

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

FACULTAD DE CIENCIAS

INSTITUTO DE ESTADÍSTICA



CONTRASTE DE MÉTODOS PARA PREDECIR EL
ALZA O BAJA DEL BITCOIN EN RELACIÓN AL
VALOR DEL EURO

Trabajo de Titulación

AUTORES

Osiel Briones
Constanza Estay
Simón Rojas

PROFESORA GUÍA

Natalia Bahamonde

Valparaíso, 2022

ÍNDICE

1.- INTRODUCCIÓN	4
1.1.- Descripción inicial	4
1.2.- Presentación del problema	4
1.3.- Objetivos	4
1.3.1.- Objetivo general	4
1.3.2.- Objetivos específicos	5
2.- HISTORIA DEL BITCOIN Y LAS CRIPTOMONEDAS	5
2.1.- ¿Qué son las criptomonedas?	5
2.2.- ¿Cuándo nacen las criptomonedas?	5
2.3.- Historia de las criptomonedas	5
2.4.- El Bitcoin	6
3.- ESTADO DEL ARTE	6
3.1.- ¿Qué se ha hecho?	6
3.2.- ¿Cómo se ha hecho?	7
3.3.- ¿Qué se hará?	8
4.- BASES DE DATOS	9
4.1.- Descripción de las variables	9
4.2.- Análisis factorial	10
5.- ANALISIS DE LAS SERIES DE TIEMPO	11
5.1.- Descripción del análisis	11
5.2.- Características de las series de tiempo financieras	12
5.3.- Heterocedasticidad condicional autorregresiva	12
5.4.- Hechos estilizados	14
5.5.- Test normalidad	16
5.6.- Limitaciones sobre las series de tiempo	16
5.7.- Serie de tiempo Ethereum	17
5.8.- Resumen de características para ajuste de un modelo GARCH a Ethereum	18
5.9.- Modelamiento de GARCH (p, q) para Bitcoin	19
5.10.- Predicción GARCH (1,1) en Bitcoin	20
6.- ANÁLISIS DE MÉTODOS DE CLASIFICACIÓN	22
6.1.- Descripción de los métodos de clasificación	22
6.2.- Análisis de resultados de métodos de clasificación en la predicción del Bitcoin	23
6.3.- Verificación de supuestos mediante test de medianas	26

6.4.- Análisis de resultados de métodos de clasificación en la predicción de 7 criptomonedas	27
6.5.- Resultados de métodos de clasificación en la predicción de 7 criptomonedas diferentes	30
6.6.- Análisis “tiempo real” de los modelos de clasificación	31
7.- CONCLUSIÓN	32
7.1.- Observaciones finales	32
7.2.- ¿Cómo continuar este trabajo?	32
8.- BIBLIOGRAFÍA	34

ÍNDICE DE GRÁFICOS

Gráfica 1: Análisis factorial	10
Gráfica 2: Serie de tiempo Bitcoin Cierre	11
Gráfica 3: Rentabilidad Bitcoin	12
Gráfica 4: Indicadores ACF & PACF	13
Gráfica 5: Cuadrados & Valor Absoluto	14
Gráfica 6: Cuantil & Caja	15
Gráfica 7: ACF de 360 datos y 180 datos.	17
Gráfica 8: Serie de tiempo ETH Cierre	18
Gráfica 9: Características de la rentabilidad ETH	19
Gráfica 10: Medias de predicciones por escenario	26
Gráfica 11: Resultados de los métodos en 7 criptomonedas y diferentes escenarios	30

ÍNDICE DE TABLAS

Tabla 1: Predicción GARCH (1,1) en Bitcoin	20
Tabla 2: Comparación de resultados de GARCH	21
Tabla 3: Aciertos del escenario 1	24
Tabla 4: Aciertos del escenario 2	24
Tabla 5: Aciertos del escenario 3	25
Tabla 6: Aciertos del escenario 4	25
Tabla 7: Aciertos escenario 2 para 7 criptomonedas	28
Tabla 8: Aciertos escenario 3 para 7 criptomonedas	28
Tabla 9: Aciertos escenario 4 para 7 criptomonedas	29
Tabla 10: Resultados predicciones datos uno a uno	31
Tabla 11: Aciertos del escenario 1 con variable Índice del Euro	33
Tabla 12: Aciertos del escenario 1 con las variables seleccionadas	33

1.- INTRODUCCIÓN

1.1.- Descripción inicial

En la actualidad, el comercio de criptomonedas está muy a la vanguardia, una criptomoneda es un activo digital que emplea un cifrado criptográfico para garantizar la seguridad de este activo. En los últimos años ha aumentado exponencialmente la cantidad de gente que se anima a participar e invertir en este mercado a pesar de lo volátil que es. Se han realizado varios estudios para poder predecir los valores que tomarán, predecir su subida o bajada y también registrar comportamientos repetitivos en sus ciclos. El tema de las criptomonedas y los estudios realizados inspiró y animó al equipo a contrastar diferentes métodos que predicen la subida o bajada de criptomonedas, esto se realizará mediante el seguimiento de los valores históricos del bitcoin principalmente, y también los de otras criptomonedas. Se trabajará con modelos de clasificación de aprendizaje automático y con el modelo GARCH de series de tiempo. Para ello, se demostrará y definirán los criterios óptimos (los cuales entreguen mejores resultados de predicción) de estos modelos para así contrastarlos siendo expuesto cada modelo en su mejor versión, con el fin de dejar en evidencia los beneficios y las debilidades de cada uno de estos ya que es de suma importancia obtener información sobre el comportamiento de los activos y tener conocimiento del mercado para así poder adentrarse o seguir participando activamente en este negocio.

1.2.- Presentación del problema

El principal problema es que las criptomonedas al ser muy volátiles son muy difíciles de predecir si irán al alza o a la baja. Se sabe de métodos de análisis de gráficos los cuales son muy útiles, pero lo que se busca es encontrar un método más científico para resolver esta problemática. Dentro de esto se encontraron diferentes factores que afectan a la hora de evaluar los datos, ya sean, en los modelos predictivos, y/o en el análisis de series de tiempo, por lo que, entender cuanto y como afectan estos factores a nuestros datos, es el principal problema.

1.3.- Objetivos

1.3.1.- Objetivo general

- Contrastar métodos de clasificación por aprendizaje automático y modelo GARCH a la hora de predecir el alza o baja de una criptomoneda.

1.3.2.- Objetivos específicos

- Demostrar y definir los criterios óptimos en el uso de los modelos de clasificación.
- Examinar pros y contras de los modelos de clasificación y series de tiempo.

- Obtener la precisión de los modelos de clasificación al predecir el alza o baja de criptomoneda.
- Analizar heterocedasticidad de la serie temporal que compone el valor histórico de una criptomoneda.
- Verificar si se cumplen las condiciones para utilizar el modelo GARCH.
- Obtener la precisión del modelo GARCH al predecir el alza o baja de criptomoneda.

2.- HISTORIA DEL BITCOIN Y LAS CRIPTOMONEDAS

2.1.- ¿Qué son las criptomonedas?

Una criptomoneda es un activo digital, que utiliza un cifrado criptográfico, que es la base principal de la seguridad de datos y también tienen como objetivo garantizar su titularidad y asegurar la integridad de las transacciones, también controlar la creación de unidades adicionales, es decir, hacer copias de estas. Las criptomonedas no existen de forma física, y se almacenan en una cartera digital llamadas hardware wallets.

2.2.- ¿Cuándo nacen las criptomonedas?

Se podría decir que el Bitcoin (BTC) es el origen y máximo representante de las criptomonedas. Este nació a raíz del artículo de Satoshi Nakamoto (hasta ahora desconocido), publicado en 2008. Al comienzo no tuvo mucho auge, ya que, solo algunos muy interesados de la tecnología lo conocían y era considerado casi inútil o sin valor. Sin embargo, su valor ha alcanzado cifras nunca imaginadas para este tipo de activo.

2.3.- Historia de las criptomonedas

El origen de las criptomonedas ha transcurrido un largo camino en el que han participado científicos, matemáticos y personas con mucha visión de futuro, todos estos han contribuido a que hoy en día estas monedas digitales sean tan usadas. Además de Satoshi Nakamoto, creador del Bitcoin, encontramos otros nombres como David Chaum y Wei Dai, a quienes podemos considerar los precursores de las criptomonedas. David Chaum (criptógrafo estadounidense), en la década de los 80. Más específicamente en el año 1983, desarrolló un primer sistema criptográfico denominado e-Cash. Fue concebido como una especie de dinero electrónico criptográfico anónimo o un sistema de efectivo electrónico. Y se utilizó como sistema de micro pagos en un banco de EE. UU. de 1995 a 1998. Luego, en 1995, desarrolló un nuevo sistema llamado DigiCash, que utilizaba la criptografía para mantener confidenciales los datos de quienes realizaban las transacciones económicas. De hecho, se puede decir que así fue como nacieron las criptomonedas. Wei Dai (ingeniero informático especialista en criptografía) el año 1998, hizo público un ensayo en el que presentó el concepto de “b-money”, un sistema de efectivo electrónico distribuido y anónimo. Dai, en su ensayo sobre “b-money” especifica lo indispensable de mantener la contabilidad colectiva con protocolos criptográficos, que se encargarían de autenticar las transacciones y, a la vez, servirían como garantía de que se mantuvieran organizadas. Con esta propuesta Dai se adelanta y da los primeros pasos en el camino hacia lo que hoy en día conocemos como la tecnología Blockchain.

2.4.- El Bitcoin

En el 2008, en medio de una crisis financiera mundial, se desconoce si una persona o un grupo de personas, tras el seudónimo de Satoshi Nakamoto, publicó un documento técnico de lo que llamó Bitcoin en el sitio web de la fundación P2P (Peer to Peer). En el documento se mencionó que las tantas veces citado Bitcoin, era una nueva visión de dinero electrónico. De esta manera, se presenta por primera vez el concepto de Bitcoin al mundo y es así como nacen las criptomonedas. Tras esto el 03 de enero del 2009 se lanza, la primera criptomoneda oficial (BTC), como parte del bloque llamado “Génesis”.

Este formato de transmisión de punto a punto indica que es un sistema de pago descentralizado. Lo que quiere decir que, a diferencia de las otras monedas tradicionales de circulación legal llamadas dinero fiduciario, Bitcoin no tiene un emisor centralizado, sino que se genera mediante cálculos basados en algoritmos específicos de los nodos de la red. De esta manera, puede circular por todo el mundo, estar en cualquier ordenador conectado y cualquiera puede participar en su fabricación o “minado”. No importa dónde se encuentre, cualquiera puede extraer, comprar, vender o recibir Bitcoins o “altcoins” (monedas alternativas).

Luego de la aparición del Bitcoin, cuyo objetivo inicial era convertirse en una moneda digital con la que se podrían hacer compras por internet, el mercado de criptoactivos tuvo un alza de forma muy rápida y lo sigue teniendo. Este mercado dio lugar al nacimiento de muchas otras criptomonedas. Algunas de ellas son: Bitcoin, Litecoin, Ripple, Dogecoin, Monero, Ethereum, Polkadot, Solana, entre otras.

3.- ESTADO DEL ARTE

3.1.- ¿Qué se ha hecho?

Como se ha explicado anteriormente, las criptomonedas son activos digitales que funcionan como un medio de intercambio basado en sistemas de criptografía que aseguran las transacciones, controlan la creación de monedas y verifican la transferencia segura de activos. La diferencia esencial que existe entre las criptomonedas del resto de monedas es que las primeras están basadas en el principio de control descentralizado, es decir, no es necesaria la existencia de un organismo central que controle la moneda y legitimar su valor. El análogo a este organismo dentro del contexto de criptomonedas es el sistema mismo de encriptación, comúnmente conocido como Blockchain. Dada la necesidad de prever de alguna manera el valor de transacción de las criptomonedas, se han desarrollado informes que aplican técnicas de aprendizaje supervisado con el objetivo de hacer predicciones de los valores de las criptomonedas, tratando de predecir el valor de estas, sin que la alta volatilidad a la que se ve expuesto el mercado interfiera de forma significativa en el resultado.

Otra forma de poder predecir la volatilidad sobre las criptomonedas es a partir de series de tiempo, con más detalles utilizaremos el modelo GARCH. Se realizó un trabajo con series de tiempo [6], en el cual comparó los modelos EGARCH y GARCH con las criptomonedas Bitcoin y Ethereum, las mismas que vamos a utilizar.

3.2.- ¿Cómo se ha hecho?

Hay estudios [8], donde se ha utilizado el método de Random Forest, que es una técnica de reducción de la dimensionalidad. Donde se generan múltiples árboles y cada árbol da una clasificación (vota por una clase). Y el resultado es la clase con mayor número de votos en todo el bosque (Forest). Los estudiantes de la Universidad Politécnica de Madrid [8], analizaron las 7 criptomonedas con mayor capitalización de mercado mediante este método para intentar predecir el valor de las criptomonedas. La ventaja que deja hacer predicciones y anticiparse a los movimientos del mercado es de gran interés por los beneficios económicos que puede conllevar. Los inversores y economistas tienen especial interés por estas técnicas, debido a que, esta herramienta es de gran ayuda para la toma de decisiones. La técnica antes planteada pretende la creación de modelos predictivos del precio de cierre a partir de unas características seleccionadas, una vez realizada la predicción se evalúan los resultados obtenidos en el conjunto de validación, y, este procedimiento comprendió tres etapas: en primer lugar, investigaron sobre el método ya citado, y, además, sobre el origen, evolución y auge que han tenido las criptomonedas, en segundo lugar, buscaron datos históricos y se construyeron modelos de predicción utilizando para ello Random Forest, que utilizaron para predecir los precios de cierre del conjunto de prueba, ellos realizaron una comparación entre los modelos propuestos en función de la cantidad de información obtenida de cada criptomoneda, y, en tercer lugar, analizaron los resultados y los utilizaron para predecir los valores de las criptomonedas estudiadas de manera más precisa, además, se analizaron estos resultados, simulando diferentes intervalos, y se ha realizado una comparación de los resultados, para saber que método es escogido para cada criptomoneda, esto tomando el peso que tiene cada variable de entrada de cada criptomoneda para instantes anteriores a la predicción, y archivos temporales conducen a mejores resultados. En otro estudio, realizado por los estudiantes de la Universidad Politécnica de Valencia, escuela de ingeniería en telecomunicaciones [1], quisieron abordar el tema de aprendizaje automatizado para desarrollar una estrategia de compraventa de bitcoins. Ellos mediante el seguimiento del tipo de cambio crearon un modelo de clasificación aprendizaje automático para determinar la alteración del próximo cambio en función de algún indicador financiero. Para esto examinaron valores históricos del Bitcoin (BTC). Utilizaron algunos métodos de aprendizaje automático para predecir si el tipo de cambio iba a aumentar o disminuir en el siguiente periodo de cotización. Basándose en esto, ellos propusieron involucrar áreas del aprendizaje automático supervisado y no supervisado, centrándose en la tarea de clasificación. Y emplearon los siguientes métodos para encontrar el modelo más eficiente: Redes Neuronales, Regresión Logística, K - Vecindad más Cercana, Máquina de Soporte Vectorial, Árbol de Decisión, Bosque Aleatorio y modelos de Naive Bayes. Para comparar estos métodos ellos consideraron métricas de exactitud, precisión, exhaustividad, puntuación F1 y tiempo de ejecución. Hicieron un análisis comparativo para quedarse con el método de clasificación más eficiente, el cual se utiliza finalmente como núcleo de la estrategia de comercio. Y luego de definir el modelo de negociación basado en el aprendizaje automático, presentaron varios estudios de pruebas retrospectivas para

supervisar el comportamiento de la estrategia. Realizaron pruebas retrospectivas con los tipos de cambio de varias criptodivisas. Por último, ellos recomiendan un programa de inversión que opera la estrategia definida basada en el modelo de clasificación en tiempo real. Para la serie de tiempo se ellos utilizaron distintos tiempos según la moneda de estudio, en el caso de Bitcoin 6 años, a partir de 8 de Julio de 2010 hasta el 7 de Junio de 2016, con 2.151 datos, y Ethereum con una venta de tiempo de un año aproximadamente, 8 de Agosto de 2015 hasta el 7 de Junio de 2016, con 311 datos. Demostrando la importancia de una cantidad mínima de datos varía según la moneda. De esta manera utilizaron distintos modelos GARCH, el original y EGARCH, comparando cada uno de sus resultados.

3.3.- ¿Qué se hará?

Teniendo datos históricos sobre estas “monedas” se generó una base de datos que refleja lo que se busca, ya que como datos se tenía el valor de apertura (valor con el que la moneda comienza el día) y el valor de cierre (el ultimo valor que tuvo durante el día) con estos datos se generará una variable llamada “Aumento” la cual representa porcentajes de subida o bajada en ese periodo, a su vez se creará otra variable llamada “Subidanext” la cual será una variable binaria la cual mostrará si la siguiente observación, es decir el día siguiente aumentó o disminuyó el valor del activo, con esta base de datos entrenaremos los modelos de clasificación con distintos parámetros buscando así, los que produzcan el mejor comportamiento en ellos siendo mejor mientras mayor asertividad de predicción tengan. Se evaluarán pros y contras de cada modelo de clasificación y del modelo GARCH. Analizando el comportamiento de la criptomoneda, si es similar a la de una serie financiera, con el fin de contrastar los requerimientos y diversas condiciones en las cuales podemos o no utilizar el modelo; Se analizará la heterocedasticidad de la serie temporal que resulta de los valores históricos del Bitcoin y otras criptomonedas. Se obtendrán porcentajes de precisión de predicción de cada modelo para finalmente contrastarlos y así exponer tanto los beneficios como las falencias de cada modelo. Lo llamativo del trabajo realizado es lo diferente que es a otros, comenzando que se utilizaron los precios del Bitcoin en relación al Euro, la mayoría de estudios que se hacen de criptomonedas se realizan utilizando el Dólar como comparación, por lo que se decidió ocupar otro activo el cual también tiene un nivel de importancia alto dentro del mundo para así poder entregar un estudio con mayor diversificación dentro del mundo financiero. También el trabajo consiste en contrastar diferentes metodologías lo cual no se había hecho en otros trabajos. Además se trabajó cada metodología con

distintos escenarios (trabajo de cada metodología con distintos parámetros) esto para así poder realizar de mejor forma el contraste.

4.- BASES DE DATOS

4.1.- Descripción de las variables

Para definir las variables que se utilizarán se realizó un análisis factorial con las variables que son de fácil acceso público, las cuales son “Date”, “Open”, “Close”, “High”, “Low”, “Volumen” y “Market cap”. Siendo “Date” la fecha de la observación, “Open” el valor de apertura, “Close” el valor de cierre, “High” el valor más alto de la observación, “Low” el valor más bajo de la observación, “Volumen” indica la cantidad de transacciones de esta moneda que se ha generado y “Market cap” es el índice de la capitalización de mercado de la moneda. Además, se generaron dos nuevas variables las cuales son “Aumento” y “Subidanext”, “Aumento” representa el valor porcentual respecto al valor de inicio de la diferencia entre las variables Open y Close, la variable “Subidanext” será representada por 0 y 1 cuando el valor de “Aumento” de la siguiente observación es negativo y positivo respectivamente.

$$Aumento(x_1) = \frac{Close(x_1) - Open(x_1)}{Open(x_1)} * 100$$

$$Subidanext(x_1) = 0 \quad si \quad Aumento(x_2) \leq 0$$

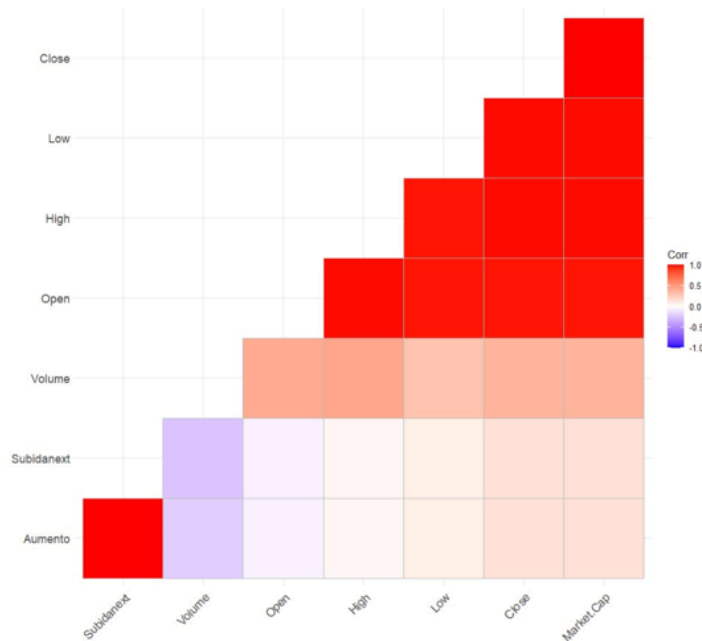
$$Subidanext(x_1) = 1 \quad si \quad Aumento(x_2) > 0$$

Estas variables se crearon para satisfacer el objetivo planteado el cual se relación con la subida o bajada del Bitcoin, siendo así “Subidanext” nuestra variable objetivo ya que esta representa binomialmente la subida o bajada de la moneda en cuestión. Habiendo mostrado las variables existentes y la variable objetivo, se procedió a realizar el análisis factorial. La base de datos que se ocupará es la data histórica del Bitcoin, esta fue extraída de “Yahoo! Finanzas” [5]. Esta base tiene una frecuencia diaria y va desde el 1 de enero del 2017 hasta el 10 de septiembre del 2021.

4.2.- Análisis factorial

Se procedió a realizar un análisis factorial con el fin de encontrar las variables con las cuales se trabajará de mejor manera en los modelos.

ANÁLISIS FACTORIAL



Gráfica 1: Análisis factorial

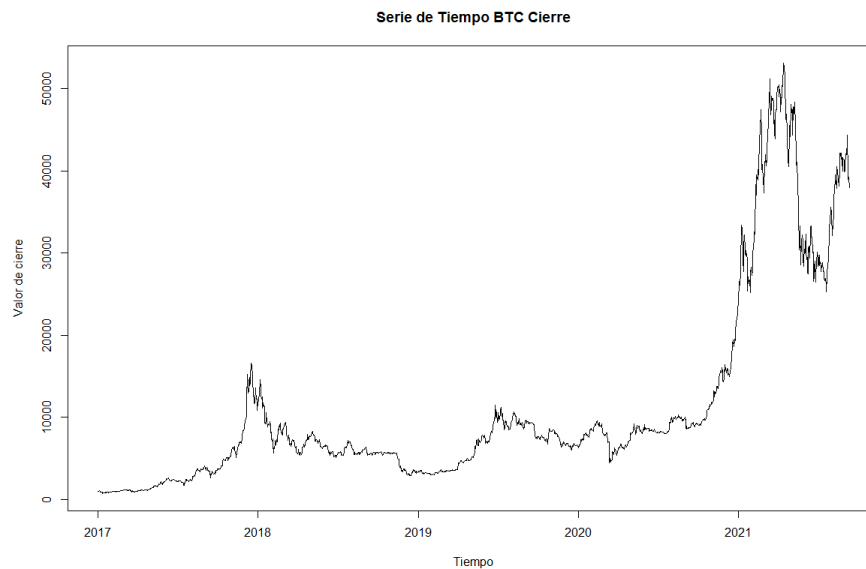
Como podemos observar las variables “Open”, “Close”, “High”, “Low” están fuertemente correlacionadas entre sí, si bien no se ve una relación directa de estas con nuestra variable objetivo “Subidanext” esta ultima fue creada a partir de dos de ellas como fue indicado anteriormente. Se puede observar que la variable “Volumen” no esta muy relacionada con las demás y para efectos de evitar añadir una variable poco relacionada al modelo, fue descartada. La variable “Market cap” si bien tiene alta relación con las variables mencionadas al principio, no se tomará en cuenta debido a que no aporta a la hora de predecir sino a la hora de analizar, y su alta relación se debe a que esta variable está formada en parte por la variable “Close” multiplicada por la cantidad de monedas circulantes, esto causaría un sobre entrenamiento a los modelos por lo que se decide dejar fuera. Quedando así de forma definitiva las siguientes variables: “Date”, “Open”, “Close”, “High”, “Low” y “Aumento”, además obviamente como variable objetivo la variable “Subidanext”.

Quedando así una base con una frecuencia diaria y la cual va desde el 1 de enero del 2017 hasta el 9 de septiembre del 2021. Cabe recalcar que esta base se irá seccionando respecto al periodo que se desee utilizar, siendo esto explicado en ítems siguientes.

5.- ANÁLISIS DE LAS SERIES DE TIEMPO

5.1.- Descripción del análisis

Una serie de tiempo es una sucesión de datos, la cual permite observar el comportamiento de una variable con respecto a un espacio tiempo. Para este caso interesa el comportamiento de la volatilidad de la moneda Bitcoin. Antes se ha de observar el comportamiento del valor de cierre de la moneda Bitcoin a lo largo del tiempo, con una frecuencia en días, la cual inicia en el día 1 de enero del 2017 hasta el 10 de septiembre del 2021. La cantidad de datos es 1715 observaciones.



Gráfica 2: Serie de tiempo Bitcoin Cierre

Puesto que suponemos tener datos financieros nos interesamos en su rentabilidad a lo largo del tiempo para poder realizar su predicción, debido a la no estacionariedad de la serie trabajamos con su logaritmo de rentabilidades. Para nuestro proyecto tomaremos las rentabilidades positivas como ganancias, y las rentabilidades negativas como pérdidas, de este modo podremos realizar predicción. Hay un alza significativa en el año 2021 llegando a un tener 5 veces de su valor original.

Conversión a log rentabilidad

$$\text{Rentabilidad : } r_t = X_t - X_{t-1}$$

$$R_t = \log r_t$$

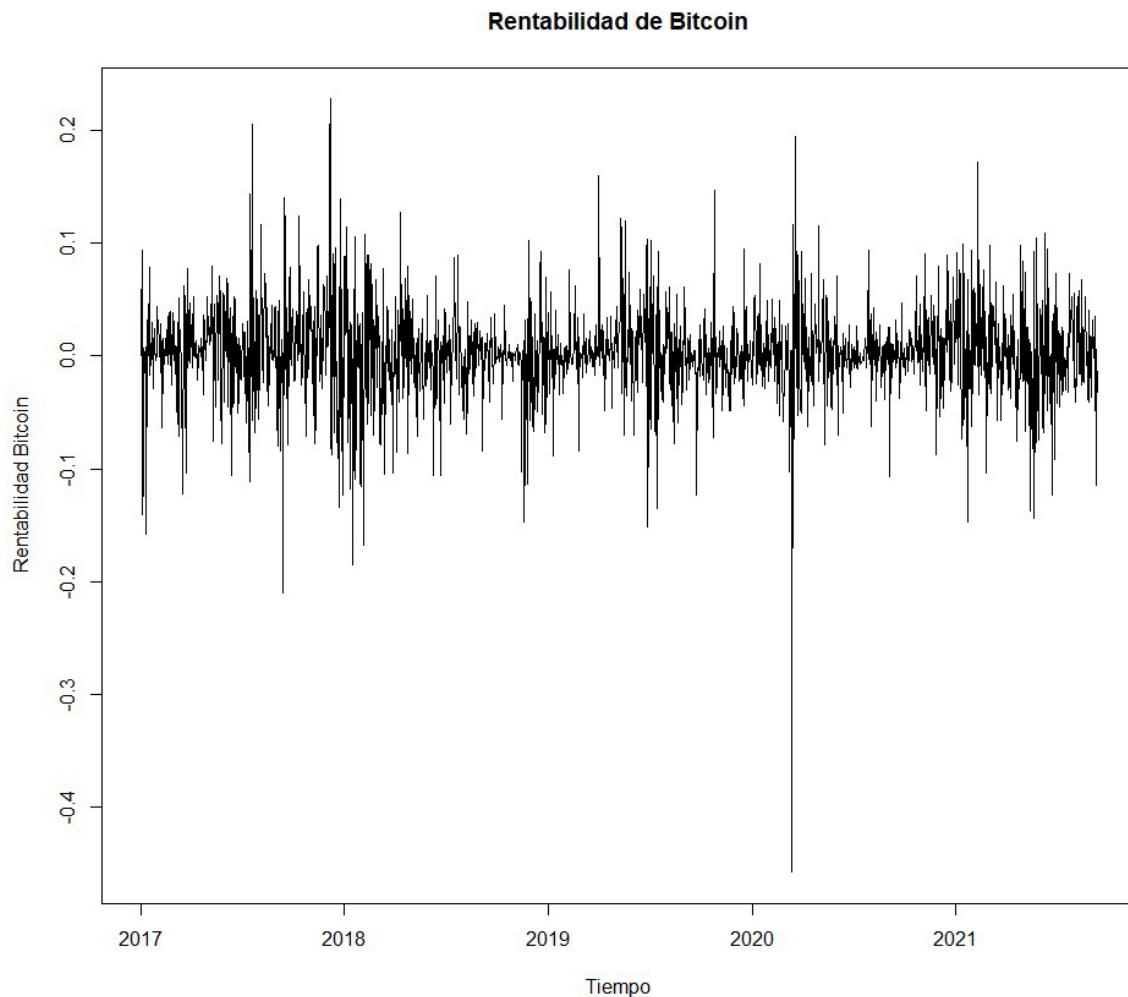
$$R_t = \log\left(\frac{X_t}{X_{t-1}}\right)$$

5.2.- Características de las series de tiempo financieras

Las series de tiempo financieras tienen ciertos tipos de principios básicos que comparten todas estas. Las más reconocidas son volátiles, no gaussianas, presencia de observaciones atípicas, extremos pesados, no correlacionadas linealmente, correlaciones en momentos de orden superior. Es por esto que hemos decidido verificar estas condiciones con el comportamiento de la criptomoneda Bitcoin y Ethereum.

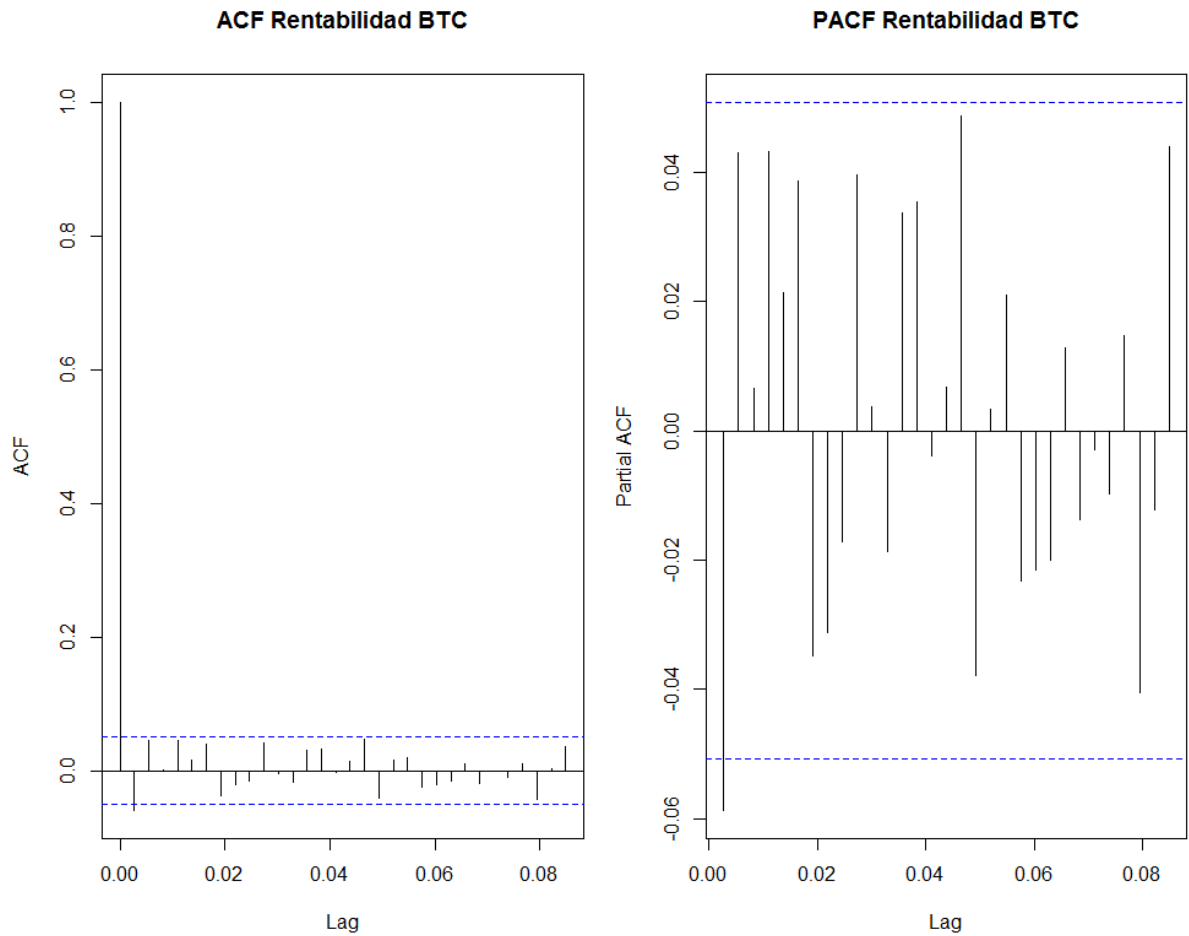
5.3.- Heterocedasticidad condicional autorregresiva

La heterocedasticidad es un indicador de la varianza de los datos, cuando hay presencia de heterocedasticidad decimos que existe una alta varianza, es decir, que los datos difieren entre unos y otros con una magnitud alta. En nuestro caso es relevante puesto que esta es una característica de las series financieras, se verá si las criptomonedas se comportan de manera similar, y comprobar así si el modelo GARCH o algunas de sus extensiones resulta útil para modelar sus características.



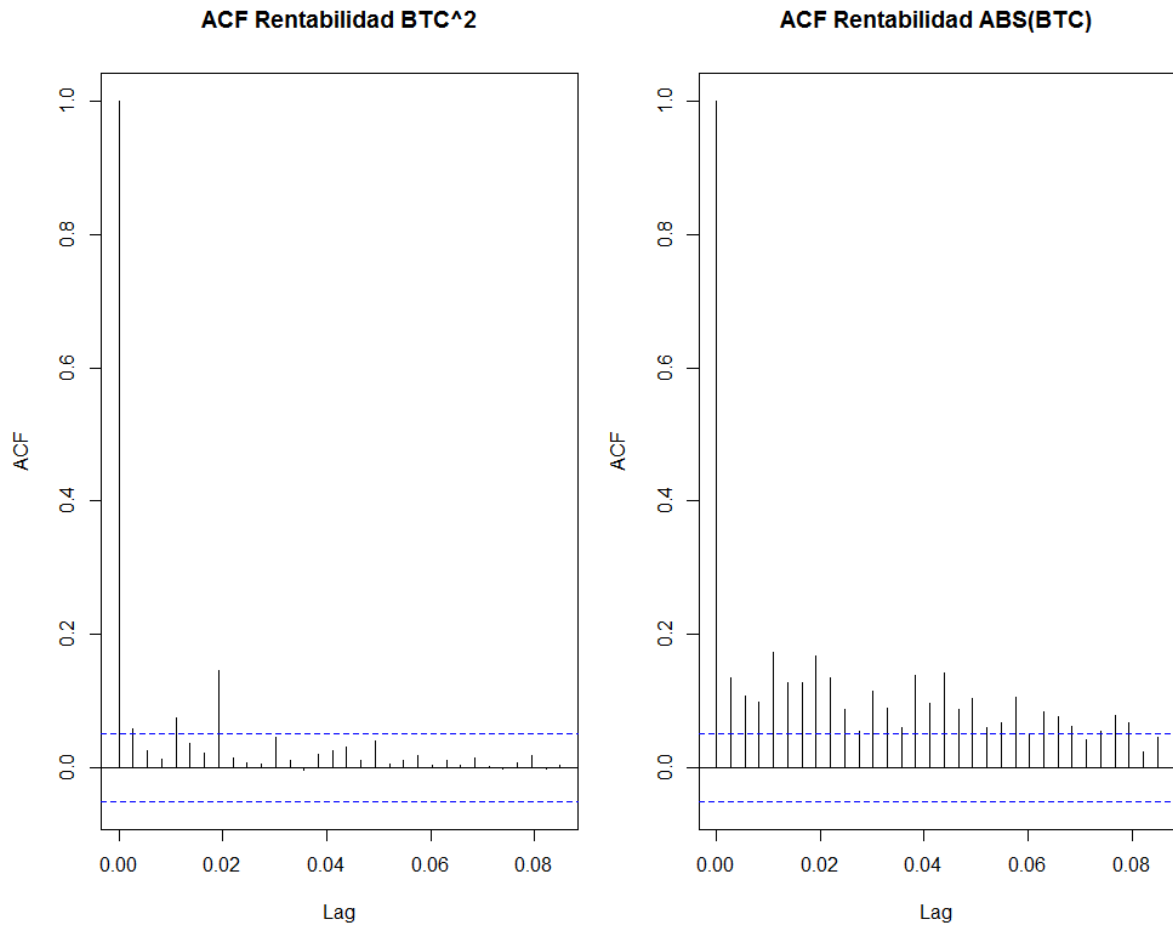
Gráfica 3: Rentabilidad Bitcoin

La presentada Gráfica 3 señala que existe una fuerte volatilidad de la serie, presentando una varianza no constante a lo largo del tiempo, lo que nos dice que no es estacionaria. Este indicador nos muestra a priori una característica que presentan otras series de tiempo financieras. Debido a la pandemia, la causante de pesimismo y miedo sobre el futuro ingreso de las personas promovió la actividad de esta criptomoneda. Produciendo así el efecto de Joseph, el cual nos señala que después de periodos de alta fluctuación existe una ventana de tiempo en la que esta es baja, para así poder realizar una futura predicción a través del tiempo.



Gráfica 4: Indicadores ACF & PACF

La Gráfica 4, muestra que no existe suficiente información de manera lineal para nuestra serie de tiempo, en otras palabras, hay ausencia de correlación lineal. La serie podría ser no lineal y de estructura de dependencia de orden superior, lo cual es otra propiedad en el área financiera.

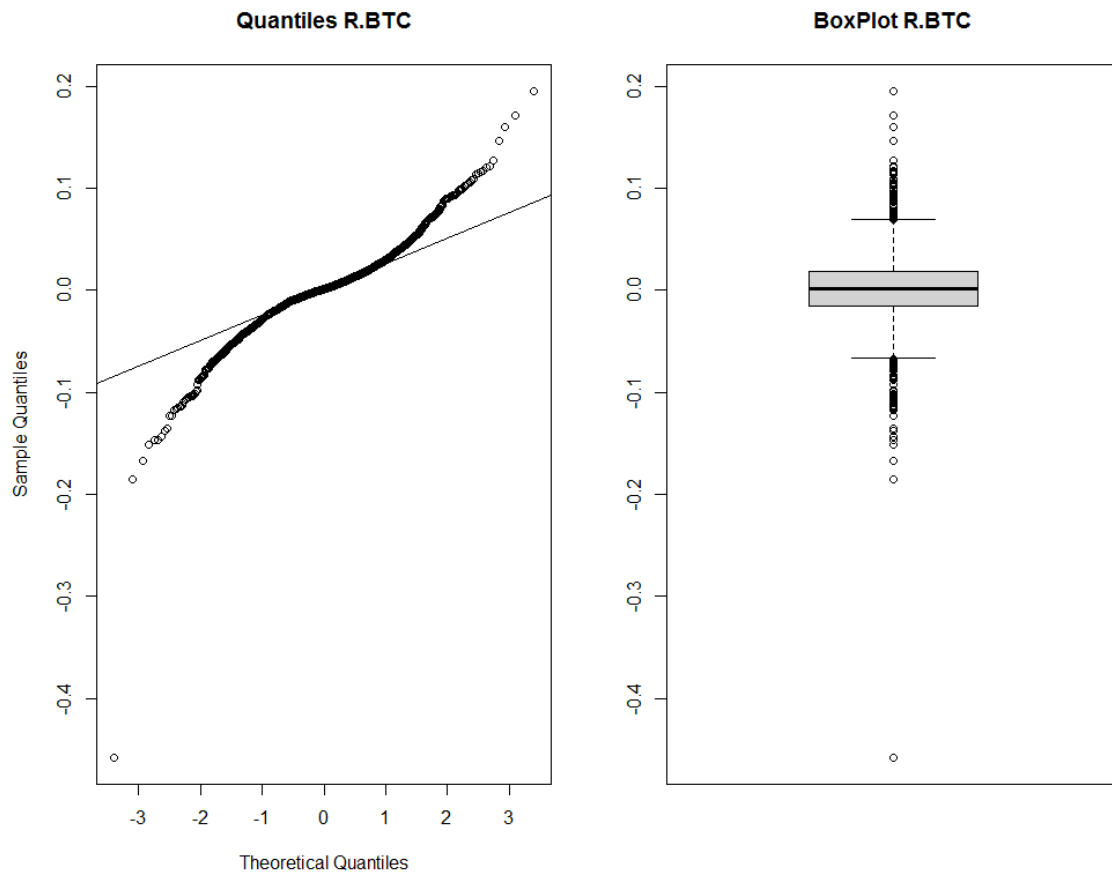


Gráfica 5: Cuadrados & Valor Absoluto

Como se puede observar en la Gráfica 5, la log rentabilidad de la serie presenta una estructura de dependencia no lineal y de orden superior, indicada por tener rezagos significativos de las ACF fuera de las bandas de confianza en la rentabilidad de modo absoluto.

5.4.- Hechos estilizados

Los hechos estilizados en los datos del Bitcoin nos demuestran que los datos no son gaussianos, con presencia abundante de datos atípicos y los extremos son más pesados como se evidencia en ambos gráficos presentados en la Gráfica 6.



Gráfica 6: Cuantil & Caja

Por ejemplo, en la Gráfica 6, con la gráfica de los cuantiles de los datos, comprobamos que este al tener muchos puntos que sobresalen de la recta generan extremos pesados. No se comporta como una distribución normal, pues en las distribuciones normales sus puntos convergen en la recta. Ahora la segunda gráfica comprobamos la gran presencia de datos atípicos, con los puntos que sobresalen de la caja con bigotes.

Con esto se comprueba otros comportamientos necesarios de las series financieros: la no normalidad, observaciones atípicas y extremos pesados.

5.5.- Test normalidad

Por otra parte, existe el estadístico de Jarque-Bera que contrasta la hipótesis nula de normalidad de la serie comparando la simetría y la curtosis de la muestra con los valores de la distribución normal. Se define como:

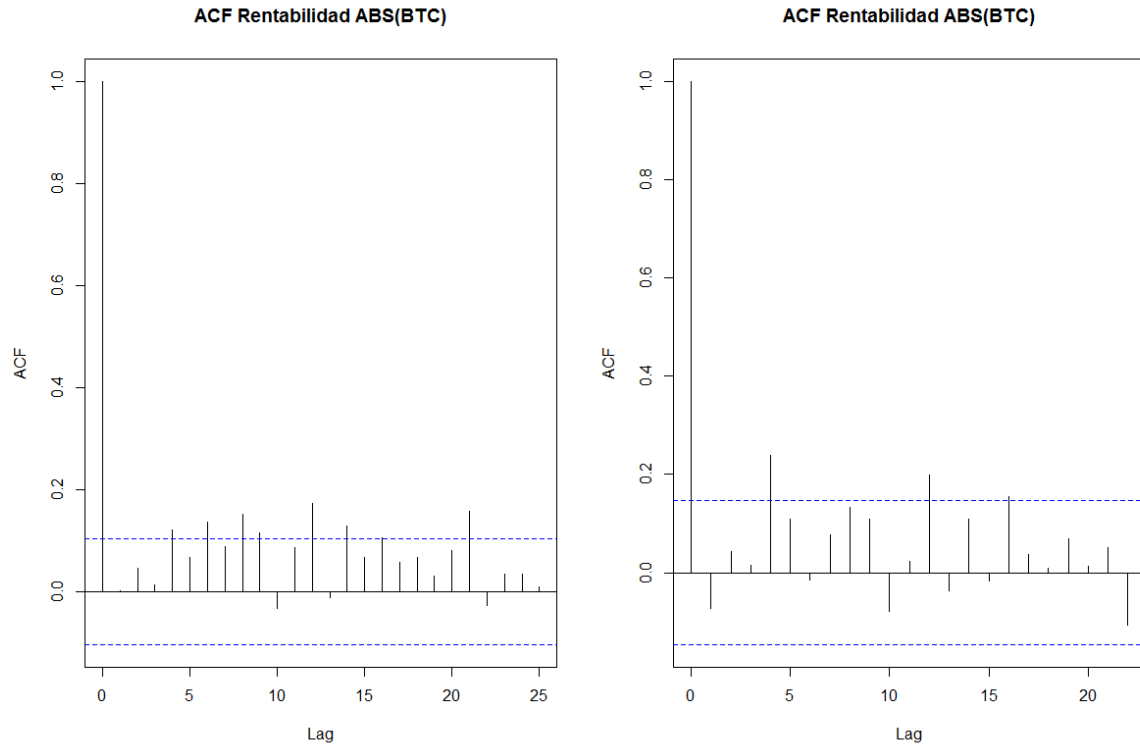
$$JB = \frac{N - k}{6} \left[S^2 + \frac{(K - 3)^2}{4} \right]$$

Donde S es la simetría de la muestra, K su curtosis y k representa en número de coeficientes utilizados para crear la serie. El estadístico Jarque-Bera sigue una distribución χ^2 con 2 grados de libertad. Para nuestro caso se rechaza la hipótesis nula, confirmando de otro modo la no normalidad.

$$H_0 = X \sim N \quad H_1 \neq X \sim N$$

5.6.- Limitaciones sobre las series de tiempo

Una de las principales limitaciones de las series de tiempo es que son necesarias bases de datos extensas para que estas tengan un desempeño adecuado. Realizando así una réplica del mismo trabajo anterior, pero para efectos del tema de las limitaciones en su cantidad de datos. El contraste se produce con una ventana de tiempo de un año (11.09.2020 – 10.09.2021) con 365 datos, y otra de 6 meses, (11.03.2020 – 10.09.2020) con 180 datos, esto va en la línea de la limitación comentada al comienzo, en la siguiente grafica se puede apreciar lo ocurrido.

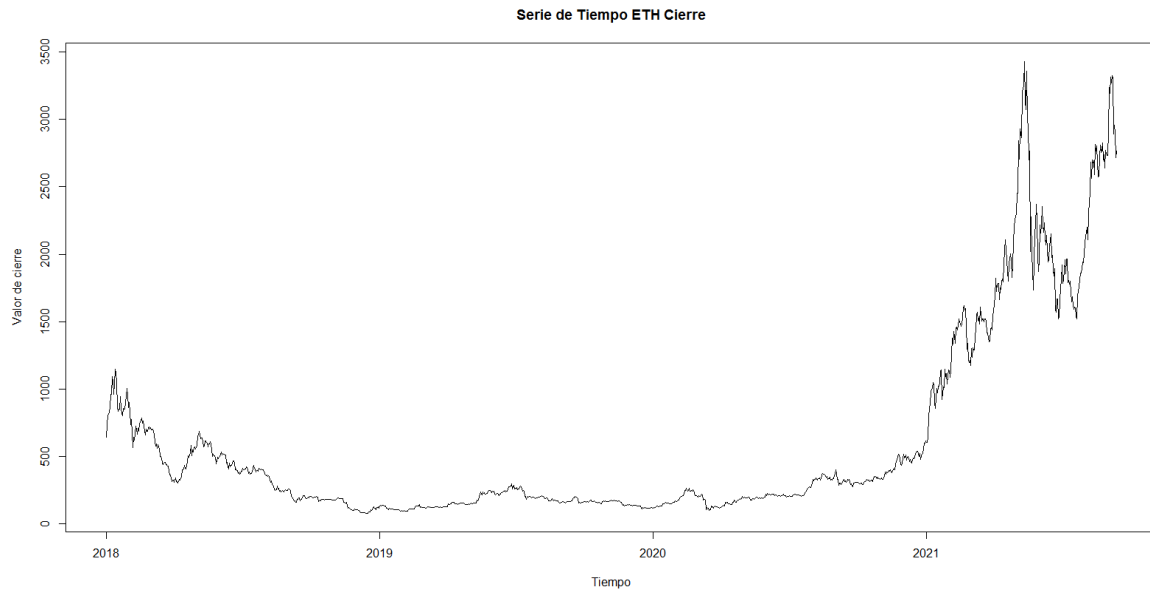


Gráfica 7: ACF DE 365 datos y 180 datos.

Como se puede observar, la Gráfica 7 sobre la autocorrelación de orden superior no se evidencia una clara estructura de dependencia de orden superior, haciendo no valida la metodología de análisis empleada anteriormente para poder definir un modelo GARCH. De esta manera se puede confirmar que un requerimiento mínimo para poder realizar un ajuste de modelo correcto sobre la serie de tiempo es necesario una robusta cantidad de datos. Como se vio en el correcto uso de la serie principal.

5.7.- Series de tiempo Ethereum

Para comparar el comportamiento de otra criptomoneda similar a Bitcoin se ha decidido tomar Ethereum (ETH), su valor de cierre, esta criptomoneda se originó el 30 de Julio del 2015, pero para efectos de mejor análisis se ha decidido utilizar a partir del 1 de enero del 2018 hasta el 10 de septiembre de 2021.

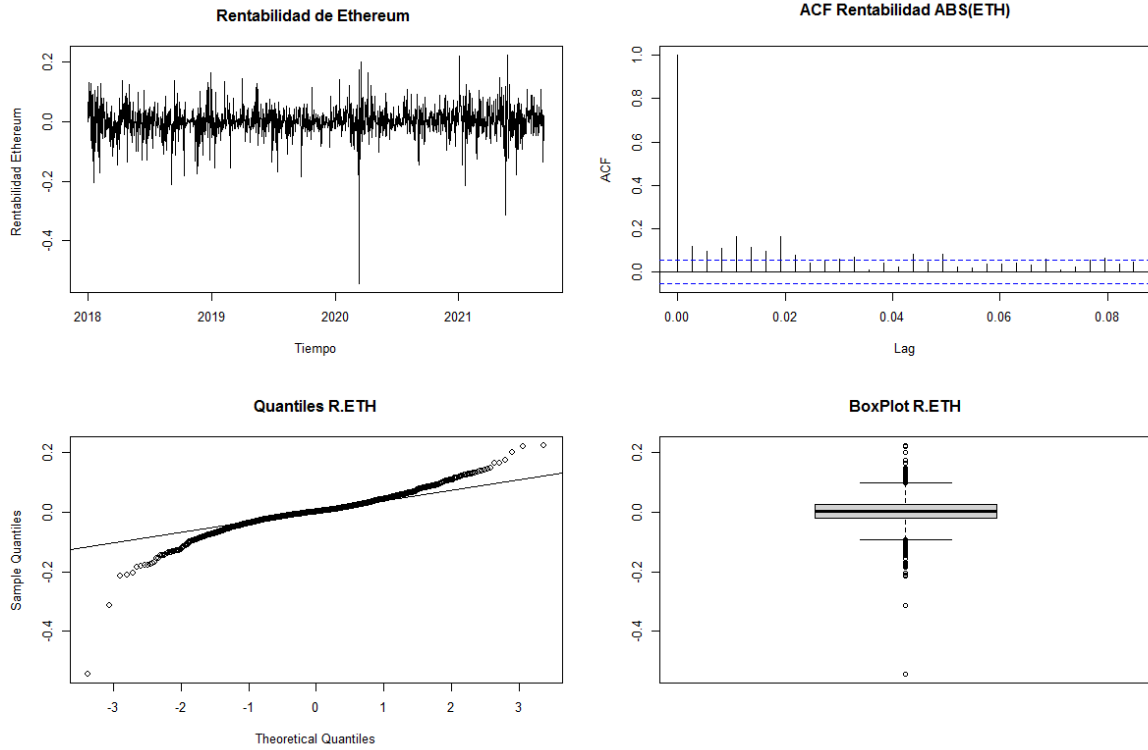


Gráfica 8: Serie de tiempo ETH Cierre

La ilustración 7 nos muestra que tiene el pico más alto en el año 2021, al igual que la moneda Bitcoin, Ahora vamos a analizar el ajuste de esta moneda usando modelos GARCH.

5.8.- Resumen de características para ajuste de un modelo GARCH a Ethereum

Siguiendo la misma línea que la moneda anterior, se mostrarán las gráficas más importantes para comprobar las características necesarias para utilizar el modelo.



Gráfica 9: Características de la rentabilidad ETH

Viendo la Gráfica 9, al observar en el primer cuadro de la log rentabilidad de Ethereum, se puede verificar la presencia de heterocedasticidad de la serie. Luego en el segundo cuadro de la ACF de la log rentabilidad absoluta de Ethereum, se visualizan rezagos significativos, lo cual indica presencia de dependencia de orden superior, y finalmente se observan características no gaussianas con los últimos gráficos de abajo, puesto que este tiene variados puntos que sobre salen de la línea qq-norm y la presencia excesiva de datos atípicos en el gráfico de bigotes. De esta manera podemos concluir que Ethereum también puede ser modelada por un GARCH.

5.9.- Modelamiento de GARCH (p,q) en Bitcoin

Originado de un modelo ARCH que se consideraba anteriormente para predecir los retornos del intercambio entre el dólar y euro. Para mantener el modelo simple Bollerslev (1986) propone una extensión generalizada del modelo ARCH (GARCH). Para los retornos logarítmicos de una serie r_t , el modelo GARCH (p, q) se describe como [7]:

$$r_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

Finalmente podemos ver qué tipo de modelo GARCH es necesario utilizar en nuestra criptomoneda Bitcoin y nuestros valores de (p, q) para el modelo es de (3,2), puesto que estos parámetros fueron los que pasaron la prueba de Jarque-Bera en el modelo. De esta manera tenemos que:

$$r_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \alpha_2 \sigma_{t-2}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \beta_3 \sigma_{t-3}^2$$

Donde r_t es la rentabilidad y σ_t^2 es la volatilidad. Así finalmente se puede observar la predicción realizada. Su estimación es:

	Coefficiente	Error estándar	T valor
α_0	0.0002765	0.0000272	10.17
α_1	0.3419362	0.0386542	8.85
α_2	0.1243195	0.0457611	2.72
β_1	0.0144233	0.0834050	0.17
β_2	0.0667787	0.0489709	1.36
β_3	0.4095367	0.0405063	10.11

Tabla 1: estimación de parámetros GARCH(3,2)

EGARCH

Para superar una debilidad del modelo GARCH en las series financieras, Nelson (1991) propone el modelo EGARCH(m,s). Para poder permitir efectos asimétricos entre retornos positivos y negativos, de forma que:

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^m \alpha_i \frac{|a_{t-i}| + \gamma_i a_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^s \beta_j \ln \ln(\sigma_{t-j}^2)$$

Donde, un a_{t-i} positivo contribuye $\alpha_i(1 + \gamma_i)|\varepsilon_{t-i}|$ para log volatilidad, un a_{t-i} negativo resulta $\alpha_i(1 - \gamma_i)|\varepsilon_{t-i}|$, donde $\varepsilon_{t-i} = a_{t-i}/\sigma_{t-i}$. El parámetro γ_i entonces significa el efecto apalancamiento de a_{t-i} .

Para nuestro caso utilizaremos un EGARCH(1,1) para ver difiere de nuestro modelo original, resultando el modelo en:

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1(|\varepsilon_{t-1}| + \gamma_1 \varepsilon_{t-1}) + \beta_1 \ln \ln(\sigma_{t-j}^2)$$

Sus estimaciones son:

	Coeficiente	Error estándar	T valor
α_0	0.000469	0.001000	0.46871
α_1	-0.069092	0.016546	-4.17570
β_1	0.923170	0.013833	66.73454
γ_1	0.177487	0.029701	5.97574

Tabla 2: estimación de parámetros EGARCH(1,1)

5.10.- Predicción GARCH en Bitcoin

A continuación, se muestra el análisis del comportamiento con la extensión exponencial sobre el modelo GARCH para la predicción de la rentabilidad de la criptomoneda Bitcoin con un método bootstrap. De esta manera se optó por 10 días, es decir del 11 de septiembre del 2021 hasta el 20 de septiembre del mismo año, también es importante añadir que en el entrenamiento se tomó la base a partir del 2018 ya que los valores del 2017 al ser muy bajos generaban valores perdidos. Comparando si su rentabilidad es positiva o negativa. Partiendo con la rentabilidad original de la serie de tiempo, y luego con los distintos tipos de modelos. Con respecto si es positiva o negativa la rentabilidad tenemos que:

FECHA	ORIGINAL	GARCH(3,2)	EGARCH(1,1)
11/09/2021	Positiva	Positiva	Positiva
12/09/2021	Negativa	Positiva	Positiva
13/09/2021	Positiva	Negativa	Positiva
14/09/2021	Positiva	Positiva	Positiva
15/09/2021	Negativa	Negativa	Positiva
16/09/2021	Negativa	Negativa	Positiva
17/09/2021	Positiva	Negativa	Positiva
18/09/2021	Negativa	Negativa	Positiva
19/09/2021	Negativa	Negativa	Positiva
20/09/2021	Negativa	Negativa	Positiva

Tabla 3: Predicción GARCH en Bitcoin

Se puede notar que, a pesar de ser otros tipo de extensión, EGARCH se ve solo signos positivos, a diferencia de GARCH(3,2). Notamos una precisión del 70% con respecto a la rentabilidad original en el caso de GARCH(3,2), y EGARCH(1,1) un 40%. De igual modo se conoce el valor exacto de la rentabilidad original para los siguientes días, y de esta manera también comparando los comportamientos de las extensiones, se observó que:

FECHA	ORIGINAL	GARCH(3,2)	EGARCH(1,1)
11/09/2021	0.01912	0.02222908	0.04523
12/09/2021	-0.02397	0.03693241	0.04542
13/09/2021	0.04687	-0.73521945	0.04559
14/09/2021	0.02137	0.94869187	0.04575
15/09/2021	-0.00361	-0.02051009	0.04590
16/09/2021	-0.00748	-0.37598358	0.04603
17/09/2021	0.02111	-0.30703252	0.04615
18/09/2021	-0.02161	-0.45326532	0.04627
19/09/2021	-0.0979	-0.14898738	0.04637
20/09/2021	-0.05114	-0.01526210	0.04647

Tabla 4: Comparación de resultados de GARCH

Podemos notar que los valores entre el tipo de extensión exponencial varían de la rentabilidad original, de modo que no recomendamos utilizar este método para poder predecir el valor exacto. Recomendamos utilizar la extensión original del modelo GARCH (3,2) con el ruido normal para poder predecir de mejor modo.

6.- ANÁLISIS DE MÉTODOS DE CLASIFICACIÓN

6.1.- Descripción de los métodos de clasificación

Se seleccionaron cinco métodos de clasificación sin algún criterio en específico. A continuación, presentaremos los modelos y se entregara una breve descripción de que consiste cada uno, también se indicara el análisis de ventajas y desventajas de estos.

SUPPORT VECTOR MACHINE (SVM)

La idea básica de la máquina de vectores de soporte (SVM) es encontrar un hiperplano separando dos clases con los márgenes más amplios. Los puntos más cercanos al hiperplano óptimo se denominan vectores de soporte. SVM utiliza estos vectores de soporte para encontrar el hiperplano óptimo el cual proporciona la clasificación. Para clasificar datos no lineales, el modelo convierte los datos originales a una dimensión superior utilizando funciones de

kernel no lineales como el polinomio kernel y el kernel de función de base radial. Algunas de sus ventajas son que puede clasificar problemas complejos usando diferentes núcleos, es eficaz tanto en espacios de grandes dimensiones como en casos donde el número de dimensiones es mayor que el número de muestras también es muy eficiente en memoria debido a que utiliza un subconjunto de puntos de entrenamiento en la función de decisión. Por otro lado, una desventaja es que, si el número de características es mucho mayor que el número de muestras, se debe evitar el exceso de ajuste al elegir las funciones del Kernel.

ÁRBOLES DE DECISIÓN

La idea básica del árbol de decisión es dividir los datos en subconjuntos recursivos para lograr que cada subconjunto contenga registros casi homogéneos. El modelo forma un árbol siendo así una estructura en la cual cada vértice representa una prueba para un atributo. El criterio para seleccionar como se divide es separar el conjunto de datos tanto como sea posible. Los bordes que salen de los vértices, representan los posibles resultados de las decisiones. Una ventaja de este modelo es que es fácil de interpretar y maneja tanto atributos numéricos como atributos nominales y su desventaja es que el método solo funciona bien con pocas variables y que su desempeño puede ser más débil si hay interacciones demasiado complejas en la base de datos.

K NEAREST NEIGHBOR (KNN)

El algoritmo KNN representa cada registro de entrenamiento como un punto de datos en el n -dimensional espacio, donde n es el número de atributos del registro. Durante la clasificación, la entrada del modelo siempre consta de k vecinos más cercanos del registro a clasificar. La proximidad métrica, que se utiliza ampliamente para encontrar los vecinos más cercanos, son la distancia euclidiana o distancia de manhattan. En la clasificación KNN, la salida será la clase más común entre los k vecinos más cercanos. Dentro de sus ventajas tenemos que es un algoritmo simple y fácil de aplicar ya que no es necesario crear un modelo, formular hipótesis suplementarias ni configurar muchos parámetros, además que es un algoritmo polivalente ya que sirve tanto para clasificación como para regresión.

Su desventaja es que se vuelve mas lento a medida que aumenta el numero de observaciones o el número de variables independientes.

REGRESIÓN LOGÍSTICA

La regresión logística es un método estadístico efectivo para analizar conjuntos de datos que tienen al menos una variable independiente. Usando la función logística, podemos medir la relación entre una o más variables independientes y la variable categóricamente dependiente mediante la estimación de probabilidades. Este resultado, muestra la probabilidad de que el registro pertenezca a una clase dada. Usando este método, La salida se puede interpretar como una probabilidad. Como ventaja tiene que es resistente al ruido sin embargo su desventaja es que requiere de mayor cantidad de datos para una operación estable.

NAIVE BAYES

El clasificador Naive Bayes asume valores de atributo que son independientes entre sí y no existen relaciones de dependencia entre las columnas de la base de datos. Supone que hay m clases denotadas por $C_1, C_2 \dots C_n$. Para un registro X , el modelo predice que X pertenece a una clase C_i dada de la siguiente manera. $P(C_i|x) = P(X|C_i)P(C_i)/P(X)$. Calculando con probabilidades condicionales, encuentre la clase para la cual esta probabilidad es más alto. Como ventaja tenemos que es fácil y rápido de calcular, pero su desventaja es que la suposición del método Naive Bayes a menudo no se cumple en el conjunto de datos, y en ausencia de independencia de las variables, el clasificador no da un resultado preciso.

6.2.- Análisis de resultados de métodos de clasificación en la predicción del Bitcoin

Para explicar la forma en que se hicieron estos análisis, en primer lugar, se trabajó con parte de la base de datos inicial, comenzando el 01 de enero del año 2017, y terminando el 10 de septiembre del año 2021. Estos análisis fueron hechos sobre distintos escenarios del Bitcoin, que son los siguientes:

Escenario 1: Aplicación de base de datos histórica, que comprende 1.715 datos.

Escenario 2: Aplicación de base de datos anual, con 360 datos, y para obtener el resultado de este escenario se utilizó el promedio de 5 años, desde 2017 a 2021.

Escenario 3: Aplicación de base de datos semestral, que contempla 180 datos, que, para este escenario, como en el anterior se utilizó el promedio de 10 periodos de 180 datos, desde el 2017 al 2021.

Escenario 4: Aplicación de base de datos de 3 meses, de 90 datos, que también se utilizó el promedio de 20 periodos, considerando los primeros 3 meses de cada semestre, por 5 años, desde el 2017 al 2021.

Y a continuación, se muestran los resultados de los distintos niveles de aciertos, que tuvieron los métodos de clasificación a la hora de predecir el alza o la baja de esta criptomoneda.

Escenario 1: Acierto data histórica

MÉTODOS	ACIERTO (P)
SVM	0.63
Arbol de decisión	0.62
KNN	0.63
R. logística	0.63
Naive Bayes	0.63

Tabla 5: Aciertos del escenario 1

Los resultados en la tabla muestran que, para el escenario 1 no hay diferencia significativa entre los modelos que se pueda apreciar visualmente.

Escenario 2: Media de aciertos de la data histórica con ventanas de tiempo de 1 año

MÉTODOS	ACIERTO (P)
SVM	0.65
Arbol de decisión	0.53
KNN	0.61
R. logística	0.63
Naive Bayes	0.59

Tabla 6: Aciertos del escenario 2

Los resultados en la tabla muestran que, para el escenario 2, visualmente los métodos más efectivos para predecir son Máquina de Soportes Vectoriales seguido por Regresión logística.

Escenario 3: Media de aciertos de la data histórica con ventanas de tiempo de 6 meses

MÉTODOS	ACIERTO (P)
SVM	0.64
Arbol de decisión	0.55
KNN	0.61
R. logística	0.64
Naive Bayes	0.59

Tabla 7: Aciertos del escenario 3

Los resultados en la tabla muestran que, para el escenario 3, visualmente los métodos más efectivos para predecir son tanto Máquina de Soportes Vectoriales como Regresión logística.

Escenario 4: Media de aciertos de la data histórica con ventanas de tiempo de 3 meses

MÉTODOS	ACIERTO (P)
SVM	0.65
Arbol de decisión	0.55
KNN	0.59
R. logística	0.66
Naive Bayes	0.60

Tabla 8: Aciertos del escenario 4

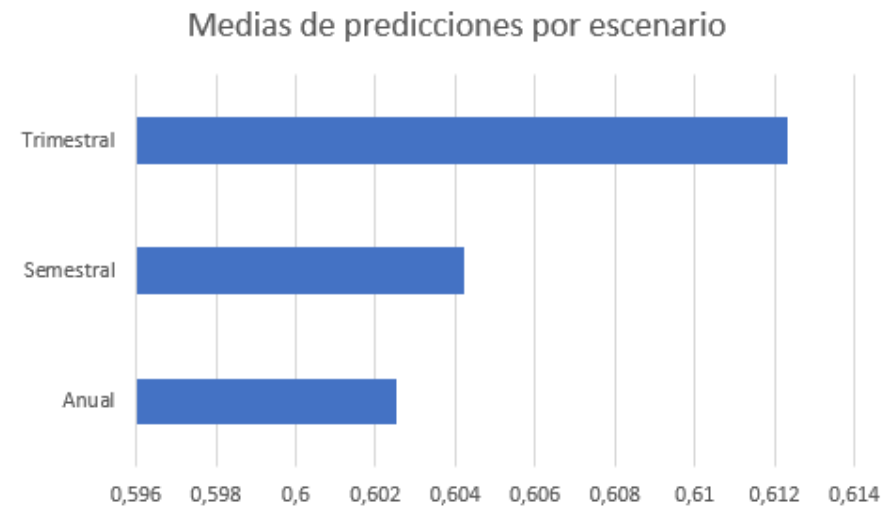
Los resultados en la tabla muestran que, para el escenario 4, visualmente los métodos más efectivos para predecir son Regresión logística seguido por Máquina de Soportes Vectoriales, siendo así los resultados más altos obtenidos de los cuatro escenarios.

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS

Resumiendo el proceso, lo que se hizo fue ir probando una base de datos de Bitcoin, en distintos escenarios, junto a esto las bases disminuían en tamaño, debido a que todas las variables son diarias, se utilizaron códigos en R, de 5 modelos de clasificación, utilizando la variable en factor de 0 y 1, que representaba si bajaba o subía el valor de la criptomoneda respectivamente, del día siguiente, y se obtuvieron resultados más satisfactorios para el Escenario 4 para Bitcoin, que era el que contemplaba la menor cantidad de datos, como se pudo notar en las tablas presentadas anteriormente.

6.3.- Verificación de supuestos mediante test de medianas

Al realizar las tablas del punto 6.2 se fue calculando predicciones para todo el tiempo que cubre la base de datos pero utilizando lapsos de tiempo distintos para cada entrenamiento, tomando todos esos resultados por escenario se creó una base con sus respectivos valores de predicción separados anualmente, semestralmente y trimestralmente, las medias resultantes dan como resultado los valores expuestos en la siguiente grafica.



Gráfica 10: Medias de predicciones por escenario

Como se puede observar, gráficamente se ve que en promedio se ven mejores resultados cuando se utilizan 3 meses para entrenar al modelo, pero esto debe ser verificado, por lo que se decidió verificar mediante un test. Primero se verifico la normalidad de las variables “anual”, “semestral” y “trimestral” resultando los siguientes p-valor:

P – valor

Anual: 0.0255 < 0.05

Semestral: 0.09306 > 0.05

Trimestral: 0.02 < 0.05

Rechazándose así la normalidad para las variables “anual” y “trimestral”

A continuación, se procedió a realizar test de mediana de Kruskal-Wallis para ver si habían diferencias significativas en las predicciones dentro de estos periodos de entrenamiento para los modelos. Dándonos como resultados para cada comparación los siguientes p-valor:

P – valor

Anual vs Semestral : 0.5694 > 0.05

Anual vs Trimestral: 0.2765 > 0.05

Semestral vs Trimestral: 0.1242 > 0.05

NO rechazándose así las hipótesis de igualdad de medianas, indicando así, no haber diferencias significativas entre los desempeños de predicción en los distintos escenarios. Sin embargo siguiendo el análisis grafico se continuara prefiriendo el escenario de 3 meses como el con mejor desempeño de los tres escenarios.

6.4.- Análisis de resultados de métodos de clasificación en la predicción de 7 criptomonedas

En esta parte del análisis de los métodos de clasificación, lo hicimos con 7 criptomonedas diferentes, que son las siguientes: Bitcoin, Ethereum, Polkadot, Monero, Solana, Binance coin y Dogecoin, los datos históricos fueron bajados de la página “Yahoo Finanzas”, ya que, es una buena fuente en lo que son criptomonedas, porque hay bases extensas, y mucho material para trabajar, y también utilizamos la variable “Subidanext”, que es la que buscamos predecir mediante los distintos tipos de métodos utilizados con anterioridad. Quisimos analizar los resultados de métodos de clasificación, en otras 7 criptomonedas diferentes al Bitcoin, con el objetivo de reafirmar los resultados obtenidos anteriormente, o bien ir en busca de nuevas respuestas. Para esto, también trabajamos con distintos escenarios, conformados de la misma manera en que se hicieron los escenarios para Bitcoin.

RESULTADOS OBTENIDOS

Escenario 2 (7 criptomonedas): Base de datos que contempla “1 año”, con 365 datos, y para obtener el resultado de este escenario se utilizó el promedio de 4 años, desde 2018 a 2021.

MÉTODOS vs CRIPTOS	BTC	ETH	DOT	XMR	SOL	BNB	DOGE
SVM	0.65	0.58	0.50	0.55	0.56	0.50	0.42
Arbol de decisión	0.53	0.57	0.53	0.51	0.46	0.52	0.51
KNN	0.61	0.51	0.53	0.51	0.52	0.48	0.55
R. logística	0.63	0.61	0.57	0.58	0.58	0.56	0.54
Naive Bayes	0.59	0.59	0.56	0.51	0.54	0.52	0.47

Tabla 9: Aciertos escenario 2 para 7 criptomonedas

En la tabla de aciertos del periodo de 1 año, tenemos que en la mayoría de las criptomonedas el método que mejor acierta a las predicciones es el de Regresión Logística, con un promedio de aciertos entre todas las monedas analizadas de 0,58.

Escenario 3 (7 criptomonedas): Base de datos semestral, que contempla 180 datos, que, para este escenario, como en el anterior se utilizó el promedio de 8 periodos de 180 datos, desde el 2018 al 2021.

MÉTODOS vs CRIPTOS	BTC	ETH	DOT	XMR	SOL	BNB	DOGE
SVM	0.64	0.54	0.61	0.51	0.49	0.59	0.35
Arbol de decisión	0.55	0.46	0.46	0.46	0.43	0.45	0.46
KNN	0.61	0.55	0.57	0.48	0.45	0.45	0.54
R. logística	0.64	0.62	0.63	0.56	0.58	0.59	0.53
Naive Bayes	0.59	0.54	0.59	0.43	0.51	0.55	0.53

Tabla 10: Aciertos escenario 3 para 7 criptomonedas

Luego para el periodo de 6 meses, el método que mejor acierta nuevamente es Regresión Logística, con un promedio de nivel de acierto entre todas las criptomonedas de 0,59.

Escenario 4 (7 criptomonedas): Base de datos de 3 meses, de 90 datos, que también se utilizó el promedio de 8 periodos, considerando los primeros 3 meses de cada semestre, por 4 años, desde el 2018 al 2021.

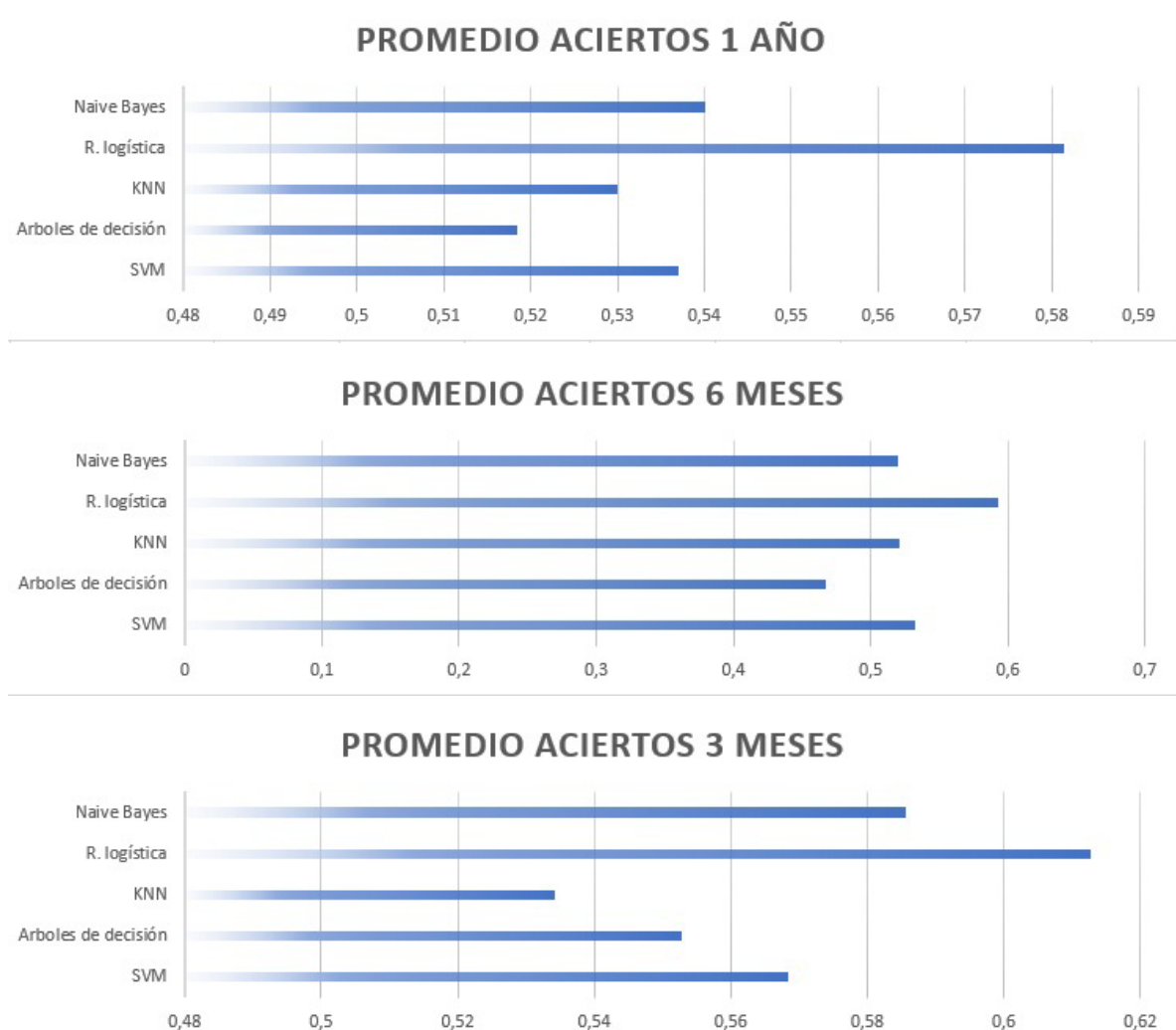
MÉTODOS vs CRIPTOS	BTC	ETH	DOT	XMR	SOL	BNB	DOGE
SVM	0.65	0.76	0.62	0.43	0.55	0.36	0.61
Arbol de decisión	0.55	0.67	0.50	0.57	0.64	0.42	0.52
KNN	0.59	0.62	0.62	0.57	0.41	0.39	0.54
R. logística	0.66	0.66	0.63	0.54	0.62	0.62	0.56
Naive Bayes	0.60	0.62	0.73	0.57	0.64	0.42	0.52

Tabla 11: Aciertos escenario 4 para 7 criptomonedas

Finalmente, para el periodo de 3 meses, tenemos que el método con mayor nivel de aciertos en promedio, entre todas las criptomonedas analizadas, es el de Regresión Logística, con un 0,613 de probabilidad de acierto, seguido por el de Naive Bayes, con un nivel de acierto de un 0,586, al ver la tabla podemos notar que hay varios métodos en este periodo que tuvieron un nivel de predicción alto, pero que como se dijo antes en promedio, sigue siendo el mejor Regresión Logística, porque, aun así su acierto es más elevado que los demás.

Además, se debe mencionar, que en el análisis anterior en que solo se tomaba en cuenta el Bitcoin, arrojó el mayor nivel de aciertos de los métodos en el periodo de 3 meses, que, en los demás periodos, lo que, con este nuevo análisis con 7 criptomonedas lo reafirma, nos dice que los métodos de clasificación en general funcionan de mejor manera con el escenario de 90 datos.

6.5.- Resultados de métodos de clasificación en la predicción de 7 criptomonedas diferentes



Gráfica 11: Resultados de los métodos en 7 criptomonedas y diferentes escenarios

En estos gráficos se puede observar de forma más clara, que en general los mejores métodos de clasificación para predecir son Regresión Logística y Naive Bayes, y que el periodo con el que los métodos tienen más efectividad para predecir es el de 3 meses, llegando a niveles de acierto de 0,63 aproximadamente.

6.6.- Análisis “tiempo real” de los modelos de clasificación

Para poder llevar la aplicación de las observaciones realizadas a partir de nuestros resultados anteriores a una situación lo más apegada a la realidad, se decidió realizar un experimento simulando una situación real la cual es ocupar 3 meses de entrenamiento para predecir uno a uno los valores diarios dentro de un periodo de 50 días, utilizándolo en 3 monedas distintas (Bitcoin, Ethereum y Polkadot) y con 5 modelos de clasificación (SVM, Árbol de decisión, KNN, Regresión logística y Naive Bayes). Los resultados son los siguientes:

MÉTODOS vs CRIPTOS	BTC	ETH	DOT
SVM	0.74	0.52	0.54
Arbol de decisión	0.48	0.56	0.56
KNN	0.72	0.54	0.50
R. logística	0.52	0.12	0.40
Naive Bayes	0.74	0.52	0.54

Tabla 12: Resultados predicciones datos uno a uno

Se pueden observar resultados muy buenos en el Bitcoin lo que lleva a pensar que se realizó de manera correcta los análisis y las decisiones de qué parámetros usar a la hora de utilizar los modelos de predicción. Se puede notar una baja en las otras monedas eso puede ser a causa de que nos enfocamos principalmente en los análisis del Bitcoin y a partir de esa moneda se fue formando el trabajo.

7.- CONCLUSIÓN

7.1.- Observaciones finales

En conclusión, se han cumplido todos los objetivos específicos planteados en el informe, se demostró y con eso se definieron los criterios óptimos para el uso de modelos de clasificación, se obtuvieron los resultados de precisión de los modelos de clasificación, se analizó la heterocedasticidad de la serie temporal, junto a esto se verificó que se cumplen las condiciones para utilizar el modelo GARCH y los resultados de este. Con esta información podemos analizar los pros y contras de cada método para así tomar mejores decisiones a la hora de elegir que camino tomar si se busca predecir el alza o baja del Bitcoin. Las ventajas y desventajas quedan evidenciadas en el informe por lo que pasa a ser a libre uso del lector que métodos utilizar pero principalmente se recomienda ya sea Naive Bayes o Support Vector Machine (SVM) ya que fueron los con mejores resultados y de manera estable a lo largo del estudio decidiendo utilizar periodos de 3 meses para entrenar los modelos y también se recomienda utilizar el método GARCH ya que a pesar de tener un poco menos de porcentaje de asertividad este utiliza la data completa lo que nos da un panorama mas amplio del comportamiento de la criptomoneda y puede estar preparado a cambios mas bruscos debido a la alta heterocedasticidad que maneja. Cabe recalcar que queda por hacer ya que siempre se va innovando en los métodos y siempre se pueden encontrar mejores parámetros debido al comportamiento de este mercado por lo que nunca se podrá llegar a una respuesta concreta, pero sí queda evidenciado que estos modelos y parámetros que se definieron funcionan mucho mejor que la media de forma constante.

7.2.- ¿Cómo continuar este trabajo?

A lo largo del trabajo fueron surgiendo nuevas ideas y recomendaciones las cuales por tiempo y/o falta de recursos se fueron descartando, como la de incluir el index del Euro, variable la cual nos indica el valor en torno a si mismo, para esto se requería hacer Web Scrapping, técnica la cual no estábamos familiarizados, sin embargo se realizó un pequeño análisis con un extracto de esta variable en un escenario de 3 meses donde se mantenía la efectividad de la predicción con nuestras variables iniciales como muestra la tabla 11, pero esto no podía ser afirmado sin mayores análisis con una data más robusta de la variable. Por otro lado, también se podría aumentar la cantidad de fechas predichas por el modelo GARCH ya que podría cambiar si se intenta predecir un mayor numero de observaciones futuras. En resumen, este trabajo aporta mucha información útil a la hora de trabajar con estos modelos de series de tiempo y de clasificación por aprendizaje automático, pero está lejos de ser toda la información adquirible en este ámbito, ya que tal como los datos financieros y de criptomonedas cambian, los modelos también lo hacen y el usuario debe adaptarse a las diferentes condiciones en las que quiere trabajar, aun así, este aporte es un paso importante para comenzar a guiarse por donde afrontar los datos y así llegar a mejores resultados.

MÉTODOS	ACIERTO (P)
SVM	0.57
Arbol de decisión	0.55
KNN	0.52
R. logística	0.54
Naive Bayes	0.56

Tabla 13: Aciertos del escenario 1 con variable Índice del Euro

MÉTODOS	ACIERTO (P)
SVM	0.63
Arbol de decisión	0.62
KNN	0.63
R. logística	0.63
Naive Bayes	0.63

Tabla 14: Aciertos del escenario 1 con las variables seleccionadas

8.- BIBLIOGRAFÍA

- [1] Noémi Uzonyi (2011) Introduction to Data Mining (2020-2021) Cryptocurrencies price prediction by using machinelearning and deep neural networks.
- [2] Ryan Farell (2015) An Analysis of Cryptocurrency Industry. Wharton Research Scholars.
- [3] Xavier Martinez Barta (2018) RLM VI. Heterocedasticidad.
- [4] Raúl Arellano Lopez (2022) Análisis de eficiencia del mercado de Bitcoin. Kaggle.
- [5] Bitcoin (2022) Yahoo Finance.
- [6] Antonio Martinez Jenkins (2016) Modelo GARCH para la volatilidad de las criptomonedas Bitcoin y Ether. Predicción en Economía – Master de Investigación – UNED
- [7] Ruey S. Tsay (2013) An Introduction Tox Analysis Of Financial Data With R. Wiley
- [8]Rafael Carrión (2020) Trabajo de Fin de Grado, Prediccion de Precios de Criptomonedas con Random Forest