fuente: MetaTrader 4 (Pepperstone).

backtesting es:

“Se refiere a una metodología que se utiliza para verificar y diagnosticar la eficiencia y los resultados de un modelo en particular. Esta metodología intenta determinar si las conclusiones del modelo o los cálculos de las variables estimadas son acertados. Como su nombre lo indica, backtesting significa evaluar hacia atrás, por esta razón, el principal objetivo es saber cuáles habrían sido las conclusiones de un modelo en periodos anteriores y compararlo con datos históricos.”

La estrategia a seguir es verificar si las predicciones realizadas permiten saber cuál es la orden que genera mayor ganancia o beneficio.

- Si los valores pronosticados están por debajo del valor real de la fecha anterior se recomendaría hacer una venta.
- Si los valores pronosticados están por arriba del valor real de la fecha anterior se recomendaría hacer una compra.

Una de los beneficios que permite usar la metodología de backtesting es que se puede probar el caso contrario de la estrategia a seguir y determinar cuál sería la pérdida para el caso de que se tome la decisión incorrecta.

Para la simulación de la estrategia se generaron 2 pronósticos más para la criptomoneda bitcoin, una para el 28 de mayo de 2018 y otra para el 21 de octubre de 2018, con la finalidad de poder probar varias veces el modelo, los pronósticos se realizaron para los próximos 5 días.

Antes de mostrar los resultados de la simulación, es importante tener conocimiento de algunos términos propios del trading usados para operar.

Stop Loss: es el precio hasta el cual se está dispuesto a perder si es que existe esa pérdida. Si la operación alcanza esa cifra se cierra dicha operación.

TakeProfit: es el valor máximo de ganancias al cual puede llegar una operación, esto no implica que se pueda cerrar la operación antes de que el precio haya llegado a ese valor de TakeProfit.

Risk: es el dinero en \$ que se está arriesgando en la operación.

4.5.1. Primera simulación de la estrategia de trading backtesting

A continuación se presenta el pronóstico generado por el modelo para la primera simulación:

fecha	valor de predicción
28/5/2018	7376,809 \$
29/5/2018	7391,094 \$
30/5/2018	7405,378 \$
31/5/2018	7419,663 \$
1/6/2018	7433,947 \$

Figura 4.50: Valores de predicción para la primera simulación.

Fuente: Elaboración propia.

La primera prueba se hace a partir del 28 de mayo de 2018, para ello se observa el valor real del día anterior el cual es, 7362.63\$ la predicción sugiere que el valor estará subiendo, pues los valores de la predicción utilizada están por arriba de 7362.63\$. Esto indicaría que la operación que generaría mayor beneficio es una compra.

Para verificar la sospecha de que una operación de compra es la que generaría mayor beneficios, se presenta una simulación para los próximos 5 días con la metodología de backtesting. En la siguiente imagen se muestra los resultados para simulación.

tipo de operación	fecha de entrada	max valor de ganancia	fecha de ganancia	hora de ganancia	max valor de pérdida	fecha de pérdida	hora de pérdida
compra	28/5/2018	1482,7 \$	31/5/2018	0:00:02	630,14 \$	29/5/2018	13:21:12
venta	28/5/2018	529,02 \$	29/5/2018	13:21:04	1011 \$	31/5/2018	0:00:10
			valor de entrada	valor de salida			
			7232,54 \$	7568,58 \$			

Figura 4.51: Resultados de la primera simulación.

Fuente: Elaboración propia.

De la figura 4.51 primera simulación se puede concluir que el modelo hizo el trabajo que se esperaba, pues si se observa la ganancia para la operación de compra es de 1482.7\$, esta ganancia se logró el cuarto día de simulación, el 31/5/218 a los 2 segundos de haber comenzado el día. También se observa que a pesar de que se logró obtener ganancia en la operación en algún momento de ella se estaba perdiendo y la mayor pérdida fue de 630.14\$ el día 29/5/2018 después de 13 horas 21 minutos y 12 segundos de ese día.

Si no se hubiera seguido la estrategia planteada y entrar en una operación de venta la perdida era de todo el capital invertido que eran 1000\$. En algún momento de la operación de venta se obtuvieron ganancias y de haber salido de la operación el 29/05/2018 a las 13 horas 21 minutos y 4 segundos se obtendría la ganancia de máxima para la operación de 529.02\$.

A continuación se muestran los valores utilizados para la primera simulación.

La figura 4.52 muestra los valores usados para la simulación, se observa el valor que está en riesgo Risk el cual es de 1000\$ además de los valores de Stop Loss y TakeProfit los cuales son 7020\$ y 7550\$ respectivamente, se muestra el momento exacto en el cual se obtuvo la ganancia máxima.

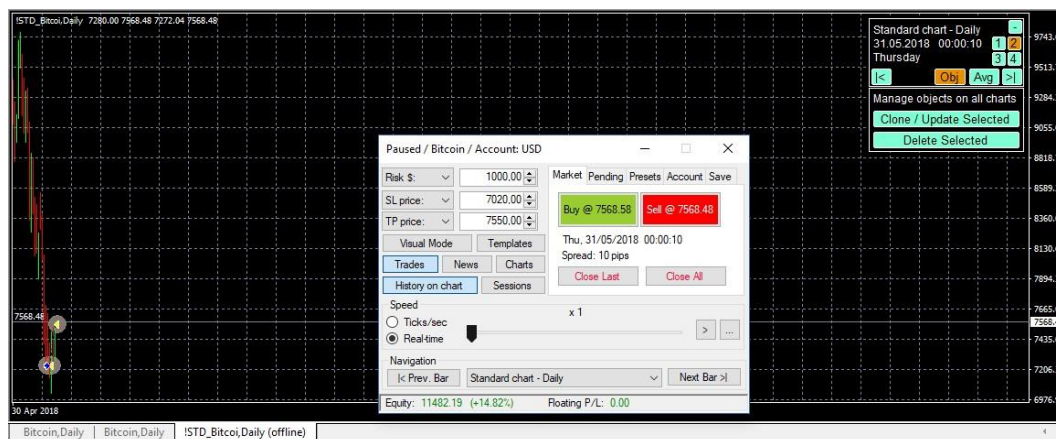


Figura 4.52: Valores usados para la primera simulación.

Fuente: Elaboración propia.

4.5.2. Segunda simulación de la estrategia de trading backtesting

A continuación se presenta el pronóstico generado por el modelo para la segunda simulación:

Fecha	Valor de predicción
22/10/2018	6509,129 \$
23/10/2018	6508,988 \$
24/10/2018	6508,848 \$
25/10/2018	6508,707 \$
26/10/2018	6508,566 \$

Figura 4.53: Valores de predicción para la segunda simulación.

Fuente: Elaboración propia.

La segunda prueba se hace a partir del 21 de octubre de 2018, para ello se observa el valor real del día anterior el cual es, 6590.56\$ la predicción sugiere que el valor estará en caída, pues los valores de la predicción utilizada están por debajo de 6590.56\$. Esto indicaría que la operación que generaría mayor beneficio es una venta.

Para verificar la sospecha de que una operación de venta es la que generaría mayor beneficios, se presenta una simulación para los próximos 5 días con la metodología de backtesting. En la siguiente imagen se muestra los resultados para simulación.

tipo de operación	fecha de entrada	max valor de ganancia	fecha de ganancia	hora de ganancia	max valor de pérdida	fecha de pérdida	hora de pérdida
Venta	22/10/2018	570,67 \$	23/10/2018	0:02:16	0	24/10/2018	0:07:59
Compra	22/10/2018	0			1048\$	23/10/2018	0:02:16
		valor de entrada	valor de salida				
		6429,01 \$	6358,77 \$				

Figura 4.54: Resultados de la segunda simulación.

Fuente: Elaboración propia.

De la imagen de la segunda simulación se puede concluir que el modelo hizo el trabajo que se esperaba, pues si se observa la ganancia para la operación de venta es de 570.67\$, esta ganancia se logró el segundo día de simulación, el 29/10/218 a los 2 minutos y 16 segundos de haber comenzado el día. También se observa que durante la simulación de los 5 días no se entró en pérdida esto se debe a que el valor nunca superó el valor con el cual entramos a la operación.

Si no se hubiera seguido la estrategia y entrar en una operación de compra la pérdida era de todo el capital invertido que eran 1000\$.

A continuación en la figura 4.55 se muestran los valores utilizados para la segunda simulación.

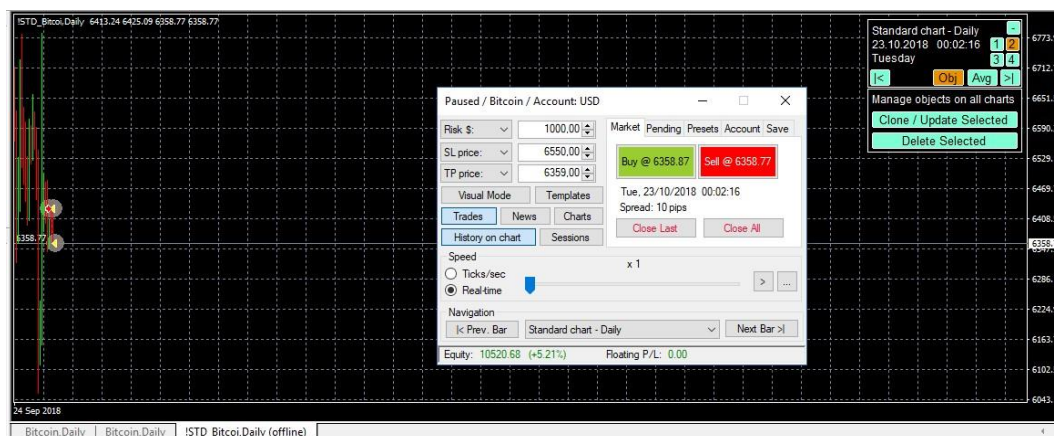


Figura 4.55: Valores usados para la segunda simulación.

Fuente: Elaboración propia.

De la imagen de los valores usados para simulación, se observa el valor que está en riesgo Risk el cual es de 1000\$ además de los valores de Stop Loss y TakeProfit los cuales son 6550\$ y 6359\$ respectivamente, se muestra el momento exacto en el cual se obtuvo la ganancia máxima.

4.5.3. Tercera simulación de la estrategia de trading backtesting

Para la tercera prueba de simulación de los modelos se realizó para la criptomoneda Ethereum, a continuación se muestran los pronósticos usados para la simulación.

fecha	valor Real	Valor Pronosticado
29/1/2019	\$ 105,60	\$ 105,37
30/1/2019	\$ 108,91	\$ 104,16
31/1/2019	\$ 107,06	\$ 102,95
1/2/2019	\$ 107,61	\$ 101,74
2/2/2019	\$ 110,43	\$ 100,53

Figura 4.56: Valores de predicción para la tercera simulación.

Fuente: Elaboración propia.

La tercera prueba se hace a partir del 29 de enero de 2019 y con la criptomoneda Ethereum, para ello se observa el valor real del día anterior el cual es, 106.36\$ la predicción sugiere que el valor estará en caída, pues los valores de la predicción utilizada están por debajo de 106.36\$. Esto indicaría que la operación que generaría mayor beneficio es una venta.

Para verificar la sospecha de que una operación de venta es la que generaría mayores beneficios, se presenta una simulación para los próximos 5 días con la metodología de backtesting. En la siguiente imagen se muestra los resultados para simulación.

tipo de operación	fecha de entrada	max valor de ganancia	fecha de ganancia	hora de ganancia	max valor de pérdida	fecha de pérdida	hora de pérdida
Venta	28/1/2019	3607,32	28/1/2019	17:01:18	0		
Compra	28/1/2019	0			2825,83	28/1/2019	17:01:18
		valor de entrada	valor de salida				
		110,09	101,01				

Figura 4.57: Resultados de la tercera simulación.

Fuente: Elaboración propia.

www.bdigital.ula.ve

De la figura 4.57 de la tercera simulación se puede concluir que el modelo hizo el trabajo que se esperaba, pues si se observa la ganancia para la operación de venta es de 3607.32\$, esta ganancia se logró el primer día de simulación, el 28/01/2019 a las 17 horas 1 minutos y 18 segundos de haber comenzado el día. También se observa que durante la simulación de los 5 días no se entró en pérdida esto se debe a que el valor nunca superó el valor con el cual entramos a la operación.

Si no se hubiera seguido la estrategia y entrar en una operación de compra la pérdida era de todo el capital invertido que eran 1000\$.

A continuación se muestran los valores utilizados para la tercera simulación.

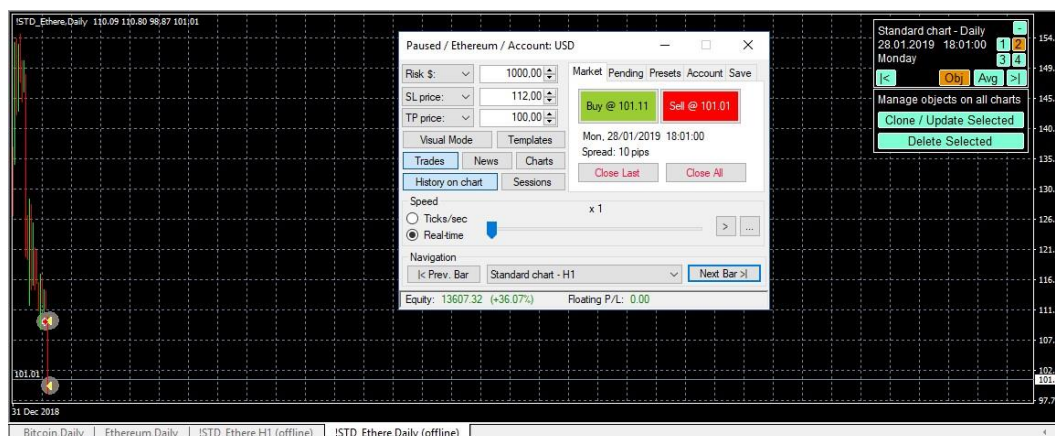


Figura 4.58: Valores usados para la tercerasimulación.

Fuente: Elaboración propia.

En la figura 4.58 se muestran los valores usados para simulación, se observa el valor que está en riesgo Risk el cual es de 1000\$ además de los valores de Stop Loss y TakeProfit los cuales son 112\$ y 100\$ respectivamente, se muestra el momento exacto en el cual se obtuvo la ganancia máxima.

4.5.4. Observación de las simulaciones

Después de observar las 3 simulaciones se puede decir que los modelos a corto plazo están cumpliendo con lo esperado y las estrategias usadas brindaron beneficios. Cuando se está haciendo trading es recomendable ser pacientes y confiar en la estrategia trazada, se puede entrar en pérdida en algún momento de la operación tal como se observó en la primera simulación pero si la estrategia está bien diseñada se tiene que tener confianza de que la operación tomará el rumbo que se espera.

Capítulo 5

Conclusiones y Recomendaciones

En el presente trabajo se logró desarrollar un modelo para el análisis de las criptomonedas, tal como lo expresaron los resultados del capítulo anterior, se demuestra la efectividad de los modelos ARIMA en la predicción de eventos futuros, logrando pronosticar precios y tendencias de las mismas.

En cuanto al análisis de series de tiempo, los pronósticos más efectivos o que mejor ajuste presentaron fueron para el caso del Bitcoin con un 88% de efectividad respectivamente. Sin embargo, no muy lejos de los modelos anteriores, se presentó un pronóstico aceptable para el Ethereum y Dash, ya que ambos comportamientos fueron predichos con un 83% y 82% de efectividad. En el caso del Bitcoin, al ser la moneda más importante e influyente en el mercado de los criptoactivos debido a la mayor capitalización, es la criptomoneda más sensible al comportamiento humano, anuncios, noticias, o divulgaciones afectan positiva o negativamente la confianza de los inversionistas, comerciantes y cualquier persona poseedor de criptodivisas, haciéndola así en una de las más volátiles. Por otra parte, algo que llamó la atención fue que el Ethereum a pesar de ser la moneda más joven en nuestro estudio, es la moneda con el rendimiento más estable de nuestro trabajo, lo cual fue confirmado al evaluar su coeficiente de variación siendo el menor entre todos. Además, su pronóstico se ajustó a los valores de rendimiento real con un 17% de error, lo cual no fue malo en comparación a los otros. Por otra parte,

El análisis de reglas de asociación nos mostró asociaciones interesantes que pudiesen estar presentes en algún momento, pero el hecho de que ocurra dicho acontecimiento es muy poco probable debido al soporte que arroja. A partir de lo obtenido en este estudio podemos concluir que la regla más importante no fue la que arrojó el modelo directamente, sino que a través de un análisis más profundo de los resultados arrojados por el algoritmo Apriori, se pudo observar que ninguna

criptomoneda afecta directamente al Bitcoin. Sin embargo, el *BTC* sí afecta de manera directa a las demás criptomonedas.

Un comportamiento particular que es necesario a destacar, es que no existen reglas que impliquen el crecimiento de una criptomoneda dado por la caída de rendimiento de otra criptomoneda, o viceversa, tomando en cuenta los casos poco comunes mostrados en los resultados del capítulo 4, donde se observa como la subida de alguna de las criptodivisas en particular no influye directamente sobre otra. Por esta razón, se concluye con certeza que el rendimiento de las criptomonedas estudiadas sube conjuntamente o baja conjuntamente en la mayoría de los casos.

En general, el método de aprendizaje automático nos mostró la correlación que existe entre las diversas criptomonedas estudiadas a lo largo de este trabajo. Mediante un antecedente y un consecuente (causa y efecto) resultante a partir de las observaciones que recibió el algoritmo y su posterior aprendizaje, el cual luego de analizar las distintas asociaciones que se formaban un número de veces determinado, proporcionaba un soporte y confiabilidad por cada regla permitiéndonos de esa forma tener una base sólida acerca de las relaciones más relevantes que pueden ocurrir.

El estudio realizado en el presente trabajo de grado nos muestra que es posible aproximar o pronosticar aceptablemente el comportamiento futuro de precios que tendrán las criptomonedas, mediante el uso de estos modelos de análisis, así como también, medir cuan correlacionadas se encuentran éstas, facilitando así la toma de decisiones de inversión en las distintas criptodivisas.

Finalmente los análisis realizados en este proyecto son recomendables y útiles para todo aquel inversionista, comerciante o minador de criptomonedas, dado que con la ayuda de estas herramientas podrían determinar cuál es el mejor momento para comprar o vender sus criptoactivos, cuándo predecir un decremento o alza en el precio. Además, pueden ser capaces de detectar o chequear el comportamiento de otras criptodivisas y como eso puede afectar su inversión, bien sea positiva o negativamente, aportándole al inversionista un panorama claro sobre lo que puede hacer dependiendo del cambio en

las relaciones observadas, resultando fundamental para así evitar pérdidas y en lugar de eso, optimizar sus ganancias.

Se recomienda tener una cartera de inversión diversificada, ya que la mayoría de veces sería mejor invertir en un criptoactivo con menor precio pero con **mejor rendimiento**, o dicho de otra manera, con un rendimiento más estable a lo largo del tiempo, ya que las fluctuaciones bruscas generadas en mayor medida por algunas puede ocasionar bien sea un aumento o pérdida significativo, por lo cual, no es recomendable arriesgarse a invertir por esas. Aferrándonos a los resultados obtenidos, la principal moneda para manejar su inversión podría ser principalmente Bitcoin, dado que es la que se logra ajustar con mayor efectividad. Otra posible opción podría ser apostar por el Ethereum, debido a su estabilidad en cuanto a rendimiento se trata, además de tener un menor precio que las anteriores, lo cual es un dato que podría aumentar el interés que existe por parte de los inversionistas de invertir en ella.

Es importante mencionar para finalizar, que el estudio realizado puede ser aplicado a un conjunto de datos que cuenten con características similares a los que se presentan en este trabajo, como por ejemplo, las variables deben ser cuantitativas y las observaciones deben seguir una frecuencia de 365 días.

Se puede concluir que es posible construir distintos modelos ARIMA para hacer predicciones de las diferentes criptomonedas, apoyados en estos pronósticos hacer una estrategia de trading, a lo largo del análisis de los resultados de los modelos propuestos podemos ver como estos no se ajustan de la mejor manera, la razón principal es que hay variables externas que afectan el comportamiento del valor del precio de las criptomonedas, no solo dependen de los datos históricos, de hecho se dificulta conseguir un patrón de comportamiento en cualquier escala de temporalidad. Según lo estudiado, ha habido anuncios o noticias que han afectado positiva o negativamente el valor del precio, estas noticias han sido de índoles:

- Social.

- Político.
- Tecnológico.
- Económico.

Por último, se puede decir que, con las series de tiempo es posible conseguir modelos ARIMA que permitan diseñar estrategias de trading, y así brindar a los inversionistas un plan de inversión más confiable en el cual se puedan confiar.

Para trabajos futuros, se recomienda trabajar con métodos de pronósticos a largo plazo u otro tipo de técnica que puedan generar una mejor predicción, debido a que los modelos ARIMA tienden a converger en una media constante y no logran predecir con efectividad las fuertes fluctuaciones que se presentan en la serie. Por otra parte, se puede enfocar el estudio a otro tipo de monedas como las de curso legal (dólar, euro, libras, pesos, entre otras), para ver si presentan alguna tendencia o estacionalidad que pueda ser detectada y poder deducir un comportamiento futuro a partir de esto.

Bibliografía

- Acosta, J. (2018). *“Estudio de Criptomonedas Mediante Series de Tiempo y Redes Neuronales”*. Trabajo especial de grado. Escuela de Matemáticas. Facultad de Ciencias. Universidad Central de Venezuela, pp. 3-63. Caracas, Venezuela.
- Bitcoin (2019). *“¿Qué es Bitcoin?”*. Sitio web. Recuperado el 12 de octubre del 2019 de <https://bitcoin.org/es/faq#que-es-bitcoin>
- Brooks, C. (2002). *“Introductory econometrics for finance”*. Cambridge University Press. USA.
- Cantó, J. (2018). *“Todo lo que debes saber acerca de las monedas virtuales”*. Blog tecnológico Bilib. Recuperado el 13 de octubre del 2019 de <https://www.bilib.es/actualidad/blog/noticia/articulo/todo-lo-que-debes-saber-acerca-de-las-monedas-virtuales/>
- Caurin, J. (2018). *“Dash”*. Economía simple. Recuperado el 15 de octubre del 2019 de <https://www.economiasimple.net/criptomonedas/dash>
- Cotizacioncriptomonedas. (2019). *“¿Qué es Litecoin?”*. Sitio web. Recuperado el 14 de octubre del 2019 de <https://www.cotizacioncriptomonedas.net/litecoin/que-es-litecoin/>
- Criptotendencia. (2017). *“¿Qué es un Satoshi?”*. Sitio web. Recuperado el 14 de octubre del 2019 de <https://criptotendencia.com/2017/06/17/que-es-un-satoshi/>
- Diario Gestión. (2019). *“¿Qué es el blockchain y cómo funciona?”*. Gestión Economía. Recuperado el 14 de octubre del 2019 de <https://gestion.pe/economia/blockchain-funciona-nnda-nnlt-258823-noticia/>

Fernández, S. (2008). “*Series Temporales: Modelo ARIMA*”. Departamento de Economía Aplicada. Facultad de Ciencias Económicas y Empresariales. Universidad Autónoma de Madrid. Recuperado el 24 de octubre del 2019 de <http://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>

Goicochea, A. (2009). “*CRISP-DM, Una metodología para proyectos de Minería de Datos*”. Recuperado el 08 de octubre del 2019 de <https://anibalgoicochea.com/2009/08/11/crisp-dm-una-metodologia-para-proyectos-de-mineria-de-datos/>

González, J. (2018). “*¿Qué es Ether?*”. Recuperado el 14 de octubre del 2019 de <https://autoincremental.es/criptomonedas/que-es-ether/>

González, R. (2019). “*¿Qué es R y RStudio?*”. Recuperado el 20 de septiembre del 2019 de <https://gonzalezgouveia.com/que-es-r-y-rstudio/>

Granell, M. (2018). “*Criptomonedas: ¿Qué es el Ripple?*”. Rankia. Recuperado el 14 de octubre del 2019 de <https://www.rankia.cl/blog/analisis-ipsa/3941094-criptomonedas-que-ripple>

Mills, T. (1991). “*Time Series Techniques for Economists*”. Cambridge University Press. USA.

Moffatt, M. (2019). “*What is the Augmented Dickey-Fuller Test?*”. ThoughtCo. Sitio web. Recuperado el 22 de octubre del 2019 de <https://www.thoughtco.com/the-augmented-dickey-fuller-test-1145985>

- Mompeán, Ó. (2016). “*Tendencia estocástica de una serie*”. Sitio web. Recuperado el 23 de octubre del 2019 de http://rstudio-pubs-static.s3.amazonaws.com/371571_8395ad57601b4758a167d85562fc26b9.html
- Mota, B., and Núñez, J. (2019). “*Estimación de la distribución multivariada de los rendimientos de los tipos de cambio contra el dólar de las criptomonedas Bitcoin, Ripple y Ether*”. Revista Mexicana de Economía y Finanzas Nueva Época. Universidad Autónoma Metropolitana and Tecnológico de Monterrey. Vol. 14, No 3, pp. 447-45. México.
- Nájera, J. (2018). “*Relaciones en el comportamiento de los precios de las criptomonedas: un análisis econométrico a través de modelos VAR y VEC*”. Estocástica: Finanzas y Riesgo. Departamento de Economía, División de Ciencias Sociales y Humanidades. Universidad Autónoma Metropolitana – Unidad Azcapotzalco, vol. 9, número 1, pp. 33-36. Ciudad de México, México. Recuperado el 08 de agosto del 2019 de <http://estocastica.azc.uam.mx/index.php/re/article/view/111/89>
- Nuevo Financiero. (2017). “*Litecoin, la irrupción de la criptomoneda nacida como complemento de Bitcoin*”. Sitio web. Recuperado el 14 de octubre del 2019 de <https://nuevofinanciero.com/litecoin-criptomoneda-complemento-bitcoin/>
- Ochoa, I. (2019). “*Aprendizaje Automático*”. Sitio web. Recuperado el 26 de septiembre del 2019 de <https://igorochoa.net/2019/09/09/que-es-machine-learning-aprendizaje-automatiko/>
- Ordinas, M. (2017). “*Las criptomonedas: ¿oportunidad o burbuja?*”. Informe mensual de estrategia de BancaMarch. Recuperado el 08 de agosto del 2019 de <https://www.bancamarch.es/recursos/doc/bancamarch/20170109/2017/informe-mensual-de-estrategia-octubre-2017.pdf>

- OroyFinanzas. (2014). “*Definición Criptomoneda: ¿Qué es una criptomoneda?*”. Sitio web. Recuperado el 24 de septiembre del 2019 de <https://www.oroynanzas.com/2014/10/que-es-criptomoneda/>
- Porro, A. (2018). “*Modelos de análisis de criptomonedas basado en aprendizaje automático y series temporales*”. Trabajo especial de grado. Escuela de Matemáticas. Facultad de Ciencias. Universidad Central de Venezuela, pp. 4-82. Caracas, Venezuela.
- Rosales, R. (2008). “*EJC 22: Metodología Box – Jenkins*”. Econometría 2 – Econ. 3301. Semestre II – 08. Facultad de Economía. Universidad de los Andes. Mérida, Venezuela. Recuperado el 08 de octubre del 2019 de https://economia.uniandes.edu.co/files/profesores/ramon_rosales_alvarez/docs/econometria2/Salidas%20y%20Ejercicios/EJC202220Metodologa20Box20-20Jenkins.pdf
- Rouse, M. (2017). “*Blockchain*”. Sitio web. Recuperado el 23 de septiembre del 2019 de <https://searchdatacenter.techtarget.com/es/definicion/Blockchain>
- Ruiz, D. (2016). “*Medidas de posición no central (Cuantiles)*”. Manual de Estadísticas. Capítulo II: Características de una Distribución de Frecuencias. Universidad Pablo de Olavide. Recuperado el 28 de octubre del 2019 de <http://www.eumed.net/cursecon/libreria/drm/1d.htm>
- Seux, J. (2015). “*Teoría y práctica de econometría financiera*”. Editorial Digital. Vol. 2, México.
- Significados (2017). “*Significado de Criptografía*”. Sitio web. Recuperado el 24 de septiembre del 2019 de <https://www.significados.com/criptografia/>
- Sims, C. (1980). “*Macroeconomic and Reality*”. Econométrica. Journal of the Econometric Society.

Suárez, M. (2014). “*Métodos de suavizamiento y pronóstico para series de tiempo*”.

Sitio web. Recuperado el 15 de octubre del 2019 de

<https://www.monografias.com/trabajos87/metodos-suavizamiento-y-pronostico-series-tiempo/metodos-suavizamiento-y-pronostico-series-tiempo.shtml>

Villagarcía, T. (2017). “*Series Temporales*”. Sitio web. Recuperado el 15 de octubre del 2019 de

http://www.est.uc3m.es/esp/nueva_docencia/leganes/ing_industrial/estadistica_industrial/doc_grupo1/archivos/Apuntes%20de%20series.pdf

Wikipedia. (2019). “*Criterio de Información de Akaike (AIC)*”. Sitio web. Recuperado el 25 de octubre del 2019 de

https://es.wikipedia.org/wiki/Criterio_de_informaci%C3%B3n_de_Akaike

WoltersKluwer (2019). “*Modelos VAR y VEC*”. Diccionario Empresarial. Recuperado el 23 de septiembre del 2019 de

http://diccionarioempresarial.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAAAAAEAMtMSbF1jTAAASMTczNLtbLUouLM_DzbsMz01LySVABNW3FFIAAAAA==WKE