

El análisis de series temporales es utilizado comunmente para hacer pronósticos de precios, incluyendo el de las criptomonedas. Implementar modelos que hagan predicciones precisas no es una tarea sencilla, especialmente para mercados tan volátiles. Se desarrolla un modelo híbrido ARIMA-LSTM que es evaluado(RECM) contra otros modelos ya utilizados(ARIMA, LSTM) para pronosticar el precio de criptomonedas. Se obtuvieron resultados que demuestran que el modelo híbrido se aproxima mejor que los otros modelos al valor real de los precios.

Capítulo 1

Introducción

En este trabajo se presenta la investigación e implementación de un modelo híbrido de series temporales utilizado para hacer pronósticos del precio de criptomonedas (Bitcoin y Litecoin). El desarrollo del proyecto implica la implementación de dos modelos (ARIMA y LSTM) que se han utilizado actualmente para hacer los pronósticos y se evalúan los resultados obtenidos por dichos modelos contra un tercer modelo híbrido ARIMA-LSTM que a su vez se implementa.

1.1 Contexto del Problema

Las criptomonedas son monedas digitales que operan de forma descentralizada, permitiendo que los usuarios lleven a cabo transacciones sin que éstas dependan de un intermediario. Las criptomonedas, y en especial Bitcoin que es la más popular, atraen cada vez más la atención haciendo que muchas personas opten por invertir en ellas y que ciertos negocios comiencen a aceptarlas como forma de pago [1].

Las criptomonedas se describen como rápidas, convenientes, libres de impuesto, e incluso revolucionarias [1], sin embargo, también son conocidas por su alta volatilidad debido a distintos factores como el tener un mercado relativamente pequeño, baja liquidez, gran cantidad de especulaciones respecto a su futuro, entre otros. Esta naturaleza volátil, hace que predecir los precios de criptomonedas sea una tarea complicada [2].

1.2 Motivación

La alta volatilidad de las criptomonedas hace que sean una inversión bastante riesgosa, sin embargo, actualmente se llevan a cabo más de 250, 000 transacciones de criptomonedas al día [3]. El desplome del valor de Bitcoin más reciente ocurrió el 12 de Marzo, reportando pérdidas de prácticamente el 50% para los inversionistas en tan solo un día [4]. A partir de datos como éstos, nace la motivación para realizar este trabajo, que busca explorar la posibilidad de acercarnos a pronósticos más precisos y así reducir el riesgo al invertir en criptomonedas.

1.3 Objetivo del trabajo

El objetivo de este trabajo es determinar si los pronósticos del precio de criptomonedas hechos por el modelo híbrido ARIMA-LSTM son más certeros que los pronósticos de los modelos ya hechos anteriormente.

1.4 Estructura del trabajo

Este trabajo está conformado por 7 capítulos que detallan la investigación hecha y posteriormente la implementación de los modelos.

En el capítulo 2 se habla del Estado del Arte. Se explican conceptos relevantes para el proyecto en general, así como otras soluciones propuestas para la problemática.

En el capítulo 3 se describe más a detalle la problemática actual y del porqué es importante seguir buscando soluciones alternas.

En el capítulo 4 se habla a profundidad de la propuesta de solución, explicando tanto los aspectos y conceptos relevantes para su desarrollo, como el funcionamiento de los modelos a implementar.

En el capítulo 5 se define la metodología de evaluación que se utiliza, así como la justificación para la elección de la misma.

En el capítulo 6 se muestran los resultados obtenidos al evaluar cada modelo. Se aprecian visualmente los pronósticos de precios hechos por cada modelo y los precios reales para compararlos.

En el capítulo 7 se presentan las conclusiones basadas en los resultados obtenidos, se retoman puntos importantes del trabajo y se plantean posibles trabajos futuros teniendo ya una base.

Capítulo 2

Estado del arte

En este capítulo se describen conceptos relevantes en el área. Posteriormente se hace mención a los modelos ARIMA y LSTM, que se han implementado para hacer pronósticos del precio de las criptomonedas, así como los resultados que obtuvieron. Analizar estos trabajos permite encontrar un área de oportunidad a la cual se puede contribuir.

2.1 Contexto del área

2.1.1 Series de tiempo

2.1.1.1 Definición

Una serie de tiempo es un conjunto secuencial de datos medidos típicamente en tiempos sucesivos. Las medidas que se toman durante cierto periodo se organizan en orden cronológico [5]. Se define matemáticamente una serie de tiempo como:

$$\{X_1, X_2, \dots, X(t)\}, t = 1, 2, 3, \dots$$

t: Representa el tiempo transcurrido

X(t): Es el valor del dato medido durante un tiempo en específico.

Las series de tiempo se denominan univariantes si los registros que se tienen a lo largo del tiempo son de una sola variable, pero si los registros son de más de una variable entonces se dice que es multivariante. [5, 6] Un ejemplo de serie de tiempo univariante es el medir la temperatura de alguna máquina de forma periódica cada minuto, por otro lado si se midiera tanto la temperatura como la vibración y las revoluciones, se estaría hablando de una serie de tiempo multivariante.

2.1.1.2 Visualización

Para visualizar los datos de una serie de tiempo, generalmente éstos se grafican trazando las observaciones con respecto al tiempo. Esto permite tener una primera aproximación y entendimiento sobre el cómo se comporta la variable a lo largo del tiempo [5].

La figura 2.1 muestra la visualización de una serie de tiempo, donde se aprecian los precios de Bitcoin (en dólares) desde Enero del 2015 hasta Septiembre del 2018. El eje x representa el tiempo, y el eje y el precio de la criptomoneda.

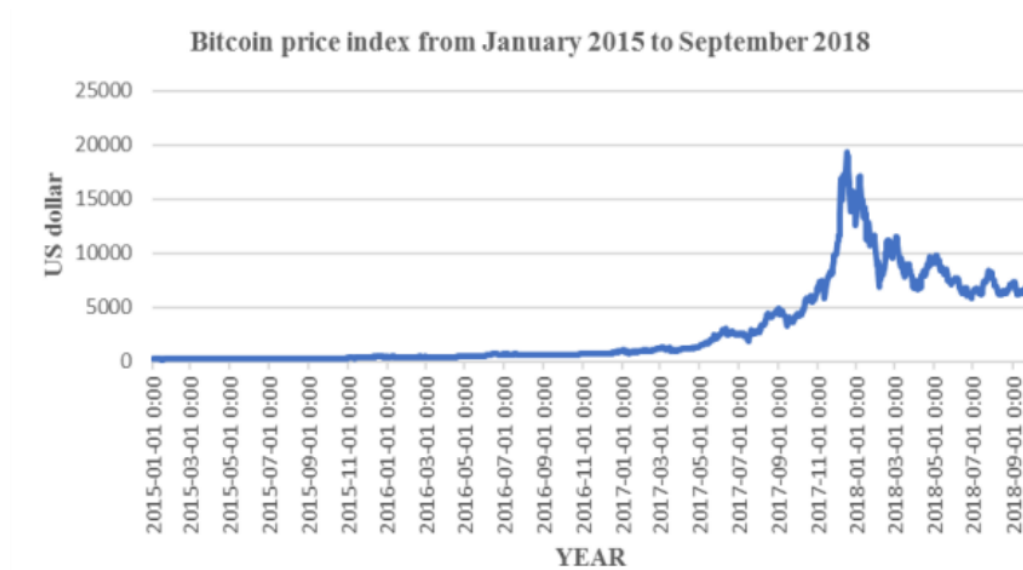


Figura 2.1 Visualización de precios históricos de Bitcoin [11].

2.1.1.3 Análisis y Pronóstico de series de tiempo

El pronóstico de series de tiempo consiste en recolectar datos históricos de alguna variable de interés para posteriormente ser analizados y desarrollar un modelo que comprenda las relaciones que existen en los datos y que posiblemente no sean notorias. El modelo después es utilizado para hacer pronósticos del comportamiento que se espera que tenga la variable recolectada [7, 8].

2.2 Trabajos Relevantes

2.2.1 ARIMA

2.2.1.1 Descripción del modelo

De todos los métodos que existen actualmente para hacer pronósticos de series de tiempo, el modelo autorregresivo integrado de promedio móvil (ARIMA) es uno de los más utilizados. Este modelo tiene una característica relativamente sencilla a grandes rasgos, y que a su vez representa una deficiencia: ARIMA asume que los valores futuros y los valores históricos se relacionan linealmente, cuando la realidad es que la mayoría de los datos de series temporales presentan una relación no lineal [9].

2.2.1.2 Pronósticos del precio de Bitcoin

El modelo ARIMA es una elección muy popular para realizar pronósticos de criptomoneda, especialmente a corto plazo. Como se describe en [10], se implementó este modelo utilizando los datos históricos de Bitcoin desde Mayo del 2016, hasta finales de Marzo del 2018. Los datos obtenidos contenían información sobre el precio de apertura y cierre, el alto y bajo de cada día, entre otros [10]. Para preprocesar los datos, se removieron todas las columnas a excepción de la fecha y el precio de cierre, se utilizó este dato porque es el que genera más interés para los inversionistas [11].

Para implementar el modelo ARIMA, se necesitan definir 3 valores (p, d, q) utilizados para la construcción del mismo, y ya que seleccionar los valores que encajan mejor en el modelo no es una tarea sencilla, el desarrollo se realizó en el lenguaje de programación R, que cuenta con una función que encuentra los valores óptimos de forma automática [10].

Las pruebas se realizaron pronosticando los precios de cierre desde el 13 de Diciembre del 2017, hasta el 27 de Marzo del 2018, y posteriormente evaluando los precios arrojados por el modelo contra los precios reales de las mismas fechas utilizando la raíz del error cuadrático medio (RECM). La evaluación resultó en un error de 593.80 [10].

2.2.2 LSTM

2.2.2.1 Descripción del modelo

Las Long Short-Term Memory Networks (LSTM) son un tipo de redes neuronales compuestas por una capa de entrada, una o más capas escondidas, y una capa de salida. Lo que diferencia a las LSTM de las redes neuronales convencionales, es la estructura de sus capas escondidas que les permite recordar información por largos periodos de tiempo [12, 13]. Dentro de la capa escondida de una LSTM se encuentran celdas de memoria, que a su vez cuentan con 3 tipos de compuertas: La compuerta de entrada, la de salida y la de olvido. A grandes rasgos estas compuertas mantienen y ajustan el estado de la celda, es decir, la información que ésta contiene [12].

Este modelo a diferencia del anterior, es capaz de ajustarse a relaciones no lineales en las secuencia de datos, lo que permite generar pronósticos más confiables cuando son a plazos más largos [13].

2.2.2.2 Pronósticos del precio de Bitcoin

Wu, C.-H., et al. [12] implementaron una LSTM utilizando los precios de Bitcoin desde el inicio del 2018, hasta el 28 de Julio del mismo año. El preprocesamiento de los datos fue

muy similar al mencionado previamente, quedándose únicamente con la fecha y el precio de cierre para el entrenamiento de los modelos y los pronósticos a generar.

Para realizar las pruebas se utilizaron los últimos 71 registros de las series de tiempo, teniendo un total de 138 muestras para entrenar al modelo. La raíz del error cuadrático medio fue una de las metodologías de evaluación usadas, resultando en un error total de 247.33 [12].

2.2.3 ARIMA-LSTM

2.2.3.1 Descripción del modelo

El modelo híbrido ARIMA-LSTM combina ambos modelos aprovechando las ventajas que cada uno ofrece, buscando lograr pronósticos más precisos. Primero se utiliza el modelo ARIMA para ajustar las relaciones lineales en los datos y posteriormente el modelo LSTM se encarga de las relaciones no lineales [9].

La secuencia de pasos para hacer pronósticos de series de tiempo con el modelo ARIMA-LSTM se representa en la figura 2.2. A grandes rasgos, los pasos son los siguientes:

- 1: Entrenar el modelo ARIMA con los datos históricos de las series de tiempo.
- 2: Obtener la secuencia residual resultante (errores en los pronósticos de entrenamiento).
- 3: Ajustar el modelo LSTM utilizando los residuos de ARIMA.
- 4: Hacer pronóstico con el modelo ARIMA del paso 1.
- 5: Hacer pronóstico con el modelo LSTM del paso 3.
- 6: Sumar ambos resultados para obtener el pronóstico final [14].

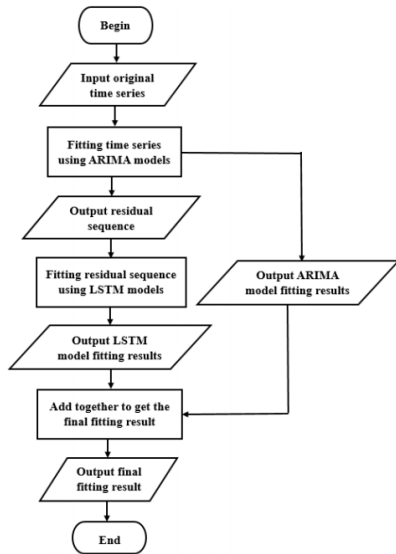


Figura 2.2 Pasos ARIMA-LSTM [14]

2.2.3.2 Pronósticos hechos con ARIMA-LSTM

Wang, Z., & Lou, Y. propusieron un modelo de series temporales ARIMA-LSTM para pronosticar el nivel de agua promedio de una estación hidrológica ubicada en el río Chuhe. Para ponerlo a prueba, desarrollaron los modelos de forma independiente y de forma híbrida para ver si en conjunto conseguían mejores resultados. Sus experimentos mostraron que las ventajas de ambos modelos se complementan, y logran pronósticos más precisos [9].

Liu, J., Tan, X., & Wang, Y. desarrollaron el modelo ARIMA-LSTM para pronosticar el envejecimiento del software para servicios en la nube. Su propuesta (muy similar a la anterior) buscaba comprobar si se podían obtener mejores resultados usando el modelo híbrido. Para comprobarlo, obtuvieron registros diarios del Centro de Datos de Google Cloud, teniendo acceso al uso del CPU y la memoria de más de 1600 máquinas. Después de evaluar los tres modelos, la raíz del error cuadrático medio de ARIMA-LSTM fue la menor de todas, concluyendo que se puede reducir el error y hacer mejores pronósticos con el modelo híbrido [14].

2.3 Conclusión

Actualmente existen varios métodos utilizados para pronosticar el precio de criptomonedas. Los trabajos previamente mencionados muestran el uso de distintos modelos de series temporales, así como las evaluaciones de los mismos. El punto más interesante a notar es el resultado que se obtiene al comparar el modelo híbrido ARIMA-LSTM contra ellos mismos de forma individual, ya que se han hecho pronósticos con un menor margen de error.

A pesar de los resultados favorables, el modelo ARIMA-LSTM no se ha utilizado comúnmente para hacer pronósticos de series de tiempo, y en particular, no se ha utilizado para pronosticar el precio de criptomonedas, por lo que no se puede afirmar aún si es una mejor opción o no que los modelos ya propuestos.

Capítulo 3

En este capítulo se habla acerca de la gran volatilidad de las criptomonedas y del cómo ha afectado a inversionistas. Posteriormente se habla acerca de un área de oportunidad que se presenta como una alternativa para posiblemente acercarse a pronósticos más precisos del precio de las criptomonedas.

Objetivo de Investigación

3.1 Volatilidad de las criptomonedas

La alta volatilidad de las criptomonedas es un hecho que hace que la mayoría de las personas tengan cuidado al invertir en ellas, pero esto no siempre ha sido así. Uno de los principales momentos en los que Bitcoin fue el centro de atención ocurrió en el 2017, año en el que se registró un aumento desmesurado en los precios de dicha criptomoneda. Esto atrajo a un gran número de inversionistas con poca experiencia aprovechando la “oportunidad de su vida”, sin saber que en poco tiempo tendrían pérdidas masivas de hasta un 96% con respecto a su inversión inicial en algunos casos [15].

Las fluctuaciones drásticas en el precio de Bitcoin siguen siendo algo común al día de hoy; el desplome del valor de Bitcoin más reciente ocurrió el 12 de Marzo, reportando pérdidas de prácticamente el 50% para los inversionistas en tan solo un día [4]. Esto no ha impedido que las criptomonedas sigan siendo una opción popular de inversión, actualmente se llevan a cabo más de 250, 000 transacciones de criptomonedas al día [3].

3.2 Trabajos anteriores (área de oportunidad)

HABLAR DE LOS MODELOS QUE SE HAN USADO ARIMA Y LSTM Y QUE EN SI NO HAY UN “PROBLEMA” POR ASÍ DECIRLO CON SU IMPLEMENTACIÓN NI CON LA METODOLOGIA DE EVALUACION QUE USARON, PERO HAY QUE HACER ENFASIS EN QUE LOS RESULTADOS DE ARIMA-LSTM FUERON POSITIVOS EN LOS CASOS EN LOS QUE SE HA PROBADO CONTRA ARIMA Y LSTM Y AÚN NO SE HA HECHO LA PRUEBA PARA CRIPTOMONEDAS

A partir de esto se detecta un área de oportunidad, ya que es posible que el modelo híbrido también sea una mejor alternativa que ARIMA y que LSTM, misma que no se ha

puesto a prueba aún.

3.3 Objetivo de Investigación

Es importante seguir buscando alternativas que permitan acercarse a pronósticos más precisos y así reducir el riesgo al invertir en criptomonedas.

El área de oportunidad detectada, permite definir un objetivo de investigación claro: Determinar si los pronósticos del precio de criptomonedas hechos por el modelo híbrido ARIMA-LSTM son más certeros que los pronósticos de los modelos ARIMA y LSTM.

Capítulo 4

Propuesta de Solución

Como solución para cumplir con el objetivo de investigación planteado, primero se obtuvieron datos históricos de las criptomonedas Bitcoin y Litecoin, posteriormente se desarrollaron los tres modelos a comparar (ARIMA, LSTM, ARIMA-LSTM), entrenándolos con una parte de los datos obtenidos, y finalmente se hicieron pronósticos que después podrán ser evaluados contra los datos no utilizados aún. Todo el proceso se llevó a cabo usando el lenguaje de programación Python.

4.1 Datos

4.1.1 Obtención de datos históricos

Los datos históricos a utilizar tanto para entrenar como probar los modelos fueron obtenidos de CoinMarketCap, que cuenta con datos relevantes de las criptomonedas día a día. Antes el sitio web ofrecía la opción de descargar estos datos pero actualmente ya no está disponible, por lo que se tuvo que acceder a la página directamente desde Python para obtener los datos necesarios. [16, 17]

Date	Open*	High	Low	Close**	Volume	Market Cap
May 17, 2020	9374,93	9823,00	9349,55	9670,74	40.084.250.663	177.745.404.470
May 16, 2020	9333,24	9564,20	9260,69	9377,01	36.164.766.408	172.340.956.579
May 15, 2020	9734,29	9755,83	9261,40	9328,20	48.158.802.327	171.436.341.789
May 14, 2020	9271,33	9793,27	9255,03	9733,72	56.426.907.637	178.881.641.547
May 13, 2020	8805,39	9317,88	8805,39	9269,99	45.558.144.023	170.352.710.071

Figura 4.1 Muestra de datos históricos de Bitcoin en dólares [16].

4.1.2 Preprocesamiento de datos

El preprocesamiento de datos se realizó como en el estado del arte, desechando todas las columnas a excepción de la fecha y el precio de cierre. El resultado del preprocesamiento es una serie de tiempo univariante que se define como:

$$\{X_1, X_2, \dots, X(t)\}, t = 1, 2, 3, \dots$$

t: Fecha dada por Día, Mes y Año

X(t): Precio de cierre de la criptomoneda en una fecha en particular

4.2 Implementación de modelos

Es prácticamente imposible encontrar modelos que se hayan entrenado y evaluado utilizando las mismas fechas, y aunque se diera la casualidad de encontrarlos, éstos no estarían evaluados con los datos históricos más recientes. Por estas razones la propuesta de solución implica implementar los tres modelos utilizando los mismos datos para que los resultados obtenidos sean más acertados.

AQUÍ HABLAR DE LAS FECHAS USADAS PARA ENTRENAR Y PARA EVALUAR

4.2.1 ARIMA

La implementación de este modelo es relativamente rápida, ya que las series de tiempo se pueden usar de forma directa para entrenar al modelo y que éste se ajuste a los datos.

Así como se mencionó en el estado del arte, no es una tarea sencilla elegir los tres valores que encajan mejor para construir el modelo ARIMA, y a diferencia de R, Python no cuenta con una función capaz de encontrarlos automáticamente. Como la cantidad de datos obtenidos no es elevada, es posible encontrar los valores que ajustan al modelo de la mejor forma probando todas las combinaciones posibles y evaluando su desempeño, logrando esto en un tiempo no muy elevado.

El modelo seleccionado fue ARIMA(4, 1, 4) y éste fue utilizado para compararse contra los modelos LSTM y ARIMA-LSTM.

4.2.2 LSTM

A diferencia del modelo ARIMA, las LSTM no se pueden entrenar de forma directa con tan solo tener la serie de tiempo.

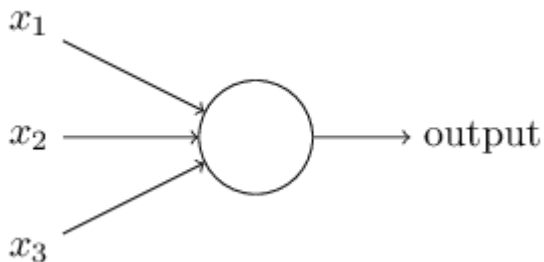


Figura 4.2: Estructura de un Perceptrón [18]

La figura 4.2 muestra la estructura más básica de una red neuronal, pero es suficiente para entender el problema que presentan los datos actualmente. Las redes neuronales reciben

un cierto de número de datos de entrada para al final tener uno o más datos de salida (en este caso uno, ya que los pronósticos serán únicamente del precio de cierre) [18], pero actualmente la serie de tiempo sólo tiene un valor, el precio de cierre, registrado para cada día, entonces... ¿Cómo se determina qué valores se usan como entrada y cuál es la salida esperada para entrenar al modelo?

4.2.2.1 Ventana deslizante

Dada una secuencia de números representados por series de tiempo, es posible reestructurarlos para hacer que tengan el formato que espera la red neuronal (cierto número de datos de entrada y un dato de salida). Esto se logra utilizando los valores de periodos de tiempo anteriores como datos de entrada, y el periodo siguiente como dato de salida. El número de periodos anteriores a tomar por cada fecha depende de un ancho de ventana que se define previo a la reestructuración de los datos [19]. Las siguientes figuras muestran un ejemplo de este proceso aplicado a los precios de Bitcoin.

FIGURA 4.3.JPG

Muestra de serie de tiempo del precio de Bitcoin

FIGURA 4.4.JPG

Ventana deslizante aplicada con ancho de ventana = 2

Como se puede apreciar, las primeras fechas que se encuentran dentro del ancho de la ventana, tienen al menos un valor anterior desconocido, por lo que estas fechas fueron descartadas por completo. Para esta implementación se utilizó un ancho de ventana de **N**.

4.2.2.1 Pronósticos

Para hacer pronósticos, lo primero que se hizo fue construir el modelo haciendo uso de la librería Keras, que permite redes neuronales de forma fácil y rápida [13]. CAPA ESCONDIDA CON N NODOS Y QUE RECIBE COMO ENTRADA LOS PRECIOS CORRESPONDIENTES A LOS VALORES ANTERIORES.

4.2.3 ARIMA-LSTM

Capítulo 7

Conclusiones

TRABAJOS FUTUROS:

Yo use el closing Price porque es como el estado del arte

Maybe un EDA para checar si es chido usar algo más que el precio de cierre

Bibliografía

- [1] M. Rahouti, K. Xiong and N. Ghani, "Bitcoin Concepts, Threats, and Machine-Learning Security Solutions," in *IEEE Access*, vol. 6, pp. 67189-67205, 2018.
- [2] Asante, S. Are Bitcoins price predictable? Evidence from machine learning techniques using technical indicators. 2019. pp. 2.
- [3] Beck et al. Sensing Social Media Signals for Cryptocurrency News . 2019. pp. 1-2.
- [4] Bovaird, C. (2020). *Bitcoin Lost Roughly 50% Of Its Value In A Day*. Mayo 4, 2020, de Forbes Sitio web: <https://www.forbes.com/sites/cbovaird/2020/03/12/bitcoin-lost-roughly-50-of-its-value-in-a-day>
- [5] Adhikari, R; Agrawal R. An Introductory Study on Time Series Modeling and Forecasting. pp. 12-13.
- [6] John H. Cochrane, "Time Series for Macroeconomics and Finance", Graduate School of Business, University of Chicago, spring 1997.
- [7] G.P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting", *Information Sciences* 177 (2007), pp: 5329–5346.
- [8] G.P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing* 50 (2003), pp: 159–175.
- [9] Wang, Z., & Lou, Y. (2019). Hydrological time series forecast model based on wavelet de-noising and ARIMA-LSTM. 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC).
- [10] Yenidogan, I., Cayir, A., Kozan, O., Dag, T., & Arslan, C. (2018). Bitcoin Forecasting Using ARIMA and PROPHET. 2018 3rd International Conference on Computer Science and Engineering (UBMK).
- [11] Anupriya, & Garg, S. (2018). Autoregressive Integrated Moving Average Model based Prediction of Bitcoin Close Price. 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT).
- [12] Wu, C.-H., Lu, C.-C., Ma, Y.-F., & Lu, R.-S. (2018). A New Forecasting Framework for Bitcoin Price with LSTM. 2018 IEEE International Conference on Data Mining Workshops (ICDMW).
- [13] S. Tandon, S. Tripathi, P. Saraswat and C. Dabas, "Bitcoin Price Forecasting using LSTM and 10-Fold Cross validation," 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 323-328.
- [14] Liu, J., Tan, X., & Wang, Y. (2019). CSSAP: Software Aging Prediction for Cloud Services Based on ARIMA-LSTM Hybrid Model. 2019 IEEE International Conference on Web Services (ICWS).
- [15] Kaplan, M. (2018). Bitcoin crash: This man lost his savings when cryptocurrencies plunged. Mayo 16, 2020, de CNN Sitio web: <https://money.cnn.com/2018/09/11/investing/bitcoin-crash-victim/index.html>
- [16] <https://coinmarketcap.com/currencies/bitcoin/historical-data/>

[17] <https://coinmarketcap.com/currencies/litecoin/>

[18] Perceptron

[19] Libro de mi amo y señor Jason <3