ANÁLISIS Y PREDICCIÓN DE PRECIOS DE ETHEREUM Y SU CORRELACIÓN CON OTROS ACTIVOS



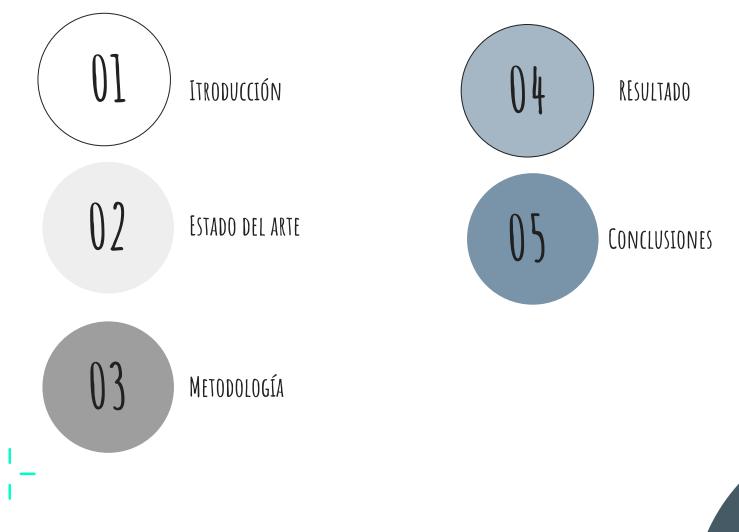
PIERO GIORGIO IT ACADEMY OCTUBRE 2023

ABSTRACT

- | -

Este proyecto se centra en el análisis de la criptomoneda Ethereum, estudiando la evolución histórica de sus precios de cierre y su correlación con otros activos financieros.

Se utilizan dos modelos de predicción, Prophet y ARIMA, en el intento de verificar si sería posible predecir movimientos en los precios futuros de la criptomoneda.





1. INTRODUCCIÓN

Es innegable que las criptomonedas están experimentando un auge espectacular y está adquiriendo una relevancia sin precedentes. Este auge es el resultado de una combinación de factores, incluyendo la tecnología subyacente, la descentralización, la seguridad, la accesibilidad y la creciente confianza de los usuarios en esta forma de dinero digital. Las criptomonedas no solo están siendo utilizadas para transacciones comerciales, sino que están dando lugar a nuevas oportunidades de inversión, innovación en tecnología financiera y cambios en la forma en que entendemos y utilizamos el dinero. Este fenómeno está transformando el panorama económico y financiero en todo el mundo.

En este contexto, Ethereum es la segunda criptomonedas más influyentes y una plataforma líder en tecnología blockchain. Lanzada en 2015, Ethereum no solo funciona como una moneda digital (Ether, o ETH) sino que también permite la ejecución de contratos inteligentes y la creación de aplicaciones descentralizadas (DApps). Esto ha hecho que Ethereum sea un punto focal en la revolución de la tecnología blockchain y la descentralización.

En este proyecto:

- Se analizarán datos históricos de Ethereum a partir de 2017 hasta el dia de hoy.
- Se comparan estos datos con los de otros activos como Bitcoin, Gold, Crude Oil, Vanguard Information Technology, Vanguard Energy ETF, CBOE Volatility Index en el intento de descubrir correlaciones entre los mismos.
- Se entrenar dos modelos muy conocidos para series temporales, Facebook Prophet y ARIMA (Autoregressive Integrated Moving Average), en el intento de verificar si es posible predecir de forma satisfactoria el precio de Ethereum.
- Finalmente, se expondrán los resultados y sus conclusiones.



2. ESTADO DEL ARTE

La predicción de los precios de Ethereum es un desafío significativo debido a su marcada volatilidad. Esta criptomoneda está sujeta a fluctuaciones extremas que pueden ser difíciles de anticipar con precisión. Esta volatilidad es el resultado de una variedad de factores, incluyendo cambios en la demanda del mercado, noticias económicas y eventos relacionados con la tecnología blockchain. A pesar de estos desafíos, existe un creciente interés en desarrollar modelos de pronóstico que puedan ayudar a los inversores y analistas a tomar decisiones informadas. Es importante señalar que en el campo de la inteligencia artificial y las criptomonedas, la información se vuelve obsoleta rápidamente debido a las actualizacion y adaptación de las estrategias y modelos utilizados para predecir los precios de Ethereum. En la red, se pueden encontrar numerosos proyectos, estudios e investigaciones que han abordado el desafío de la predicción de precios en el contexto de las criptomonedas. Estos proyectos anteriores proporcionan información valiosa sobre enfoques exitosos, tendencias en el mercado y desafíos comunes que enfrentan los analistas. Además, estos proyectos han explorado una variedad de modelos de series temporales y técnicas de análisis de datos.

Además de los modelos Prophet y ARIMA, que se explicarán en detalle en nuestra análisis, existen otros modelos que han demostrado ser relevantes para la predicción de series temporales. Entre estos modelos se incluyen:

 LSTM (Long Short-Term Memory): Una variante de las redes neuronales recurrentes (RNN), las LSTM son conocidas por su capacidad para capturar patrones temporales en datos secuenciales. La predicción se obtiene suministrando los datos de entrada a una red neuronal profunda (las N últimas lecturas de características de la serie cronológica).

Random Forests: Un enfoque basado en árboles de decisión que puede ser eficaz pará la predicción de series temporales. Un modelo Random Forest está formado por múltiples árboles de decisión individuales. Cada uno de estos árboles es entrenado con una muestra ligeramente diferente de los datos de entrenamiento, generada mediante una técnica conocida como bootstrapping.

En resumen, este proyecto busca contribuir a este campo en constante cambio, considerando múltiples enfoques y modelos para mejorar la precisión de las predicciones y abordar los desafíos inherentes a la volatilidad del mercado de las criptomonedas.

3. METODOLOGÍA

3.1 CARGA DE DATOS HISTÓRICOS Y ANÁLISIS DE DATOS

Los datos históricos se obtienen usando "yfinance", que es una biblioteca de Python que permite acceder a datos financieros de Yahoo Finance. El histórico de datos abarca desde el día 2017-11-09 hasta el día 2023-10-09. Si bien se observan cambios significativos en los datos a partir de 2019, es importante señalar que se ha decidido no excluir ningún dato para este proyecto, a pesar de la posibilidad de obtener resultados predictivos menos precisos.

```
# Using the today function data will be updated everytime the function is runned end_date = datetime.today().strftime('%Y-%m-%d')

et_price_df = yf.download('ETH-USD', start='2016-01-01', end=end_date)
```

	Open	High	Low	Close	Adj Close	Volume
Date						
2017-11-09	308.644989	329.451996	307.056000	320.884003	320.884003	893249984
2017-11-10	320.670990	324.717987	294.541992	299.252991	299.252991	885985984
2017-11-11	298.585999	319.453003	298.191986	314.681000	314.681000	842300992
2017-11-12	314.690002	319.153015	298.513000	307.907990	307.907990	1613479936
2017-11-13	307.024994	328.415009	307.024994	316.716003	316.716003	1041889984

As we can see, we have the following columns:

- · Date: The date when the trading data was recorded.
- . Open: The price at which the stock started trading when the market opened on that day.
- . High: The highest price at which the stock traded during the day.
- Low: The lowest price at which the stock traded during the day.
- . Close*: The price at which the stock was trading when the market closed on that day.
- Adj Close (Adjusted Close): This is the closing price adjusted for factors such as dividends, stock splits, and new stock offerings. It provides a more
 accurate reflection of the stock's value.
- · Volume: The number of shares of the stock that were traded during the day.

Tras realizar una EDA, se visualiza el conjunto de datos a través de un gráfico interactivo de "Plotly" para tener una idea inicial de cuál haya podido ser la evolución de Ethereum, junto a una "seasonal_decompose" que se utiliza en el análisis de series temporales para descomponer una serie temporal misma en sus componentes principales, que suelen incluir:

- Tendencia (Trend): La tendencia muestra la dirección general en la que se mueven los datos a lo largo del tiempo. Puede ser ascendente, descendente o fluctuar en una dirección específica.
- Estacionalidad (Seasonality): La estacionalidad representa patrones que se repiten a lo largo del tiempo debido a factores estacionales, como las estaciones del año, meses, semanas o días de la semana. Estos patrones se repiten de manera predecible.
- Componente Residual (Residual Component): El componente residual representa la variación aleatoria o no sistemática que no se puede atribuir a la tendencia ni a la estacionalidad. Incluye ruido y otros factores impredecibles en los datos.





3.2 ANÁLISIS DE CORRELACIÓN

En esta parte, se busca encontrar algún tipo de correlación entre Ethereum y otros assets:

- Petróleo Crudo (CL=F): Este recurso esencial, conocido por sus sustanciales fluctuaciones de precios, es el más grande y notablemente sensible a factores geopolíticos, sociales y económicos, tanto en la demanda como en la oferta. A menudo reacciona significativamente a los cambios en eventos globales.
- Oro (GC=F): El oro se considera a nivel mundial como un activo seguro. Sirve como reserva de valor y suele ser una de las fuentes de inversión preferida durante épocas de turbulencia económica y pérdida de valor de una moneda.
- Vanguard Information Technology ETF (VGT): Este indice incluye acciones de múltiple capitalización en el sector de tecnologías de la información. Vale la pena mencionar que muchas criptomonedas se basan en la tecnología blockchain, que depende en gran medida de tecnología informática avanzada. En consecuencia, los desarrollos en el sector tecnológico pueden ejercer influencia en el mercado de criptomonedas.
- Vanguard Energy Index Fund (VDE): Este índice incluye acciones de múltiple capitalización en el sector de energía, incluyendo empresas en los sectores del petróleo, gas y carbón. Los cambios en los precios de la energía y el desempeño de las empresas energéticas pueden tener implicaciones significativas tanto para la economía en general como para las carteras de inversión.
- VIX (Índice de Volatilidad): El VIX, a menudo llamado el "Índice del Miedo", mide la volatilidad del mercado y refleja el sentimiento de los inversores. No está directamente relacionado con los activos mencionados anteriormente, pero desempeña un papel crucial en la evaluación de la estabilidad general del mercado. Niveles elevados de VIX pueden indicar una mayor incertidumbre en el mercado y aversión al riesgo entre los inversores, lo que potencialmente afecta las decisiones de inversión en una amplia gama de activos. Comprender el VIX es esencial para evaluar el sentimiento del mercado y tomar decisiones de inversión informadas.

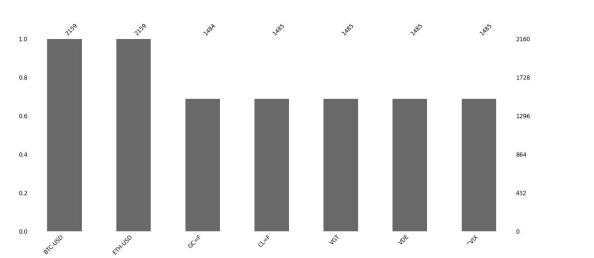
En el análisis se incluye también Bitcoin.



Se procede a crear un nuevo df con los precios de cierre de todos los activos.

Debido a que los mercados de algunos activos solo trabajan en los días entre semana, se aplica al dataset la técnica del "forward-filling", que se utilizada en el manejo de datos faltantes o valores nulos en series temporales y consiste en reemplazar los valores faltantes con el valor más reciente conocido antes del punto en el tiempo en el que se encuentra el valor faltante.

En otras palabras, se copia el último valor válido y se utiliza para llenar los espacios en blanco hacia adelante en la secuencia y finalmente se comparan los datos y se busca correlación entre los mismos.



```
1 prices filled = prices.fillna(method='ffill')
 3 # there are no previous trading days before '2017-11
 4 prices_filled = prices_filled.dropna()
 6 prices filled.info()
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 2159 entries, 2017-11-10 to 2023-10-08
Data columns (total 7 columns):
     Column
             Non-Null Count Dtype
     BTC-USD 2159 non-null float64
    ETH-USD 2159 non-null
                            float64
     GC=F
             2159 non-null
                           float64
    CL=F
             2159 non-null
                           float64
    VGT
             2159 non-null
                           float64
    VDE
             2159 non-null
                            float64
     ^VIX
             2159 non-null float64
dtypes: float64(7)
memory usage: 134.9 KB
```

Entre sus ventajas, se encuentran tanto la posibilidad de crear pronósticos interactivos sin necesidad de instalar bibliotecas adicionales ya que lo necesario se incorpora por defecto con Prophet, como su propia cross-validation.

Tal y como se ha comentado anteriormente, a diferencia de otros modelos, Prophet es especialmente user friendly.

Æl únicos requerimientos es que las columnas de input tengan los nombres "ds" para la columna temporal y "y" para los valores numéricos:

```
prophet_df = et_price_df[["Date", "Close"]]
prophet_df.rename(columns={"Date": "ds", "Close": "y"}, inplace=True)
```

1 prophet_df.head()

us	,
2017-11-09	320.884003
2017-11-10	299.252991
2017-11-11	314.681000
2017-11-12	307.907990
2017-11-13	316.716003
	2017-11-10 2017-11-11 2017-11-12

Sucesivamente, los pasos que se han seguido han sido:

- Split en train y test (puesto que el goal es el de generar una predicción a 30 días, se fija esta cantidad de días como tamaño del test set.
- Creación de un primer modelo básico y visualización de los resultados;
- Hyperparameter tuning usando la misma cross-validation de Prophet y creación de un nuevo modelo.
 - Validación de los resultados con los datos del set test.





3.3 MODELO PROPHET

Facebook Prophet es una biblioteca de código abierto desarrollada por Facebook para el análisis y pronóstico de series temporales.

Sus caracteristicas son:

- Modelo aditivo: Prophet asume que las series temporales tienen una estructura aditiva, lo que significa que los efectos de tendencia, estacionalidad y días festivos se pueden sumar para formar la predicción.
- Manejo de valores atípicos: La biblioteca tiene la capacidad de manejar valores atípicos (outliers) en los datos, lo que la hace robusta ante datos anómalos.
- Manejo de vacaciones: Prophet permite especificar días festivos y eventos que pueden afectar las series temporales, lo que es importante en aplicaciones como las ventas minoristas, donde los patrones de compra pueden variar durante las festividades.
- Componentes de tendencia y estacionalidad: La biblioteca descompone la serie temporal en sus componentes de tendencia y estacionalidad, lo que facilita la interpretación de los resultados.
- Disponible para Python y R.

Facebook Prophet se ha utilizado ampliamente en la industria y en la comunidad de ciencia de datos para realizar pronósticos precisos en una variedad de aplicaciones, desde la previsión de ventas hasta el análisis de datos de tráfico web. Es una herramienta valiosa para aquellos que necesitan realizar pronósticos basados en series temporales sin tener que profundizar en técnicas de modelado complejas.



3.4 MODELO ARIMA

ARIMA (AutoRegressive Integrated Moving Average) es un modelo estadístico utilizado para el análisis y la predicción de series temporales. Este modelo se compone de tres componentes clave:

- AutoRegressive (AR): Esta parte del modelo se refiere a los "lags" de la serie de tiempo diferenciada. Los lags son valores previos de la serie temporal que se utilizan para predecir el valor actual. La parte AR toma en cuenta la relación lineal entre el valor actual y los valores pasados. Se denomina "p".
- Integrated (I): La parte "Integrated" se refiere al número de diferenciaciones necesarias para hacer que la serie temporal sea estacionaria. La estacionariedad implica que la serie no tiene una tendencia evidente y su media y varianza son aproximadamente constantes a lo largo del tiempo. La diferenciación se utiliza para lograr esta estacionariedad. Se denomina "d".
- Moving Average (MA): Esta parte se refiere a los "lags" de los errores. Los errores son las diferencias entre los valores observados y los valores predichos por el modelo. La parte MA toma en cuenta la relación lineal entre los errores pasados y el valor actual. Se denomina "q".

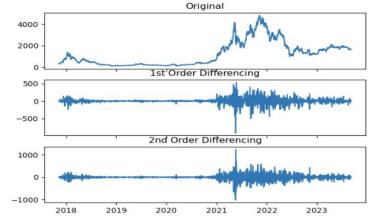
El modelo ARIMA combina estas tres componentes para modelar y predecir series temporales. Puede ajustarse a los datos históricos y utilizarse para hacer predicciones futuras en función de patrones y relaciones identificadas en los datos pasados.

Cabe destacar que los modelos ARIMA son más complejos y más sensibles que los modelos de Prophet.

Stationarity for the Original Series
ADF Statistic: -1.4950449683250049, p-value: 0.5359610826288749
KPSS Statistic: 3.8353241138629803, p-value: 0.01
Stationarity for the First Order Differenced Series

KPSS Statistic: 0.07979847923336325, p-value: 0.1
Stationarity for the Second Order Differenced Series
ADF Statistic: -17.2616851258637375, p-value: 5.935873996598548e-30
KPSS Statistic: 0.01805415873826867, p-value: 0.1

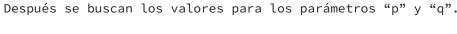
ADF Statistic: -11.594362293330766, p-value: 2.7431689455246097e-21

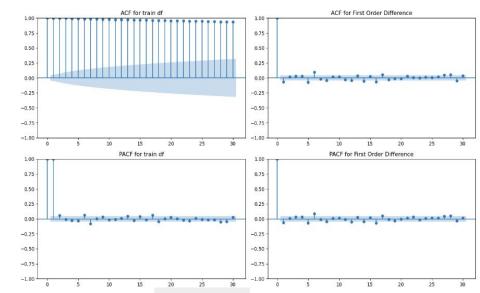


Sucesivamente:

- Se crea un primer modelo y se visualizan los resultados.
- Se usa auto_arima, que es una función proporcionada por la biblioteca pmdarima, una biblioteca que se utiliza para la implementación de modelos ARIMA.
- Se vuelven a evaluar los resultados.

Tras dividir nuestros datos en train y test, el siguiente paso en el proceso de identificación de los órdenes AR, I y MA (p,d,q) adecuados se realiza a través del análisis de la serie temporal y la evaluación de la estacionariedad, es decir que la media, la varianza y la estructura de correlación entre los datos se mantienen constantes a lo largo del tiempo. Esto llevará a encontrar el valor para el término "d".



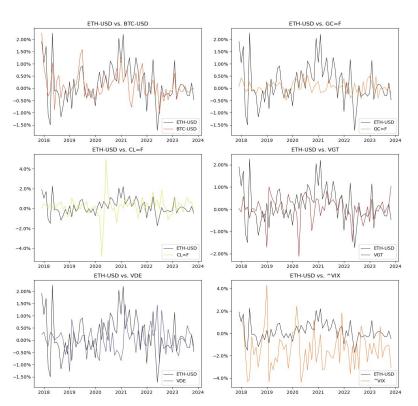


Tras verificar que, incluso buscando los mejores hiperparámetros, el modelo ARIMA no consigue adaptarse a los datos de forma satisfactoria, finalmente se aplica una transformación logarítmica a los mismos en un intento de mejorar la capacidad del modelo para capturar patrones y estructuras subyacentes en los datos, estabilizando la varianza, linealizando las tendencias y reduciendo los efectos extremos de los mismos en el modelo.

Finalmente, se vuelven a evaluar los resultados con el test set.

4. RESULTADOS

4.1 RESULTADO ANÁLISIS CORRELACIÓN



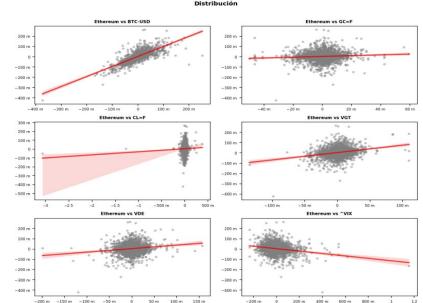


1 prices_filled.corr()['ETH-USD'].sort_values(ascending=False) ETH-USD 1.000000

BTC-USD 0.775735 VGT 0.249960 VDE 0.151326 GC=F 0.074253 0.056088 CL=F ^VIX -0.209582 Name: ETH-USD, dtype: float64

Resultado correlación entre Ethereum y los otros Assets.

Distribución



Resultados de Rendimiento Promedio Anual, Desviación Estándar y Sharpe ratio(métrica utilizada en finanzas para determinar si una inversión está generando rendimientos suficientes en comparación con el riesgo asumido.)

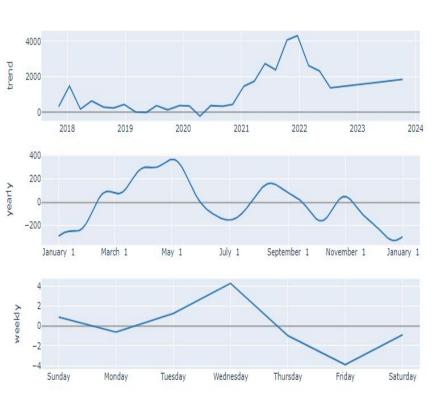
```
1 # Calculate average annualized returns
    average returns = prices filled.mean() * 365 * 100
    print('Average annualized returns (arithmetic mean) are:\n', average returns)
    # Calculate standard deviation
    standard deviation = prices filled.std() * np.sqrt(365)
    print('Standard Deviation is:\n', standard deviation)
    # Calculate Sharpe ratio
10 sharpe_ratio = average_returns / standard_deviation
11 print('Sharpe ratio:\n', sharpe ratio)
Average annualized returns (arithmetic mean) are:
BTC-USD
           49.822025
ETH-USD
          70.235111
```

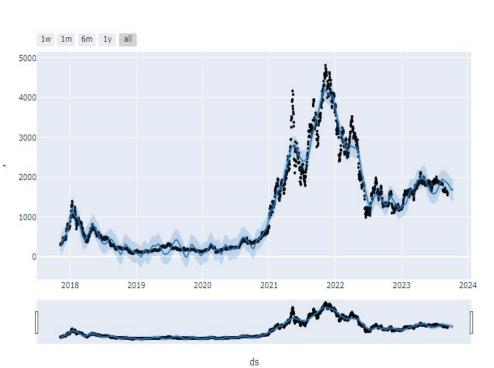
```
GC=F
           12.372950
          -24.242869
CL=F
           26.161098
VGT
           21.595637
VDF
^VIX
          -75.021184
dtype: float64
Standard Deviation is:
BTC-USD
            0.727960
ETH-USD
           0.916942
GC=F
           0.180882
CL=F
           1.481510
           0.319174
VGT
           0.408662
VDE
^VTX
           1.630761
dtype: float64
Sharpe ratio:
BTC-USD
            68,440617
ETH-USD
           76.597147
GC=F
           68.403425
CL=F
          -16.363617
VGT
           81.964942
VDE
           52.844687
^VIX
          -46.003800
dtype: float64
```

4.2 RESULTADO BEST MODEL PROPHET

RMSE: 97

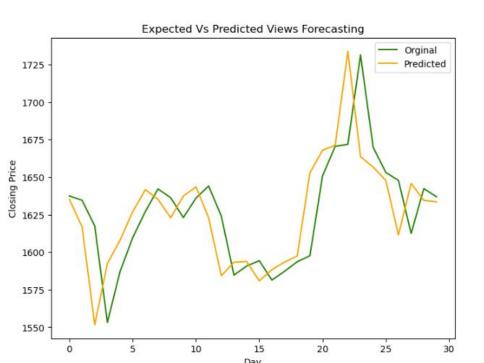
MAPE: 4.6862880443562425





4.3 RESULTADO BEST MODEL ARIMA

MAPE: 1.2790713434616692 MAE: 20.835765843413657 MSE: 2207.1565612668714 RMSE: 28.5883036910661



predicted = 1637.458177, expected = 1635.162354, error = 0.140403 % predicted = 1634.557627. expected = 1616.828857, error = 1.096515 % predicted = 1617.502738, expected = 1551.637695, error = 4.244873 % predicted = 1553,195948. expected = 1592.429443, error = 2.463751 % predicted = 1587.089261, expected = 1607.988525, error = 1.299715predicted = 1609.711859, expected = 1626.974365, error = 1.061019 % predicted = 1627.101991. expected = 1641.640381. error = 0.885601 % predicted = 1642.137523, expected = 1635.221680, error = 0.422930 % predicted = 1636.304529, expected = 1622.896729, error = 0.826165 % predicted = 1623,029226. expected = 1637,347046. error = 0.874452 % predicted = 1636.070036. expected = 1643.544678. error = 0.454788 % predicted = 1644.116903, expected = 1622.890625, error = 1.307930 % predicted = 1624.056318, expected = 1584.307007, error = 2.508940 % predicted = 1584,701085. expected = 1593,268311. error = 0.537714 % predicted = 1590.755225. expected = 1593.857788. error = 0.194657 % expected = 1580.853394, predicted = 1594.339869, error = 0.853114 % predicted = 1581.403509, expected = 1588.322876, error = 0.435640predicted = 1587.292568, expected = 1593.417236, error = 0.384373 % predicted = 1593.635761, expected = 1597.491211, error = 0.241344 % expected = 1652.882690, predicted = 1597.615933. error = 3.343659 % expected = 1667.943848, predicted = 1650.872504, error = 1.023496predicted = 1670.541204, expected = 1671.161865, error = 0.037139predicted = 1671.886430, expected = 1733.810425, error = 3,571555 % predicted = 1731.458890. expected = 1663.627563, error = 4.077314 % predicted = 1669.843431, expected = 1656.685669, error = 0.794222 % predicted = 1653.117159, expected = 1647.838135, error = 0.320361predicted = 1647.803862. expected = 1611.476440. error = 2.254294 % predicted = 1612.430328. expected = 1645.831543. error = 2.029443 % predicted = 1642.376304, expected = 1634.511230. error = 0.481188 % predicted = 1636.906247, expected = 1633.548584, error = 0.205544 %

5. CONCLUSIONES

5.1 CONCLUSIONES ANÁLISIS CORRELACIÓN

- | -

Bitcoin y Ethereum muestran una correlación positiva notablemente fuerte, lo cual concuerda con las expectativas. Al parecer en el mundo de las criptos, las fluctuaciones de Bitcoin influyen fuertemente en las otras monedas. Se puede confirmar por los datos obtenidos que estas dos principales criptomonedas comparten una relación sólida y a menudo lineal, con sus precios moviéndose frecuentemente de manera conjunta. Esta correlación se puede atribuir a las características compartidas y las tecnologías que sustentan el mercado más amplio de las criptomonedas.

Sin embargo, cuando examinamos la correlación de Ethereum con otros activos tradicionales, como el Petróleo, el Oro y los otros indicadores incluyendo el VIX, observamos que Ethereum parece operar de manera independiente en cierta medida. La ausencia de una correlación significativa parece sugerir que la dinámica de precios de Ethereum no está directamente influenciada por los factores que afectan a otros activos tradicionales ni por las fluctuaciones de los activos mismos.

5.2 CONCLUSIONES MODEL PROPHET

Además de ser un modelo relativamente fácil de usar, el resultado demuestra que el modelo Prophet es muy sólido para predecir los precios de las criptomonedas.

Con un RMSE (Error Cuadrático Medio) de 97 y un MAPE (Error Porcentual Absoluto Medio) de aproximadamente 4.69%, Prophet demuestra un desempeño razonablemente bueno y una óptima capacidad en capturar los patrones subyacentes y las tendencias en los datos de precios de criptomonedas. Ademas, el modelo nos enseña que podrían existir algunos patrones que no son evidentes a simple vista, como cierta tendencia a que los precios suban en mayo y que los mismos aumenten los miércoles y tengan una tendencia a bajar los viernes.

5.3 CONCLUSIONES MODEL ARIMA

Después de obtener resultados muy pobres, gracias a la transformación logarítmica de los datos ARIMA alcanza una performance sorprendentemente buena y superior a la performance del modelo Prophet:

- Un MAPE de aproximadamente 1.28% es bastante bajo y sugiere que las predicciones del modelo ARIMA tienen una baja variabilidad en comparación con los valores reales. Esto indica que el modelo ARIMA es capaz de hacer predicciones muy cercanas a los valores reales, con un error porcentual promedio relativamente bajo del 1.28%.
- Con un MAE de 20.84, se observa que el modelo ARIMA tiende a cometer errores absolutos promedio de aproximadamente 20.84 unidades en la escala original. Esto significa que, en promedio, las predicciones se desvían de los valores reales por alrededor de 20.84 unidades.
- Un MSE de 2207.16 indica que los errores cuadrados promedio de las predicciones del modelo son relativamente bajos, lo que sugiere que el modelo es capaz de mantener errores pequeños en la mayoría de las predicciones.
- El RMSE de 28.59 es la raíz cuadrada del MSE y proporciona una medida de la precisión en la misma escala que los datos originales. Este valor indica que, en promedio, las predicciones del modelo ARIMA se desvían de los valores reales en aproximadamente 28.59 unidades en la misma escala que los datos originales.

En resumen, los resultados del modelo ARIMA, con datos transformados logarítmicamente, muestran una buena precisión en la predicción de los precios de criptomonedas, con errores relativamente bajos. Sin embargo, es importante tener en cuenta que ARIMA es mucho más complejo de configurar que Prophet, más maquinoso y puede ser menos accesible para quienes no están familiarizados con el análisis de series temporales.



