



**LOS LIBERTADORES**  
FUNDACIÓN UNIVERSITARIA

## **Predicción del precio en dólares de la moneda digital Solana a través de métodos de machine learning**

### **Predicting Solana's Price in US dollars through machine learning**

Rodolfo Manuel Gómez Rodríguez, [rmgomezr@libertadores.edu.co](mailto:rmgomezr@libertadores.edu.co), Fundación Universitaria Los Libertadores

#### **RESUMEN**

La presente investigación se centra en el uso de modelos de predicción fundamentados en árboles de decisión, como una alternativa al uso de modelos de regresión lineal para la estimación del precio de Solana en dólares. Para tal fin se llevó a cabo el ciclo de vida de un modelo en Machine Learning (ML), realizando la comprensión de los datos, verificación de su integridad, modelado y análisis de los resultados.



Se utilizaron datos del precio de la moneda digital Solana para el último año en periodos por hora y se verificaron distintos tipos de modelos de predicción, concluyéndose que un árbol de decisión era el que mejor desempeño lograba.

Luego se realizó el análisis de la pertinencia del uso del modelo para la toma de decisiones asociadas a la predicción del próximo periodo o periodo (t+1), en donde se propone según los resultados, la necesidad de explorar el uso modelos multivariados que incluyan variables exógenas relevantes.

**Palabras clave:** Machine Learning, Predicciones, Árbol de decisiones, Monedas digitales, Solana

## **ABSTRACT**

This research aims to explore the use of prediction models based on decision trees as an alternative to the use of linear regression models for forecasting digital coin Solana un US dollars. For this purpose, it was applied the life cycle of a Machine Learning (ML) model, performing the understanding of the data, the verification of its integrity, the modeling and the results analysis.



Data about Solana digital currency price was used in a period for the last year (from march 1, 2021 to Mar 31, 2022) and different types of prediction models were verified, concluding that a decision tree was the one that achieved the best performance.

Finally, the analysis of the relevance about the usage of the model for making decisions associated with the prediction of the next period or period (t+1) was carried out, where is concluded that it's necessary to continue exploring better results for the prediction, an therefore it's suggested to apply multivariate models that consider relevant exogenous variables.

**Keywords:** Machine Learning, forecasting, decision tree, digital coins, Solana

## **INTRODUCCIÓN**

La evolución del mercado de las monedas digitales propone la necesidad de ahondar en métodos de predicción de precios que sirvan de apoyo para la toma de decisiones de inversión, en donde la metodología debe ser similar a la establecida para modelos de predicción de variables cuantitativas como los precios de acciones, bonos, tasas de cambio e inflación, entre otros.

El precio de las monedas digitales expresado en dólares se caracteriza por un comportamiento caracterizado por su volatilidad, dinamismo y no linealidad, en donde la ciencia de datos continúa explorando técnicas para comprenderlo de mejor manera (Li, 2.022).



# LOS LIBERTADORES

## FUNDACIÓN UNIVERSITARIA

Para entender dicha evolución de este mercado, es imprescindible hacer referencia inicialmente al Bitcoin, su mayor exponente hasta la actualidad. Desde el año 2.008, con la operación de Bitcoin (BTC), se ha presentado un crecimiento exponencial del valor de las operaciones con monedas digitales. Es así que para abril 30 de 2.013 la capitalización del mercado<sup>1</sup> ya correspondía a \$1.699.160.064 USD, mientras que para el 31 de marzo de 2.022 ya se contó con una capitalización de \$2.129.689.641.243 USD<sup>2</sup>, indicando un crecimiento de más de 1.253 veces en 9 años.

Cada día hay más instituciones y personas que están accediendo a negociar con este activo, ya que muchos lo consideran como un mecanismo de reserva de valor en el cual pueden preservar su patrimonio o entenderlo a su vez como un activo que les puede generar utilidades y por ende, incrementos en su ingreso.

---

<sup>1</sup> La capitalización de mercado es una medida económica que indica el valor total de todas las monedas digitales reconocidas.

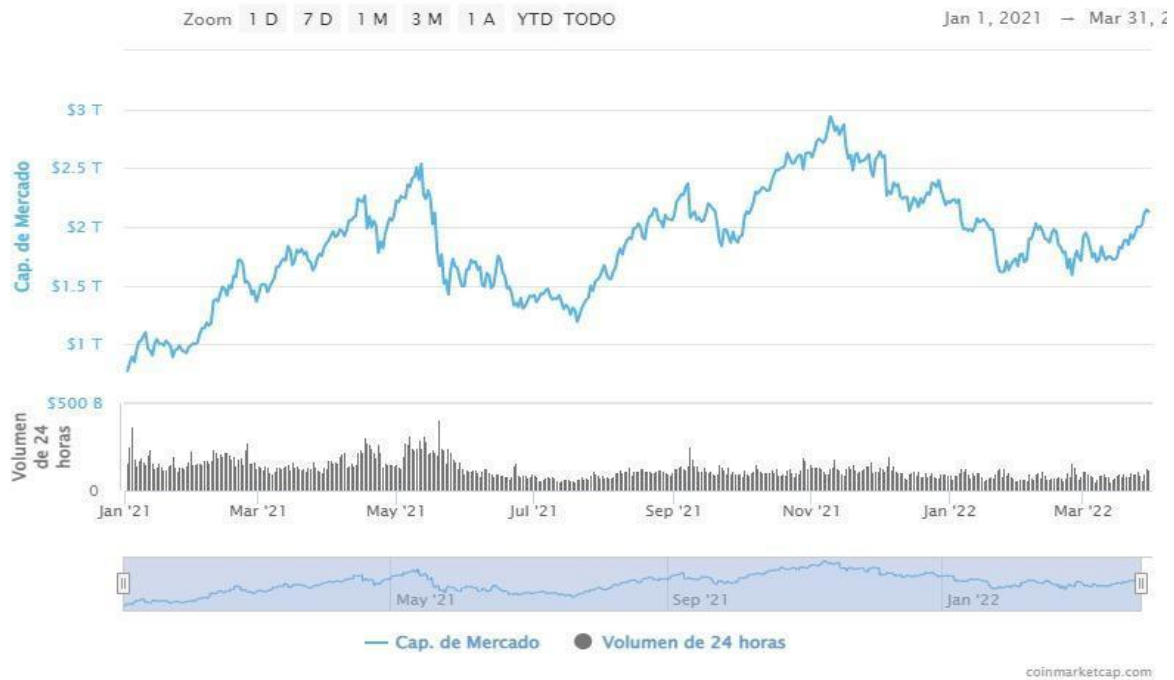
<sup>2</sup> Coin Market Cap. Capitalización total del mercado de criptomonedas, marzo 27 de 2.022).  
<https://coinmarketcap.com/es/charts/>



# LOS LIBERTADORES

## FUNDACIÓN UNIVERSITARIA

Ilustración 1 Capitalización total del mercado de criptomonedas



Fuente [www.coinmarketcap.com](http://www.coinmarketcap.com)

El crecimiento no sólo se evidencia con el valor de capitalización, también ocurrió con la oferta de productos en el mercado. Si bien se dio un inicio a esta revolución con el Bitcoin, después de doce años se puede indicar que a la fecha operan 476 plataformas para hacer intercambio de 18.497 monedas digitales (Coin Market Cap, marzo 27 de 2022).

Ahora bien, este crecimiento dinámico de monedas alternativas al Bitcoin, sugieren una tendencia decreciente del impacto de la dominancia del Bitcoin en el comportamiento de los precios. Kulal (2021), con lo cual se genera un espacio para la generación de modelos de predicción de precios específicos para distintas monedas y no plantearse que un modelo de predicción del precio del Bitcoin sea generalizable a otras monedas digitales.



# LOS LIBERTADORES

## FUNDACIÓN UNIVERSITARIA

Ilustración 2 Evolución de la dominancia del Bitcoin en el mercado de monedas digitales



Fuente [www.coinmarketcap.com](http://www.coinmarketcap.com)

En la actualidad se identifican en el mercado otros actores con propuestas que innovan frente a la propuesta inicial de Bitcoin, dentro de las que cabe mencionar como las más importantes a Ethereum, Cardano, Solana, Terra y Avalanche.

Estas opciones ofrecen diversas oportunidades de articulación con todo lo que actualmente se hace en la sociedad, tal como puede ser la compraventa de activos, la certificación de un sin número de operaciones, juegos electrónicos en los que se genera valor por la participación realizada, préstamos y titularidad de propiedad de activos entre otros.

Dentro de lo anterior, al considerarse la posibilidad de hacer intercambio de estos activos, también se construye un mercado de trading (comercio) con las mismas, como ocurre con los



mercados de valores en las bolsas bursátiles, en lo que corresponde a las monedas nativas que tiene cada una de estas plataformas.

Solana empezó a idearse en el año 2017 como plataforma de Blockchain, logrando materializarse en el mercado en marzo de 2020. Entendida como plataforma, es considerada un protocolo que facilita la creación de aplicaciones descentralizadas, siendo menos complejo su proceso de validación de transacciones al compararse con las ya existentes, Sebastián, A. (2021).

A veces los procesos de validación en las plataformas de Blockchain toman minutos y la arquitectura del protocolo de Solana disminuye ostensiblemente esta barrera que se tiene en la transabilidad de todo el sistema, el algoritmo de proof of history propone mejores rendimientos en tiempo al algoritmo de proof of stake sobre el cual operan Bitcoin y las otras representativas del mercado, por lo cual se presentó como una nueva arquitectura para una cadena de bloques con un mejor rendimiento. Yakovenko, A. (2017).

Además del anterior elemento como parte del análisis fundamental, la apreciación de la moneda digital que opera la plataforma Solana se puede observar en el siguiente gráfico, notándose que en marzo de 2020 era de \$19 USD, el 5 de noviembre de 2021 logró su máximo histórico en \$258.93 USD y a 31 de marzo de 2022 el valor a dólares fue de \$124.22 USD.



# LOS LIBERTADORES

## FUNDACIÓN UNIVERSITARIA

Ilustración 3 Evolución histórica del precio de Solana



Fuente [www.coinmarketcap.com](http://www.coinmarketcap.com)

Por otra parte, en referencia al análisis técnico, asociado al comportamiento del precio expresado en dólares, éste se centra en la detección de un modelo que nos permita predecir en el corto plazo, específicamente para sólo el periodo subsiguiente, que correspondería a la próxima hora.

Actualmente, se han propuesto modelos que se apartan de la estimación de regresores lineales o modelos ARIMA, que descomponen la regresión en el componente autoregresivo, la integración y la media móvil. La aplicación de las técnicas de Machine Learning y en especial el desarrollo de los métodos de árboles de decisión han obtenido resultados más precisos, Löning, (2.021).

Por otra parte, con la evolución de Big Data en cuanto a la disponibilidad de información y la capacidad de procesarla, se pretende explorar distintos modelos de predicción y



evolucionar en el análisis de la capacidad del que se pueda definir como el que menor error logra tener en las predicciones a realizar, Guida(2.019).

En congruencia con lo anterior, el objetivo de ésta investigación se centra en Pronosticar el tipo de cambio en dólares de la moneda digital Solana, implicando el utilizar herramientas de programación que faciliten el procesamiento de datos y la interpretación de los resultados que se obtienen, en donde se va a analizar el desempeño de diferentes modelos de predicción, y con esos resultados se procede a validar el alcance que tienen los modelos univariados en la predicción de precios.

Esto último, se hace a través de un comparativo del resultado de la predicción, con respecto al valor realmente ocurrido, por lo que se puede enmarcar como un backtesting o comprobación en el pasado.

## **REFERENTES TEORICOS**

Actualmente, al hacer referencia a los mercados de capitales, se encuentran planteamientos como los de Gómez (2.000) en los que es recurrente preguntarse acerca del comportamiento de estos mercados. En buena medida por la preocupación inherente al inversionista con respecto al riesgo y a la incertidumbre.



Una de las salidas a este tipo de cuestionamientos para la toma de decisiones ha sido abordada a través del análisis técnico para la predicción del comportamiento de precios. Gómez (2000) en la segunda parte de su libro plantea la serie de herramientas y consideraciones que pueden ser utilizadas, las cuales no pueden apartarse de un componente teórico que respalde el análisis.

Es así que resulta común identificar la proliferación de propuestas de predicción de precios fundamentadas en el trazado visual de líneas de tendencia, canales, soportes, resistencias, así como la identificación de patrones de orden gráfico en las que las figuras de cabeza y hombros, triángulos simétricos, círculos o triángulos ascendentes o descendentes, rectángulos, abanicos y bordes entre otros, se conjugan con promedios móviles y bandas de Bollinger para intentar dar información acerca del posible comportamiento de un precio.

Por otro lado, si cuenta con el análisis de series de tiempo, del cual podríamos revisar de manera detallada lo expuesto por Wooldridge (2001), en donde el cálculo de mínimos cuadrados ordinarios nos presenta una solución más acompañada de elementos teóricos si lo comparamos con el análisis técnico. En este caso, el esfuerzo se centra en identificar de qué manera lo ocurrido en el pasado nos permite hacer predicciones de un precio en el futuro.

Elementos como los rezagos o componentes de tendencia, ciclo, estacionalidad y de irregularidad o residuales, son los que toman el protagonismo para contar con mejores predicciones, Anderson (2000) y Chao (1993). El solucionar problemas que permitan probar



que la diferencia entre las predicciones y lo que realmente ocurre sea atribuido a algo que no es observable se convierte en el éxito de la investigación o el experimento que se lleva a cabo.

De lo anterior, surge la necesidad de hacer una descomposición de los datos históricos de la variable o serie de tiempo y bajo ese escudriñamiento, hacer un entendimiento de su comportamiento, mediante medias móviles, funciones de tendencia y tratamiento a casos atípicos, suavizado de los datos y en ese camino llegar a pruebas estadísticas que nos acompañen el análisis.

Aquí mismo, es importante definir si el tipo de pronóstico es de un paso hacia adelante o si la pretensión se enruta a contar con predicciones de más de un periodo, Wooldridge(2.001). Por otra parte, en el marco de los mercados de futuros y opciones, se le da utilidad a los árboles de decisión, Hull (2.002).

En los últimos años, se han resuelto unos limitantes en la capacidad de análisis de información, especialmente en lo asociado al acceso a información y la capacidad de procesamiento de esta.

Como se ha planteado previamente, se ha sostenido que en predicción de precios, se tiene como mejor opción los modelos el valor observado más recientemente, en donde lo sencillo es mejor y los modelos naive o ingenuos basados en probabilidad bayesiana son los mejores, Escobar (2.022).



Sin embargo, con el uso de bibliotecas de machine learning, se comprueba que merece la pena revisar modelos basados en árboles de decisión, bosques aleatorios o gradientes. En este caso, se ordenó por el resultado del error absoluto medio (MAE) a los modelos evaluados. El MAE se entiende como la medida del comportamiento de un pronóstico y es calculado mediante el promedio de los valores absolutos de los errores del pronóstico con respecto al valor real.

Un elemento adicional a considerar en la generación de modelos, corresponde a la decisión que se debe tomar con respecto a cuál modelo puede tener mejor desempeño en lo que a predicción se refiere. Existen diferentes tipos de modelos con los cuales se han hecho diversos análisis o propuestas para la predicción de precios, en los que estos mismos también requieren de la toma de decisiones acerca de los parámetros que utiliza. En muchos casos esto se refleja en los rezagos a considerar, circunscribirse a las relaciones lineales que se pueden detectar, a la estimación de modelos no paramétricos o la necesidad de la inclusión de variables exógenas que permitan tener mayor capacidad de proponer la aleatoriedad en los residuales.

En lo que corresponde al análisis del modelo en lo concerniente a su su robustez, se hace uso de la validación cruzada, la cual permite evaluar el modelo a partir de subconjuntos de los datos utilizados. Una de las técnicas de validación cruzada es conocida como la de k-fold (que pueden ser entendidas como pliegues, divisiones o iteraciones), en donde k veces se procede a dividir el conjunto de datos disponibles en dos: entrenamiento y validación. El

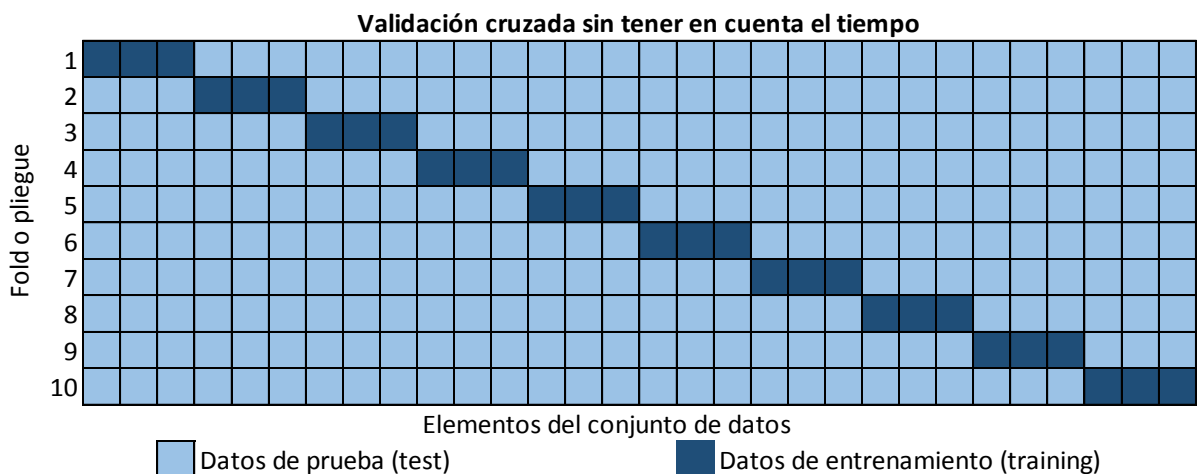


modelo se crea con los datos de entrenamiento y se generan predicciones para el subconjunto que comprenden los datos destinados a la validación.

Estos resultados se contrastan con los datos que no se utilizaron en el entrenamiento y que no hicieron parte del proceso de formulación del modelo. El contraste estadístico de los resultados de la predicción con respecto a los datos de validación son los que nos permiten evaluar la validez del modelo en términos de error, lo cual es muy útil en modelos no paramétricos.

En resumen, la validación cruzada plantea la división del conjunto de datos en diferentes subconjuntos a los que se les denomina folds, en los cuales se repite  $k$  veces el entrenamiento y la predicción, para luego evaluar de manera integrada los resultados, que en esta técnica incluyen la mayor cantidad de datos posibles como entrenamiento. (Pérez-Planells, Ll. 2015).

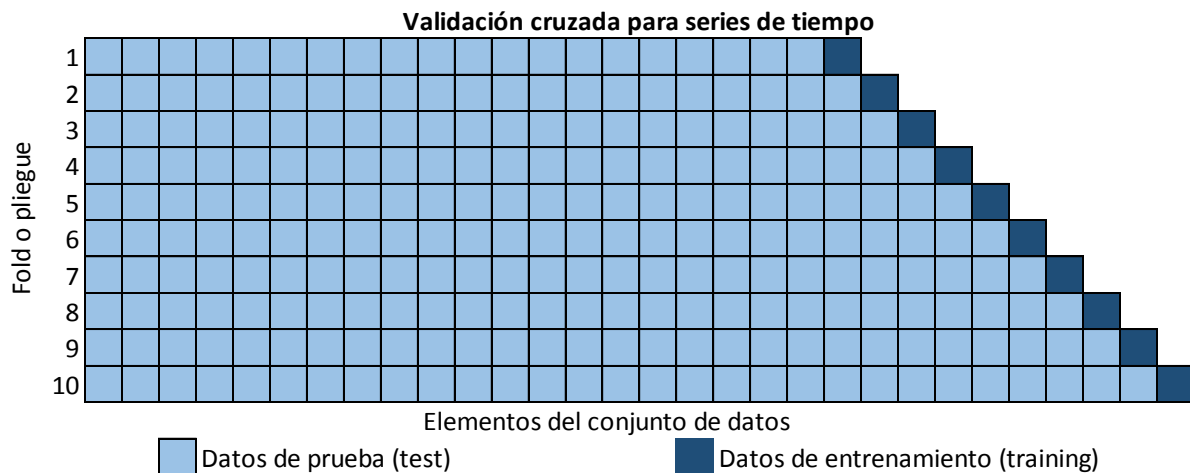
*Ilustración 1 Distribución de los subconjuntos de datos de prueba y entrenamiento para 10 fold*





Es importante considerar que para series de tiempo, esta validación cruzada no permite hacer selecciones de subconjuntos de manera aleatoria, toda vez que no se puede plantear el evaluar un predicción de un periodo específico con datos que se obtienen después en el tiempo, lo que hace relevante plantear un criterio para definir los datos de entrenamiento y de validación.

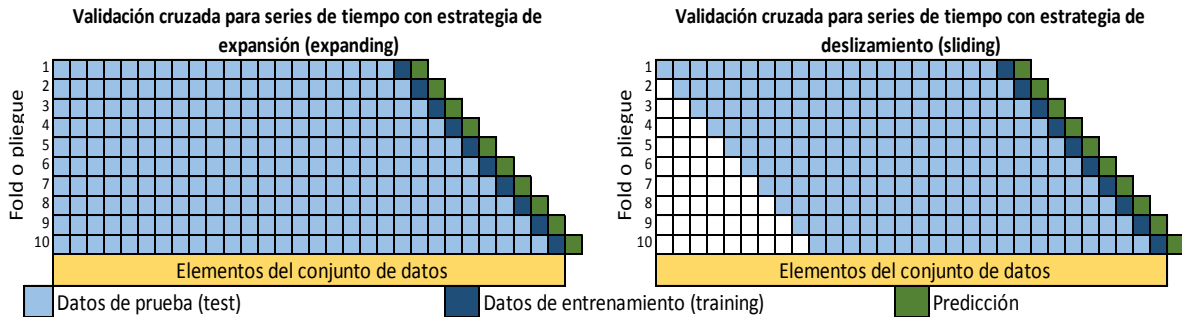
*Ilustración 2 Distribución de los subconjuntos de datos de prueba y entrenamiento para 10 fold en series de tiempo*



A su vez, es necesario tener en cuenta la estrategia con la que se seleccionan los datos para la realización de la validación cruzada entre expanding y sliding. La primera opción va incrementando el tamaño de los datos de entrenamiento, mientras que la segunda desliza en el tiempo la selección de los datos.



*Ilustración 4 Representación gráfica de la validación cruzada a través de muestras crecientes (expanding) o de desplazamiento*



Cuando se cuenta con una considerable cantidad de datos, no se esperan diferencias significativas, pues el conjunto que compone los datos de prueba no cambia de manera significativa.

## **METODOLOGÍA**

La investigación propuesta es de orden cuantitativo, toda vez que la podemos circunscribir a un componente secuencial y probatorio, Hernández (2.010). El análisis se basa en una variable cuantitativa continua, como es el precio expresado en dólares, pero cabe mencionar que los datos corresponden al precio de cierre cada hora. Cabe mencionar la presencia de un componente exploratorio, en el que se acompaña la posición y el camino recorrido de autores



recientes, orientados a la búsqueda de modelos de predicción de precios, que ven en los árboles de decisión, las técnicas de bosque aleatorio y las redes neuronales (especial las redes recurrentes Long-Short Term Memory-LSTM), la posibilidad de obtener mejores resultados en lo que hace referencia a predicciones de un paso. Li (2022).

De los resultados obtenidos, se debe plantear si es suficiente la predicción a través de un componente autoregresivo o si resulta recomendable la inclusión de variables exógenas que apoyen mejoras en los resultados de la predicción.

En referencia a los datos, se hizo la búsqueda de fuentes confiables de información, definiéndose el uso de portal lunarcrush, en cual cuenta con datos por hora del precio de variadas monedas digitales, así como otras variables que en análisis posteriores podrían ser utilizadas como variables exógenas para eventualmente explorar otras alternativas de predicción de un modelo, manteniendo la fuente.

A través de una API<sup>3</sup> se hace el cargue de la información histórica y mediante Python se obtiene un dataframe para Solana.

El análisis de la información permite identificar la existencia de datos consistentes para el precio de Solana desde el mes de Octubre de 2020, pero para efectos del presente estudio, se define que el análisis comprenda los datos desde 1 de marzo de 2021 y se realicen

---

<sup>3</sup> Una API es una interfaz de programación de aplicaciones, mediante la cual se pueden extraer datos para que otro programa computacional los pueda utilizar.



predicciones para cada hora en el mes de marzo de 2.022. A su vez, ya se cuenta con datos reales de marzo de 2.022 para hacer la comparación y la medición de errores correspondiente.

Por otra parte, se definió el uso de la biblioteca de Python denominada Pycaret, la cual ha sido diseñada para la realización de tareas asociadas al uso de modelos de machine learning. Es de destacar que permite hacer la comparación de resultados entre diferentes modelos, para así tomar la mejor decisión en el uso de los mismos. El criterio para seleccionar el mejor modelo corresponde al que genere menor error cuadrático medio dentro del número de pliegues o iteraciones que se ha establecido para hacer la validación cruzada.

Luego de hacer la selección del modelo a utilizar, se calculó el modelo y se hace la predicción para sólo un periodo posterior, es decir la siguiente hora, debido que no se evidencia que sea pertinente realizar predicciones a más periodos.

Finalmente, se calculan todas las predicciones por hora para el mes de marzo de 2.022 y se presenta un análisis de los resultados.

La información obtenida a través de la API de Lunarcrush cuenta con el 100% datos para la variable de estudio, que corresponde al precio de cierre de Solana cada hora desde el 1 de marzo de 2.021 a las 00:00:00 horas, y hasta el 31 de marzo de 2.022, contando con información de inicio de más de un año, en la que no se detectaron valores perdidos que requirieran tratamiento de calidad del dato,

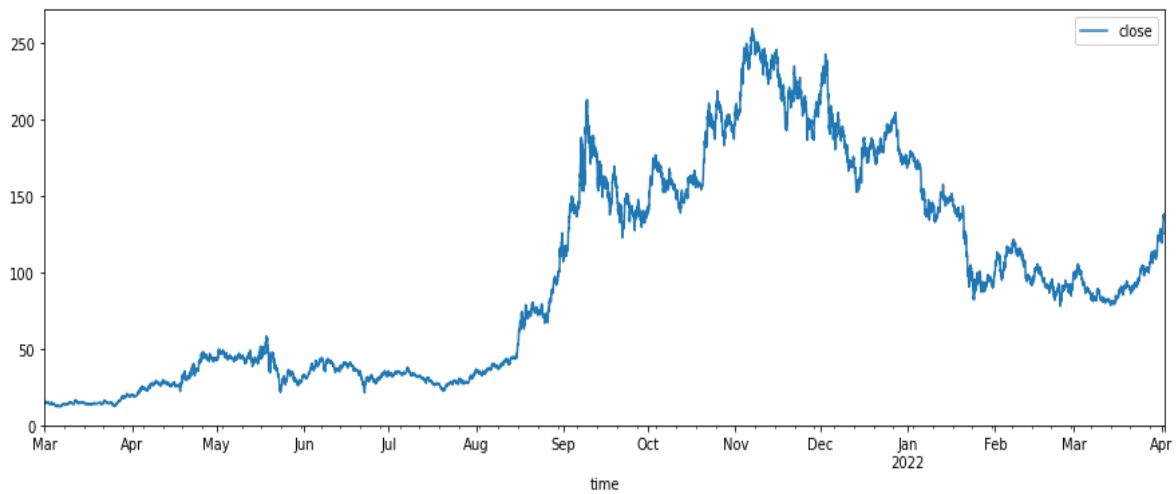


# LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Un histórico del precio de Solana expresado en dólares se puede apreciar en el siguiente gráfico:

*Ilustración 5 Comportamiento del precio de Solana entre marzo de 2.021 y marzo de 2.022*

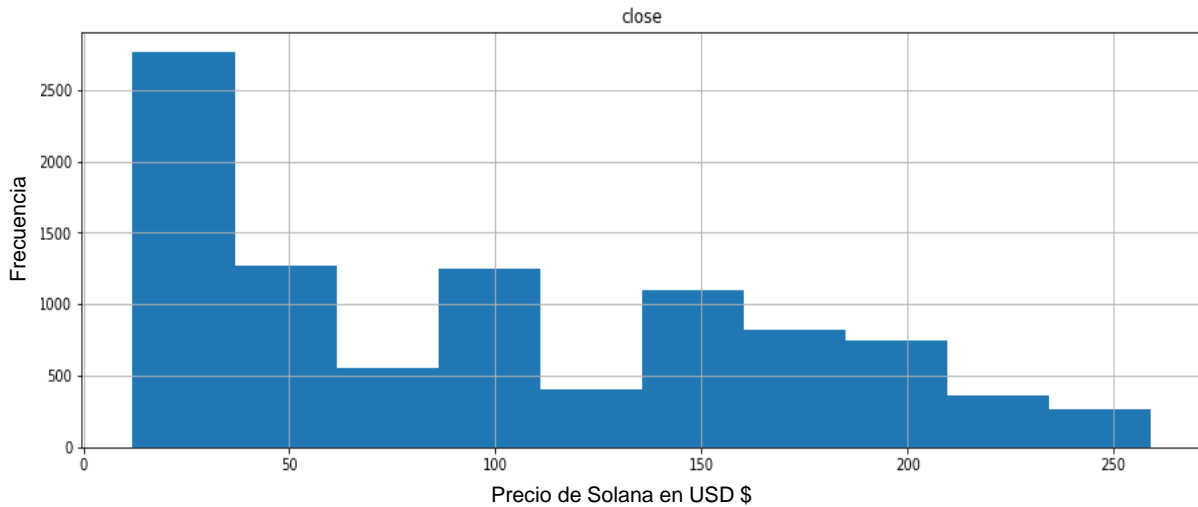


*Fuente cálculos propios con base en la información de [www.lunarcrush.com](http://www.lunarcrush.com)*

Del comportamiento histórico, se observa que el precio de Solana no sigue una distribución normal, lo cual se evidencia en el siguiente gráfico y esto se debe a que en el tiempo ha tendido a apreciarse.



*Ilustración 6 Histograma de frecuencias del precio de Solana entre marzo de 2.021 y marzo de 2.022*

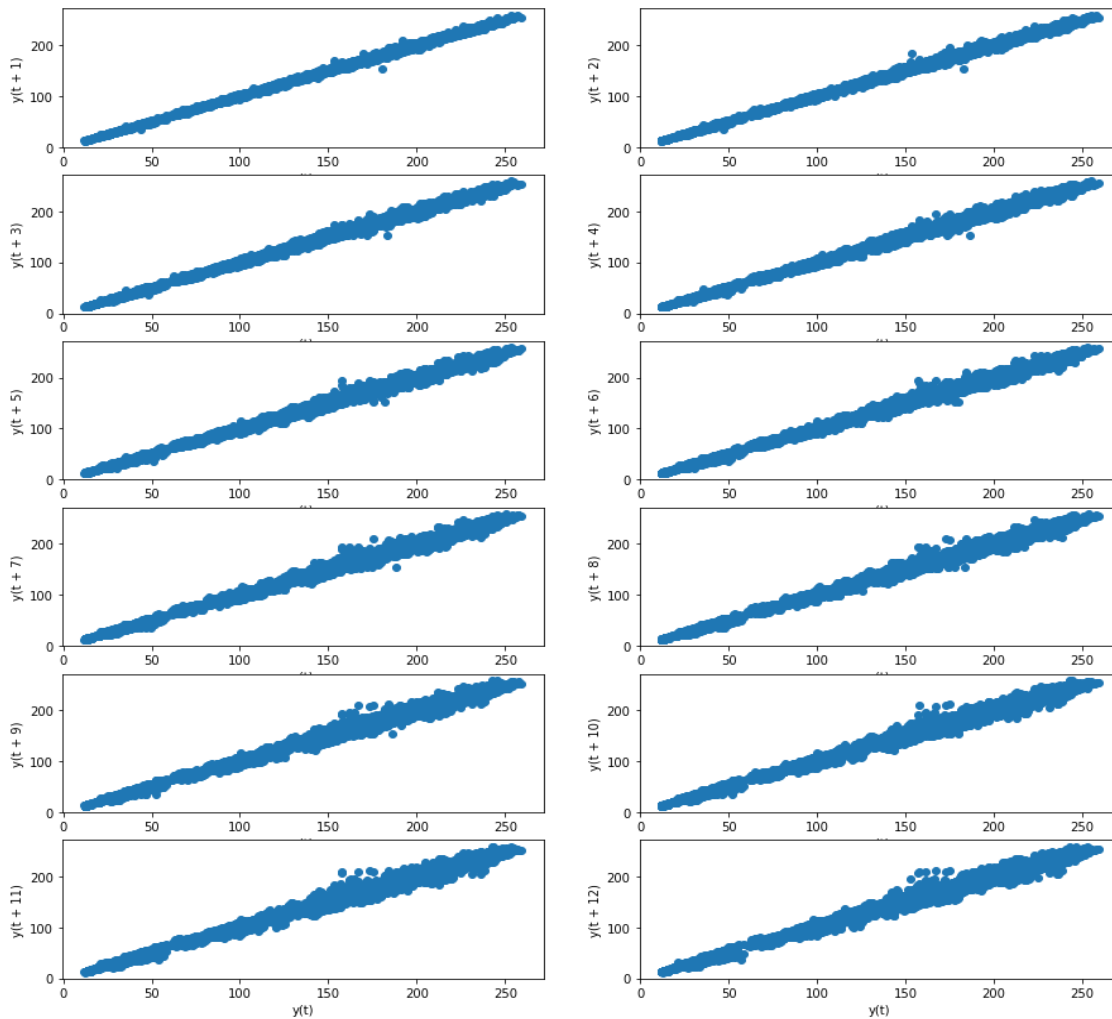


*Fuente cálculos propios con base en la información de [www.lunarcrush.com](http://www.lunarcrush.com)*

Al intentar plantear la posibilidad de una predicción a partir de componentes autoregresivos, es decir, con sus propios datos del pasado, encontramos que las relaciones con respecto a doce periodos atrás que en este caso corresponden a horas, plantean la existencia de una relación entre sus comportamientos. Esto desde el ámbito exploratorio indica que sí puede ocurrir que se cuente con predicciones pertinentes, a partir de los datos del pasado.



Ilustración 7 Relación del precio de Solana con sus rezagos de hasta 12 horas



Fuente: Cálculos propios.

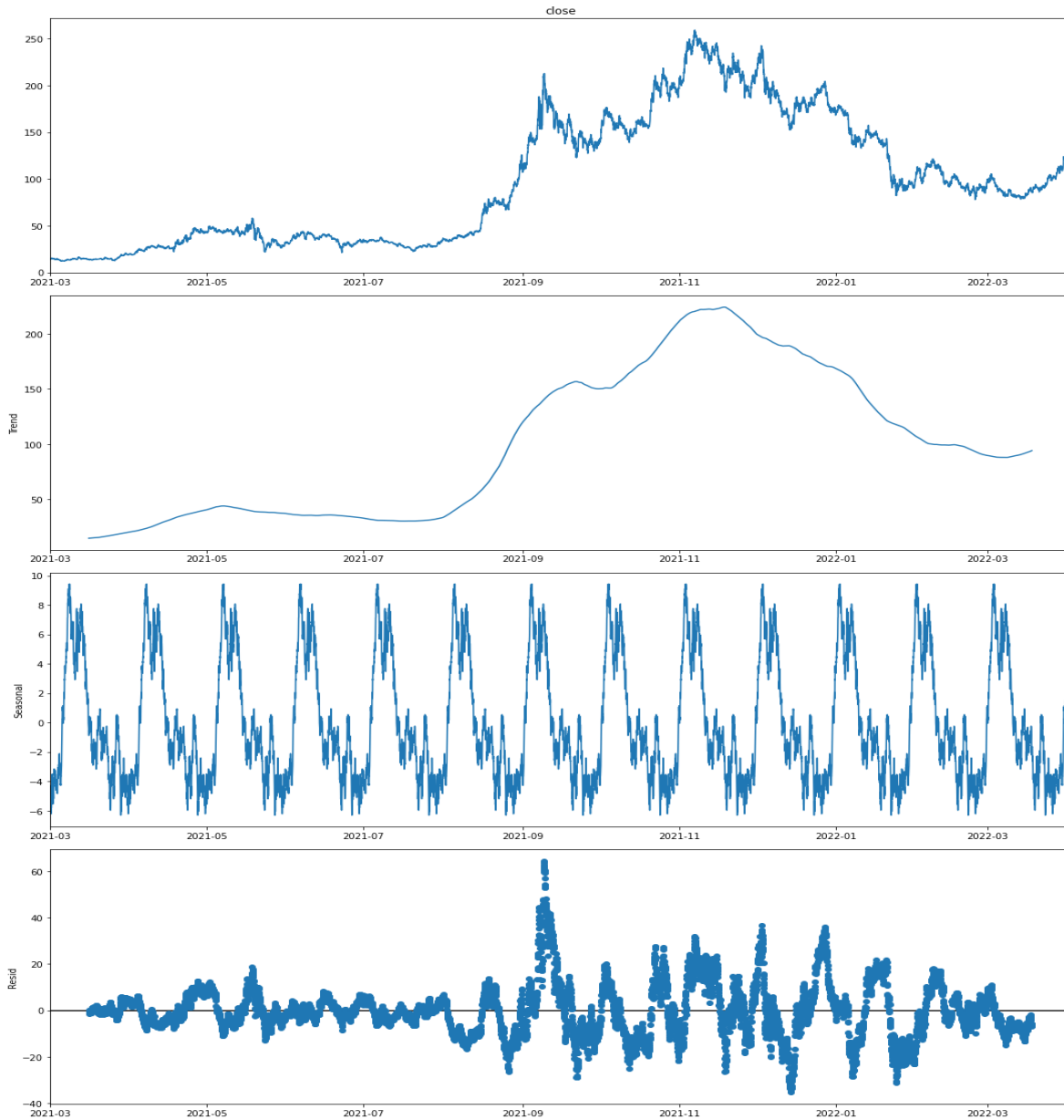
Al realizar una descomposición de la serie de datos, es posible identificar que Solana en el segundo semestre de 2021 tuvo un crecimiento sostenido, el cual ha luego pasado a un comportamiento en el que su precio ha venido decreciendo, lo cual es una situación que ha venido ocurriendo con todo el mercado, en donde se plantea la evolución de un proceso de corrección de precios, posterior al boom presentado el año pasado, en donde muchas monedas digitales, incluida el Bitcoin, obtuvieron su nivel máximo histórico.



# LOS LIBERTADORES

## FUNDACIÓN UNIVERSITARIA

Ilustración 8 Descomposición de la serie de datos del precio de Solana



Fuente: Cálculos propios.

Por otra parte, el comportamiento de Solana ya planteado permite evidenciar la dificultad de explicarlo a través de análisis de tipo lineal, como pudo notarse en el gráfico de residuales antes de septiembre, en donde se contaba con una mayor estabilidad que lo observable en el último semestre.



## **Definición de los parámetros para el análisis**

Para el análisis de los datos, es necesario plantear las consideraciones o parámetros que se van a tener en cuenta. En primera medida, luego de identificar la existencia de datos desde el año 2.020, en donde para futuros análisis multivariados es consistente la disponibilidad de información de octubre del mismo año, se ha optado por acotar la serie a un periodo de tiempo comprendido entre marzo 1 de 2.021 y marzo 31 de 2.022.

Se van a predecir los valores de Solana para cada hora en el mes de marzo de 2.022. Con este marco, se procede a verificar la existencia de datos para cada periodo y hace la preparación de los datos para procesamiento en Python, a través de Google Colaboratory que es un cuaderno de Jupyter que permite procesamiento de datos y registro de información útil para el desarrollo del análisis.

Ahora bien, en cuanto al número de pasos o pliegues con los que se hace la predicción, si bien se plantea en la literatura el uso de 10 subconjuntos o folds, en esta investigación se realizó el análisis con 12 y 72 folds, en orden a verificar si el hacer mayor división en la validación cruzada nos arrojaba diferentes resultados.

La metodología planteada y el uso de las herramientas de programación antes descritas garantiza el procesamiento y análisis 27 diferentes modelos, que a la vez responden a



tipologías diferentes, dentro de los cuales se incluyen los ARIMA, árbol de decisiones y bosques aleatorios. De esta manera es posible comparar y hacer selección del mejor modelo de predicción, de acuerdo a las capacidades de predicción, con respecto a los datos reales.

Luego de la selección de modelo, se realizaron predicciones para cada hora desde el 1 de marzo y hasta el 27 de marzo de 2022 a las 22:00. No se realizó una sola predicción para este intervalo de tiempo, sino que se realizaron 646 predicciones independientes. Con esta información se realizó un comparativo con los datos reales y se identificaron las conclusiones con respecto al objetivo de la predicción de Solana.

## RESULTADOS

De acuerdo con la información utilizada, la siguiente es la lista de modelos organizada por sus resultados de desempeño al ser evaluados:

*Tabla 1 Resultados de la evaluación de modelos para la predicción de Solana*

Código	Model	MAE	MAPE
dt_cds_dt	Decision Tree w/ Cond. Deseasonalize & Detrending	0.5572	0.0045
Croston	Croston	0.6747	0.0055
et_cds_dt	Extra Trees w/ Cond. Deseasonalize & Detrending	0.7333	0.0059
rf_cds_dt	Random Forest w/ Cond. Deseasonalize & Detrending	0.7653	0.0062
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.7702	0.0062
lasso_cds_dt	Lasso w/ Cond. Deseasonalize & Detrending	0.8132	0.0066
ada_cds_dt	AdaBoost w/ Cond. Deseasonalize & Detrending	0.8426	0.0068
gbr_cds_dt	Gradient Boosting w/ Cond. Deseasonalize & Det...	0.8724	0.0071
Theta	Theta Forecaster	0.8802	0.0071
Naive	Naive Forecaster	0.8875	0.0072

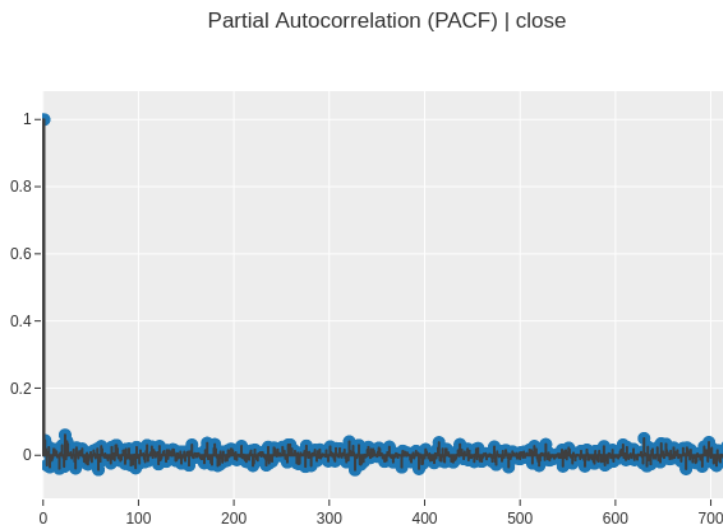
*Fuente: Cálculos propios a través de time series experiment de Pycaret.*



En conclusión, la estimación del precio de solana para la siguiente hora se realiza mediante el modelo de árbol de decisiones que incluye la desestacionalización de datos y el suavizado exponencial de la tendencia.

El resultado de las correlaciones parciales fue el siguiente:

*Ilustración 9 Gráfico de autocorrelación parcial*

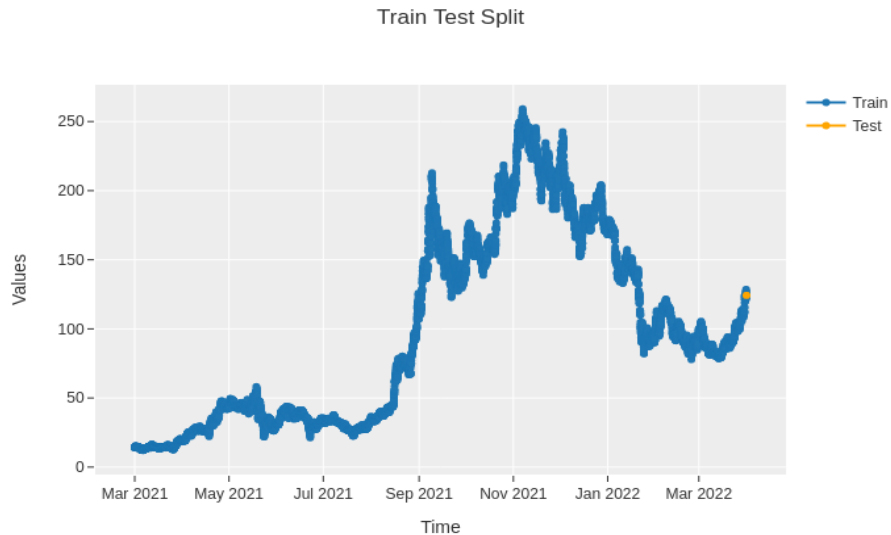


*Fuente: Cálculos propios a través de time series experiment de Pycaret.*

Como resultados del modelo, en atención a la predicción de un solo periodo, inicialmente se cuenta con la siguiente predicción para la primera hora del 1 de abril de 2.022:



*Ilustración 10 Contraste entre el entrenamiento y el valor resultante en la predicción*



*Fuente: Cálculos propios a través de time series experiment de Pycaret.*

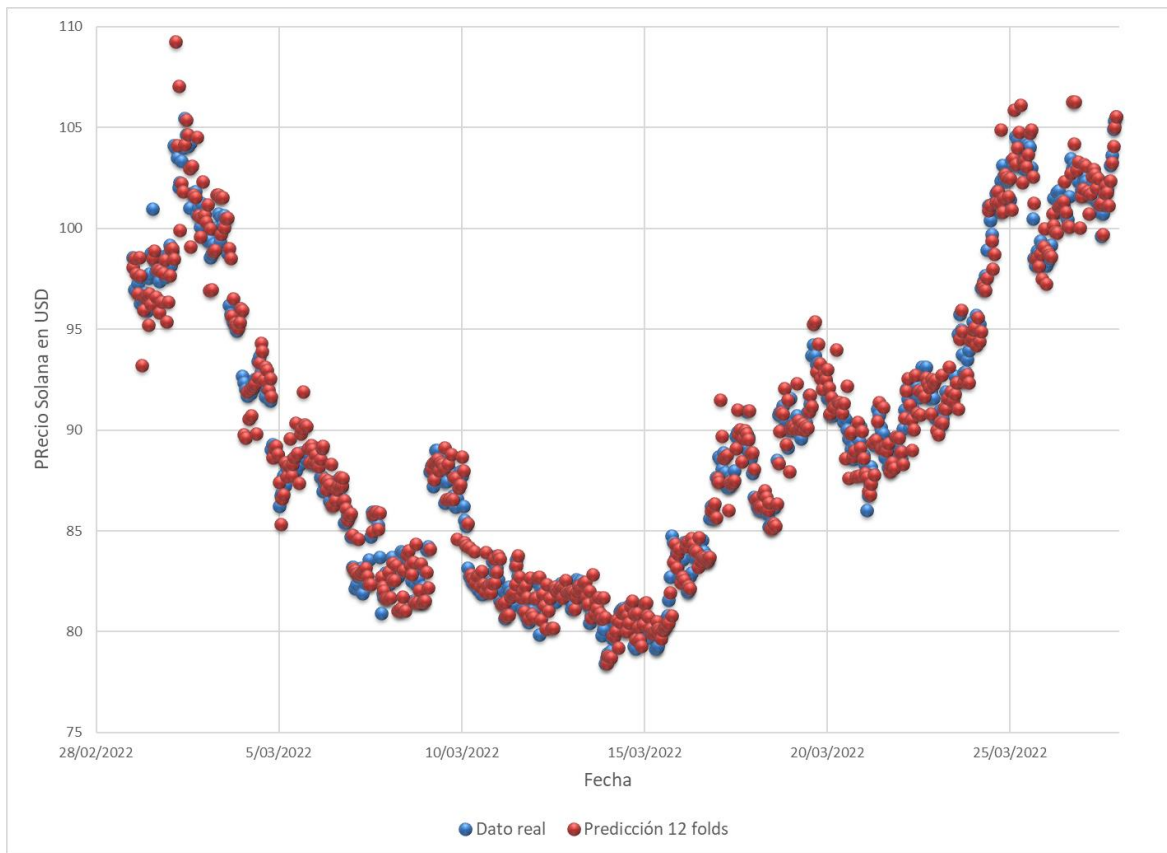
## **DISCUSIÓN DE RESULTADOS**

Ya con la elección del modelo que nos propone los mejores resultados, se pone en consideración las predicciones por intervalos de una hora que se realizaron bajo las condiciones previamente descritas.



Para el periodo comprendido entre el 1 de marzo a las 0:00:00 y el 1 27 de marzo a las 22:00, se realizó la estimación de los precios de Solana, en donde se evidencia que el MAE<sup>4</sup> en una predicción con 12 folds para ese periodo es de 0,9053<sup>5</sup>.

*Ilustración 11 Comparativo entre las predicciones y los datos reales*



*Fuente: Cálculos propios*

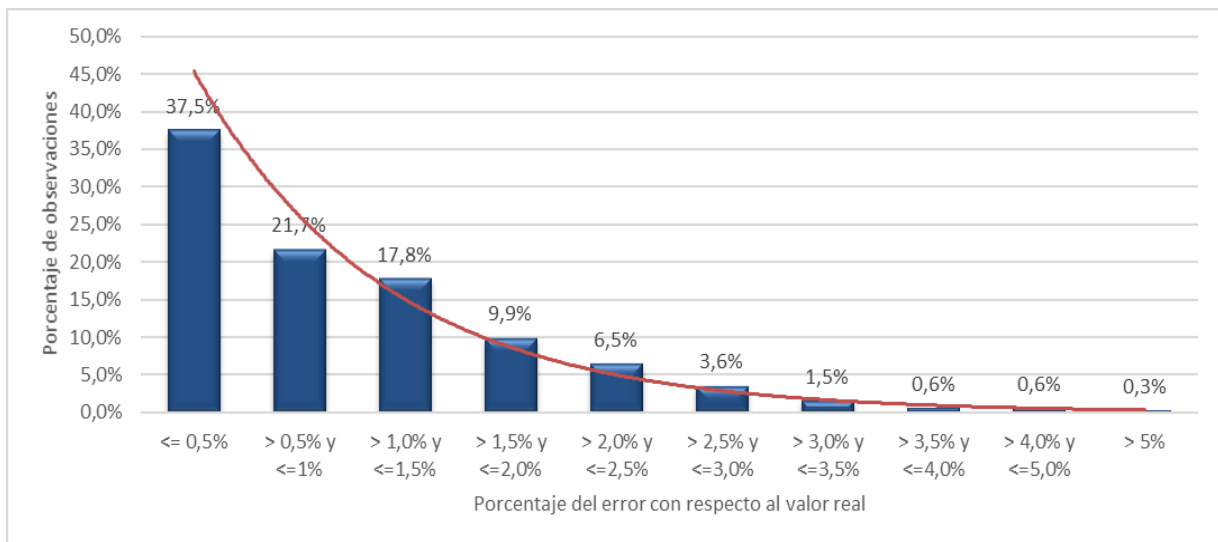
<sup>4</sup> El error absoluto medio (MAE) es el promedio de las diferencias entre el valor de la predicción y el valor real del precio de Solana en un determinado momento de medición, en términos de valores absolutos.

<sup>5</sup> También se realizó el cálculo, análisis y comparación de modelo para 72 folds, en donde el MAE fue de 0,9364 implicando que el dedicar mayores capacidades de procesamiento (seis veces más con respecto a 12 folds) no genera resultados significativamente diferentes.



En este entendimiento, se encuentra que las predicciones para cada hora en el mes de marzo de 2.022 con 12 folds, nos presentaron errores frente a los valores reales que se encuentran con la siguiente distribución:

*Ilustración 3 Distribución de las predicciones según el porcentaje de error de la predicción con respecto al valor real*

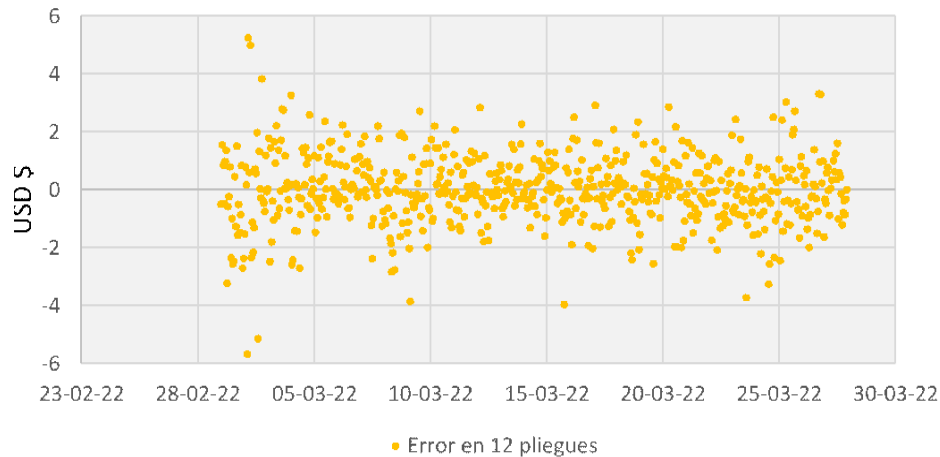


*Fuente: Cálculos propios*

En consecuencia, un 86,8% de las predicciones no se alejaron en más de un 2% del valor real, El cálculo de las diferencias entre los valores reales y las predicciones, que nos permiten identificar un error medio absoluto de 0,91 USD para el periodo contrastado.



### Distribución de errores 12 pliegues



Sin embargo, se observó que el promedio de los valores absolutos de las variaciones del precio de Solana fue de 0,6 USD y en las predicciones, el valor promedio de las variaciones de los valores absolutos fue de 1,04 USD.

Por otra parte, se proyectó un escenario en el que se identificaron las orientaciones que tenían las variaciones de precio en la predicción (sube, baja o se mantiene estable) y se contrastó con las variaciones reales del precio. Es decir, se observó la convergencia en el tipo de variación de los precios reales de Solana y las predicciones.

Para esto, se tipificaron tres estados en orden a establecer la orientación de las variaciones de los precios de un periodo a otro, tanto en los datos reales, como en las predicciones:



*Tabla 1 Descripción de los tipos de variación establecidos para el análisis de convergencia en los sentidos de variación del precio de Solana*

Tipo de variación	Descripción
Incremento	Variación del precio en dólares de Solana que corresponde a un aumento de más del 1% con respecto al periodo anterior
Estable	Las variaciones del precio en dólares de Solana no son superiores al 1% con respecto al periodo anterior.
Caída	Variación del precio en dólares de Solana que corresponde a una disminución de más del 1% con respecto al periodo anterior

En este sentido, se hace el análisis de las variaciones del precio real y estimado de solana en la serie de datos contenida entre el día 1 de marzo de 2.022 a las 00:00 y las 22:00 horas del 27 de marzo de 2.022.

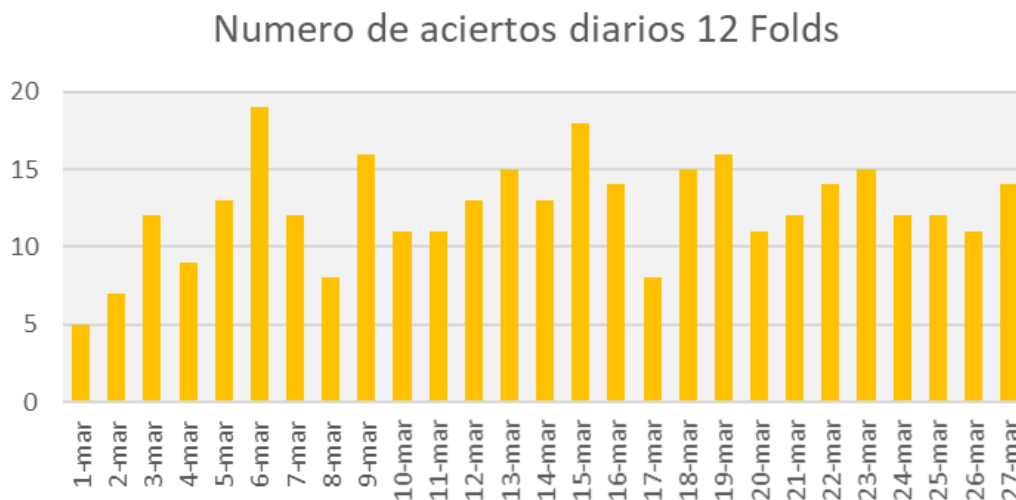
Se definió que si las variaciones del precio no eran superiores al 1% con respecto al precio del periodo anterior, se consideraba que no habían cambios significativos, por lo que variaciones al alza (incremento) o a la baja (caída). Este criterio se fundamentó en los márgenes de intermediación que tienen los portales de intercambio, en lo que supodría que no tiene sentido plantearse variaciones sino superan ese porcentaje.

En consecuencia, si para cada variación del precio de Solana por hora se contaba con la misma tipificación tanto en la serie de datos reales, como la serie de datos estimados, se asume un acierto.



El análisis nos permite encontrar que el día con más aciertos entre la variación real y la de la predicción, fue en 18 de las 24 horas en las que se hizo el contraste para el 6 de marzo de 2.021.

Ilustración 4



Si bien se tienen considerables aciertos, corresponden a una efectividad del 52% con respecto al total, por lo que desde este punto de vista, se debe entender como una aproximación que debe seguir evolucionando, porque desde esta perspectiva, no ganancias frente a la toma de decisiones basadas en el promedio.

El mismo ejercicio se ha realizado con 72 folds, tratando de identificar si con una mayor segmentación de los datos (que comprendieran entrenamientos y pruebas para cada hora durante tres días) se encontrarán algunos elementos distintos frente al análisis de 12 folds.



Los resultados no fueron mejores, el MAE fue de USD \$0,94 y el número de aciertos subió escasamente al 52.63%, habiéndose realizado un procesamiento seis veces más dispendioso que el anterior. Desde otra perspectiva, con el modelo se logra una mejora del 2,6% con respecto a la incertidumbre total, que probabilidad binomial, corresponde al valor esperado de un 50% de aciertos.

## CONCLUSIONES

- Los modelos basados en árboles de decisión dan mejores resultados. Esto sugiere que para este tipo de eventos, los modelos no paramétricos pueden tener mejores resultados que los paramétricos.
- El modelo de predicción identificado presenta niveles bajos de error, si se verifica el porcentaje de éste con respecto al valor de Solana en dólares.
- Los valores atípicos corresponden a una baja proporción del total de estimaciones realizadas.
- El uso del modelo para la toma de decisiones de inversión genera incertidumbre, dado que se obtuvo una predicción acertada del tipo de variación en el 52% de las ocasiones. Esto implica que en estrategias de inversión de corto plazo como el trading y el scalping se puedan presentar situaciones en donde se esperaba un sentido de la variación distinto al que realmente ocurrió casi en la mitad de los casos.
- No se observa overfitting, dado que el cambio en las métricas de error cuando se evalúa con 12 y 72 folds es muy similar.
- Es necesario explorar el aporte de variables exógenas como los medidores de sentimiento y expectativas frente al comportamiento de Solana, las operaciones programadas para determinado nivel de precio, el índice de dominancia del Bitcoin y comportamiento de precios de otras monedas tanto de su ecosistema como del resto del mercado.



## REFERENCIAS BIBLIOGRÁFICAS

Santa Cruz, R. and Corrêa, C., Previsión de demanda intermitente con métodos de series de tiempo y redes neuronales artificiales: Estudio de caso DYNA, 84(203), pp. 9-16, December, 2017.

Löning, Markus (2021). Machine Learning with Time Series, University College London (UCL).

Guida, T. (2019). Big data and machine learning in quantitative investment. 1st ed. Gran Bretaña: John Wiley & Sons, Ltd.

Andres, S. (2021). *Solana en Español: La guía definitiva para introducirte al mundo de las finanzas descentralizadas, Lending, Yield Farming, Dapps y dominarlo por completo*. Wonderbok Publishing.

Kumar, N., Gayathri, N., Rahman, A., & Balamurugan, B. (Eds.). (2022). *Blockchain, big data and machine learning: Trends and applications*. CRC Press.

Escobar Ortiz, J.(2022). Predicción del precio de Bitcoin con Python, cuando el pasado no se repite, disponible bajo una Reconocimiento 4.0 Internacional (CC BY 4.0) en <https://www.cienciadedatos.net/documentos/py41-forecasting-criptomoneda-bitcoin-machine-learning-python.html>

Bortoli, F., Costantino, M., Marzari, G., Negritto, R. G. (2021). Criptomonedas. Riesgos, rendimientos, valor y perspectivas a futuro. Universidad Nacional de Córdoba, Argentina.

Chen, Le (2018), Evolución del precio del Bitcoin, Análisis estadístico de series temporales. Universidad Complutense de Madrid.

Kulal Abhinandan (2021), Followness of Altcoins in the Dominance of Bitcoin-A Phase Analysis, Macro Management & Public Policies, Vol 3, Issue 3.

Pérez-Planells, Ll., Delegido, J., Rivera-Caicedo, J.P., Verrelst, J.(2015), Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos, revista de teledetección, Valencia, España.

Roche, A. (2009), Árboles de decisión y Series de tiempo, Universidad de la República, Uruguay.

Li, Yi. (Enero 25, 2022). Predicting Stock Prices Using Machine Learning. Neptuneblog. <https://neptune.ai/blog/predicting-stock-prices-using-machine-learning>.

Santana, J. C. (2006), Predicción de series temporales con redes neuronales una aplicación a la inflación colombiana, Revista Colombiana de Estadística, Vol 29, No. 1.



**LOS LIBERTADORES**  
FUNDACIÓN UNIVERSITARIA

Sanchez, L. (2020), Prediciendo el precio del Bitcoin, y más, Revista Suma de Negocios, Komrad Lorenz Editores.

Yakovenko, A. (2013), Solana- A new architecture for a high performance blockchain.