



COMILLAS
UNIVERSIDAD PONTIFICIA
John

Facultad De Ciencias Económicas y Empresariales

Big Data en la valoración de Bitcoin

Alumno: Álvaro Yáguez de Vargas
Tutor: Leandro S. Escobar Torres

MADRID | marzo de 2023

RESUMEN

El trabajo que se presenta consiste en el análisis de la valoración de Bitcoin y las variables que la afectan para la posterior elaboración de un modelo predictivo con el programa SPSS de IBM. En la primera parte del trabajo se realiza un breve estudio del contexto del trabajo explicando qué son las criptomonedas y, en particular, qué es el Bitcoin, esto ayuda a sentar la base del trabajo. Después, se estudian las aplicaciones del Big Data explicando algunos ejemplos y se lleva a cabo un breve estudio de los modelos predictivos más utilizados en el sector de las finanzas.

El modelo predictivo utilizado consiste en un modelo ARIMA (1,1,0) con logaritmos lo cual permite reducir la varianza y asimetría mejorando la precisión del modelo. Después se realiza una explicación y análisis individual de las diferentes variables midiendo los resultados de los estadísticos del t-valor y de sig. o el P-valor, lo mismo ocurre con el modelo en general, este se evalúa e interpreta basándose en FAS Residual, FAP Residual, RMSE, MAPE, MAE, BIC Normalizado, R^2 y la prueba de Ljung Box. Por último, se explican las conclusiones, las implicaciones prácticas del trabajo y las limitaciones y recomendaciones para el futuro. Al final del trabajo se encuentra la bibliografía y los anexos utilizados a lo largo de la explicación.

ABSTRACT

This Project that is being presented consists of the analysis of Bitcoin's valuation and the variables that affect it for the posterior elaboration of a predictive model with the program SPSS by IBM. In the first part of the project, the context is briefly explained by describing what cryptocurrencies are, and in particular what Bitcoin is; this helps to set the base for a better understanding of the project. After, the applications of Big Data are studied and there are some examples, then there is a short study of the most common predictive models used in finance.

The predictive model used consists of an ARIMA (1,1,0) with logarithms which allows a reduction of the variation and asymmetry improving the precision of the model. After this, an explanation and an individual analysis of each variable measuring the results of t-value and P-value was carried out. The same is done with the predictive model, it was evaluated based on Residual FAS, Residual FAP, RMSE, MAPE, MAE, Normalized BIC, R^2 and the Ljung Box test. The last part of the project consists of a conclusion, an explanation of the practical implications of the project and the limitations and recommendations or the future. At the end of the document there is a bibliography and the figures used for the explanations.

PALABRAS CLAVE

Bitcoin, Big Data, Modelo Predictivo, Predicción, Valoración Bitcoin, Criptoactivos, Criptomonedas, ARIMA, SPSS.

KEY WORDS

Bitcoin, Big Data, Predictive Model, Prediction, Bitcoin Valuation, Criptoassets, Cryptocurrencies, ARIMA, SPSS

ÍNDICE

RESUMEN	2
ABSTRACT	3
PALABRAS CLAVE	3
KEY WORDS	3
ANEXOS	5
1. Introducción:	6
1.1 Contexto y justificación del trabajo:.....	6
1.2 Objetivos del trabajo:	9
1.3 Preguntas de investigación	10
1.4 Hipótesis de investigación.....	10
2. Revisión de la literatura:	11
2.1 Definición de Big Data.....	11
2.2 Aplicación del Big Data en finanzas	12
2.3 ¿Qué son las criptomonedas?	14
2.4 La valoración de Bitcoin:	16
2.5 Revisión de los modelos predictivos más utilizados en finanzas	18
3. Metodología:	20
3.1 Descripción de los datos utilizados y su fuente.....	20
3.2 Proceso de preprocesamiento de los datos	22
3.3 Descripción del modelo predictivo utilizado y su justificación	23
3.4 Evaluación del modelo predictivo	25
4. Resultados y discusión:	29
4.1 Análisis exploratorio de datos.....	29
4.2 Resultados del modelo predictivo	33
4.3 Discusión de los resultados obtenidos.....	37
5. Conclusiones y recomendaciones:	38
5.1 Conclusiones obtenidas	38
5.2 Limitaciones del estudio y recomendaciones para futuras investigaciones.....	39
5.3 Implicaciones prácticas del estudio	40
6. Bibliografía	42
7. Anexos	44

ANEXOS

Figura 1: Gráfico de predicciones y datos reales. Elaboración Propia	44
Figura 2 Gráfico FAS residual y FAP residual. Elaboración Propia	44
Figura 3 Tablas Estadísticos de evaluación. Elaboración Propia	44
Figura 4 Resultado por semana. Elaboración Propia.....	45
Figura 5: Resultados por cada variable. Elaboración Propia.....	46

1. Introducción:

1.1 Contexto y justificación del trabajo:

En la actualidad, el mercado de criptomonedas se ha convertido en un tema de gran interés para inversores, empresas, gobiernos y la comunidad financiera en general. Desde la creación de Bitcoin en 2009, la cantidad de transacciones en el mercado de las criptomonedas ha crecido de forma exponencial, según los datos de Blockchain.com (Blockchain.com, 2023) el día 17 de enero de 2009 se registraban las primeras transacciones de bitcoin logrando un total de 109 transacciones mientras que en febrero de 2022 se llegó a cerca de 290.000 transacciones diarias. Esto representa el gran aumento en interés, uso y por tanto importancia de las criptomonedas y como ha crecido la adopción de estas entre el público general.

En un comienzo, el uso de bitcoin se veía restringido a una parte pequeña de la sociedad entre ellos a informáticos bien informados y la comunidad de videojuegos que implementó las criptomonedas como método de pago en los propios videojuegos e incluso como recompensa dentro de estos en algunos casos. Tras la creación de Bitcoin hubo un “boom” de las criptomonedas y a día de hoy, según criptonoticias.com (Vanci, 2022), en febrero de 2022 existían más de 20.000 criptomonedas diferentes en el mundo, mostrando que el uso de las mismas se ha abierto notablemente al público logrando así el objetivo de las mismas, ya que cuanto mayor sea el número de usuarios mayor funcionalidad y seguridad tienen las criptomonedas gracias al efecto “network”.

Se ha decidido realizar el trabajo sobre Bitcoin en particular puesto que, no solo fue la primera criptomoneda lanzada al público, sino que se considera la base de las criptomonedas y, de entre 11.143 criptomonedas monitoreadas por coingecko.com (CoinGecko.com, 2023), Bitcoin goza de una capitalización de mercado de cerca del 45% del total, dándole un puesto de grandísima relevancia en el mercado de las criptomonedas y convirtiéndola en la referencia en el sector. En el gráfico puesto a disposición por la página web anteriormente

mencionada, se puede observar como la capitalización de mercado de todas las criptomonedas reflejadas siguen la misma tendencia, por lo que se puede asumir que los eventos afectan al mercado de las criptomonedas de forma muy similar independientemente del tipo de moneda y su capitalización de mercado.

En septiembre de 2021 El Salvador se transformó en el primer país en convertir el Bitcoin en una moneda legal en curso y fue seguida de la República Centroafricana en abril de 2022. Este acontecimiento resalta, de nuevo, el aumento en interés y el gran impacto que podría tener la criptomoneda en el futuro de la sociedad en su conjunto. Al prescindir de los gobiernos para la supervisión de las divisas, se crea un verdadero mercado libre en el cual la cantidad de uso, el interés de los usuarios y la oferta de la criptomoneda son los que dan valor a Bitcoin; y al existir un número limitado de bitcoins, se pasa de una economía inflacionaria a una economía deflacionaria, lo cual puede tener numerosos desenlaces en función de las políticas adoptadas por los diferentes órganos de gobierno.

También son muchas las empresas que han pasado a aceptar Bitcoin y otras criptomonedas como formas de pago como por ejemplo Norwegian, Expedia, Microsoft y Starbucks, lo cual marca una tendencia hacia la adopción de las criptomonedas. Por otra parte, también son muchos los bancos que se suman a la causa ofreciendo servicios de enseñanza a los clientes menos familiarizados, mediante los cuales facilitan el entendimiento de las criptomonedas, su funcionamiento y asesoramiento de inversiones en los mismos. Además, comienzan a existir cajeros de bitcoin, los cuales permiten extraer dinero en la divisa *fiat* correspondiente tras el intercambio de Bitcoins. Empieza también a ser común ver diferentes negocios que ofrecen servicios relacionados con Bitcoin y otras criptomonedas populares, entre estos servicios enseñan a las personas menos afines a la tecnología sobre la funcionalidad de las criptomonedas, ayudan a realizar compras y ventas de estas y a gestionar las carteras de los clientes aconsejando acerca de qué hacer con sus inversiones. Estos últimos

negocios ayudan a difundir la funcionalidad de las criptomonedas y por tanto es una oportunidad para aumentar la popularidad de estas.

Tras la crisis actual del mercado financiero, generada por la quiebra de los bancos Silicon Valley Bank y First Republic, el valor de bitcoin ha visto un incremento de alrededor del 13%, los inversores que poseen depósitos de valor en bancos pequeños o medianos han traspasado dichos depósitos a carteras de criptomonedas particularmente de Bitcoin, esto muestra como la sociedad empieza a desconfiar de los bancos y comienza a adoptar las criptomonedas como los nuevos depósitos de valor que podrían sustituir parcialmente a los bancos más pequeños. Al ser Bitcoin un activo deflacionario, los inversores lo consideran una buena decisión trasladar sus inversiones a carteras de bitcoin.

Hoy en día existen múltiples herramientas que ayudan a digerir la información encontrada en enormes bases de datos, dichas herramientas ayudan a su filtración, interpretación y posterior uso; aumentando de esta manera la cantidad de información que puede ser utilizada tanto por individuos, como por empresas e incluso el gobierno, ayudando de esta manera, a que todos los anteriores puedan tomar mejores decisiones. Las aplicaciones del Big Data son amplias y se verán algunos ejemplos posteriormente.

El uso del Big Data en la valoración de criptoactivos puede ser una herramienta valiosa para los inversores que buscan tomar decisiones informadas ya que les permitiría identificar patrones, analizar el sentimiento del mercado, monitorear riesgos e identificar oportunidades entre otras muchas cosas. El objetivo de este trabajo es desarrollar un modelo predictivo utilizando técnicas de Big Data y el programa SPSS de IBM para la valoración de criptoactivos. Dicho modelo predictivo estará basado en los datos del valor de Bitcoin desde 2019 hasta 2021, los cuales se tomarán como base para el modelo y los datos de 2022 serán utilizados para evaluar la exactitud del modelo.

1.2 Objetivos del trabajo:

El objetivo principal del trabajo es desarrollar un modelo predictivo utilizando técnicas de Big data y SPSS para la valoración de criptoactivos, esto implicará la creación de un modelo de regresión que tenga en cuenta las variables identificadas y que pueda predecir el valor de Bitcoin en función de esos factores. Los objetivos específicos son identificar y recopilar las variables más relevantes para la valoración de Bitcoin, preprocesar los datos necesarios para el análisis, aplicar técnicas de análisis predictivo para predecir los precios de Bitcoin y evaluar la precisión del modelo comparando las predicciones con los valores reales de 2022. La precisión buscada en el modelo será de un valor por cada semana del año, que, a la hora de evaluar la precisión, será comparado con los valores reales de las correspondientes semanas del año 2022.

Por otra parte, se busca la capacidad de analizar el modelo predictivo basándose en los diferentes estadísticos puestos a disposición por el programa utilizado, SPSS, y la interpretación de estas. Se apunta a conseguir datos de las variables relevantes para la valoración de la criptomoneda provenientes de fuentes fiables y que estén en un formato que permita su importación al propio programa para su posterior análisis. Los datos se obtendrán de fuentes fiables y se contrastarán entre diferentes bases de datos para comprobar la veracidad de estos y de esta forma asegurar el correcto funcionamiento y precisión del modelo predictivo creado.

En el caso de este trabajo la cantidad de datos utilizados será inferior a lo que los expertos considerarían como Big data, dadas las limitaciones encontradas tanto en el equipo disponible para la realización del trabajo como en la cantidad de datos accesibles; por lo tanto se considerará que el modelo realizado será una simulación en pequeña escala de los resultados que se podrían conseguir si la cantidad de datos fuera mayor y si el equipo contase con la potencia necesaria para el procesamiento de los mismos. Esto quedará posteriormente reflejado en las conclusiones y se tendrá en cuenta a la hora de valorar la precisión del modelo.

Se utilizará el programa Excel para la filtración de los datos ya que el programa SPSS permite importar documentos de dicho formato. Además, la mayor parte de los datos accesibles coinciden con dicho formato y gracias a las fórmulas de Excel, la filtración y ordenación de estos se hará más rápida y eficazmente.

1.3 Preguntas de investigación

La pregunta principal que se intentará responder en el trabajo es ¿Cómo se pueden utilizar las técnicas de Big Data y SPSS para desarrollar un modelo predictivo para la valoración de criptoactivos?; esta pregunta se responderá junto a otras 5 preguntas de investigación, que son: ¿Qué variables son más relevantes para la valoración de Bitcoin? ¿Qué fuentes deben ser utilizadas para obtener los datos? ¿Qué estadísticos han de ser utilizados para comprobar tanto la bondad de ajuste como la precisión del modelo? ¿Qué tan precisos son los resultados del modelo predictivo en la valoración de Bitcoin? ¿Qué variantes tiene mayor peso en la valoración de Bitcoin?

1.4 Hipótesis de investigación

Las expectativas del trabajo son las de crear un modelo predictivo funcional a corto o medio plazo, el cual sea capaz de predecir los valores futuros de Bitcoin interpretando y analizando las variaciones y las posibles tendencias. Dicho modelo funcionará de forma más precisa si los datos son actualizados semanalmente, es por eso por lo que se arrastrarán los datos, esto significa que los datos empleados para la semana 2 serán aquellos correspondientes a la semana 1 puesto que son los únicos que se tendrán al alcance del usuario; de esta forma se previene que un posible error del modelo vaya en aumento exponencial con el paso del tiempo

Se espera que las variables más relevantes para la valoración de criptoactivos sean aquellas relacionadas con la oferta y demanda de Bitcoin, así como con las condiciones del mercado. Además, se espera que el modelo predictivo

desarrollado utilizando técnicas de Big Data y SPSS sea preciso en la valoración de Bitcoin dejando espacio a los posibles acontecimientos inesperados que puedan darse a lo largo del año. Cabe destacar que es posible que, de todas las variables introducidas en el modelo, exista alguna de ellas cuya relevancia para la valoración de Bitcoin sea insuficiente como para ser tomada en cuenta en el modelo, esto se reflejará con los resultados del t-test y el valor de sig.

Se espera que, una vez obtenidas las predicciones del modelo y se puedan comparar con los valores reales, se pueda observar una semejanza principalmente en las tendencias, aunque los valores de las predicciones no sean exactos es de gran interés que el modelo sea capaz de predecir las fluctuaciones al alza y a la baja del valor de Bitcoin, mostrando de esta forma que es capaz de relacionar las variaciones de las diferentes variables con los cambios en la valoración de Bitcoin; la medida en la que los relaciona puede ser modificada ya sea con la introducción de más datos en un futuro o con una calibración del modelo.

2. Revisión de la literatura:

2.1 Definición de Big Data

El Big Data, según powerdata.es se define como *“conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales”* (PowerData, s.f.). El Big Data se caracteriza por las 4 Vs: Volumen, Variedad, Velocidad y Veracidad.

El volumen hace referencia a la cantidad de datos que se manejan, estos datos son generados automáticamente ya sea por máquinas, redes, algoritmos o interacciones en sistemas como redes sociales. Existen infinitas fuentes de datos y la capacidad de recoger dichos datos aumenta con las innovaciones, la más reciente y posiblemente la más importante es la inteligencia artificial, que no

solo es capaz de recoger la información sino de interpretarla, analizarla e incluso clasificarla en diferentes categorías para facilitar su uso. Dicha herramienta se ve utilizada a diario por grandes empresas, algunas de las cuales trabajan en crear algoritmos abiertos al público que permiten acceder a las fuentes de datos capturadas por la inteligencia artificial, generando de esta forma nuevos datos como puede ser la cantidad de búsquedas sobre un tema o el interés mostrado por los usuarios.

En cuanto a variedad, a diferencia de hace años cuando los datos a los que se tenía acceso eran limitados y por tanto filtrados, hoy en día las fuentes de datos son innumerables y por ello hay una variedad incalculable de datos que pueden ser usados para el análisis. Es tanta la variedad que es extremadamente complicado trabajar con los datos si no son clasificados con anterioridad ya sea a través de algoritmos o inteligencias artificiales. Existen algoritmos que son capaces, de forma autónoma, de recopilar información sobre un tema específico de plataformas como redes sociales o blogs.

La velocidad hace referencia a la capacidad con la que los ordenadores y procesadores de hoy en día pueden analizar dichos datos, gracias a los avances tecnológicos cada día es más fácil hacer uso de enormes cantidades de datos en menos tiempo, se estima que Google procesa más de 20 petabytes (20.000 terabytes) al día. (PowerData, s.f.)

La veracidad, a la hora de realizar estudios, es fundamental; los datos que se extraen y que son utilizados para dichos estudios deben ser verdaderos y fiables. El sesgo, el ruido y la posible alteración son elementos que pueden interferir con los resultados de estudio. Para realizar un estudio han de utilizarse datos limpios y analizar las bases de datos para deshacerse de posibles “datos sucios”.

2.2 Aplicación del Big Data en finanzas

Para hacer uso del Big Data en el ambiente financiero es importante la validación de los datos, este debe ser considerado el primer paso antes de comenzar

cualquier estudio para verificar la fiabilidad del mismo. En cuanto a las posibles aplicaciones del Big Data en finanzas, los más destacables son: Análisis de riesgos, análisis de mercado, marketing y análisis del consumidor y detección de fraudes.

El análisis de riesgos es una parte fundamental de la gestión financiera. El Big Data permite a los inversores y gestores de riesgos identificar patrones y tendencias en los datos históricos de mercado y en la información actual del mercado. Esto ayuda a los inversores a evaluar el riesgo potencial de sus inversiones y a tomar decisiones más informadas. El Big Data supone una gran revolución puesto que nunca antes se había tenido acceso a tanta cantidad de información de forma tan rápida y en el mejor de los casos organizada. Además, gracias a diferentes programas como puede ser SPSS, se pueden buscar correlaciones y patrones entre los datos para facilitar la interpretación de los mismos y de esta manera identificar patrones.

En cuanto al análisis de mercado, el Big Data permite a los analistas de mercado comprender mejor los patrones y tendencias de los mercados financieros. Los analistas pueden utilizar herramientas de Big Data para recopilar y analizar datos de múltiples fuentes, incluyendo datos económicos, datos de precios de acciones y datos de redes sociales, entre otros; esto les permite comprender mejor los movimientos del mercado para poder tomar mejores decisiones con más y mejor información.

Las empresas hacen uso de dichos datos para tomar decisiones informadas acerca de sus estrategias comerciales, se pueden obtener datos tanto de fuentes internas, como la propia página web de la compañía, como externa, de redes sociales o bancos de datos. El Big data también puede utilizarse en el marketing y análisis del consumidor para comprender mejor los comportamientos de estos. Las empresas financieras pueden utilizar datos de transacciones de clientes y datos de redes sociales para comprender mejor las necesidades y deseos de sus clientes. Este comportamiento les permite adaptar sus ofertas de productos y

servicios para satisfacer mejor las necesidades de sus clientes y aumentar su lealtad; pero también les puede servir para identificar y adelantarse a nuevas tendencias y llevar a cabo las inversiones pertinentes.

Los fraudes afectan a un gran número de empresas financieras y es por ello por lo que el Big Data se utiliza cada vez más para detectar fraudes financieros. Los bancos y otras instituciones financieras pueden utilizar el Big Data para analizar patrones de transacciones y detectar transacciones inusuales o sospechosas que pueden indicar fraude o actividades ilegales.

Otra de las aplicaciones consiste en saber la calificación crediticia a tiempo real de los clientes, hoy en día las calificaciones de crédito se actualizan cada mes aproximadamente (White, 2023), por tanto, los bancos no tienen acceso completo a todo el historial del prestatario. Con la tecnología de Big Data, esta información será actualizada constantemente y por tanto, los bancos tendrán acceso a la puntuación en tiempo real. (Trelewicz, 2017)

2.3 ¿Qué son las criptomonedas?

“Una criptomoneda es una divisa que utiliza la criptografía para generar una forma de dinero “codificado” que se sustenta en la tecnología digital y no depende de la intervención de un organismo centralizado para su funcionamiento ni regulación” (Ronco & Callejo, 2020). Existen diferentes tipos, entre ellas destaca Bitcoin como primera criptomoneda sacada al público y tras la cual se crearon numerosas alternativas, dichas alternativas se llaman Altcoins; el propio nombre sugiere que son monedas alternativas. Por otra parte existen las llamadas memecoins, la definición de las mismas según materalbitcoin.com es *“ una criptomoneda que se basa en el concepto de un meme y que nace como alternativa de inversión de dinero fácil”* (Rivadulla, 2022), estas monedas experimentaron un boom después de que grandes figuras como Elon Musk las mencionaran en sus redes sociales y se observasen exponenciales subidas de precio en las mismas, Tesla llegó a aceptar la memecoin Dogecoin como método de pago para sus vehículos.

Los aspectos fundamentales que diferencian a las criptomonedas de las monedas tradicionales o fiduciarias son: la falta de organismo regulador, la escasez, la ausencia de intermediarios y la inmediatez. Todo lo anterior ha significado un cambio que no es plenamente aceptado por la sociedad por el momento; las criptomonedas por tanto actualmente se consideran un activo para inversores especulativos, aunque recientemente tras la crisis bancaria se ha visto un incremento considerable del precio de Bitcoin, lo cual indica la falta de confianza por parte de un gran número de inversores en los bancos de tamaño pequeño o mediano, que tras ver lo ocurrido con bancos como Silicon Valley Bank o First Republic, han trasladado sus depósitos de ahorro a Bitcoin; convirtiendo a la criptomoneda en una caja de ahorro para estos usuarios. La naturaleza deflacionaria de la criptomoneda es una de las motivaciones de los inversores en cuanto a esta toma de decisión, al existir un número limitado de Bitcoins, 21.000.000, a medida que se generan las criptomonedas, el precio de las mismas incrementa hasta llegar al máximo.

Bitcoin, al igual que el resto de las criptomonedas, no tiene un órgano regulador, las regulaciones son públicas e inamovibles y los protocolos se establecieron en la fecha de su creación. Estos protocolos son, por ejemplo, el “halving”, lo cual es un procedimiento por el cual, cada 210.000 bloques creados en la Blockchain (esto corresponde a 4 años) se divide entre dos la recompensa de los mineros cada vez que resuelven un bloque; otro protocolo es el máximo número de bitcoins, que está estipulado que serán 21.000.000. Esto hace que bitcoin sea una moneda deflacionaria, con el paso del tiempo y el aumento de la demanda, el valor de bitcoin sube dado que la oferta tiene un máximo establecido. (Maldonado, 2020)

El funcionamiento de bitcoin y de su red de Blockchain no requiere de intermediarios, las transacciones son anónimas, pero sin perder la confianza gracias al proof-of-work. Los usuarios no han de conocerse entre ellos para realizar transacciones, con el número de cartera es suficiente. Cabe destacar que

todas las transacciones son públicas y cualquiera puede descargarlas, pero en el registro todas las transacciones son anónimas y no son rastreables. Todas las transacciones son procesadas directamente por la red de Bitcoin.

Al contrario que las transacciones ordinarias a través de los bancos, las criptomonedas gozan de transacciones inmediatas o casi, en el caso de bitcoin, las transacciones se realizan cada 10 minutos que es lo que se tarda en generar un nuevo bloque; pero esto supone un gran aumento en rapidez respecto a los bancos si se trata de operaciones entre diferentes bancos, en caso de transacciones interbancarias la red de Bitcoin tiene una desventaja. Además, la creación de métodos de pago como Bizum también amenazan la ventaja de la que gozaba Bitcoin en cuanto a velocidad de transacción.

Actualmente la criptomoneda bitcoin sufre de un problema de escalabilidad, las transacciones no son instantáneas ya que los bloques se cierran cada diez minutos aproximadamente y la cantidad de transacciones registradas por bloque es de 2048 (Pastor, 2016). La media de transacciones por segundo es de 7 lo cual es un número muy pequeño como para ser considerada la criptomoneda como una moneda de cambio, es decir, la red de Bitcoin no tiene la infraestructura necesaria como para soportar la cantidad de transacciones que podrían realizarse si fuese considerado un método de pago mundialmente, la red colapsaría. Es por eso por lo que se está trabajando en una mejora de la red y la razón por la que existen otras criptomonedas que se diferencian de Bitcoin en este aspecto, por ejemplo, se espera el lanzamiento de Ethereum 2.0 la cual sería capaz de tramitar hasta 100.000 transacciones por segundo (Worldcoin, 2022), superando ampliamente las 7 que Bitcoin ofrece actualmente.

2.4 La valoración de Bitcoin:

Al ser una moneda pública, la cual es considerada un activo por muchos inversores, a la que cualquier persona puede tener acceso, es difícil determinar qué factores en concreto le dan el valor a Bitcoin. El valor de bitcoin depende principalmente de: la ley de oferta y demanda, el sentimiento de los inversores,

la capitalización, la cantidad de bitcoins en el mercado, la usabilidad de la moneda y la complejidad de minado.

Como bitcoin es una moneda que puede ser adquirida por cualquier persona, el valor de esta queda en gran medida determinada por la dinámica de la oferta y la demanda. Al haber un número máximo de bitcoins existentes, la oferta tiene un límite, y a medida que vaya subiendo la demanda el precio lo hará también, esto se conoce como una moneda deflacionaria. Esto no significa que no haya Bitcoins suficientes para todos ya que un Bitcoin puede quedar dividida hasta en 100 millones de partes denominadas "Satoshi" (George, 2023).

El sentimiento de los inversores es lo que mueve la oferta y la demanda, dentro del sentimiento de los inversores podemos encontrar muchos factores que afectan: las tendencias en redes sociales, las noticias sobre la moneda, los proyectos que surjan en la Blockchain y la usabilidad de esta. Existen algoritmos capaces de contabilizar dicho sentimiento permitiendo medirlo y por tanto utilizarlo como variable en el modelo predictivo.

La capitalización de bitcoin representa cerca del 45% de la capitalización total de las criptomonedas (CoinGecko.com, 2023), es decir el valor conjunto de todos los Bitcoins representa más de una tercera parte del valor total del conjunto de todas las criptomonedas. Este factor afecta al valor de Bitcoin ya que le proporciona seguridad a los inversores y una estabilidad difícil de conseguir con otras criptomonedas más inestables e inseguras con menos estructuras. A pesar de esta aparente dominancia frente a otras criptomonedas, la volatilidad de Bitcoin es un factor importante a tener en cuenta a la hora de invertir en ella, ya que sufre variaciones en el precio importantes en un tiempo muy reducido, está en la razón por la que les resulta tan atractiva a algunos inversores, sobre todo aquellos que especulan acerca de dichas variaciones de precio.

El número de bitcoins en el mercado afecta directamente a la oferta de estas por lo que entra dentro de la ley de oferta y demanda afectando de esta manera al

precio de la criptomoneda. El número máximo de Bitcoins está estipulado en 21.000.000 pero dicha cantidad no está disponible en el mercado, son los mineros los encargados de minar las criptomonedas y se calcula que hasta 2140 según Cointelegraph.com (Jimenez, 2019) no se acabará de minar completamente.

Una moneda ha de poder ser utilizada, y en este caso la usabilidad de las criptomonedas afecta a su precio, cuantos más establecimientos, gobiernos y entidades acepten bitcoin como método de pago, más popular se hará la misma y más demanda habrá, aumentando de esta forma el precio. Actualmente la usabilidad de Bitcoin está aumentando gracias a la adopción por parte de las empresas de la criptomoneda como método de pago, e incluso la aparición de cajeros que intercambian Bitcoins por divisas *fiat*.

Desde la creación de bitcoin el número de mineros se ha visto exponencialmente aumentado, a medida que ganaba popularidad la moneda, aumentaban las inversiones en minado, pero a causa del “halving” cada vez es menos rentable minar, cada vez requiere más energía y la recompensa es menor. Esto también hace que la oferta sea menor por tanto se aprecia el efecto deflacionario de la moneda. Esta complejidad puede ser medida con el llamado Hashrate, consiste en medir el número de intentos de descifrar el bloque que hace cada ordenador por segundo.

2.5 Revisión de los modelos predictivos más utilizados en finanzas

Desde hace unos años, se ha visto un aumento en la popularidad y la utilización de Big Data y “Machine learning” lo cual ha permitido llevar a cabo la creación de múltiples tipos de modelos que son utilizados en el sector de las finanzas. Entre estos modelos destacan: análisis de regresión, modelos de series temporales, redes neuronales artificiales, árboles de decisión y modelos de riesgo.

El análisis de regresión es un modelo predictivo utilizado para analizar la relación entre dos o más variables, *“permite realizar proyecciones y pronósticos de una variable dependiente explicada por una o más variables independientes”*. (González, 2017) En finanzas, el análisis de regresión se utiliza para predecir el precio de las acciones de una empresa basándose en variables como el desempeño financiero de la empresa, los cambios en las tasas de interés, y otros factores relevantes. (Galán, 2020)

Los modelos de series temporales son modelos que se utilizan para predecir movimientos futuros de los mercados financieros basados en datos históricos. Estos modelos se basan en la idea de que los movimientos del mercado son cíclicos y que las tendencias pasadas pueden ser utilizadas para predecir el comportamiento futuro del mercado. (López, 2018)

Las redes neuronales artificiales *“Son algoritmos matemáticos que emulan la estructura de funcionamiento del cerebro humano.”* (Rospide) que se utilizan para analizar grandes cantidades de datos y encontrar patrones y tendencias. En finanzas, las redes neuronales artificiales se utilizan para predecir el precio de las acciones de una empresa basándose en una amplia variedad de variables.

Los árboles de decisión son modelos predictivos que se utilizan para analizar datos y tomar decisiones basadas en ellos. En finanzas, los árboles de decisión se utilizan para predecir el comportamiento del mercado y las tendencias basándose en datos históricos y otros factores relevantes.

Los modelos de riesgo son modelos predictivos que se utilizan para evaluar el riesgo financiero y el rendimiento de una empresa o una inversión. Estos modelos se utilizan para predecir los posibles riesgos y recompensas de una inversión y ayudan a los inversores y gestores de riesgos a tomar decisiones informadas. (Izquierdo, 2000)

3. Metodología:

3.1 Descripción de los datos utilizados y su fuente

Para el desarrollo del modelo predictivo para la valoración de Bitcoin, se ha recopilado un conjunto de datos presuntamente relevantes. Los datos seleccionados son: valor de BTC-USD, valor máximo y mínimo de BTC-USD, volumen de transacciones, hashrate, datos de tendencias y datos de sentimiento. Estos datos son seleccionados puesto que, tras el estudio de la valoración de Bitcoin, las principales variables que afectan al precio de bitcoin son las expresadas en los datos. También influye en la decisión la disponibilidad de los datos, puesto que muchos datos no son públicos y no pueden ser utilizados.

Los datos del valor de BTC-USD fueron descargados de la base de datos de Investing.com (Investing.com, 2023), dichos datos reflejaban el valor semanal de Bitcoin desde el año 2019 hasta el 2022. En esta misma base de datos quedaban también reflejados los valores máximos y mínimos semanales, el volumen de transacciones y se incluía el porcentaje de variación respecto a la semana anterior. Está será la base del modelo, ya que se utilizarán los datos de 2019, 2020 y 2021 para que el modelo tenga una base de la cual pueda estudiar e identificar los diferentes patrones y los efectos que pueden tener las diferentes variables y sus movimientos con respecto a la variación del precio de Bitcoin a lo largo de los tres años. Los datos del precio de 2022 serán excluidos del modelo dejándole al modelo el trabajo de predecir dichos precios y sucesivamente se compararán las predicciones del modelo con los datos reales para valorar la precisión real del modelo. Estos datos serán comparados gráficamente para poder ver de forma visual las semejanzas y diferencias de los valores predichos y los reales, pudiendo de esta forma percibir si el modelo es capaz de predecir las fluctuaciones y las tendencias de la criptomoneda tanto en la cantidad de varianza como en el momento en el que ocurren.

Para obtener los datos del hashrate desde 2019 hasta 2022, se utilizó la base de datos de Ycharts.com (YCharts, 2023) la cual proporcionaba datos diarios de *“la cantidad de operaciones computacionales que un minero, o más bien una red de mineros en su conjunto, es capaz de realizar para resolver los acertijos criptográficos que son necesarios para generar nuevas criptomonedas.”* (Santander, s.f.). Esta variable es relevante ya que si el esfuerzo requerido para la resolución del acertijo es elevado y se mantiene una recompensa que no lo es, la cantidad de mineros se verá reducida y por tanto la cantidad de operaciones realizables disminuirá, a su vez aumentarán las comisiones que los mineros obtienen por asegurar la realizaciones de diferentes operaciones; al introducir esta variable en el modelo se consigue estudiar si dichos sucesos son, o no, relevantes en cuanto a la variación del precio de la criptomoneda.

Google trends fue la herramienta utilizada para los datos de tendencia, esta herramienta permite acceder a las búsquedas de una palabra. Los datos van en función del valor máximo en el periodo elegido, siendo el punto más alto aquel marcado con un 100 y los datos restantes obtienen su valor en proporción a su diferencia con el dato mayor. Esto crea una variable que se muestra en forma de escala permitiendo determinar los altibajos en las búsquedas relacionadas con Bitcoin y permitiendo de esta manera interpretar si el aumento o disminución del número de búsquedas está relacionado con los cambios de precio de la criptomoneda. (Google, 2023)

Por último, se obtuvieron datos del sentimiento sobre Bitcoin desde la herramienta de augmento.ai (Augmento, 2023), esta herramienta utiliza la inteligencia artificial para recopilar información de las redes sociales Twitter, Reddit y BitcoinTalk, y a través de su algoritmo clasifica las publicaciones en una serie de categorías. Los datos están expresados como el número de publicaciones analizadas y clasificadas según las diferentes categorías. De todas las diferentes categorías se escogieron un total de 7 en cada una de las redes sociales siendo estas: “bearish”, “bullish”, “bad news”, “Good news”, “Positive”, “Optimistic” y “Negative”. Con la inclusión de estas variables se permite estudiar

el efecto de los diferentes sentimientos que las personas publican en las diversas redes sociales con respecto a la variación del precio de Bitcoin.

3.2 Proceso de preprocesamiento de los datos

El primer paso que hubo de realizarse fue clasificar los datos temporalmente de la misma manera puesto que había datos en intervalos de horas, días y semanas. Se decidió aplicar las semanas como medida temporal puesto que era la medida más grande y encajaba con el objetivo del proyecto, el cual consistía en predecir el valor semanal de Bitcoin en el año 2022, de esta manera se podrán comparar los resultados obtenidos con los valores reales. El proceso para clasificar los datos por semanas consistió en colocar un número en cada celda que correspondiese a la misma semana, los números van desde el 1 hasta el 208 contabilizando así todas las semanas de los 4 años desde el 2019 hasta el 2022, esto permitiría poder aplicar fórmulas como sumar.si o promedio.si en Excel para facilitar el procesamiento de los datos.

Los datos clasificados por semanas se ordenaron por orden ascendente numerando las semanas del 1 al 208; mientras que los datos clasificados por días y por horas fueron alterados. Para los datos de varianza, se calculó manualmente mediante la fórmula $((\text{Semana } 2 / \text{Semana } 1) - 1)$, esta variable será de gran ayuda ya que el modelo ARIMA (1,1,0) hace uso de la muestra anterior al dato predicho por tanto al introducir esta variable se podrá también mostrar al modelo cómo ha evolucionado el valor respecto a la semana anterior a la del dato mostrado. Para poner un ejemplo, los datos observados en la semana 3 corresponde a la varianza existente entre el valor de la semana 2 y aquel de la semana 1, se hace de esta forma ya que en un caso real este sería el dato al que el usuario tendría acceso en ese preciso momento.

Para los datos sacados de las redes sociales Twitter, BitcoinTalk y Reddit, las cuales mostraban publicación relativas a los sentimientos tanto negativos como positivos acerca de Bitcoin, se llevó a cabo la suma de las publicaciones contabilizadas por cada semana. Esto se hizo a través de la fórmula sumar.si en

Excel, gracias a la fórmula, se sumaron las 168 horas que hay en cada semana agrupando todas las sumas en su correspondiente semana.

En cuanto a los valores de Bitcoin semanales, se sacaron los datos del valor semanal, los cuales reflejan los últimos valores de Bitcoin por cada semana desde 2019 hasta 2021, sacando también aquellos de 2022 para su posterior comparación y validación del modelo predictivo. Esta base de datos también incluía los datos de máximos semanales, mínimos semanales, varianza y volumen de transacciones; éstos últimos también fueron utilizados en el modelo para la predicción.

Una vez filtrados y ordenados los diferentes datos fueron puestos en conjunto en un documento de Excel único el cual posteriormente sería importado al programa. Es importante destacar, que para crear un modelo realista se tomó la decisión de trastocar los datos de tal forma que los datos de sentimiento, volumen de transacción, el hashrate y la varianza que se le atribuye a la semana 2 corresponderían a aquellos de la semana 1; esto se hizo con la intención de que fuese más realista ya que en una situación real en la semana 2, por ejemplo, solo se tendría acceso a los datos de la semana previa y no a los datos de la propia semana 2; si queremos la predicción de la semana actual solo podremos introducir los datos de los sucesos que ya han acontecido véase los datos de las semanas previas. Es por esto que en la base de datos se puede observar que en los datos de la semana 1 tan solo aparece el valor de Bitcoin para esa semana y en aquellos correspondientes al resto de semanas todos los datos están completos a excepción de aquellos de la semana 158 y adelante donde, con el fin de que el programa los prediga, faltan los propios del valor de la criptomoneda.

3.3 Descripción del modelo predictivo utilizado y su justificación

Se decidió hacer uso del modelo ARIMA (p, d, q) para la predicción de los valores futuros de Bitcoin en base a las diferentes variables elegidas. Este modelo fue elegido ya que puede ser utilizado para tratar bases de datos no estacionarias,

las cuales son convertidas en estacionarias mediante una o más diferenciaciones; el número de diferenciaciones queda reflejada y será explicada más adelante en el documento. Según estadística.net (Fernández, s.f.), los modelos autorregresivos AR permiten predecir los valores de la variable dependiente a partir de sus valores previos más un término de error, por su parte los modelos de medias móviles MA describen una serie temporal estacionaria y pueden predecir el valor de la variable dependiente a partir de la variable aleatoria de cada momento. Con el conjunto de los tres procesos anteriores, la transformación de los datos determinada como "I", los modelos autorregresivos nombrados "AR" y los modelos de medias móviles expresados como "MA" se constituye el modelo nombrado ARIMA (p, d, q)

El Modelo ARIMA (1,1,0) ha sido el utilizado en este caso, con ello se determina que los datos utilizados son aquellos de la diferencia de primer grado, lo cual permite al modelo convertir los datos en estacionarios, esto queda reflejado en el segundo 1 es decir (p, 1, q). En segundo lugar, la "p" determina el número de muestras utilizadas, en este caso al ser p=1 se hace uso de solo la muestra anterior al dato que se quiere predecir para calcular la predicción; si fuese p=2 se haría uso de las dos muestras anteriores. Al tratar con datos semanales y contar con 208 semanas en total, se considera que el modelo es de memoria corta y es por eso por lo que se le atribuye el valor q=0. El modelo es elegido por el propio programa SPSS que tiene en cuenta los parámetros de estacionalidad de los datos, estacionariedad y diferencias entre otras y tras este estudio recomienda que valores son óptimos para los parámetros de p, d y q. (García, s.f.)

El modelo predictivo utilizado consiste en un modelo ARIMA con transformación logarítmica; puesto que los datos utilizados para la predicción tienen una distribución asimétrica. Se implementan los logaritmos para de esta forma corregir la distribución, la función de los logaritmos es reducir la varianza y asimetría mejorando la precisión y facilitando la posterior interpretación de los

datos. Se dice que los datos tienen una función asimétrica cuando la media, moda y mediana son coincidentes.

Para la elaboración del modelo predictivo se ha hecho uso de la herramienta SPSS de IBM, la cual tras la introducción de una base de datos es capaz de realizar múltiples operaciones. En el caso del modelo predictivo en cuestión, se ha utilizado la herramienta de predicciones, la cual analiza las variables introducidas en la base de datos y tras configurar el tipo de modelo y los parámetros es capaz de generar el modelo y mostrar diferentes métricas para la evaluación del modelo.

A lo largo del trabajo se ha experimentado con diferentes modelos en los cuales se procedió a testear con la introducción y retirada de diferentes variables y comparando los modelos resultantes con los anteriores buscando aquel que mejores resultados produjese. Tras múltiples intentos y comparaciones de los estadísticos FAS Residual, FAP Residual, RMSE, MAPE, MAE, BIC Normalizado, R^2 y el test de Ljung Box, se obtuvo el modelo final que puede verse en la figura 1: Gráfico de predicciones y datos reales.

3.4 Evaluación del modelo predictivo

Para la evaluación del modelo se han utilizado diferentes métricas proporcionadas por el propio programa SPSS además de los datos reales del año 2022. Las diferentes métricas serán utilizadas para comparar los diferentes modelos realizados hasta encontrar aquel que mejor ajuste y precisión tenga, el cual será el elegido y sobre el cual se va a trabajar a lo largo del documento. Las métricas proporcionadas y utilizadas en este caso son: FAS Residual, FAP Residual, RMSE, MAPE, MAE, BIC Normalizado, R^2 y la prueba de Ljung Box.

El FAS Residual (Fully Anticipated Selection) muestra la bondad de ajuste utilizada en los modelos de regresión, la cual mide la cantidad de información perdida al eliminar variables del modelo y se utiliza para evaluar si la selección

de variables del modelo es adecuada. Cuanto menor sea el valor de FAS residual, mejor ajustado estará el modelo. (IBM)

El FAP Residual “Fully Anticipated Prediction” es utilizado para medir la cantidad de información perdida al eliminar variables del modelo y es utilizada para valorar si la selección de variables utilizadas en el modelo es adecuada en términos de precisión predictiva. En cuanto a la interpretación de esta métrica, cuanto menor sea el valor del FAP residual, mejor ajustado estará el modelo y por ello la precisión será mayor. (IBM)

El RMSE “Root Mean Squared Error” es una medida de error cuadrático medio que es utilizada para evaluar la precisión de los modelos predictivos. El método para obtener el RMSE consiste en calcular la diferencia entre los valores dados y los valores predichos por el modelo, esta diferencia es elevada al cuadrado, después se hace una media y finalmente se hace la raíz cuadrada dejando el RMSE. Para la interpretación de esta métrica se suele recomendar compararla entre diferentes modelos, en su cómputo global, cuanto menor sea el resultado mayor precisión tendrá el modelo. El RMSE está expresado en la misma unidad que la variable dependiente, por lo que si el RSME es menor que el rango de la variable se considera que el modelo tiene una buena precisión, si al contrario, el valor es notablemente mayor, se entiende que el modelo no es adecuado. (IBM)

El MAPE (Mean Absolute Percentage Error) consiste en una media de error porcentual, es utilizada para evaluar la precisión de un modelo predictivo. La forma de calcularlo es mediante la diferencia en porcentaje de los valores observados frente a los predichos por el modelo y haciendo seguidamente el promedio y se expresa en porcentaje. La interpretación depende del rango de valores de la variable dependiente. Para la interpretación de esta métrica se recomienda la comparación entre diferentes modelos; para la correcta comparación es importante tener en cuenta que un valor inferior representa una mayor precisión del modelo. Esta métrica se muestra menos eficaz en los casos

en los que la variable dependiente tiene valores cercanos a 0 ya que la división resultaría en un resultado considerablemente alto dificultando la interpretación de este. (IBM)

El MAE (Mean Absolute Error) es una medida de error absoluto medio que es utilizado para evaluar la precisión de un modelo predictivo. Se calcula mediante la diferencia absoluta entre los valores observados y los valores predichos por el propio modelo y se expresa en las mismas unidades que la variable dependiente. La utilidad de esta métrica consiste en la ausencia de dirección a la hora de calcularse; esto quiere decir que no se tiene en cuenta la dirección del error sino la magnitud de este. Para la interpretación del resultado se tiene en cuenta que cuanto menor sea el MAE, mayor precisión presentará el modelo; puede compararse directamente con el rango de la variable dependiente para de esta forma, teniendo en cuenta el rango, valorar si el resultado es significativo o en su defecto es considerablemente inferior al rango. Al usar esta métrica es importante tener en cuenta que la interpretación se ve dificultada cuando los valores de la variable dependiente son o muy grandes o considerablemente pequeños, ya que al contrario que el MAPE, esta variable no se muestra en términos porcentuales. (IBM)

El BIC normalizado (Bayesian Information Criterion) es una medida de complejidad del error utilizada para comparar y seleccionar modelos. Esta métrica considera el tamaño muestral y la cantidad de variables predictoras, para su cálculo, se divide el BIC del modelo entre el tamaño de la muestra y se añade un término que dependerá del número de variables independientes. La unidad de medida utilizada para expresar la métrica consiste en una probabilidad relativa, cuanto menor sea el valor del BIC normalizado, mayor será la probabilidad de que el modelo sea preciso y produzca buenas predicciones. En términos generales se recomienda la comparación con diferentes modelos para escoger aquel con menor resultado; habiendo dicho eso, cabe destacar que la forma anterior es compleja puesto que, si los modelos tienen diferente número de variables independientes, la fórmula para calcular el BIC normalizado no

coincidirá y los resultados serán difícilmente comparables; aun así, prevalece la regla de que el modelo con menor BIC normalizado será el preferido. (IBM)

El R^2 es una medida de ajuste que se define como la proporción de la variabilidad de la variable endógena explicada por la variabilidad de las variables explicativas (Zurita, y otros, 2021). El valor de esta métrica se encuentra siempre entre 0 y 1, siendo 1 el resultado al que se aspira puesto que significa que el modelo explica toda la variabilidad de la variable dependiente. El resultado de R^2 del modelo puede variar en función del número de variables introducidas, así como de la escala de medición de la variable dependiente. Esta medida no puede ser utilizada para medir la precisión del modelo, en su defecto es utilizada para medir la relevancia de las diferentes variables y la proporción de la variabilidad de la variable dependiente que puede ser explicada por estas. (IBM)

La métrica de la Ljung-Box consiste en una prueba de bondad de ajuste y es utilizada para determinar si una serie temporal tiene autocorrelación y con ello se puede determinar también si una serie temporal es estacionaria o aleatoria. La hipótesis nula de la prueba Ljung-Box es que no hay autocorrelación en la serie temporal; si el valor p es menor que 0,05, se rechaza la hipótesis nula. Si al contrario el valor p es alto, no se rechaza y se concluye que los valores de los datos son independientes. En la tabla proporcionada por el programa se muestran 3 métricas: “Estadísticos”, “DF” y “Sig.” En este caso “estadístico” y “sig.” Son los propios de la Ljung-box y “DF” representa el grado de libertad de las variables (Degrees of freedom). En nuestro modelo es de gran interés que la métrica “sig.” sea lo más cercano a 1 posible. (IBM)

Tanto para el FAS residual como para el FAP residual, el programa proporciona dos gráficos, los cuales se interpretan asegurándose que las barras azules se encuentran dentro de las líneas negras (figura 2) y las barras azules no deben de seguir ningún patrón aparente o claro. Para el resto de las métricas, el programa proporciona tablas con los respectivos valores de cada métrica.

4. Resultados y discusión:

4.1 Análisis exploratorio de datos

Para el apartado del análisis exploratorio de los datos se procederá a dividir los datos en 7 grupos que serán los siguientes: Hashrate, Varianza, Volumen de transacciones, Google trends, Variables relacionadas con Twitter, variables relacionadas con la red social Reddit y, por último, variables relacionadas con BitcoinTalk. Se procede de esta manera puesto que, dentro de los grupos de Twitter, Reddit y BitcoinTalk, las variables relacionadas coinciden entre sí por lo que se compararán en conjunto tras un breve análisis individual de cada una de ellas; esto nos permitirá evaluar el contenido de las redes sociales en su conjunto y determinar cuáles de ellas se consideran relevantes y cuales, en su defecto, no tiene mayor relevancia para el correcto funcionamiento del modelo predictivo en cuestión. Las métricas que serán utilizadas para valorar la utilidad de las variables son la “t” y la “sig.” La variable “t” hace referencia a los resultados del “t-test” donde se buscan valores alejados de 0; “sig.” Hace referencia al p-valor, en el cual se buscan valores lo más cercanos a 0 para el rechazo de la hipótesis nula. Las tablas con los valores por cada variable quedan reflejadas en la figura 3; se puede ver un resumen completo de los resultados obtenidos por cada una de ellas en dicha figura, en la cual se subrayan de color amarillo aquellas que obtienen un resultado suficiente como para ser consideradas como relevantes para su uso en el modelo predictivo.

En primer lugar, se analizarán los resultados obtenidos de la variable “hashrate”. Como ya sabemos, para generar nuevas criptomonedas en la red de Bitcoin, han de resolverse acertijos criptográficos, dichos acertijos tienen una importante complejidad y han de ser resueltos por ordenadores; el hashrate es la medida utilizada para medir el poder computacional del conjunto de mineros y de esta forma comprobar la seguridad de la propia criptomoneda. Los datos obtenidos con respecto a esta variable son los siguientes: $t=-0,045$ y $\text{sig.}=0.964$; viendo esos datos queda claro que la utilidad de la variable en cuestión nos es adecuada y las

variaciones en esta variable, al contrario de lo que se pensaba en un inicio, no son relevantes en el modelo predictivo para explicar los cambios del precio de Bitcoin.

En segundo lugar, se considerarán los valores obtenidos para la variable "Varianza"; esta variable muestra la varianza respecto a la semana anterior; estos datos son fácilmente accesible por el usuario y podrían ser introducidos en el modelo semanalmente para el correcto funcionamiento de este. por eso los datos introducidos en el modelo son aquellos de la varianza real. Los resultados obtenidos respecto a esta variable son $t=94,7$ y $\text{sig.}= 0,000$; estos datos son excelentes y muestran como la variable en cuestión es de gran relevancia en el modelo. La explicación para la gran utilidad de la variable puede venir del hecho de que los valores son los reales correspondientes en todos los casos a aquellos de la semana previa, como se ha explicado anteriormente, los datos de las variables independientes de la semana 2 corresponden a aquellos de la semana 1 los de la semana 3 son aquellos de la semana 2 y sucesivamente.

En tercer lugar, se analizarán los resultados de la variable "Vol.", esta variable muestra, en millones, el volumen de transacciones semanales; en otras palabras, puede ser utilizado para evaluar la cantidad de uso que se hace de la criptomoneda, lo cual pareciera ser, a priori, un indicativo de la fuerza de la criptomoneda. Los datos obtenidos no fueron los esperados; se obtuvieron valores de $t=-0,034$ y $\text{sig.}=0,973$, dando a entender que la variable en cuestión no es relevante para el modelo predictivo puesto que los valores de t son demasiado cercanos a 0 y los valores de sig. demasiado altos.

La próxima variable en ser analizada será la correspondiente con Google trends, esta variable expresa la cantidad de búsquedas relacionadas con Bitcoin realizadas por los usuarios mundialmente en el periodo del estudio, desde 2019 hasta 2022; se marca con un valor de 100 la semana con más búsquedas y el resto de las semanas reciben una nota en proporcional respecto a esa semana.

Los datos obtenidos para estas variables son $t=-1,235$ y $\text{sig.}=0,219$; estos valores muestran de nuevo que la variable no debe ser utilizada en el modelo.

La variable "Twitter" será la siguiente en ser analizada; en esta variable se han introducido todas aquellas relacionadas con la red social, estas serán analizadas individualmente primero y posteriormente se hará una evaluación general de la relevancia de la red social en su conjunto. La primera variable corresponde a las publicaciones categorizadas como "bearish", lo cual indica que el usuario considera que la criptomoneda va a reducir su valor, los resultados obtenidos son $t=-0,280$ y $\text{sig.}= 0,780$ lo cual indica que la variable no es relevante. La siguiente variable es la de "bullish" significa lo contrario que la anterior, que el usuario considera que el valor de Bitcoin se verá aumentado en un futuro, los valores obtenidos son $t=-0,306$ y $\text{sig.}0,760$ lo cual vuelve a indicar que la no utilización es la decisión correcta. Para la variable "bad news" los datos obtenidos son $t=-1,528$ y $\text{sig.}=0,129$, resultados que muestran que la variable en cuestión no aporta utilidad en el modelo predictivo llevado a cabo. En cuanto a la variable "Good news" los resultados cambian con respecto a los anteriores, se obtiene $t=2,733$ y $\text{sig.}=0,007$ mostrando que esta variable es relevante para el modelo y debe ser utilizada. Para la variable "positive" se obtienen los valores de $t=1,580$ y $\text{sig.}=0,117$ siendo ambos valores claros indicadores de que no ha de utilizarse la variable en el modelo; lo mismo ocurre con los valores de la variable "optimistic" para la cual $t=-0,048$ y $\text{sig.}=0,961$. Por último, se estudia los resultados de la variable "negative", los cuales indican que la variable es relevante para la predicción del valor de Bitcoin obteniendo los resultados de $t=-2,317$ y $\text{sig.}=0,022$. En su cómputo general, el conjunto de variables de la red social Twitter se muestran irrelevantes para el modelo predictivo a excepción de "negative" y "Good news" por lo que se concluye que las publicaciones en Twitter no son buena muestra de lo que le ocurrirá al precio de Bitcoin en el futuro.

La siguiente variable que será evaluada es "Reddit", igual que pasa con Twitter, se analizarán brevemente las diferentes variables relacionadas con la red social

Reddit y tras ese breve análisis se procederá a realizar un análisis general de la red social para evaluar la relevancia de esta en la valoración de Bitcoin. La primera variable que se analizará es “bearish” la cual obtiene los resultados de $t=-0,319$ y $\text{sig.}= 0,750$, estos resultados muestran la falta de relevancia de la variable en el modelo predictivo. La siguiente variable será “bullish” para la cual, $t=0,455$ y $\text{sig.}=0,650$ muestran de nuevo que esta variable no debe ser introducida para la predicción. La variable “bad news” logra los resultados de $t=1,191$ y $\text{sig.}=0,236$ que otra vez dan a entender que la variable no debe ser tenida en cuenta en el modelo creado. La próxima variable, “Good news”, obtiene resultados de $t=-1,142$ y $\text{sig.}= 0,256$ de nuevo, mostrando que no es una variable con suficiente significación para ser considerada en el modelo; lo mismo ocurre con la variable “positive” que obtiene valores de $t=-0,858$ y $\text{sig.}=0,392$. “Optimistic” es la siguiente variable que analizar y los datos muestran valores de $t=0,114$ y $\text{sig.}=0,909$ que indica una falta de relevancia de dicha variable en el modelo. Por último, se estudian los resultados obtenidos para la variable “negative” que son $t=1,480$ y $\text{sig.}=0,141$ siendo insuficientes de nuevo para poder considerar la introducción de la variable en el modelo por su falta de relevancia. En su conjunto, tras el análisis previo, queda como evidente el hecho de que la red social Reddit y los datos obtenidos de la misma no han de ser introducidos en el modelo para la predicción del valor futuro de la criptomoneda Bitcoin, puesto que todas las variables sin excepción muestran resultados insatisfactorios para las métricas t y sig. obteniendo resultados demasiado cercanos a 0 para la t y demasiado elevados para la variable sig.

Para la variable de BitcoinTalk se procederá de la misma forma que las anteriores redes sociales, realizando un breve análisis de las variables relacionadas con esta y luego llevando a cabo un análisis general de la relevancia de la red social en su conjunto para la valoración de la criptomoneda. La variable “bearish” obtiene valores de $t=0,009$ y $\text{sig.}=0,993$, dejando dicha variable fuera del modelo por la falta de relevancia de esta. La variable “bullish” consigue resultados de $t=0,218$ y $\text{sig.}=0,828$ dando a entender la falta de relevancia de esta y por tanto su exclusión del modelo. “Bad news” será la siguiente variable en ser analizada, los

resultados obtenidos por esta variable son $t=0,171$ y $\text{sig.}=0,828$, de nuevo prescindiendo de la variable en el modelo predictivo. La variable “Good news” muestra datos de $t=-0,620$ y $\text{sig.}=0,536$ siendo ambos insuficientes para su consideración. “Positive” obtiene datos de $t=-1,163$ y $\text{sig.}=0,247$ dejándola también fuera del modelo; lo mismo ocurre con la variable “optimistic” que obtiene datos de $t=0,348$ y $\text{sig.}=0,728$. La última variable analizada es “negative” que consigue datos de $t=1,1445$ y $\text{sig.}=0,151$, resultados que muestran que la variable no debe ser tenida en cuenta en el modelo. Como es tendencia con las redes sociales en el caso de este modelo predictivo, las variables introducidas se muestran irrelevantes para la creación del modelo predictivo, obteniendo valores insuficientemente lejanos a 0 en la t y excesivamente altos para sig. ; aunque en la hipótesis se remarcaba que las variables anteriores parecerían relevantes, tras sus estudio y análisis se ha demostrado lo contrario.

4.2 Resultados del modelo predictivo

Los gráficos representativos de los resultados del modelo predictivo se encuentran en la figura 1 quedan también reflejados los mismos semana a semana en una tabla en la figura 4. Como se puede observar en la figura 1, los pronósticos del modelo en cuestión (línea azul) tienen una gran semejanza con los valores reales de Bitcoin (línea roja). Se puede observar que a medida que pasa el tiempo la precisión del modelo se reduce dejando lugar a cada vez más errores en el mismo. A pesar de la pérdida de precisión del modelo con el paso del tiempo, las tendencias que se observan en los datos reales están reflejados con gran precisión por el modelo, lo cual serviría para que el usuario pudiese predecir posibles altibajos y por ende se podrían tomar decisiones de compra o venta.

Para el análisis del modelo y de sus métricas se hará en todo momento utilizando los valores medios de dichas métricas, las cuales son: FAS Residual, FAP Residual, RMSE, MAPE, MAE, BIC Normalizado, R^2 y la prueba de Ljung-Box. Se puede observar en la figura 3 una tabla con los datos relativos a cada uno de estos estadísticos.

La primera métrica, FAS residual será analizada junto al FAP residual, para los cuales el programa SPSS genera un gráfico para cada una el cual se puede observar en la figura 2. Para el análisis de ambas gráficas es importante recordar que se buscan los valores más cercanos a 0 posibles. Como se puede observar en el gráfico todos los valores se encuentran dentro de las líneas negras que son utilizadas para reflejar los extremos que se permiten para considerar que los valores son correctos y que el modelo predictivo se ajusta correctamente. En el caso del FAS, no se observa ningún patrón aparente ni simétrico lo cual significa que los valores son positivos para el estudio, todos los valores se encuentran entre las líneas negras y se observan diferentes valores más y menos próximos al 0, como es normal. En el caso del FAP residual, el gráfico observado es similar a aquel del FAS residual, no se observa ninguna simetría en los datos obtenidos y todos los valores se encuentran dentro de los límites con variación en los valores, todo lo anterior indica que el modelo está correctamente ajustado. Como resumen de ambas métricas se puede observar que la cantidad de información perdida en el modelo predictivo es aceptable y por tanto el ajuste del modelo en términos generales es correcto.

En el caso del RSME, la medida de error cuadrático medio es utilizada para medir la precisión del modelo mediante el cálculo de la diferencia entre los valores dados y los valores predichos, la cual es elevada al cuadrado y posteriormente se calcula la media. Para el modelo en cuestión, el resultado de dicha métrica es de 217,534, el cual, para ser interpretado, ha de ser comparado con diferentes modelos; esto fue realizado en el proceso de pruebas entre diferentes modelos antes de la elección definitiva. Otra forma de evaluar el resultado es comparándolo con el rango de la variable dependiente, el cual tiene un valor mínimo de 3502,5 en el periodo estudiado y un máximo de 64398,6 lo cual significa que el rango de la variable es la diferencia entre ambos valores, el cual resulta ser 60896,1. El rango de la variable es notablemente inferior al resultado del RMSE dejando en evidencia que la precisión del modelo es de buena calidad y por tanto puede ser considerado como fiable.

El MAPE es la siguiente métrica estudiada, la cual se calcula mediante la diferencia en porcentaje de los valores observados y aquellos predichos por el modelo, sobre la cual se hace un promedio. De nuevo se recomienda la comparación con otros modelos, la cual ha sido llevada a cabo en la fase de preparación del trabajo. Una vez comparados los resultados con diferentes modelos, es importante resaltar que un valor menor es lo que se busca; en este caso el valor del MAPE es de 0,574%, esto es un valor considerablemente pequeño lo cual indica que la buena precisión del modelo. Cabe destacar, que, aunque no sea el caso, la evaluación de esta métrica podría verse afectada si los valores utilizados en el modelo son significativamente cercanos a 0.

El estadístico MAE se entiende como la media de error absoluto, esto permite identificar la magnitud del error sin tener en cuenta la dirección de este. Se calcula mediante la diferencia absoluta entre los valores observados y los valores predichos por el modelo predictivo, la medida coincide con aquella utilizada en la variable dependiente, en este caso la unidad son dólares americanos. El resultado obtenido ha de compararse de nuevo con el rango de la variable que se busca predecir, cabe recordar que el rango era de 60896,1. El resultado obtenido para el MAE es de 111,62 lo cual es significativamente inferior al rango de la variable que considera el precio semanal de Bitcoin por lo que el modelo goza de precisión suficiente. De nuevo cabe destacar que la interpretación de dicha métrica queda dificultada si son utilizados valores próximos a 0 o valores muy elevados, no es el caso es este trabajo.

La próxima variable en interpretarse será el BIC normalizado; dicho estadístico consiste en una medida de complejidad del modelo y es utilizada para comparar y seleccionar modelos. El estadístico considera el tamaño muestral y la cantidad de variables; se calcula dividiendo el BIC entre el tamaño muestral y añadiendo un término que depende de la cantidad de variables. La unidad en la que se expresa el resultado es una probabilidad relativa, cuanto menor sea, mayor será la probabilidad de que la precisión del modelo sea buena. En este caso el

resultado es 11,639, lo cual indica que el modelo es pobre ya que un valor cercano al 0 es aquel que se busca. Esta métrica nos indica que la precisión del modelo no es aceptable. Se podría atribuir este resultado al bajo número de variables del modelo, principalmente porque el término que se añade en función del número de variables será bajo.

El R^2 explica la proporción de variabilidad de la variable dependiente que puede ser explicada por la variabilidad de las variables independientes; se buscan valores cercanos a 1, los cuales indican que la variabilidad de la variable endógena queda totalmente explicada por aquella de las variables explicativas. En el caso de nuestro modelo, los resultados son óptimos, siendo $R^2=1$. Esto nos explica como las predicciones se basan por completo en las variables independientes como debería ser en un modelo predictivo. Esta métrica nos explica que la relevancia de las variables independientes introducidas es óptima.

Por último, se estudiarán los resultados de la métrica de Ljung-Box, esta prueba es utilizada para determinar la bondad de ajuste del modelo e indica si la serie temporal tiene autocorrelación. En nuestro modelo el resultado de sig. es 0,821, lo cual es cercano a 1 que es lo que interesa en este caso, ese resultado indica que las variables son independientes y que el ajuste del modelo es correcto y por tanto se puede hacer uso del modelo.

De todos los estadísticos utilizados, tan solo aparece uno que diagnostica que el modelo no es óptimo, el BIC normalizado. Dada la difícil interpretación de este, se ha tomado la decisión de descartar dicho estadístico y basarse en el resto para el estudio, puesto que las pruebas restantes, y por tanto la mayoría de ellas, muestran que el modelo está adecuadamente ajustado y que goza de una precisión correcta para su utilización; además tras observar las predicciones y observar la gran semejanza con los datos reales, queda visible que los resultados del modelo predictivo son excepcionales.

4.3 Discusión de los resultados obtenidos

Tras observar el gráfico obtenido de SPSS, se puede ver una clara similitud tanto en las tendencias como en los propios valores predichos con los valores reales de Bitcoin lo cual nos demuestra que el modelo es funcional y a su vez preciso en sus predicciones. De entre todas las variables introducidas en el modelo, tan solo existen tres variables realmente significativas en el modelo, las cuales son % varianza, Twitter “Good news” y Twitter “Negative”.

La primera variable relevante, % varianza, muestra la varianza en el valor de bitcoin entre la semana anterior y la previa, por ejemplo, si estamos prediciendo el valor de la semana 3, la varianza es aquella entre el valor de la semana 2 y la semana 1. Este retardo en la varianza puede explicar la falta de precisión que experimenta el modelo, la cual se ve que va en aumento a medida que pasa el tiempo, pero sin dejar de seguir las tendencias del valor real lo cual es una gran ventaja ya que el uso de este modelo predictivo está pensado de tal forma que a medida que pasen las semanas se actualicen los datos y de esta forma se corrija las posibles imprecisiones del modelo. En la última semana, como se puede observar en la figura 4, la diferencia entre el precio real y el precio predicho es de alrededor de 3300\$, lo cual, teniendo en cuenta la cantidad de datos introducidos, representa una precisión aceptable.

Por parte de la variable independiente Twitter Good news, representa el número de publicaciones recopiladas en una semana que están relacionadas con bitcoin y que hablan de posibles buenas noticias para la criptomoneda. Los datos son proporcionados por la página web augmento.ai que consiste en una inteligencia artificial que recopila los datos y siguiendo un algoritmo clasifica las publicaciones dentro de las diferentes categorías. Dicha variable tiene aparente relevancia en el modelo, se puede atribuir esta relevancia al aumento y decrecimiento del número de publicaciones en momentos de subida de precio o decrecimiento de este. Lo mismo ocurre con la variable de Twitter negative, cuyos datos provienen de la misma fuente y cuya relevancia se explica de la misma forma.

5. Conclusiones y recomendaciones:

5.1 Conclusiones obtenidas

El modelo llevado a cabo consiste en un modelo a pequeña escala, las posibilidades son mucho mayores si se dispone de los datos y el equipo necesario para manejar tales cantidades de datos. En términos de las variables introducidas, se puede observar como el modelo se basa principalmente en la variable % varianza, la cual crea el pequeño desajuste del modelo con respecto a los datos reales. Dicha variable no debería incluirse en modelos futuros puesto que si falla un valor la varianza dejará de ser la real y por tanto se corre el riesgo de que el modelo tenga fallos exponenciales. En el caso de nuestro modelo se han introducido de tal forma que se simula la introducción de los datos semana a semana lo cual corrige las posibles imperfecciones en la precisión del modelo facilitando las predicciones correctas del mismo.

El objetivo del trabajo consistía en crear un modelo de la forma anteriormente explicada, un modelo que funcione correctamente a base de su actualización continua; esto reduce la utilidad del modelo y tan solo lo hace funcional para predicciones en el corto plazo por lo que las ventajas de las cuales se habla anteriormente en el documento son reducidas. El modelo aún serviría para asistir a inversores en su toma de decisión, pero tan solo podría ser utilizado para decisiones a corto o medio plazo.

Como se describió en la hipótesis, existen variable cuya relevancia no fue suficiente como para que fueran tenidas en cuenta en el modelo; al contrario de lo que se supuso, la cantidad de variables no relevantes es considerablemente mayor de lo que se esperaba en un primer instante. No era de esperar que la variable % varianza obtuviese un valor tan importante para el modelo, aun así, la función del modelo coincide con aquella descrita en la hipótesis y en los objetivos: crear un modelo cuya precisión sea buena en el corto y medio plazo que requiera de una actualización de los datos semanalmente para su correcto y mejor funcionamiento.

5.2 Limitaciones del estudio y recomendaciones para futuras investigaciones

Los resultados obtenidos en el estudio han sido generalmente satisfactorios dadas las condiciones en las cuales han sido realizadas las pruebas. Para un correcto funcionamiento del modelo predictivo, las bases de datos utilizadas deberían haber sido más amplias, haciendo uso de todos los datos desde la creación de la criptomoneda en 2009 hasta la fecha actual; en el caso que se presenta el manejo de dicha cantidad de datos resultó imposible dado el equipo disponible. Además, para una correcta evaluación, el modelo habría de ser evaluado a lo largo de un periodo más largo, haciendo uso de todos los datos e introduciendo los datos nuevos en función de su generación pudiendo sustituir las predicciones erróneas del modelo por aquellas reales y exigiendo tan solo una predicción del modelo, haciendo lo anterior se vería el funcionamiento en tiempo real. Además, la cantidad de datos accesibles de forma gratuita queda enormemente limitada puesto que las compañías que hacen recogida de muchos de los datos relevantes exigen un pago previo a la descarga y/o visualización de los datos.

Por otra parte, al haberse realizado por semanas se reduce la precisión del modelo dejando lugar a una gran posibilidad de error, esto fue hecho de esta manera puesto que una pequeña selección los datos a los que se tuvieron acceso se encontraban organizados semanalmente y la obtención de datos diarios no fue posible. Al existir eventos que ocurren en un día particular y cuyo efecto puede durar tan solo unos días, la fluctuación de precio en una semana puede haber sido considerable pero no quedar realmente reflejada en la predicción del modelo; por otra parte, al requerir el modelo de una actualización semanal de las variables, se reduce la utilidad del modelo puesto que no podrá ser utilizado para intercambios en un largo plazo, sino que será más efectivo para una inversión a corto o medio plazo.

Para futuras investigaciones se recomienda la elección de nuevas variables, se podrían considerar indicadores que comparen el cambio de divisas fiduciarias

frente a Bitcoin. También se podrían introducir variables que contabilicen el número de mineros activos, las comisiones que se cobran por los mineros por las diferentes transacciones realizadas y el número de transacciones realizadas por la red de bitcoin. En el mercado de las criptomonedas se pueden observar tendencias que afectan a un gran número de criptomonedas, algunas sufren los efectos antes que otras por lo que introducir variables con los valores de las diferentes criptomonedas, el top 50 en capitalización de mercados, por ejemplo, podría ayudar a identificar dichas tendencias antes de que tengan efectos sobre el valor de Bitcoin. Otras variables que se podrían insertar en el modelo para aumentar su precisión serían aquellas que midan la ratio de adopción de la criptomoneda, ya sea por parte de las empresas como de países, el número de “wallets” también podría ser utilizado para ver el número de usuarios y su crecimiento o decrecimiento a lo largo del tiempo y como afecta esto al valor de Bitcoin.

5.3 Implicaciones prácticas del estudio

La implicación práctica más evidente que brinda el modelo es la capacidad de, mediante la introducción y periódica actualización de diferentes variables seleccionadas, predecir el valor de Bitcoin; esto puede suponer grandes ventajas para aquellos usuarios que deseen invertir en la criptomoneda ya que les permitiría tomar decisiones con fundamento. Relativo a los inversores las ventajas son evidentes, no solo se pueden prevenir pérdidas de las inversiones con el correcto uso del modelo, sino que se pueden llegar a tener grandes ganancias una vez refinado el modelo. Además, se podría utilizar un modelo similar con una mayor precisión en fondos de inversiones, dicho modelo se podría considerar un activo en la empresa y podría ser utilizado para atraer a un gran número de inversores.

En términos de los comercios, también existen aplicaciones en los mismos, el modelo permitiría a las empresas planificar sus estrategias comerciales relacionadas con el mercado de las criptomonedas gracias a las predicciones del modelo. Al haberse introducido variables que están directamente relacionadas y

viendo la aparente efectividad del modelo; estas decisiones serían tomadas de manera informada sin necesidad de ser experto en la materia; además este modelo puede ser tomada como base y luego con conocimientos adicionales se podrá decidir si confiar en el modelo o en su defecto tomar una decisión diferente a la que el modelo proponga. Para compañías que acepten el Bitcoin como método de pago, el modelo sería de gran utilidad para poder ajustar sus precios de forma que no sean demasiado bajos como para crearles pérdidas ni demasiado elevados como para no ser atractivos para los potenciales clientes del producto ofrecido.

Las autoridades reguladoras también podrían hacer uso del modelo para la supervisión y regulación del mercado de las criptomonedas; si se predice un gran aumento del valor de la criptomoneda, se podrían tomar medidas para evitar la manipulación del mercado. Esta implicación desafiaría la motivación principal de la criptomoneda y su regulación, la intención de las criptomonedas y la base de estas consiste en la falta de un gobierno regulador o agentes intervinientes en las operaciones, ese rol lo suplen los mineros, los cuales se encargan, a cambio de una recompensa o comisión, de que las operaciones se lleven a cabo y de la comprobación y legitimación de estas a lo largo de toda la red blockchain de Bitcoin.

6. Bibliografía

- Augmento. (26 de Febrero de 2023). *Augmento.ai*. Obtenido de Augmento.ai:
<https://www.augmento.ai>
- Blockchain.com. (2 de Marzo de 2023). *Blockchain.com*. Obtenido de Confirmed Transactions Per Day: <https://www.blockchain.com/es/explorer/charts/n-transactions>
- CoinGecko.com. (2 de Marzo de 2023). *CoinGecko.com*. Obtenido de Gráficos de capitalización de mercado de criptomonedas global:
<https://www.coingecko.com/es/global-charts>
- Corporation, I. (s.f.). *IBM SPSS - Predicciones 28*. Armonk.
- Fernández, S. d. (s.f.). *Portal de Estadística Aplicada*. Obtenido de Series Temporales, Modelo Arima Metodología de BOX-JENKINS:
<https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>
- Galán, J. S. (1 de Julio de 2020). *Economipedia*. Obtenido de Análisis de regresión:
<https://economipedia.com/definiciones/analisis-de-regresion.html?nab=1>
- García, I. R. (s.f.). *The Machine Learners*. Obtenido de Introducción a las series temporales (2): modelo ARIMA con R:
<https://www.themachinelearners.com/series-temporales-arima/#comment-62>
- George, B. (6 de Enero de 2023). *Yahoo Finanzas*. Obtenido de Qué es y cómo funciona un satoshi, la unidad más pequeña de bitcoin:
<https://es.finance.yahoo.com/noticias/funciona-satoshi-unidad-pequeña-bitcoin-183242326.html>
- González, H. A. (10 de Diciembre de 2017). *Revista Electrónica de Investigación en Ciencias Económicas*. Obtenido de Abriendo Camino al Conocimiento :
<http://revistacienciaseconomicas.unan.edu.ni/index.php/REICE>
- Google. (25 de Febrero de 2023). *Google Trends*. Obtenido de Bitcoin:
<https://trends.google.com/trends/explore?date=2019-01-01%202023-03-27&q=bitcoin&hl=en>
- IBM. (s.f.). *IBM SPSS - Predicciones 28*. Armonk, NY.
- Investing.com. (13 de Enero de 2023). *Investing.com*. Obtenido de Historical Data Bitcoin: <https://www.investing.com/crypto/bitcoin/historical-data>
- Izquierdo, A. J. (2000). Modelos estadísticos del riesgo y riesgo de los modelos estadísticos. *Resvista de metodología de Ciencias Sociales*, 101-129.
- Jimenez, D. (19 de Octubre de 2019). *Cointelegraph*. Obtenido de Expertos indican que el último Bitcoin será minado en el año 2140:
<https://es.cointelegraph.com/news/experts-indicate-that-the-last-bitcoin-will-be-mined-in-the-year-2140>
- López, J. F. (8 de Noviembre de 2018). *Economipedia*. Obtenido de Serie temporal:
<https://economipedia.com/definiciones/serie-temporal.html?nab=1>
- Maldonado, J. (25 de Febrero de 2020). *Bit2me Academy*. Obtenido de Minería Bitcoin ¿Cómo se crea un bloque?: <https://academy.bit2me.com/mineria-bitcoin-como-se-crea-un-bloque/#:~:text=En%20el%20sistema%20blockchain%2C%20los,proceso%20ca da%2010%20minutos%20aproximadamente>

- Pastor, J. (20 de Octubre de 2016). *Bit2me Academy*. Obtenido de ¿Qué es la escalabilidad de Bitcoin?: <https://academy.bit2me.com/que-es-escalabilidad-de-bitcoin/>
- PowerData. (s.f.). *PowerData.es*. Obtenido de Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad: <https://www.powerdata.es/big-data>
- Rivadulla, S. (19 de Abril de 2022). *Materialbitcoin.com*. Obtenido de Los memecoins: Los nuevos activos digitales: <https://materialbitcoin.com/blog/memecoins-que-son/>
- Ronco, V., & Callejo, C. (2020). *Criptomonedas para dummies*. Barcelona: Grupo Planeta.
- Rospide, J. (s.f.). UNA revisión sobre redes neuronales artificiales en finanzas. *Sociedad Argentina Docentes en Administración Financiera*, 265-274.
- Santander, B. (s.f.). *Banco Santander Glosario*. Obtenido de ¿Qué es el Hashrate?: <https://www.bancosantander.es/glosario/hashrate>
- Trelewicz, J. Q. (2017). Big Data and Big Money The Role of Data in the Financial Sector. *IT Trends*, 8-10. Obtenido de Big Data and Big Money.
- Vanci, M. (12 de Octubre de 2022). *Criptonoticias*. Obtenido de ¿Cuántas criptomonedas existen?: <https://www.criptonoticias.com/cryptopedia/cuantas-criptomonedas-existen/>
- White, A. (31 de Enero de 2023). *CNBC*. Obtenido de Constantly checking your credit score? Here's how often it updates: <https://www.cnbc.com/select/how-often-does-your-credit-score-update/>
- Worldcoin. (29 de Noviembre de 2022). *Worldcoin*. Obtenido de What's Ethereum 2.0? A Complete Guide: <https://worldcoin.org/articles/whats-ethereum-2-0>
- YCharts. (26 de Marzo de 2023). *YCHARTS*. Obtenido de Bitcoin Network Hash Rate: https://ycharts.com/indicators/bitcoin_network_hash_rate
- Zurita, C. M., Álvarez Fernández, C., Borrás Palá, F., Budría Rodríguez, S., Curto González, T., & Escobar Torres, L. S. (2021). *Modelos cuantitativos para la Economía y la Empresa en 101 ejemplos*. Madrid: EV Services.

7. Anexos

Figura 1: Gráfico de predicciones y datos reales

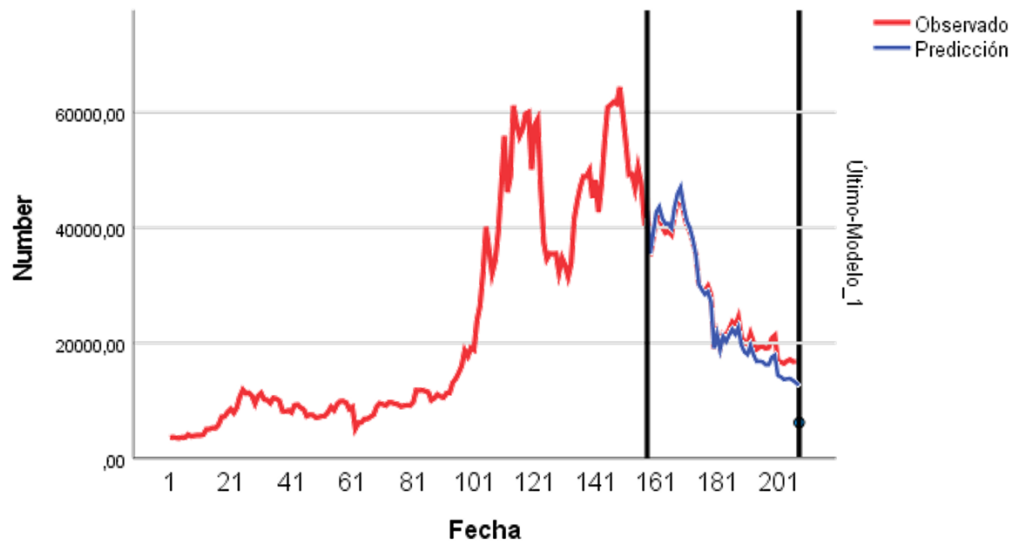


Figura 1: Gráfico de predicciones y datos reales. Elaboración Propia

Figura 2: FAS residual y FAP residual

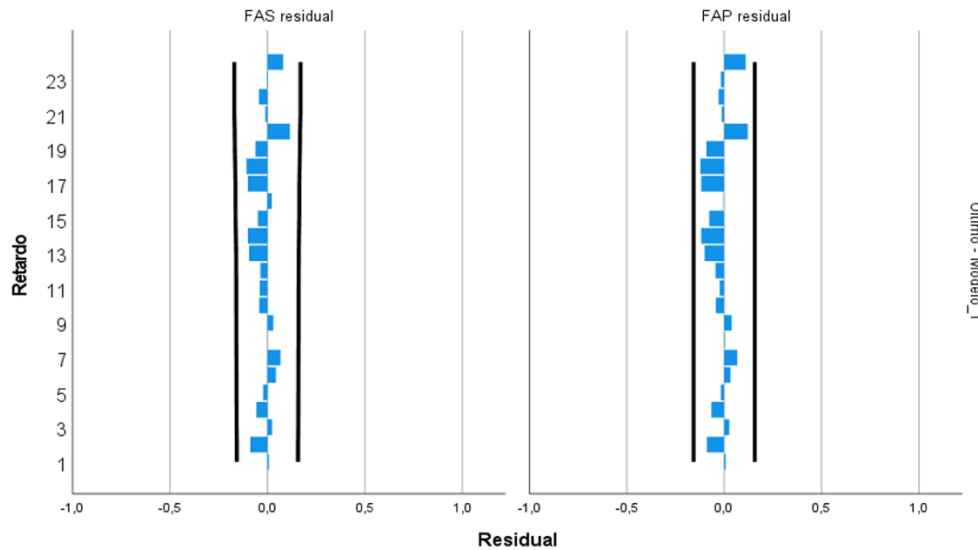


Figura 2 Gráfico FAS residual y FAP residual. Elaboración Propia

Figura 3: Tablas Estadísticas de evaluación

Estadístico de ajuste	Ajuste del modelo											
	Media	SE	Mínimo	Máximo	Percentil							
					5	10	25	50	75	90	95	
R cuadrado estacionaria	0,989		0,989	0,989	0,989	0,989	0,989	0,989	0,989	0,989	0,989	0,989
R cuadrado	1,000		1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RMSE	217,534		217,534	217,534	217,534	217,534	217,534	217,534	217,534	217,534	217,534	217,534
MAPE	0,574		0,574	0,574	0,574	0,574	0,574	0,574	0,574	0,574	0,574	0,574
MaxAPE	10,902		10,902	10,902	10,902	10,902	10,902	10,902	10,902	10,902	10,902	10,902
MAE	111,620		111,620	111,620	111,620	111,620	111,620	111,620	111,620	111,620	111,620	111,620
MaxAE	884,379		884,379	884,379	884,379	884,379	884,379	884,379	884,379	884,379	884,379	884,379
BIC normalizado	11,639		11,639	11,639	11,639	11,639	11,639	11,639	11,639	11,639	11,639	11,639

Figura 3 Tablas Estadísticas de evaluación. Elaboración Propia

Figura 4: Resultados por semana

Modelo	Último -Modelo_1			
	Predicción	UCL	LCL	Valor real
158	43003,54	44015,78	42008,73	43097,00
159	35520,31	36757,02	34314,84	35075,20
160	39053,57	40749,40	37410,84	38170,80
161	42732,90	44897,25	40647,19	41412,10
162	43615,08	46101,48	41229,92	42205,20
163	41629,96	44242,68	39133,89	40090,30
164	40667,73	43435,49	38033,47	39115,50
165	40723,84	43696,56	37904,38	39395,80
166	39968,00	43071,29	37034,32	38814,30
167	43598,95	47176,21	40227,67	42233,00
168	45911,92	49871,70	42191,12	44548,00
169	47021,81	51266,31	43044,66	45811,00
170	43715,27	47830,16	39869,92	42767,00
171	41221,88	45255,60	37462,13	40382,00
172	40035,14	44096,89	36258,70	39418,00
173	38046,49	42039,26	34343,15	37650,00
174	35374,36	39206,58	31828,24	35468,00
175	30173,33	33541,58	27063,61	30080,40
176	29255,85	32615,65	26160,76	29434,60
177	28448,31	31804,56	25363,15	29027,10
178	28914,69	32414,51	25704,28	29864,30
179	27078,13	30436,78	24003,52	28403,40
180	19072,98	21494,66	16860,52	18986,50
181	21474,28	24262,60	18931,84	21489,90
182	18848,70	21349,27	16572,99	19243,20
183	21104,51	23962,80	18508,13	21587,50
184	20277,09	23078,51	17737,06	21209,90
185	21382,34	24393,66	18656,91	22460,40
186	22369,23	25578,42	19469,86	23634,20
187	21527,23	24671,41	18691,55	22944,20
188	22696,17	26069,02	19659,48	24442,50
189	19651,84	22621,71	16982,47	21138,90
190	18463,26	21299,29	15918,41	20033,90
191	18025,23	20838,08	15505,31	19831,40
192	19535,61	22631,31	16766,75	21650,40
193	17980,20	20872,28	15397,57	20113,50
194	16810,46	19553,96	14364,36	18925,20
195	16847,83	19636,61	14365,19	19311,90
196	16745,74	19556,10	14247,69	19415,00
197	16234,21	18995,60	13783,38	19068,70
198	16249,49	19050,00	13767,64	19204,80
199	17532,61	20593,22	14824,24	20809,80
200	17883,67	21044,89	15090,32	21301,60
201	14364,91	16935,40	12096,80	16795,20
202	14135,67	16695,59	11880,08	16699,20
203	13673,21	16178,55	11468,80	16456,50
204	13796,26	16353,30	11549,47	16884,50
205	13849,96	16445,95	11572,09	17127,20
206	13500,92	16059,49	11258,92	16777,10
207	13469,30	16049,57	11211,33	16837,20

Figura 4 Resultados por semana. Elaboración Propia

Figura 5: Resultados por cada variable

Parámetros del modelo ARIMA					Estimación	SE	t	Sig.
Último - Modelo_1	Último	Logaritmo natural	Constante		-0,007	0,014	-0,493	0,62
			AR	Retardo 1	0,082	0,092	0,883	0,38
			Diferencia		1			
	Hashrate (M)	Ninguna transformación	Numerador	Retardo 0	-3,457E-06	7,608E-05	-0,045	0,96
	bitcointalk_bearish	Ninguna transformación	Numerador	Retardo 0	2,852E-07	3,305E-05	0,009	0,99
	bitcointalk_bullish	Ninguna transformación	Numerador	Retardo 0	6,083E-06	2,789E-05	0,218	0,83
	bitcointalk_bad_news	Ninguna transformación	Numerador	Retardo 0	1,250E-05	7,302E-05	0,171	0,86
	bitcointalk_good_news	Ninguna transformación	Numerador	Retardo 0	-1,979E-05	3,190E-05	-0,620	0,54
	bitcointalk_positive	Ninguna transformación	Numerador	Retardo 0	-2,288E-05	1,968E-05	-1,163	0,25
	bitcointalk_optimistic	Ninguna transformación	Numerador	Retardo 0	7,310E-06	2,098E-05	0,348	0,73
	bitcointalk_negative	Ninguna transformación	Numerador	Retardo 0	1,282E-05	8,871E-06	1,445	0,15
	reddit_bearish	Ninguna transformación	Numerador	Retardo 0	-4,130E-06	1,293E-05	-0,319	0,75
	reddit_bullish	Ninguna transformación	Numerador	Retardo 0	6,382E-06	1,402E-05	0,455	0,65
	reddit_bad_news	Ninguna transformación	Numerador	Retardo 0	1,145E-05	9,614E-06	1,191	0,24
	reddit_good_news	Ninguna transformación	Numerador	Retardo 0	-7,272E-06	6,369E-06	-1,142	0,26
	reddit_positive	Ninguna transformación	Numerador	Retardo 0	-8,684E-06	1,012E-05	-0,858	0,39
	reddit_optimistic	Ninguna transformación	Numerador	Retardo 0	2,023E-06	1,771E-05	0,114	0,91
	reddit_negative	Ninguna transformación	Numerador	Retardo 0	4,126E-06	2,788E-06	1,480	0,14
	% var.	Ninguna transformación	Numerador	Retardo 0	1,013	0,011	94,765	0,00
	Vol. (Millones)	Ninguna transformación	Numerador	Retardo 0	-3,225E-07	9,431E-06	-0,034	0,97
	Bitcoin: (Worldwide)	Ninguna transformación	Numerador	Retardo 0	0,000	0,000	-1,235	0,22
	twitter_bearish	Ninguna transformación	Numerador	Retardo 0	-5,819E-06	2,075E-05	-0,280	0,78
	twitter_bullish	Ninguna transformación	Numerador	Retardo 0	-4,695E-06	1,537E-05	-0,306	0,76
	twitter_bad_news	Ninguna transformación	Numerador	Retardo 0	-1,297E-05	8,489E-06	-1,528	0,13
	twitter_good_news	Ninguna transformación	Numerador	Retardo 0	1,042E-05	3,811E-06	2,733	0,01
	twitter_positive	Ninguna transformación	Numerador	Retardo 0	2,013E-05	1,274E-05	1,580	0,12
	twitter_optimistic	Ninguna transformación	Numerador	Retardo 0	-1,431E-06	2,953E-05	-0,048	0,96
	twitter_negative	Ninguna transformación	Numerador	Retardo 0	-1,862E-05	8,037E-06	-2,317	0,02

Figura 5 Resultados por cada variable. Elaboración Propia