



**DATA SCIENCE  
MACHINE LEARNING**

**DOCUMENTO DE TRABAJO  
-ENTREGA FINAL-  
BITCOIN**



**CODERHOUSE**



**COMISION 31490**



**LUCIANO BIZIN**



**21 OCTUBRE 2022**

---





# ÍNDICE

1

## **Presentación de la investigación**

Introducción al estudio del BTC

1-2

## **Problema de investigación**

Análisis fundamental y técnico

2

## **Preguntas y objetivos de investigación**

Preguntas de investigación y objetivos

3-4

## **Estructura de desglose de trabajo**

Etapas de la investigación

5-8

## **Data acquisition y data wrangling**

Etapas de la investigación

8-18

## **Análisis univariado, bivariado y multivariado de los diversos data sets**

Análisis de variables y discriminación

18-23

## **Data sets para modelar con ML**

Data set finales para modelar

23-42

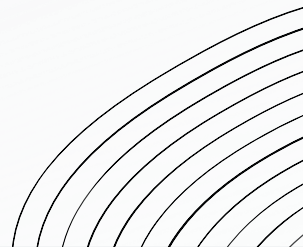
## **Modelado de ML**

Data set finales para modelar

42

## **Metadata**

[Link a metadata](#)



## Documento de trabajo

### *Aplicando modelos de ML a BTC*

#### 1. Presentación de la investigación

En este estudio se aplican diversos modelos de machine learning al precio intra-diario de cierre de Bitcoin (+1 día) en el período 2010-2022, a fin de determinar cuál es el mejor algoritmo supervisado capaz de predecir con mejor previsibilidad la fluctuación del valor de este cripto-activo. El análisis es cuestión se diseñó como un problema de clasificación (no de regresión) y según capacidades de procesamiento de un ordenador portátil de uso personal (i.e. 8 GB RAM, Intel Core i5).

#### 2. Problema de investigación

El precio de todo activo financiero es un fenómeno de tipo multicausal que escapa a la mera concepción de la economía racionalista (mayor beneficio al menor costo sobre el fundamento de la eficiencia como principio regulador del mercado). Esto significa que cualquier intento de aproximación hermenéutica explicativa al dinamismo de un bien económico (cualquiera sea su naturaleza) conlleva la necesidad de recurrir a la ejecución de diversos tipos de aproximaciones:

- *Análisis fundamental*

Este es un tipo de análisis introducido en el mundo económico por Benjamin Graham y David Dodd en 1934, con el objetivo de poder estimar cuál es el valor real de un bien, título o acción mediante la comparación de balances y del valor real del mercado del producto, bien, compañía o servicio. En este caso particular, no se llevará a cabo este tipo de investigación dado que escapa al marco general de este proyecto. Todo aquel interesado en el Bitcoin podría remitirse a su [“white paper”](#) (documento explicativo).

- *Análisis técnico*

Este tipo de análisis complementa y se ve complementado por el análisis fundamental. En síntesis, el análisis técnico, introducido por Charles Henry Dow a finales del siglo XIX, busca detectar la tendencia del precio de un activo según las reglas del mercado (el análisis fundamental se centraba en el estudio de la entidad emisora o portadora del activo). Para lograr un buen conocimiento del dinamismo del precio del bien, servicio o producto en cuestión se suele recurrir a la identificación-análisis de tendencias de precio, indicadores de momentum, volumen, patrones de velas japonesas, volatilidad, entre otros (e.g. ciclicidad, estacionariedad, residualidad, auto-correlación, etc.).

Entendiendo el nivel introductorio de esta primera aproximación al análisis del dinamismo de bajas y subas de precio del BTC, solo se llevará adelante el análisis técnico del mismo. Se entiende, además, que este problema de estudio posee a su vez otras dos complejidades que deben ser explicitadas en este documento.

La primera radica en que la investigación de todo activo financiero implica el abordaje de series temporales y la identificación de sus principales elementos. La segunda problemática se relaciona con la multicausalidad del movimiento del precio del BTC (de por sí muy volátil) y la necesidad de una alta capacidad de procesamiento computacional. Por eso, en consideración de ambas cuestiones, se diseñó un proyecto de clasificación en el que se analizan diversas bases de datos por separado, y finalmente, una base de datos con las variables con mayor correlación con el precio de BTC, la variación de su precio y la variable target.

### 3. Preguntas y objetivos de investigación

Dada la multi-dimensionalidad del problema de investigación que se aborda se proponen las siguientes preguntas orientadoras del proyecto.

- *Preguntas de investigación*

- I. ¿Cuáles son las diversas variables que influyen en el dinamismo del precio de BTC?
- II. ¿Son capaces los modelos supervisados de clasificación de machine learning que se pueden ejecutar en computadoras doméstica de predecir con cierto grado suficiente de previsibilidad (>50%) el movimiento del precio de este criptoactivo?
- III. En caso de que no se puedan correr en ordenadores domésticos, ¿es posible reducir la cantidad de variables y registros procesados mediante algún modelo de machine learning no supervisado que pueda ser ejecutado en computadoras domésticas, y a su vez resulte en predicciones de subidas-bajadas de precio con una previsibilidad superior al 50/55% de verdaderos positivos?

- *Objetivos general*

- I. Diseñar diversos modelos de machine learning capaces de predecir con cierto grado de previsibilidad positiva (>50/55%) el dinamismo del precio de Bitcoin en el corto plazo, es decir, si el precio de este criptoactivo subirá o bajará en la temporalidad estudiada (+1 día).

- *Objetivos específicos*

- I. Detectar diversos factores (multi-dimensionales) que afectan el precio de BTC.
- II. Analizar las variables de los factores detectados a fin conocer el modo mediante el cual influyen la variación del precio de este criptoactivo.
- III. Diseñar diversos modelos de machine learning capaces de predecir con cierto grado de previsibilidad positiva (>50/55%) a corto plazo (+1 día), el movimiento alcista o bajista del precio de BTC.

### 4. Estructura de Desglose de Trabajo

En esta sección se presenta el Esquema de Desglose de Trabajo (EDT) de esta investigación. Según los estándares internacionales del project management (i.e. PMBOK 6) cada etapa de una EDT debe dividirse en unidades organizacionales menores codificadas bajo el formato ID:

i) Cuentas control (1.1., 1.2. y 1.3, etc...); ii) Paquetes de trabajo (1.1.1., 1.1.2., 1.1.3., 1.2.1., 1.2.2., 1.2.3., etc,...), y iii) Tareas de trabajo concretas (1.1.1.1., 1.1.1.2., 1.1.1.3., etc...).

En este caso concreto solo se desarrollarán las dos primeras unidades organizacionales dado el carácter macro de esta sección. Cada una de las unidades que se describirán debajo han sido proyectadas según los siguientes elementos/criterios: a) duración de las unidades de trabajo; b) interrelación entre las tareas diseñadas; c) metodología de análisis y ejecución de cada unidad y sub-unidad; e d) hitos tentativos de informes y/o entrega de productos.

### *Etapas de la EDT*

Etapas 1. Investigación preliminar

Etapas 2. Data acquisition

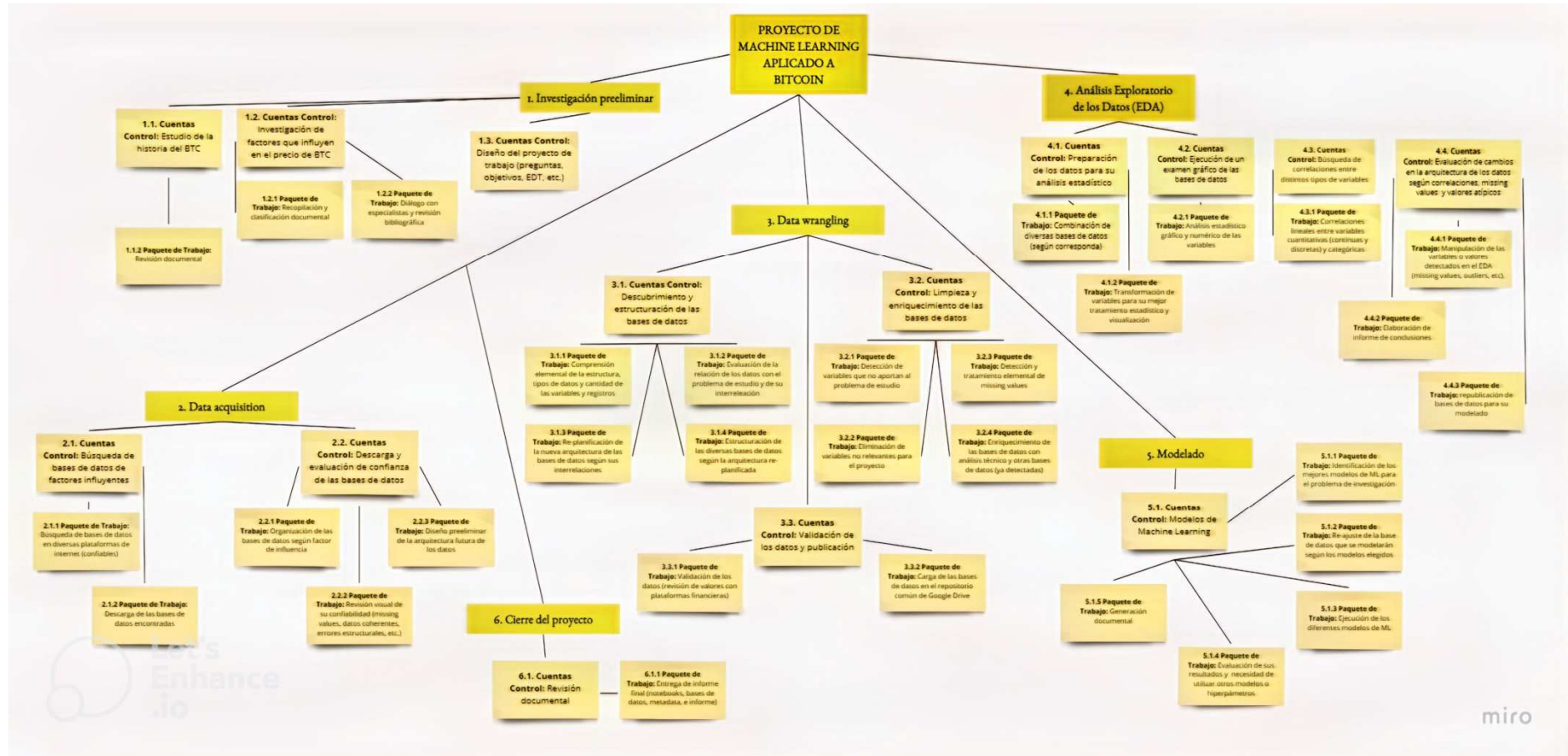
Etapas 3. Data wrangling

Etapas 4. Análisis Exploratorio de los Datos

Etapas 5. Modelos de Machine Learning

Etapas 6. Cierre del proyecto

## Estructura de Desglose de Trabajo (Gráfico)



Link original (permiso de visualización): [https://miro.com/app/board/uXjVPeYRDtw=?share\\_link\\_id=33358300391](https://miro.com/app/board/uXjVPeYRDtw=?share_link_id=33358300391)

La imagen original ha sido procesada para mejorar su calidad por: <https://letsenhance.io/boost>



## 5. Data acquisition y data wrangling

El proyecto cuenta con numerosos data set relacionados con diversos factores económicos que influyen en el dinamismo del precio de BTC. Si bien el detalle específico de cada base de datos se encuentra explicitado en el documento titulado “Metadata.docx” (en el que se explicita la fecha de creación, tipo de datos, diccionario, ubicación del archivo, notebooks, ubicación del notebook, bases de datos sin tratamiento, y base de datos finales analizadas, etc), a continuación se ofrece una descripción general de las mismas.

- **Base de datos: BTC\_final**

*Fuente de los datos en bruto:* <https://www.investing.com/crypto/bitcoin/historical-data>

*Descripción del data set:* se extrajeron los datos más antiguos que la página web permite descargar (julio de 2010) con fecha máxima, día en que se descendieron los datos (20 de julio de 2022). En el notebook respectivo (BTC\_final.ipynb) se pueden observar los comentarios de las variables originales y como estas fueron tratadas oportunamente. La temporalidad del data set es de un +1 día y contaba originalmente con las siguientes variables: Date, Price, High, Low, Open, Vol., y Change %.

*Criterio de selección:* es la base principal que se concatena con el resto de los data sets.

*Tratamiento de la base de datos:*

- I. Se realizó un análisis estructural básico: tipos de datos, missing values, etc.
- II. Se modificó el nombre de la columna “Change %” por “Percentage\_diff”.
- III. Se re-diseñaron las columnas “Vol.”, y “Percentage\_diff”, por nomenclatura inapropiada, para su posterior tratamiento (signo %, y letras finales B, M, K).
- IV. Se modificó el tipo de datos (“Date”, “Vol.”, y “Percentage\_diff”).
- V. Se completaron missing values en la columna “Vol.” con un promedio de días anteriores (3 días anteriores a los valores y 3 días posteriores).
- VI. Se exportó como “csv” y se guardó en un repositorio digital (Google Drive).
- VII. Se construyó la variable “target”: 1 (precio subió) y 0 (precio bajó).

- **Base de datos: BTC\_metals**

*Fuente de los datos en bruto:* <https://www.investing.com/commodities/metals>

*Descripción del data set:* se extrajeron los datos más antiguos que permite bajar la página web (junio 2014) hasta el día en que se descendieron los datos (20 de julio de 2022). En el notebook respectivo (BTC\_metals.ipynb) se puede observar los comentarios de las variables originales y como estas fueron tratadas. La temporalidad del data set es de un 1 día. Se descendieron las bases de datos de los siguientes metales: oro, plata, cobre, paladio, aluminio y platino.

*Criterio de selección:* se eligieron estos metales porque se consideran los más importantes al momento de producir bienes de alto valor y porque en algunos casos son utilizados como reserva de valor (e.g. oro). Además en el proceso de investigación documental se adquirió el conocimiento de que había una fuerte ligazón entre el cobre y el BTC

(relación que se nota en el gráfico de correlación entre el precio de ambos activos -> análisis univariado, bivariado y multivariado).

*Tratamiento de la base de datos:*

- I. Se realizó un análisis estructural básico: tipos de datos, missing values, etc.
  - II. Se modificó el nombre de la columna “Change %” por “Percentage\_diff”.
  - III. Se eliminó la columna “Vol.” de cada data set porque no se analizará por ponerse el acento en el precio de los metales y sus diferencias diarias.
  - IV. Se re-diseñó la variable “Percentage\_diff” por nomenclatura inapropiada para su posterior tratamiento (signo %, y letras finales B, M, K).
  - V. Se modificó el tipo de datos (“Date” y “Percentage\_diff”).
  - VI. Se concatenaron todas las bases individuales de cada metal.
  - VII. Se realizó un merge con la base de datos de metales (VI) con BTC\_final.
  - VIII. Se exportó como “csv” y guardó en un repositorio digital (Google Drive).
- *Bases de datos: BTC\_ind\_trend, BTC\_ind\_trend\_ET, BTC\_ind\_mom, BTC\_ind\_mom\_ET, BTC\_ind\_vol, BTC\_ind\_vol\_ET, BTC\_ind\_volat, BTC\_patterns\_indicator, BTC\_statistic\_functions, y BTC\_cycle\_indicators*

*Fuente de los datos en bruto:* BTC\_final.csv y librería Ta-lib (link en “Metadata.docx”).

*Descripción de los data sets:* poseen las variables y registros correspondientes a la aplicación del análisis técnico a los valores de la variable “Price” de “BTC\_final.csv”.

Entre los indicadores técnicos los hay de tendencia (trend), momento (momentum), volatilidad, volumen, ciclicidad, patrones y funciones estadísticas básicas (e.g. desviación estándar, varianza, etc.) y complejas (e.g. regresiones de diverso tipos, entre otros).

Las bases de datos de igual nombre pero diferenciadas por un “\_ET” final implica el agregado de variables categóricas que refieren el momento de compra o venta del activo según estrategias básicas de trading (explicadas en los respectivos notebooks).

*Criterio de selección:* se construyeron estas bases de datos para lograr un conocimiento acabado del data set “BTC\_final.csv” en lo respectivo con sus diversos aspectos técnicos (tendencia, momentos, etc.), y para detectar las señales de compra y venta del activo según estrategias de trading básicas (¿ayudan a saber cuándo sube o baja el precio?).

*Tratamiento de la base de datos:*

- I. Se procesaron los diversos indicadores y se los agrupó por tipo (tendencia, momento, volatilidad, volumen, etc.).
- II. Se aplicaron estrategias de trading básicas y se construyeron las variables categóricas correspondientes (comprar, vender, mantener posición).
- III. Se realizó un merge con la base de datos “BTC\_final”.
- IV. Se exportó cada base de dato como “csv” y se guardaron en un repositorio digital común (Google Drive).



- **Base de datos: BTC\_Forex**

*Fuente de los datos en bruto:* <https://www.investing.com/currencies/>

*Descripción del data set:* se extrajeron los datos más antiguos que permite bajar la página web (junio 2010) hasta el día en que se descargaron los datos (20 de julio de 2022) de los pares de monedas EUR/USD (Euro y Dólar Norteamericano), USD/JPY (Yuan japonés) y USD/CHY (Yen chino). Las variables originales: “Date”, “Price”, “High”, “Low”, “Open”, “Vol.”, y “Change %”. Finalmente se realizó un merge entre todas las bases de datos de Forex y con “BTC\_final” a fin de tratar los valores en conjunto.

*Criterio de selección:* se eligieron estos pares de monedas debido a la relevancia económica de Estados Unidos, Europa, Japón (moneda fuerte en estabilidad con el Dólar) y China.

*Tratamiento de la base de datos:*

- IX. Se realizó un análisis estructural básico: tipos de datos, missing values, etc.
- X. Se modificó el nombre de la columna “Change %” por “Percentage\_diff”.
- XI. Se eliminó la columna “Vol.” de cada data set porque no se analizará por ponerse el acento en el precio de los pares de monedas y sus diferencias diarias.
- XII. Se re-diseñó la variable “Percentage\_diff” por nomenclatura inapropiada.
- XIII. Se modificó el tipo de datos (“Date” y “Percentage\_diff”).
- XIV. Se concatenaron todas las bases individuales de cada para de monedas.
- XV. Se realizó un merge con la base de datos “BTC\_final”.
- XVI. Se exportó como “csv” y guardó en un repositorio digital (Google Drive).

- **Base de datos: BTC\_US\_Finance**

*Fuente de los datos en bruto:* <https://fred.stlouisfed.org/series/FEDFUNDS> (Tasa de interés de la Reserva Federal de EEUU), <https://fred.stlouisfed.org/series/CPIAUCSL> (Tasa del índice de precios al consumidor EEUU), <https://www.investing.com/rates-bonds/u.s.-10-year-bond-yield-historical-data> (Bono a 10 años de los Estados Unidos).

*Descripción del data set:* se extrajeron los datos más antiguos que permite bajar la página web (junio 2010) hasta el día en que se descargaron los datos (20 de julio de 2022) de los datos de inflación, tasas de interés y precio de los bonos a 10 de los Estados Unidos. Finalmente se realizó una concatenación con la base de datos de “BTC\_final”.

*Criterio de selección:* se eligieron estos datos influyentes en la macroeconomía mundial a fin de conocer si tiene algún tipo de correlación y/o causalidad con el dinamismo del precio del BTC.

*Tratamiento de la base de datos:*

- I. Se realizó un análisis estructural básico: tipos de datos, missing values, etc.
- II. Se modificó el nombre de la columna relativa con los cambios porcentuales de los valores a “Percentage\_diff”.
- III. Se modificó el tipo de datos (“Date” y “Percentage\_diff”).
- IV. Se concatenaron todas las bases individuales de cada tabla de datos.

- V. Se realizó un merge con la base de datos “BTC\_final”.
- VI. Se exportó como csv y guardó en un repositorio digital (Google Drive).

## 6. Análisis univariado, bivariado y multivariado de los diversos data sets

En esta sección se describe a nivel general la etapa del Análisis Exploratorio de los Datos (EDA). Para el detalle específico del análisis se remite a los respectivos notebooks del repositorio de este proyecto (Google Drive). La metadata de este proceso se encuentra en “Metadata.docx”.

Para cada base de datos anteriormente mencionada se realizó un EDA univariado, bivariado y multivariado. En concreto se analizaron las bases de datos: BTC\_metals, BTC\_ind\_trend\_ET, BTC\_ind\_mom\_ET, BTC\_patterns\_indicator, BTC\_ind\_vol\_ET - BTC\_ind\_volat - BTC\_Statistic\_Functions (en un solo análisis conjunto), y BTC\_US\_Finance y BTC\_Forex (en otro análisis conjunto).

### *Análisis univariado*

Se debe tener en cuenta que al ser una serie de tiempo algunos análisis que se deberían ejecutar en esta etapa se realizaron en la fase bivariada en contraste con la variable “Date”. Los resultados de este análisis se pueden observar en los notebooks correspondientes (ver repositorio).

En general, se llevaron adelante las siguientes acciones:

- Conocimiento de la arquitectura de los datos (cantidad de variables y registros).
- Identificación de missing values (tratamiento si se requirió).
- Reconocimiento de tipos de datos.
- Descripción estadística básica (cuartiles, mediana, media, desviación estándar, etc.)
- Gráficos de barras e histogramas para conocer distribución de los valores por variable.
- Se ejecutó un profiling a cada data set (en los notebooks se debe borrar el # delante de cada data profiling dado que generó problemas de memoria a medida que se avanzaba en el análisis bivariado y multivariado con gráficos y procesamientos más complejos).

Librerías utilizadas: pandas, matplotlib, seaborn.

### *Análisis bivariado*

Este tipo de análisis se focalizó en las variables “Date”, “Price”, “Percentage\_diff” y “Target”, las cuales se las estudió de a pares con el resto de los features. Cabe destacar que se ejecutó un desarrollo de elementos propios de una serie de tiempo: ciclicidad, estacionariedad, tendencia, y residualidad. Los resultados específicos de este análisis se pueden observar en los notebooks.

En general, se llevaron adelante las siguientes acciones:

- Gráficos de líneas, barras, violín y dispersión para la distribución temporal de los datos.
- Se ejecutó una descomposición de la serie de tiempo en sus diversos elementos. Estos se anexaron a la base de datos en análisis (ciclos, residualidad, tendencia y estacionariedad).
- Se demostró por medio de la variación de la varianza, media y estructura de autocorrelación que las series temporales en análisis no eran estacionarias.
- Se realizaron mapas de calor para conocer la correlación entre los pares de variables.

- Se ejecutaron diversos análisis de la estructura de auto-correlación (pueden servir para determinar los hiperparámetros en los modelos de ML).

Librerías utilizadas: numpy, pandas, matplotlib, seaborn, plotly, pandas\_profiling, statsmodels, pandas.plotting, statsmodels.graphics y statsmodels.graphics.tsaplots.

### *Análisis multivariado*

Este tipo de análisis se focalizó en todas variables de los data sets. Los resultados específicos del mismo se pueden observar en los notebooks correspondientes.

En general, se llevaron adelante las siguientes acciones:

- Gráficos de dispersión entre variables con discriminación por colores de los valores de “Target”. Esto permitió corroborar los valores de correlación de los heatmaps de la fase de análisis bivariado sino también conocer la distribución de las subidas y bajadas de precio de BTC en las mismas.
- Se graficaron diversos heatmaps entre variables categóricas y numéricas (continuas y discretas) haciendo las modificaciones pertinentes (variables dummies).

Librerías utilizadas: numpy, pandas, matplotlib, seaborn, plotly, pandas\_profiling, statsmodels, pandas.plotting, statsmodels.graphics y statsmodels.graphics.tsaplots.

### *Conclusiones del análisis univariado, bivariado y multivariado*

Para mayor información se remite a la lectura y análisis de los diferentes notebooks. Aquí solo se condensan las conclusiones generalísimas correspondientes a cada uno de los data sets.

- *Análisis de BTC\_metals*

- I. Se nota un incremento notable en el precio y el volumen total de transacciones diarias, en concreto, desde los 157.3 USD (2014) que BTC tenía en 2014, este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (julio 2022), su valor se redujo a 25.000 USD. De 260 operaciones (mínimo valor registro del vol. del período 2014-2022) se alcanzó el record de 4470000000 (última fase).
- II. Al momento de modelar los datos se necesitará normalizar (según el modelo).
- III. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
- IV. Los precios de los metales presentan correlaciones positivas altas y muy altas.
- V. Los precios de BTC encuentran correlaciones positivas bajas, medias y altas con distintos metales. Se destaca la correlación con el cobre (tal como anticipó la fase de revisión bibliográfica de la etapa de investigación preliminar).
- VI. Como era de esperar la correlación entre las variables [Price] y [Trend] es muy alta. Lo mismo sucede con las variables de [Price], [Low], [High] y [Open].
- VII. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días) la auto-correlación disminuye, como era de suponer.

- VIII. Se dieron subidas y bajadas en todos los valores de estacionalidad aunque mayor densidad en valores cercanos a -30, 0, 15 y 25.
  - IX. En el gráfico de residualidad se puede observar que cuando esta alcanzó picos mínimos se dieron bajadas de precio, y por el contrario, cuando alcanzó picos máximos hubo subidas de precio. En el gráfico de correlación (heatmap) solo se percibe una relación de  $|0.23|$  (muy débil) (correlación negativa para los valores 0.0 de la variable Target y correlación positiva para sus valores 1.0).
  - X. Los mayores outliers en [Percentage\_diff] se dieron con estacionariedad muy baja, y viceversa.
  - XI. La diferencia de precio de BTC tiene un comportamiento inverso a la diferencia de precios de los metales en cuanto respecta sus respectivos máximos y mínimos. Esto podría significar que cuando los metales incrementan o decrecen en gran proporción, el BTC se toma como posible inversión o resguardo de valor (habría que estudiarlo), y viceversa. Al menos esa hipótesis tiene sentido a la inversa, es decir, cuando BTC varía poco de valor es porque el capital de los inversores va a la seguridad de los metales (oro, por ejemplo).
  - XII. A medida que la diferencia del precio del cobre es positiva, la de BTC tiende a serlo también.
- *Análisis de BTC\_ind\_trend\_ET*
    - I. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Información más precisa al respecto se puede encontrar en el notebook BTC\_ind\_trading.ipynb (link disponible en: Metadata.docx).
    - II. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el caso del precio de los 0.00 USD (2014) BTC alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo período).
    - III. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
    - IV. Al momento de modelar los datos se necesitará normalizar los valores del data set (según el modelo que se termine aplicando).
    - V. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
    - VI. Como era de esperar, cuanto más días de rolling fueron tenidos en cuenta al momento de calcular los indicadores de tendencias, más suavizada son las líneas de sus gráficos a lo largo del tiempo.
    - VII. Existe una correlación positiva muy fuerte entre los indicadores de tendencia y la variable [Price] ( $>0.90$ ).

- VIII. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días), la auto-correlación disminuye, como era de suponer.
  - IX. La variable [Target] tiene una correlación débil con los valores [Comprar] y [Vender] de las diversas variables categóricas de terminación “\_ET” ( $\leq 0.30$  y  $> 0.3$ , respectivamente).
  - X. Con respecto a la residualidad se podría pensar en una estrategia de trading en valores distintos a 0. En caso de filtrarse los valores cercanos a 0, parece haber una correlación fuerte entre [Residuals] y [Target] (positiva y negativa según los valores de Target).
  - XI. El par ["Price"] ["Residuals"] nos indica que en general, que cuando los residuos se encuentran en el rango -2500/-5000 el precio bajó, y cuando se encuentra en el otro extremo, el precio subió.
  - XII. El par ["Residuals"] ["Seasonality"] nos indica que en general, cuando la residualidad fue alta, sin importar si la estacionalidad subió o bajó, el precio de BTC subió. Por el contrario, cuando la residualidad fue muy baja, el precio bajó.
- *Análisis de BTC\_ind\_mom\_ET*
    - I. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Para mayor información ver notebook: BTC\_ind\_trading.ipynb (link disponible en el documento: Metadata.docx).
    - II. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el primer caso, desde los 0.00 USD (2010) del precio de BTC este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo período).
    - III. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
    - IV. Al momento de modelar los datos se necesitará normalizar el data set (según los modelos de ML que se apliquen para predecir).
    - V. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
    - VI. Los pares [Date] y las variables de indicadores de momento denotan la posibilidad de estacionariedad. En general, no tienen tendencia. Habría que evaluar la posibilidad de ciclicidad (quizás pueda ser útil para el proyecto).
    - VII. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días) la auto-correlación disminuye, como era de suponer.

- VIII. El método estocástico de tipo k [STOCHRSI\_fastk] parece ser más sensible a la variación de precio que el estocástico de tipo d [STOCHRSI\_fastd]. Por lo tanto, al momento del modelado se podría utilizar solo [STOCHRSI\_fastk].
- IX. Se nota una muy fuerte correlación positiva entre [Price] y [ADX\_PDM] (0.91).
- X. La variable [Percentage\_diff] tiene una correlación positiva débil (0.34) con [STOCHRSI\_fastk].
- XI. A valores más altos de [Price], la variable [RSI] subió más que lo bajó, y con [RSI] más bajos hay más bajas de precio de la variable [Price] (esto se da especialmente en valores mayores a 50). Esto podría generar una estrategia de trading que tenga el valor 50 como base de compra.
- XII. Con la variable [STOCHRSI\_fastk] acontece lo mismo que en XI, y es mucho más marcado que con respecto al [STOCHRSI\_fastd].
- XIII. Con la variable [ULTOSC] acontece lo mismo que en XI y XII (por encima del valor 50). Debajo del valor=25 sería un buen indicador de que hay que vender el activo porque el precio tiene tendencia a bajar en esa franja.
- XIV. Con valores muy bajos del indicador [WILLR] (valores negativos pero cercanos a 0), el precio de BTC subió más que lo que bajó. Con valores muy bajos (valores cercanos a -100) hubo más bajas que subidas de precio del activo en cuestión.
- XV. No se observaron correlaciones significativas de la variable [Percentage\_diff] con el resto de las variables numéricas excepto con las variables con [ROC] (positiva, débil, 0.23), [RSI] (positiva, débil, 0.24), [STOCHRSI\_fastk] (positiva, débil, 0.34), WILLR (positiva, débil, 0.25).
- XVI. Solo se observan las siguientes correlaciones significativas de la variable [Target] con las variables numéricas del data set: [RSI] (débil, positiva, 0.33), [WILLR] (débil, positiva, 0.29), [STOCHRSI\_fastk] (moderada, positiva, 0.54).
- XVII. La variable [Target] no tiene correlaciones significativas en relación con las variables categóricas del data set excepto con los registros “Comprar” de la variable [STOCHRSI\_fastk\_ET] (débil, positiva, 0.33) y con los registros “Vender” de la misma variable (débil, negativa, -0.36).
- XVIII. Con respecto a la residualidad se podría pensar en una estrategia de trading en valores distintos a 0. En caso de filtrarse los valores cercanos a 0, parece haber una correlación fuerte entre [Residuals] y [Target].
- XIX. El par ["Price"] ["Residuals"] nos indica que en general, que cuando los residuos se encuentran en el rango -2500/-5000 el precio bajó, y cuando se encuentra en el otro extremo, el precio subió.
- XX. El par ["Residuals"] ["Seasonality"] nos indica que cuando la residualidad fue alta, sin importar la estacionalidad el precio tendió a subir. Por el contrario, cuando la residualidad fue muy baja, el precio bajó.
- XXI. Se pueden deducir numerosas estrategias de trading que podrían ser modeladas:



- a. En cualquier valor de [ADX] si el valor de [ADX\_MDI] y [ADX\_MDM] se mantiene cercano a 0, el precio sube.
- b. En cualquier valor de [ADX] si el valor de [ADX\_MDI] y [ADX\_MDM] se mantiene cercano a su máximo (100), el precio baja.
- c. Si [ADX] es mayor a 10 y el [ADX\_DX] es mayor a 30/40 y menor a 75, el precio sube (caso contrario el precio tiende a bajar).
- d. Si [ADX\_MDM] y [ADX\_MDI] son menores a 20 y [ADX\_PDI] es mayor a 40, el precio sube (salvo en los extremos de ADX\_PDI).
- e. Si [ADX\_DX] es mayor a 65 y [ADX] es mayor a 35 el precio tiende a subir.
- f. Si [ADX\_DX] es mayor a 40 y [ADX\_PDI] es mayor a 35, el precio tiende a subir.
- g. Si [AroonDown] es menor que 20 y [ADX\_PDI] es mayor que 40, el precio sube.
- h. Entre [MACD] y [MOM] se podría pensar en una regresión lineal o un algoritmo de K-Nearest Neighbors. Se podría trazar una línea en la que queden muchas más bajas y compras de un lado y otro de la línea respectivamente. Lo mismo se observa con claridad entre [MACD] y [RSI].
- i. También se percibe que en los extremos del estocástico rápido (cualquiera de sus dos versiones), el precio sube o baja respectivamente a medida que lo hace el indicador. Lo mismo sucede en los extremos de [WILLR].
- j. Si el [MOM] es mayor a 0 y el [ULSTOC] es mayor a 40, el precio tiende a subir.
- k. Si el [MOM] es mayor a 0 y [RSI] mayor a 50, el precio parece subir.
- l. Entre los estocásticos se podría pensar en una regresión lineal o un algoritmo de K-Nearest Neighbors para ciertos rangos a fin de obtener zonas relativamente seguras de suba y baja de precio de BTC.
- m. Si [ADX] es mayor a 0 y [RSI] es mayor a 70, el precio tiende a subir. En caso de que [ADX] es menor a 50 y [RSI] es menor a 40, el precio tiende a bajar.
- n. Entre [RSI] y [ADX\_MDI], al igual que entre [RSI] y [ADX\_MDM] se podría pensar en una regresión lineal o un algoritmo de K-Nearest Neighbors dado que existe correlación negativa fuerte. Se podría trazar una línea en la que queden muchas más bajas/compras de un lado y otro de líneas imaginarias respectivamente.
- o. Entre [RSI] y [ADX\_PDI] se podría pensar en una regresión lineal o un algoritmo de K-Nearest Neighbors dado que existe correlación positiva fuerte. Se podría trazar una línea en la que queden muchas más bajas/compras de un lado y otro de una línea imaginaria, respectivamente.
- p. En el par RSI y los Aroon\_Oscillator podrían generarse algunos modelos de ML para predecir aumento o descenso del precio.

- *Análisis de BTC\_patterns\_indicator*

- I. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Información más precisa al respecto se puede encontrar en el notebook BTC\_ind\_trading.ipynb (link disponible en: Metadata.docx).

- II. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el primer caso, desde los 0.00 USD del precio de BTC, este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo período).
- III. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
- IV. Al momento de modelar los datos se necesitará normalizar el data set (según los modelos que se terminen eligiendo para aplicar).
- V. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
- VI. Luego de realizar un gráfico de los patrones en el tiempo ([Date]) se realizó un filtrado de variables. El análisis se continuó con las variables que se detallan a continuación. El resto de las columnas se descartó porque no mostraban valores != de 0. Variables que se mantuvieron:
  - a. [CDL3OUTSIDE]
  - b. [CDLADVANCEBLOCK]
  - c. [CDLBELTHOLD]
  - d. [CDLCLOSINGMARUBOZU]
  - e. [CDLDOJI]
  - f. [CDLDOJISTAR]
  - g. [CDLDRAGONFLYDOJI]
  - h. [CDLENGULFING]
  - i. [CDLGRAVESTONEDOJI]
  - j. [CDLHAMMER]
  - k. [CDLHANGINGMAN]
  - l. [CDLHARAMI]
  - m. [CDLHARAMICROSS]
  - n. [CDLHIGHWAVE]
  - o. [CDLHIKKAKE]
  - p. [CDLLONGLEGGEDDOJI]
  - q. [CDLLONGLINE]
  - r. [CDLMARUBOZU]
  - s. [CDLMATCHINGLOW]
  - t. [CDLRICKSHAWMAN]
  - u. [CDLSEPARATINGLINES]
  - v. [CDLSHOOTINGSTAR]
  - w. [CDLSHORTLINE]
  - x. [CDLSPINNINGTOP]
  - y. [CDLTAKURI]
  - z. [CDLXSIDEGAP3METHODS]

- VII. En un gran porcentaje de los patrones filtrados se observa su aparición a la largo de toda la serie de tiempo (gráfico -> [Pattern x] y [Price]).
- VIII. La correlación entre [Percentage\_diff] y [CDLBELTHOLD] es positiva y de 0.35 (débil). La correlación entre [Percentage\_diff] y [CDLLONGLINE] es positiva y de 0.21 (débil).
- IX. La variable [Target] tiene una correlación débil con [CDLBELTHOLD] de 0.38 y una correlación moderada con [CDLLONGLINE] de 0.40. En caso de los valores 0 y 1 de [Target] los signos de las correlaciones son negativo y positivo respectivamente si se los analiza por separado.
- X. La variable [Target] tiene una correlación débil con [CDLSHORTLINE] de 0.37 y una correlación moderada con [CDLSPINNINGTOP] de 0.39. En caso de los valores 0 y 1 de [Target] los signos de las correlaciones son negativo y positivo respectivamente si se los analiza por separado.
- XI. Luego de realizar el análisis de [Seasonality] y de los patrones filtrados arriba, se volvió a hacer una selección de aquellos patrones que muestran con claridad que al cumplirse la forma invocada no había falsos positivos o negativos. Así el nuevo listado de patrones que se pueden utilizar para modelar con ML son los siguientes:
  - a. [CDL3OUTSIDE]
  - b. [CDLADVANCEBLOCK]
  - c. [CDLBELTHOLD]
  - d. [CDLCLOSINGMARUBOZU]
  - e. [CDLDOJISTAR]
  - f. [CDLENGULFING]
  - g. [CDLHARAMI]
  - h. [CDLHARAMICROSS]
  - i. [CDLHIGHWAVE]
  - j. [CDLHIKKAKE]
  - k. [CDLLONGLINE]
  - l. [CDLMARUBOZU]
  - m. [CDLSEPARATINGLINES]
  - n. [CDLSHOOTINGSTAR]
  - o. [CDLSHORTLINE]
  - p. [CDLSPINNINGTOP]
  - q. [CDLXSIDEGAP3METHODS]

- *Análisis de BTC\_multiple\_indicators*

*Se analizó en un solo notebook las siguientes bases de datos:*

*BTC\_ind\_vol\_ET, BTC\_ind\_volat, BTC\_statistic\_functions, y BTC\_ind\_cycle\_indicators, y variables de una serie de tiempo obtenidas al momento del estudio concreto.*

- I. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Para mayor información ver notebook: BTC\_ind\_trading.ipynb (link disponible en el documento: Metadata.docx).
- II. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el primer caso, desde los 0.00 USD (2010) del precio de BTC este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo).
- III. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
- IV. Al momento de modelar los datos se necesitará normalizar el data set (según los modelos de ML que se apliquen para predecir).
- V. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
- VI. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días) la auto-correlación disminuye, como era de suponer.
- VII. Existe una correlación positiva muy fuerte entre [TSF], [Price], [Low], [High] y [Open], [Linearregresion], y [Linearintercept].
- VIII. Se nota una muy fuerte correlación positiva entre [Price] y [ATR] (volatilidad).
- IX. Se nota una fuerte correlación positiva entre [Price], [VAR], y [STDDEV].
- X. Hay una correlación negativa débil entre [Vol.] y [ADOSC] (volumen).
- XI. La residualidad y ciclicidad de BTC se encuentran fuertemente correlacionadas (no así la estacionariedad).
- XII. Se pueden deducir numerosas oportunidades para aplicar modelos de ML y de generar estrategias de trading que también podrían ser modeladas a partir de la relación entre las variables:
  - a. El par [ADOSC] - [HT\_DCPERIOD] refleja que en determinadas franjas el precio subió y en otras bajó. Ej. mientras [ADOSC] se mantuvo entre 0,4 y 1, y [HT\_DCPERIOD] entre 0 y 20, el precio de BTC subió. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC).
  - b. El par [ADOSC] - [HT\_DCPHASE] refleja que en determinadas franjas el precio subió y en otras bajó. Ej. mientras [ADOSC] se mantuvo entre 0 y 1, y [HT\_DCPHASE] entre 100 y 200, el precio de BTC subió. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC).
  - c. El par [ADOSC] - [HT\_PHASOR\_quadrature] refleja que en determinadas franjas el precio subió y en otras bajó. Ej. mientras [ADOSC] se mantuvo entre 0 y 1,5, y [HT\_PHASOR\_quadrature] entre -500 y 500, el precio de BTC subió. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC).
  - d. El par [ADOSC] - [HT\_SINE\_sine] refleja que en determinadas franjas el precio subió y en otras bajó. Ej. mientras [ADOSC] se mantuvo por encima de 0, el precio

de BTC subió. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC).

- e. El par [HT\_PHASOR\_quadrature] - [HT\_SINE\_sine] refleja que en determinadas franjas el precio subió y en otras bajó. Se podría pensar en una estrategia de trading fundada en KNN.
- f. El par [HT\_PHASOR\_quadrature] - [HT\_PHASOR\_leadsine] refleja que en determinadas franjas el precio subió y en otras bajó. Se podría pensar en una estrategia de trading fundada en KNN.
- g. El par [Linearslope] - [Linearangle] refleja que en determinadas franjas el precio subió y en otras bajó. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC) mediante algún modelo de KNN (por su modo de predecir).
- h. Se nota que la variable [Residuals] podría utilizarse con diversos modelos para predecir la subida o bajada de precio con [Price], [Percentage\_diff], [VAR], [STDDEV], [TSF], [Trend], [Seasonality] y [BTC\_cycle]. Lo mismo acontece con BTC\_cycle. Se hipotetiza que modelos de clasificación podrían ofrecer resultados alineados con los objetivos de este estudio (KNN, SVM, etc).
- i. Se nota que la variable [Residuals] podría utilizarse con diversos modelos para predecir la subida o bajada de precio con [BTC\_cycle], [HT\_PHASOR], [HP\_SINE\_sine], [HP\_SINE\_leadsine]. Se hipotetiza que modelos de clasificación podrían ofrecer resultados alineados con los objetivos generales y específicos de este estudio (KNN, SVM, etc).

- *Análisis de BTC\_US*

*Se analizó en un solo notebook las siguientes bases de datos:*

*BTC\_Forex, BTC\_US\_Finance y variables de una serie de tiempo obtenidas al momento del estudio concreto.*

- I. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Para mayor información ver los notebooks: BTC\_Forex.ipynb y BTC\_US\_Finance.ipynb (link disponible en el documento: Metadata.docx).
- II. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el primer caso, desde los 0.00 USD (2010) del precio de BTC este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo).
- III. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
- IV. Al momento de modelar los datos se necesitará normalizar el data set (según los modelos de ML que se apliquen para predecir).
- V. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
- VI. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días) la auto-correlación disminuye, como era de suponer.

- VII. Se percibe un incremento notable en las bolsas de valores de US hasta finales del año 2021.
- VIII. En lo que respecta a las diferencias de precio intradiaria en las bolsas de valores se percibe un máximo y mínimo histórico a principios del año 2020 con el inicio de la pandemia del COVID-19.
- IX. La inflación acumulada (ACPI) creció exponencialmente a partir del 2020 (pandemia).
- X. Durante la pandemia del COVID-19 se mantuvieron en valores cercanos a cero las tasas de interés de los US.
- XI. Se puede observar como el USD se empodera a partir del 2022 con las tres monedas de comparación (EUR, JPY, CNY).
- XII. Se nota una muy fuerte correlación positiva entre [Price], [Open], [High], [Low], [Price\_SP500], [Price\_Nasdaq], [Price\_DJ].
- XIII. Se nota una correlación negativa débil entre [Price] y [Price\_US\_Bond].
- XIV. Se nota una correlación positiva fuerte entre [Price] y [I\_Percentage\_diff\_ACPI].
- XV. Se nota una correlación positiva moderada entre [Price] y [M\_Percentage\_diff\_MCPI].
- XVI. Se nota una correlación negativa débil entre [Price] y [EUR\_USD].
- XVII. Se nota una correlación positiva débil entre [Price] y [USD\_JPY].
- XVIII. Se nota una correlación positiva débil entre [Vol.] y [I\_Percentage\_diff\_ACPI].
- XIX. Se nota una correlación positiva moderada entre [Trend] y [I\_Percentage\_diff\_ACPI].
- XX. Se nota una correlación muy fuerte entre los precios de las tres bolsas de EEUU.
- XXI. Entre la variable [Price] y [Percentage\_diff\_NASDAQ] se puede percibir una relación entre subidas y bajadas de precio. A mayor diferencia de NASDAQ el precio tendió a subir, mientras que a diferencia más negativa del valor de NASDAQ, el precio tendió a bajar.
- XXII. Dado que no se observan patrones definidos con claridad más que algunas zonas de densidad de bajas y ventas del precio del BTC, se podría pensar en modelos supervisados de ML de clasificación (SVM, KNN, etc.).

## 7. Data sets para modelar con ML

Los análisis univariado, bivariado y multivariado de los data sets en cuestión permitieron la reducción de variables a modelar. Esto resulta fundamental para reducir el poder cómputo necesario, favorecer el tiempo de procesamiento de los algoritmos, y obtener mejores métricas en los modelos (exactitud, precisión, sensibilidad, especificidad, y F1-score).

Los criterios de selección de las variables numéricas (continuas y discretas) y categóricas que se conservan para el modelado de ML son las siguientes:



- Que las variables guarden una correlación débil, moderada, fuerte o muy fuerte con las variables [Price], [Percentage\_diff] y/o [Target].
  - En el caso de correlaciones negativas la escala aplicada es la siguiente:  $-0.4 < x \leq -0.2$  (débil),  $-0.6 < x \leq -0.4$  (moderada),  $-0.8 < x \leq -0.6$  (fuerte), y  $-1 \leq x \leq -0.8$  (muy fuerte).
  - En el caso de correlaciones positivas la escala aplicada es contraria en signo:  $0.2 \leq x < 0.4$  (débil),  $0.4 \leq x < 0.6$  (moderada),  $0.6 \leq x < 0.8$  (fuerte),  $0.8 \leq x \leq 1$  (muy fuerte).
- Que las variables favorezcan la discriminación de subidas y bajadas de precio de BTC más allá de la correlación que exista con las variables [Price], [Percentage\_diff] y/o [Target]. Esto sería el caso de correlaciones de tipo indirectas que influyan con correlaciones fuertes y muy fuertes a una variable que posea una correlación directa con las variables [Price], [Percentage\_diff] y/o [Target].

Debajo se especifican las variables que se conservarán cuando se proceda a diseñar y ejecutar los algoritmos supervisados (y no supervisados según caso también) de ML.

- *BTC\_metals.csv*

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

- [Date]
- [Price]
- [Open]
- [High]
- [Low]
- [Vol.]
- [Target]
- [Percentage\_diff]
- [Price\_gold]-> por estimarse mundialmente como reserva de valor
- [Perc\_diff\_gold]-> por estimarse mundialmente como reserva de valor
- [Price\_copper]
- [Perc\_diff\_copper]
- [Price\_aluminium]-> por guardar correlación positiva moderada (0.34) con [Vol.].
- [Perc\_diff\_aluminium]-> por guardar correlación positiva moderada (0.34) con [Vol.].
- [Trend]
- [Residuals]

Nota: a pesar de guardar correlación fuerte con la variable [Price] se descartaron precios y diferencias porcentuales de metales dado que la mayor correlación con esta variable la lleva el cobre [Price\_copper], a saber, una de tipo positiva y muy fuerte (0.88).

- *BTC\_ind\_trend\_ET.csv*

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

- [Date]
- [Price]
- [Open]
- [High]
- [Low]
- [Vol.]
- [Percentage\_diff]
- [Target]
- [EMA\_7d]
- [HT\_TRENDLINE]
- [BBANDS\_upper\_band]
- [BBANDS\_middle\_band]
- [BBANDS\_lower\_band]
- [PSAR]
- [EMA\_7d\_ET]
- [HTTL\_ET]
- [PSAR\_ET]
- [Trend]
- [Residuals]

*Nota:* el resto de las medias móviles que no se conservaron (de distinta temporalidad) se eliminaron por una menor correlación con la variable [Price]. Además se tuvo en cuenta que la mayor auto-correlación de la serie de tiempo [Date] y [Price] ocurre con menor cantidad de días de “lag”. También se añade que se prefirió la media móvil exponencial (EMA) a la media móvil simple (SMA) por ajustarse mejor al precio dado que pondera en su fórmula el peso de los últimos valores de la serie.

- *BTC\_ind\_mom\_ET.csv*

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

- [Price]
- [Open]
- [High]
- [Low]
- [Vol.]
- [Percentage\_diff]
- [Target]
- [ADX]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [ADX\_MDI]-> se conserva por posible estrategia de trading (descrita arriba en 6.)

- [ADX\_MDM]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [ADX\_DX]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [ADX\_PDI]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [ADX\_PDM]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [Aroon\_Oscillator]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [MACD] -> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [MOM]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [ROC]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [RSI]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [STOCHRSI\_fastk]
- [ULTOSC]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [WILLR]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [STOCHRSI\_fastk\_ET]
- [Trend]
- [Residual]

*Nota:* se prefiere el método estocástico rápido de tipo k al método rápido de tipo d porque se ajusta mejor al precio de BTC ([Price]).

- ***BTC\_pattern\_indicators.csv***

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

- [Price]
- [Open]
- [High]
- [Low]
- [Vol.]
- [Percentage\_diff]
- [Target]
- [CDL3OUTSIDE]
- [CDLADVANCEBLOCK]
- [CDLBELTHOLD]
- [CDLCLOSINGMARUBOZU]
- [CDLDOJISTAR]
- [CDLENGULFING]
- [CDLHARAMI]
- [CDLHARAMICROSS]
- [CDLHIGHWAVE]
- [CDLHIKKAKE]
- [CDLLONGLINE]
- [CDLMARUBOZU]

- [CDLSEPARATINGLINES]
- [CDLSHOOTINGSTAR]
- [CDLSHORTLINE]
- [CDLSPINNINGTOP]
- [CDLXSIDEGAP3METHODS]
- [Trend]
- [Residual]

Nota: la explicación de por qué se conservaban estos patrones fue realizada arriba (6, XI).

- ***BTC\_multiple\_indicators (conjunto de bases de datos):***

*BTC\_ind\_vol\_ET, BTC\_ind\_volat, BTC\_statistic\_functions, y BTC\_ind\_cycle\_indicators, y variables de análisis de serie de tiempo (Seasonality, Trend, etc) obtenidas al momento del estudio en particular.*

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

- [Price]
- [Open]
- [High]
- [Low]
- [Vol.]
- [Percentage\_diff]
- [Target]
- [ADOSC] (BTC\_ind\_vol\_ET)
- [ATR] (BTC\_ind\_volat)
- [Linearregression] (BTC\_statistic\_functions)
- [Linearintercept] (BTC\_statistic\_functions)
- [Linearslope] (BTC\_statistic\_functions)
- [Linearangle] (BTC\_statistic\_functions)
- [VAR] (BTC\_statistic\_functions)-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [STDDEV] (BTC\_statistic\_functions) )-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [HT\_SINE\_sine] (BTC\_ind\_cycle\_indicators)-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [HT\_PHASOR\_leadsine] (BTC\_ind\_cycle\_indicators)-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [HT\_DCPHASE] (BTC\_ind\_cycle\_indicators)-> se conserva por posible estrategia de trading (descrita arriba en 6.)
- [Seasonality] (análisis de serie de tiempo)
- [Trend] (análisis de serie de tiempo)
- [Residual] (análisis de serie de tiempo)
- [BTC\_cycle] (análisis de serie de tiempo)

- *BTC\_US (conjunto de bases de datos):*

*BTC\_Forex, BTC\_US\_Finance, y variables de análisis de serie de tiempo (Seasonality, Trend, etc) obtenidas al momento del estudio en particular.*

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

- [Price]
- [Open]
- [High]
- [Low]
- [Percentage\_diff]
- [Target]
- [Price\_SP500] (BTC\_US\_Finance)
- [Price\_NASDAQ] (BTC\_US\_Finance)
- [Price\_DJ] (BTC\_US\_Finance)
- [Price\_US\_Bond] (BTC\_US\_Finance)
- [I\_Percentage\_diff\_ACPI] (BTC\_US\_Finance)
- [M\_Percentage\_diff\_MCPI] (BTC\_US\_Finance)
- [FEDfunds] (BTC\_US\_Finance)
- [Percentage\_diff\_USD\_JPY] (BTC\_Forex)
- [Seasonality] (análisis de serie de tiempo)
- [Trend] (análisis de serie de tiempo)
- [Residual] (análisis de serie de tiempo)
- [BTC\_cycle] (análisis de serie de tiempo)

## 8. Modelado de ML

En este apartado se describe la etapa 5 (Modelado) de este proyecto (EDT). Se desarrollarán modelos supervisados de ML (en primera instancia) y no supervisados (según el caso), a fin de poder cumplimentar con el objetivo general descrito arriba.

Así se espera poder predecir la subida y bajadas futuras de precio de BTC en el corto plazo y detectar con mayor exactitud (accuracy), y otras métricas, con un porcentaje de previsibilidad mayor al 50% (mera aleatoriedad).

### 8.1. *Data wrangling*

Dado que cada base de datos explicitada en “7” (en este documento) será utilizada para generar diversos modelos de ML, a fin de no repetir un mismo proceso de manipulación de variables en cada uno de los notebooks en lo que se trabajará con cada algoritmo, se procede a la creación de los data sets finales que serán modelados.

- *BTC\_metals\_fv.ipynb*

El objetivo de este notebook es reajustar el data set "BTC\_metals.csv" según las conclusiones alcanzadas al momento de llevarse a cabo su correspondiente análisis univariado, bivariado y

multivariado (ver, "Primera entrega.docx", "Metadata.docx", y/o "Análisis univariado, bivariado y multivariado de BTC\_metals.ipynb").

En concreto, se procederá a eliminar las variables originales del data set que fueron descartadas según los criterios establecidos en el punto 7 (de este documento), y se agregarán las variables [Trend] y [Residuals] (propias de una serie de tiempo). Así, el data set "BTC\_metals\_fv.csv" solo requerirá manipulaciones menores al momento de ser procesado por los diversos modelos de ML en lo que resta del proyecto. Para mayor explicación remitirse a "Metadata.docx".

*Procedimiento:* se eliminaron las variables que no se van a utilizar (ver, 7) y se adjuntaron las variables [Trend] y [Residuals].

*Base de datos:* BTC\_metals\_fv.csv

*Notebook:* BTC\_metals\_fv.ipynb

- *BTC\_ind\_trend\_ET\_fv.ipynb*

El objetivo de este notebook es reajustar el data set "BTC\_ind\_trend\_ET.csv" según las conclusiones alcanzadas al momento de llevarse a cabo su correspondiente análisis univariado, bivariado y multivariado (ver, "Primera entrega.docx", "Metadata.docx", y/o "Análisis univariado, bivariado y multivariado de BTC\_ind\_trend\_ET.ipynb").

En concreto, se procederá a eliminar las variables originales de la base de datos que fueron descartadas según los criterios establecidos en el punto 7 (de este documento) y se agregarán las variables [Trend] y [Residuals] (propias de una serie de tiempo). Así, el data set "BTC\_ind\_trend\_ET\_fv.csv" solo requerirá manipulaciones menores al momento de ser procesado por los diversos modelos de ML en lo que resta del proyecto. Para mayor explicación remitirse a "Metadata.docx".

*Procedimiento:* se eliminaron las variables que no se van a utilizar (ver, 7) y se adjuntaron las variables [Trend] y [Residuals].

*Base de datos:* BTC\_ind\_trend\_ET\_fv.csv

*Notebook:* BTC\_ind\_trend\_ET\_fv.ipynb

- *BTC\_ind\_mom\_ET\_fv.ipynb*

El objetivo de este notebook es reajustar el data set "BTC\_ind\_mom\_ET.csv" según las conclusiones alcanzadas al momento de llevarse a cabo su correspondiente análisis univariado, bivariado y multivariado (ver, "Primera entrega.docx", "Metadata.docx", y/o "Análisis univariado, bivariado y multivariado de BTC\_ind\_mom\_ET.ipynb").

En concreto, se procederá a eliminar las variables originales de la base de datos que fueron descartadas según los criterios establecidos en el punto 7 (de este documento) y se agregarán las variables [Trend] y [Residuals] (propias de una serie de tiempo). Así, el data set "BTC\_ind\_mom\_ET\_fv.csv" solo requerirá manipulaciones menores al momento de ser procesado por los



diversos modelos de ML en lo que resta del proyecto. Para mayor explicación remitirse a "Metadata.docx".

*Procedimiento:* se eliminaron las variables que no se van a utilizar (ver, 7) y se adjuntaron las variables [Trend] y [Residuals].

*Base de datos:* BTC\_ind\_mom\_ET\_fv.csv

*Notebook:* BTC\_ind\_mom\_ET\_fv.ipynb

- *BTC\_US\_fv.csv*

Esta base de dato fue creada en la última parte del notebook "Análisis univariado, bivariado y multivariado de BTC\_US\_Finance.ipynb" según las conclusiones alcanzadas al momento de llevarse a cabo los análisis univariado, bivariado y multivariado de los data set BTC\_Forex.csv y BTC\_US\_Finance.csv (ver "Primera entrega.docx", "Metadata.docx").

En concreto, se eliminaron las variables originales de la base de datos que fueron descartadas según los criterios establecidos en el punto 7 (de este documento). Así, "BTC\_US\_fv.csv" solo requerirá manipulaciones menores cuando sea procesado por los diversos modelos de ML en lo que resta del proyecto. Para mayor explicación remitirse a "Metadata.docx".

*Procedimiento:* se eliminaron las variables que no se van a utilizar (ver 7).

*Base de datos:* BTC\_US\_fv.csv

*Notebook:* Análisis univariado, bivariado y multivariado de BTC\_US\_Finance.ipynb

- *BTC\_multiple\_indicators\_fv.csv*

Esta base de dato fue creada en la última parte del notebook "Análisis univariado, bivariado, multivariado BTC\_multiple\_indicators.ipynb" según las conclusiones alcanzadas al momento de llevarse a cabo los análisis univariado, bivariado y multivariado de los data set BTC\_ind\_vol\_ET.csv, BTC\_ind\_volat.csv, BTC\_statistic\_functions.csv y BTC\_ind\_cycle\_indicators.csv (ver "Primera entrega.docx", "Metadata.docx").

En concreto, se eliminaron las variables originales de las bases de datos que fueron descartadas según los criterios establecidos en el punto 7 (de este documento). Así, el data set "BTC\_multiple\_indicators\_fv.csv" solo requerirá manipulaciones menores cuando sea procesado por los diversos modelos de ML en lo que resta del proyecto. Para mayor explicación remitirse a "Metadata.docx".

*Procedimiento:* se eliminaron las variables que no se van a utilizar (ver 7).

*Base de datos:* BTC\_multiple\_indicators\_fv.csv

*Notebook:* Análisis univariado, bivariado y multivariado de BTC\_multiple\_indicators.ipynb

- *BTC\_pattern\_indicators\_fv.csv*

Esta base de dato fue creada mediante el notebook "BTC\_pattern\_indicators\_fv.ipynb" según las conclusiones alcanzadas al momento de llevarse a cabo el análisis univariado, bivariado y multivariado de BTC\_patterns\_indicator.csv ("Primera entrega.docx", "Metadata.docx").

En concreto, se eliminaron las variables originales de la base de datos que fueron descartadas según los criterios establecidos en el punto 7 (de este documento). Así, el data set "BTC\_pattern\_indicators\_fv.csv" solo requerirá manipulaciones menores cuando sea procesado por los diversos modelos de ML en lo que resta del proyecto.

*Procedimiento:* se eliminaron las variables que no se van a utilizar (ver 7).

*Base de datos:* BTC\_pattern\_indicators\_fv.csv

*Notebook:* BTC\_pattern\_indicators\_fv.ipynb

- *BTC\_various\_indicators\_fv.csv*

Esta base de dato fue creada mediante el notebook "BTC\_various\_indicators\_fv.ipynb" según las conclusiones alcanzadas al momento de llevarse a cabo el análisis univariado, bivariado y multivariado de todas las bases de datos (ver, "Primera entrega.docx", "Metadata.docx").

El criterio de selección de las variables de este data set ha sido la correlación (> 70%) de las variables analizadas con respecto a las columnas "Price", "Percentage\_difference", y "Target" en cada base de datos. Solo se descartó la base de datos BTC\_pattern\_indicators\_fv.csv por no presentar alta correlación con las variables en cuestión.

Así "BTC\_various\_indicators\_fv.csv" solo requerirá manipulaciones menores cuando sea procesado por los diversos modelos de ML en lo que resta del proyecto.

*Base de datos:* BTC\_various\_indicators\_fv.csv

*Notebook:* BTC\_various\_indicators\_fv.ipynb

## 8.2. Modelos (Decision Tree, Random Forest, KNN, Logistic Regression y SVM)

Una cuestión importante a destacar es el hecho de que se la variable [Target] fue retrasada un registro (shift (-1)) a fin de intentar predecir si el precio futuro inmediato subirá o bajará. Se debe recordar que se está trabajando con la predicción de precios futuros de una serie de tiempo y que por lo tanto es necesario realizar este tipo de modificación.

### 8.2.1. Algoritmo de clasificación (Decision Tree + BTC\_metals\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Decision Tree (sklearn). En un conjunto de modelos se eliminaron los 12 registros NaN en las variables [Trend] y [Residuals] (la explicación se encuentran sobre el mismo notebook) y en el otro conjunto de modelos se los preservó. Al final del notebook se puede observar la manipulación de los hiperparámetros del modelo a fin de mejorar las métricas de predicción:

- Los mejores modelos encontrados fueron los siguientes:
  - Eliminación de NaN (DecisionTreeClassifier):
    - % de aciertos sobre el set de entrenamiento: 0.6736081597960051
    - % de aciertos sobre el set de evaluación (accuracy): 0.6723259762308998
    - Precision Score of the classifier is: 0.6911764705882353
    - Recall Score of the classifier is: 0.632996632996633
    - F1 Score of the classifier is: 0.6608084358523726
    - AUC for our classifier is: 0.7045166274618329
  - Conservación de NaN (HistGradientBoostingClassifier()):
    - % de aciertos sobre el set de entrenamiento: 0.9742593491986401
    - % de aciertos sobre el set de evaluación (accuracy): 0.6455266138165345
    - Precision Score of the classifier is: 0.6778947368421052
    - Recall Score of the classifier is: 0.6680497925311203
    - F1 Score of the classifier is: 0.6729362591431557
    - AUC for our classifier is: 0.7007326083132418

*Notebook:* Algoritmo de clasificación (Decision Tree + BTC\_metals).ipynb

<https://drive.google.com/file/d/1EvlkQaLTYu5362Mw16rLjn7S4mnQeS8Z/view?usp=sharing>

### 8.2.2. *Algoritmos de clasificación (Random forest + BTC\_metals\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Random forest (sklearn). Se eliminaron los 12 registros NaN en las variables [Trend] y [Residuals] (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ()):
  - % de aciertos sobre el set de entrenamiento: 0.6893203883495146
  - % de aciertos sobre el set de evaluación (accuracy): 0.6500566251415628
  - Precision Score of the classifier is: 0.6812933025404158
  - Recall Score of the classifier is: 0.6330472103004292
  - F1 Score of the classifier is: 0.6562847608453838
  - AUC for our classifier is: 0.6950731260485175

*Notebook:* Algoritmos de clasificación (Random forest + BTC\_metals\_fv).ipynb

<https://drive.google.com/file/d/1HY7oMJwti-1ScVB03rtw363PbZO16DKU/view?usp=sharing>

### 8.2.3. *Algoritmos de clasificación (KNN + BTC\_metals\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo KNN (sklearn). Se eliminaron los 12 registros NaN en las variables [Trend] y [Residuals] (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ()):

- % de aciertos sobre el set de entrenamiento: 0.6786407766990291
- % de aciertos sobre el set de evaluación: 0.5877689694224235
- Precision Score of the classifier is: 0.601593625498008
- Recall Score of the classifier is: 0.648068669527897
- F1 Score of the classifier is: 0.6239669421487604
- AUC for our classifier is: 0.6140117948559607

*Notebook:* Algoritmos de clasificación (KNN + BTC\_metals\_fv).ipynb

<https://drive.google.com/file/d/15VpLgDRtGuLKRYlbdFKWUtNOPlhM-P3q/view?usp=sharing>

#### 8.2.4. *Algoritmos de clasificación (Logistic Regression + BTC\_metals\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Logistic Regression (sklearn). Se eliminaron los 12 registros NaN en las variables [Trend] y [Residuals] (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ()):
  - % de aciertos sobre el set de entrenamiento: 0.6572815533980583
  - % de aciertos sobre el set de evaluación es (accuracy): 0.6511891279728199
  - Accuracy of the classifier is: 0.652321630804077
  - Precision Score of the classifier is: 0.6377816291161178
  - Recall Score of the classifier is: 0.7896995708154506
  - F1 Score of the classifier is: 0.7056567593480345
  - AUC for our classifier is: 0.7133366268358704

*Notebook:* Algoritmos de clasificación (Logistic Regression + BTC\_metals\_fv).ipynb

[https://drive.google.com/file/d/1yWYMxh8zCi2Mp3epRFMc\\_Tci4L9cZQPb/view?usp=sharing](https://drive.google.com/file/d/1yWYMxh8zCi2Mp3epRFMc_Tci4L9cZQPb/view?usp=sharing)

#### 8.2.5. *Algoritmos de clasificación (SVM + BTC\_metals\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo SVM (sklearn). Se eliminaron los 12 registros NaN en las variables [Trend] y [Residuals] (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ()):
  - % de aciertos sobre el set de evaluación para train: 0.6839422259983008
  - % de aciertos sobre el set de evaluación para test (accuracy): 0.6400679117147708
  - Accuracy of the classifier is: 0.6400679117147708
  - Precision Score of the classifier is: 0.6183206106870229
  - Recall Score of the classifier is: 0.7967213114754098
  - F1 Score of the classifier is: 0.6962750716332379

*Notebook:* Algoritmos de clasificación (SVM + BTC\_metals\_fv).ipynb

<https://drive.google.com/file/d/1q7dDJBu1pYKwBnO4fTD8EJ26nSPb7n6o/view?usp=sharing>

### 8.2.6. Algoritmo de clasificación (Decision Tree + BTC\_ind\_trend\_ET\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Decision Tree (sklearn). En un conjunto de modelos se eliminaron los registros NaN y en el otro conjunto de modelos se los preservó. Al final del notebook se puede observar la manipulación de los hiperparámetros del modelo a fin de mejorar las métricas de predicción:

- Los mejores modelos encontrados fueron los siguientes:
  - Eliminación de NaN (DecisionTreeClassifier):
    - % de aciertos sobre el set de entrenamiento: 0.6715856481481481
    - % de aciertos sobre el set de evaluación (accuracy): 0.6701388888888888
    - Accuracy of the classifier is: 0.6701388888888888
    - Precision Score of the classifier is: 0.699724517906336
    - Recall Score of the classifier is: 0.5906976744186047
    - F1 Score of the classifier is: 0.6406052963430012
    - AUC for our classifier is: 0.7084583645911477
  - Conservación de NaN (HistGradientBoostingClassifier()):
    - % de aciertos sobre el set de entrenamiento: 0.909742587161942
    - % de aciertos sobre el set de evaluación (accuracy): 0.6671732522796353
    - Precision Score of the classifier is: 0.6341829085457271
    - Recall Score of the classifier is: 0.6855753646677472
    - F1 Score of the classifier is: 0.6588785046728972
    - AUC for our classifier is: 0.755325389593376

*Notebook:* Algoritmo de clasificación (Decision Tree + BTC\_ind\_trend\_ET\_fv).ipynb  
<https://drive.google.com/file/d/1wVhE-awW1R7cxbhiLNMMyz5ISTKUWYbt7/view?usp=sharing>

### 8.2.7. Algoritmos de clasificación (Random forest + BTC\_ind\_trend\_ET\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Random forest (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.6636904761904762
  - % de aciertos sobre el set de evaluación (accuracy): 0.6651234567901234
  - Precision Score of the classifier is: 0.6836363636363636
  - Recall Score of the classifier is: 0.5911949685534591
  - F1 Score of the classifier is: 0.6340640809443507
  - AUC for our classifier is: 0.7061249285305888

*Notebook:* Algoritmos de clasificación (Random forest + BTC\_ind\_trend\_ET\_fv).ipynb  
<https://drive.google.com/file/d/1UYNo9-nbtjwhx-jl--Mb-7EGz61okFUv/view?usp=sharing>

#### 8.2.8. *Algoritmos de clasificación (KNN + BTC\_ind\_trend\_ET\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo KNN (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran en el notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 1.0
  - % de aciertos sobre el set de evaluación es (accuracy): 0.6396604938271605
  - Precision Score of the classifier is: 0.6248153618906942
  - Recall Score of the classifier is: 0.6650943396226415
  - F1 Score of the classifier is: 0.6443259710586443
  - AUC for our classifier is: 0.7078783114160472

*Notebook:* Algoritmos de clasificación (KNN + BTC\_ind\_trend\_ET\_fv).ipynb

[https://drive.google.com/file/d/1HFiWT\\_vXYUGQaWb5-bMF5PLzzZOKi5Pj/view?usp=sharing](https://drive.google.com/file/d/1HFiWT_vXYUGQaWb5-bMF5PLzzZOKi5Pj/view?usp=sharing)

#### 8.2.9. *Algoritmos de clasificación (Logistic Regression + BTC\_ind\_trend\_ET\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Logistic Regression (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran en el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.6732804232804233
  - % de aciertos sobre el set de evaluación es (accuracy): 0.6628086419753086
  - Accuracy of the classifier is: 0.6628086419753086
  - Precision Score of the classifier is: 0.7112526539278131
  - Recall Score of the classifier is: 0.5267295597484277
  - F1 Score of the classifier is: 0.6052393857271906
  - AUC for our classifier is: 0.736878216123499

*Notebook:* Algoritmos de clasificación (Logistic Regression + BTC\_ind\_trend\_ET\_fv).ipynb

[https://drive.google.com/file/d/1\\_p0LmLBNkQrTybgKyWfzJdDaEGspFe3/view?usp=sharing](https://drive.google.com/file/d/1_p0LmLBNkQrTybgKyWfzJdDaEGspFe3/view?usp=sharing)

#### 8.2.10. *Algoritmos de clasificación (SVM + BTC\_ind\_trend\_ET\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo SVM (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):



- % de aciertos sobre el set de evaluación para train: 0.6953125
- % de aciertos sobre el set de evaluación para test (accuracy): 0.6296296296296297
- Precision Score of the classifier is: 0.6676557863501483
- Recall Score of the classifier is: 0.5196304849884527
- F1 Score of the classifier is: 0.5844155844155844

*Notebook:* Algoritmos de clasificación (SVM + BTC\_ind\_trend\_ET\_fv).ipynb

<https://drive.google.com/file/d/1oKZQgnb4mSZvERKNVL-yHXt4ceFwsFC2/view?usp=sharing>

#### 8.2.11. *Algoritmos de clasificación (Decision Tree + BTC\_ind\_mom\_ET\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Decision Tree (sklearn). En unos modelos se preservaron los NaN y en otros se eliminaron estos registros (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (Decision Tree):
  - % de aciertos sobre el set de entrenamiento: 0.6720905553590378
  - % de aciertos sobre el set de evaluación (accuracy): 0.6710439921208142
  - Precision Score of the classifier is: 0.6841294298921418
  - Recall Score of the classifier is: 0.6
  - F1 Score of the classifier is: 0.6393088552915767
  - AUC for our classifier is: 0.7009733871802837

*Notebook:* Algoritmo de clasificación (Decision Tree + BTC\_ind\_mom\_ET\_fv).ipynb

[https://drive.google.com/file/d/1AkjI\\_QfMi\\_yG6rXYOih-mYAMmpGj6aSR/view?usp=sharing](https://drive.google.com/file/d/1AkjI_QfMi_yG6rXYOih-mYAMmpGj6aSR/view?usp=sharing)

#### 8.2.12. *Algoritmos de clasificación (Random forest + BTC\_ind\_mom\_ET\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Random forest (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.6571428571428571
  - % de aciertos sobre el set de evaluación (accuracy): 0.660536398467433
  - Precision Score of the classifier is: 0.6271409749670619
  - Recall Score of the classifier is: 0.7484276729559748
  - F1 Score of the classifier is: 0.6824372759856631
  - AUC for our classifier is: 0.7308147897453252

*Notebook:* Algoritmos de clasificación (Random forest + BTC\_ind\_mom\_ET\_fv).ipynb

<https://drive.google.com/file/d/1F-lByjyh-OuOZW9lB9bAS2DDs95nxSM3/view?usp=sharing>

#### 8.2.13. Algoritmos de clasificación (KNN + BTC\_ind\_mom\_ET\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo KNN (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran en el notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 1.0
  - % de aciertos sobre el set de evaluación es: 0.5417624521072797
  - Precision Score of the classifier is: 0.5279411764705882
  - Recall Score of the classifier is: 0.5644654088050315
  - F1 Score of the classifier is: 0.5455927051671733
  - AUC for our classifier is: 0.5599740530783766

*Notebook:* Algoritmos de clasificación (KNN + BTC\_ind\_mom\_ET\_fv).ipynb

<https://drive.google.com/file/d/1F-lByjyh-OuOZW9lB9bAS2DDs95nxSM3/view?usp=sharing>

#### 8.2.14. Algoritmos de clasificación (Logistic Regression + BTC\_ind\_mom\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Logistic Regression (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran en el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.6357963875205255
  - % de aciertos sobre el set de evaluación es (accuracy): 0.6444444444444445
  - Precision Score of the classifier is: 0.6369426751592356
  - Recall Score of the classifier is: 0.6289308176100629
  - F1 Score of the classifier is: 0.6329113924050632
  - AUC for our classifier is: 0.7142712769457842

*Notebook:* Algoritmos de clasificación (Logistic Regression + BTC\_ind\_mom\_ET\_fv).ipynb

<https://drive.google.com/file/d/1F-lByjyh-OuOZW9lB9bAS2DDs95nxSM3/view?usp=sharing>

#### 8.2.15. Algoritmos de clasificación (SVM + BTC\_ind\_mom\_ET\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo SVM (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de evaluación para train: 0.7419540229885058
  - % de aciertos sobre el set de evaluación para test: 0.632183908045977

- Accuracy of the classifier is: 0.632183908045977
- Precision Score of the classifier is: 0.6336405529953917
- Recall Score of the classifier is: 0.6307339449541285
- F1 Score of the classifier is: 0.632183908045977

*Notebook:* Algoritmos de clasificación (SVM + BTC\_ind\_mom\_ET\_fv).ipynb

<https://drive.google.com/file/d/1oKZQgnb4mSZvERKNVL-yHXt4ceFwsFC2/view?usp=sharing>

#### 8.2.16. *Algoritmos de clasificación (Decision Tree + BTC\_US\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Decision Tree (sklearn). En unos modelos se preservaron los NaN y en otros se eliminaron estos registros (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (Decision Tree ( )):
  - % de aciertos sobre el set de entrenamiento: 0.6889269406392694
  - % de aciertos sobre el set de evaluación (accuracy): 0.6883561643835616
  - Precision Score of the classifier is: 0.7019704433497537
  - Recall Score of the classifier is: 0.6521739130434783
  - F1 Score of the classifier is: 0.6761565836298932
  - AUC for our classifier is: 0.7265889294892177

*Notebook:* Algoritmo de clasificación (Decision Tree + BTC\_US\_fv).ipynb

[https://drive.google.com/file/d/1PnHW89DzookC6syIwu4144JR\\_3toQZ6e/view?usp=sharing](https://drive.google.com/file/d/1PnHW89DzookC6syIwu4144JR_3toQZ6e/view?usp=sharing)

#### 8.2.17. *Algoritmos de clasificación (Random forest + BTC\_US\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Random forest (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.6894977168949772
  - % de aciertos sobre el set de evaluación (accuracy): 0.6971080669710806
  - Precision Score of the classifier is: 0.683076923076923
  - Recall Score of the classifier is: 0.6981132075471698
  - F1 Score of the classifier is: 0.6905132192846034
  - AUC for our classifier is: 0.7540154635349993

*Notebook:* Algoritmos de clasificación (Random forest + BTC\_US\_fv).ipynb

<https://drive.google.com/file/d/10X04KT9HuSxiCRZxpq60dOSp0SvJN8-O/view?usp=sharing>

#### 8.2.18. Algoritmos de clasificación (KNN + BTC\_US\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo KNN (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran en el notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 1.0
  - % de aciertos sobre el set de evaluación es (accuracy): 0.58675799086758
  - Precision Score of the classifier is: 0.5885797950219619
  - Recall Score of the classifier is: 0.6320754716981132
  - F1 Score of the classifier is: 0.6095526914329036
  - AUC for our classifier is: 0.6482834270236173

*Notebook:* Algoritmos de clasificación (KNN + BTC\_US\_fv).ipynb

<https://drive.google.com/file/d/1AKvrw5qKrvfnaYtRjUefsZsY5JlyHisD/view?usp=sharing>

#### 8.2.19. Algoritmos de clasificación (Logistic Regression + BTC\_US\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Logistic Regression (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran en el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.6441617742987606
  - % de aciertos sobre el set de evaluación es (accuracy): 0.6468797564687976
  - Precision Score of the classifier is: 0.6191135734072022
  - Recall Score of the classifier is: 0.7028301886792453
  - F1 Score of the classifier is: 0.658321060382916
  - AUC for our classifier is: 0.712278065342016

*Notebook:* Algoritmos de clasificación (Logistic Regression + BTC\_US\_fv).ipynb

[https://drive.google.com/file/d/11Yqd6\\_CRYNbQ7GPUTpo0\\_V7kHPGZbIBz/view?usp=sharing](https://drive.google.com/file/d/11Yqd6_CRYNbQ7GPUTpo0_V7kHPGZbIBz/view?usp=sharing)

#### 8.2.20. Algoritmos de clasificación (SVM + BTC\_US\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo SVM (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ( )):
  - % de aciertos sobre el set de evaluación para train: 0.8125
  - % de aciertos sobre el set de evaluación para test (accuracy): 0.6472602739726028

- Precision Score of the classifier is: 0.6722488038277512
- Recall Score of the classifier is: 0.6203090507726269
- F1 Score of the classifier is: 0.6452353616532721

*Notebook:* Algoritmos de clasificación (SVM + BTC\_US\_fv).ipynb

[https://drive.google.com/file/d/10uDQaQB6aiU5X8fzLyiTslT\\_MSbG0KhI/view?usp=sharing](https://drive.google.com/file/d/10uDQaQB6aiU5X8fzLyiTslT_MSbG0KhI/view?usp=sharing)

#### 8.2.21. *Algoritmos de clasificación (Decision Tree + BTC\_multiple\_indicators\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Decision Tree (sklearn). En unos modelos se preservaron los NaN y en otros se eliminaron estos registros (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (Decision Tree ()):
  - % de aciertos sobre el set de entrenamiento: 0.6873219373219374
  - % de aciertos sobre el set de evaluación (accuracy): 0.6851851851851852
  - Precision Score of the classifier is: 0.6743554952510177
  - Recall Score of the classifier is: 0.6780354706684857
  - F1 Score of the classifier is: 0.6761904761904762
  - AUC for our classifier is: 0.7297730150418472
  - El modelo de HistogramGradientBoostClassifier () fue:
    - % de aciertos sobre el set de entrenamiento: 0.9485174323884001
    - % de aciertos sobre el set de evaluación (accuracy): 0.7120060790273556
    - Precision Score of the classifier is: 0.6859375
    - Recall Score of the classifier is: 0.7115072933549432
    - F1 Score of the classifier is: 0.6984884645982498
    - AUC for our classifier is: 0.7969569864798752

*Notebook:* Algoritmo de clasificación (Decision Tree + BTC\_multiple\_indicators\_fv).ipynb

<https://drive.google.com/file/d/1WpDOY8KjQNBbhJCYSzBQ7Lfe8J9nATN/view?usp=sharing>

#### 8.2.22. *Algoritmos de clasificación (Random forest + BTC\_multiple\_indicators\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Random forest (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ()):
  - % de aciertos sobre el set de entrenamiento: 0.691468253968254
  - % de aciertos sobre el set de evaluación: 0.6967592592592593
  - Precision Score of the classifier is: 0.68
  - Recall Score of the classifier is: 0.7216981132075472
  - F1 Score of the classifier is: 0.700228832951945
  - AUC for our classifier is: 0.7604381074899944

*Notebook:* Algoritmos de clasificación (Random forest + BTC\_multiple\_indicators\_fv).ipynb  
[https://drive.google.com/file/d/1c8WWoBa5\\_ZaE182XNhGh2UkrgXVIZdi7/view?usp=sharing](https://drive.google.com/file/d/1c8WWoBa5_ZaE182XNhGh2UkrgXVIZdi7/view?usp=sharing)

#### 8.2.23. Algoritmos de clasificación (KNN + BTC\_multiple\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo KNN (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran en el notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.5783730158730159
  - % de aciertos sobre el set de evaluación es (accuracy): 0.5625
  - Precision Score of the classifier is: 0.5642458100558659
  - Recall Score of the classifier is: 0.47641509433962265
  - F1 Score of the classifier is: 0.5166240409207161
  - AUC for our classifier is: 0.5953187535734705

*Notebook:* Algoritmos de clasificación (KNN + BTC\_multiple\_indicators\_fv).ipynb  
<https://drive.google.com/file/d/1dZt62XrHkJG2XOHMxJB2Wg4XG8FqvB9C/view?usp=sharing>

#### 8.2.24. Algoritmos de clasificación (Logistic Regression + BTC\_multiple\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Logistic Regression (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran en el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.644510582010582
  - % de aciertos sobre el set de evaluación es (accuracy): 0.6620370370370371
  - Precision Score of the classifier is: 0.6426512968299711
  - Recall Score of the classifier is: 0.7012578616352201
  - F1 Score of the classifier is: 0.6706766917293233
  - AUC for our classifier is: 0.7396607585286831

*Notebook:* Algoritmos de clasificación (Logistic Regression + BTC\_multiple\_indicators\_fv).ipynb  
<https://drive.google.com/file/d/1QVjvRvFlxN7Y-SoUszkVdIIZ2Vtp8DVk/view?usp=sharing>

#### 8.2.25. Algoritmos de clasificación (SVM + BTC\_multiple\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo SVM (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):

- % de aciertos sobre el set de evaluación para train: 0.6472800925925926
- % de aciertos sobre el set de evaluación para test (accuracy): 0.6192129629629629
- Precision Score of the classifier is: 0.667741935483871
- Recall Score of the classifier is: 0.4780600461893764
- F1 Score of the classifier is: 0.557200538358008

*Notebook:* Algoritmos de clasificación (SVM + BTC\_multiple\_indicators\_fv).ipynb  
[https://drive.google.com/file/d/1XWSX6V3YO2PMWY1hkG\\_mjLVMrS-A4Z37/view?usp=sharing](https://drive.google.com/file/d/1XWSX6V3YO2PMWY1hkG_mjLVMrS-A4Z37/view?usp=sharing)

#### 8.2.26. Algoritmos de clasificación (Decision Tree + BTC\_pattern\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Decision Tree (sklearn). En unos modelos se preservaron los NaN y en otros se eliminaron estos registros (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (Decision Tree ()):
  - % de aciertos sobre el set de entrenamiento: 0.6718417047184171
  - % de aciertos sobre el set de evaluación (accuracy): 0.6703196347031963
  - Precision Score of the classifier is: 0.6776859504132231
  - Recall Score of the classifier is: 0.6153846153846154
  - F1 Score of the classifier is: 0.6450344149459193
  - AUC for our classifier is: 0.6906518531377485
- El mejor modelo de Histogram Gradient Boosting Classifier con varios patrones fue:
  - % de aciertos sobre el set de entrenamiento: 0.8963831867057673
  - % de aciertos sobre el set de evaluación (accuracy): 0.6808510638297872
  - Precision Score of the classifier is: 0.6513056835637481
  - Recall Score of the classifier is: 0.6871961102106969
  - F1 Score of the classifier is: 0.668769716088328
  - AUC for our classifier is: 0.7587477364050983

*Notebook:* Algoritmo de clasificación (Decision Tree + BTC\_pattern\_indicators\_fv).ipynb  
[https://drive.google.com/file/d/1DEy4xmwZaIpeIrrlY\\_Kk5bG82fg\\_cIEC/view?usp=sharing](https://drive.google.com/file/d/1DEy4xmwZaIpeIrrlY_Kk5bG82fg_cIEC/view?usp=sharing)

#### 8.2.27. Algoritmos de clasificación (Random forest + BTC\_pattern\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Random forest (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ()):
  - % de aciertos sobre el set de entrenamiento: 0.6797129810828441
  - % de aciertos sobre el set de evaluación (accuracy): 0.6666666666666666
  - Precision Score of the classifier is: 0.6464497041420119
  - Recall Score of the classifier is: 0.6871069182389937
  - F1 Score of the classifier is: 0.6661585365853658



- AUC for our classifier is: 0.7293985733103282

*Notebook:* Algoritmos de clasificación (Random forest + BTC\_pattern\_indicators\_fv).ipynb  
<https://drive.google.com/file/d/1yZjVypgfdlNt26bYRXZvFL7lVnI5u0nG/view?usp=sharing>

#### 8.2.28. *Algoritmos de clasificación (KNN + BTC\_pattern\_indicators\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo KNN (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran en el notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ()):
  - % de aciertos sobre el set de entrenamiento: 0.7840834964122635
  - % de aciertos sobre el set de evaluación es (accuracy): 0.5616438356164384
  - Precision Score of the classifier is: 0.567651632970451
  - Recall Score of the classifier is: 0.5738993710691824
  - F1 Score of the classifier is: 0.5707584050039093
  - AUC for our classifier is: 0.624577929908536

*Notebook:* Algoritmos de clasificación (KNN + BTC\_pattern\_indicators\_fv).ipynb  
[https://drive.google.com/file/d/1m-YfNDUL4XvvtZN0\\_zoK9mASp4hWz0H3/view?usp=sharing](https://drive.google.com/file/d/1m-YfNDUL4XvvtZN0_zoK9mASp4hWz0H3/view?usp=sharing)

#### 8.2.29. *Algoritmos de clasificación (Logistic Regression + BTC\_pattern\_indicators\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo Logistic Regression (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran en el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ()):
  - % de aciertos sobre el set de entrenamiento: 0.6395955642530985
  - % de aciertos sobre el set de evaluación es (accuracy): 0.6537290715372908
  - Precision Score of the classifier is: 0.6630630630630631
  - Recall Score of the classifier is: 0.5786163522012578
  - F1 Score of the classifier is: 0.6179680940386231
  - AUC for our classifier is: 0.7115081352850596

*Notebook:* Algoritmos de clasificación (Logistic Regression + BTC\_pattern\_indicators\_fv).ipynb  
[https://drive.google.com/file/d/1uzf6wJR0c7uwcJiFx5zrlhKOItOvwq\\_7/view?usp=sharing](https://drive.google.com/file/d/1uzf6wJR0c7uwcJiFx5zrlhKOItOvwq_7/view?usp=sharing)

#### 8.2.30. *Algoritmos de clasificación (SVM + BTC\_pattern\_indicators\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo SVM (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ()):
  - % de aciertos sobre el set de evaluación para train: 0.6349885844748858
  - % de aciertos sobre el set de evaluación para test: 0.6038812785388128
  - Precision Score of the classifier is: 0.6790540540540541
  - Recall Score of the classifier is: 0.44370860927152317
  - F1 Score of the classifier is: 0.5367156208277704

*Notebook:* Algoritmos de clasificación (SVM + BTC\_pattern\_indicators\_fv).ipynb  
[https://drive.google.com/file/d/1-DINfl\\_OD1Z-49GEtw-vqrPjyZn8pni-/view?usp=sharing](https://drive.google.com/file/d/1-DINfl_OD1Z-49GEtw-vqrPjyZn8pni-/view?usp=sharing)

#### 8.2.31. Algoritmos de clasificación (Decision Tree + BTC\_various\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Decision Tree (sklearn). En unos modelos se preservaron los NaN y en otros se eliminaron estos registros (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (Decision Tree ()):
  - % de aciertos sobre el set de entrenamiento: 0.6872521246458924
  - % de aciertos sobre el set de evaluación (accuracy): 0.6867572156196944
  - Precision Score of the classifier is: 0.6949924127465857
  - Recall Score of the classifier is: 0.731629392971246
  - F1 Score of the classifier is: 0.7128404669260701
  - AUC for our classifier is: 0.7052093462054915
- El mejor modelo de Histogram Gradient Boosting Classifier con varios patrones fue:
  - % de aciertos sobre el set de entrenamiento: 0.9767329132331556
  - % de aciertos sobre el set de evaluación: 0.6384180790960452
  - Precision Score of the classifier is: 0.6788008565310493
  - Recall Score of the classifier is: 0.6509240246406571
  - F1 Score of the classifier is: 0.6645702306079665
  - AUC for our classifier is: 0.6957425732357888

*Notebook:* Algoritmo de clasificación (Decision Tree + BTC\_various\_indicators\_fv).ipynb  
<https://drive.google.com/file/d/1UwltypsN46Ev9V-rtqn8WyzsfTRBwrz0/view?usp=sharing>

#### 8.2.32. Algoritmos de clasificación (Random forest + BTC\_various\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Random forest (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ()):
  - % de aciertos sobre el set de entrenamiento: 0.6854368932038835
  - % de aciertos sobre el set de evaluación (accuracy): 0.6432616081540203

- Precision Score of the classifier is: 0.6751740139211136
- Recall Score of the classifier is: 0.6244635193133047
- F1 Score of the classifier is: 0.6488294314381271
- AUC for our classifier is: 0.6827507950720968

*Notebook:* Algoritmos de clasificación (Random forest + BTC\_various\_indicators\_fv).ipynb  
<https://drive.google.com/file/d/1UwltvpsN46Ev9V-rtqn8WyzsfTRBwrz0/view?usp=sharing>

### 8.2.33. Algoritmos de clasificación (KNN + BTC\_various\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo KNN (sklearn). Se eliminaron los registros NaN en las variables del data set (la explicación se encuentran en el notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando GridSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.5902912621359223
  - % de aciertos sobre el set de evaluación es (accuracy): 0.579841449603624
  - Precision Score of the classifier is: 0.5849731663685152
  - Recall Score of the classifier is: 0.7017167381974249
  - F1 Score of the classifier is: 0.6380487804878049
  - AUC for our classifier is: 0.6017898127849651

*Notebook:* Algoritmos de clasificación (KNN + BTC\_various\_indicators\_fv).ipynb  
<https://drive.google.com/file/d/1tn7iQQCTH-LFcNbYRgkvCVq2eP0myaP9/view?usp=sharing>

### 8.2.34. Algoritmos de clasificación (Logistic Regression + BTC\_various\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo Logistic Regression (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran en el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de entrenamiento: 0.7067961165048544
  - % de aciertos sobre el set de evaluación es (accuracy): 0.6636466591166478
  - Precision Score of the classifier is: 0.6825053995680346
  - Recall Score of the classifier is: 0.6781115879828327
  - F1 Score of the classifier is: 0.6803013993541444
  - AUC for our classifier is: 0.7427620135651136

*Notebook:* Algoritmos de clasificación (Logistic Regression + BTC\_various\_indicators\_fv).ipynb  
[https://drive.google.com/file/d/1AYj479p2rIEgc4Iyp6xIelv-wXs\\_L9Gv/view?usp=sharing](https://drive.google.com/file/d/1AYj479p2rIEgc4Iyp6xIelv-wXs_L9Gv/view?usp=sharing)

#### 8.2.35. Algoritmos de clasificación (SVM + BTC\_various\_indicators\_fv).ipynb

Se llevó adelante diversos modelos supervisado de ML de tipo SVM (sklearn). Se eliminaron los registros NaN del data set (la explicación se encuentran sobre el mismo notebook). Al final del notebook se puede observar la manipulación de hiperparámetros a fin de mejorar las métricas de predicción:

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV ( )):
  - % de aciertos sobre el set de evaluación para train: 0.7200509770603228
  - % de aciertos sobre el set de evaluación para test (accuracy): 0.66553480475382
  - Precision Score of the classifier is: 0.6656441717791411
  - Recall Score of the classifier is: 0.7114754098360656
  - F1 Score of the classifier is: 0.6877971473851031

*Notebook:* Algoritmos de clasificación (SVM + BTC\_various\_indicators\_fv).ipynb

<https://drive.google.com/file/d/1ZSEA8hKkTEa9lQ0G0uS7dYjCna7utIto/view?usp=sharing>

### 8.3. Modelos de boosting

Dado que en 8.2, salvo con excepción del modelo de boosting HistGradientBoostingClassifier (), solo fueron diseñados y ejecutados modelos simples de machine learning, en este apartado se llevan ejecutan modelos de Adaboost, Gradient Boosting, XG Boost, y Light GBM para solo algunas de las bases de datos (elegidas por criterios de mayor valor de correlación de variables con los features “Price”, “Percentage\_difference”, y “Target”).

#### 8.3.1. Algoritmos de boosting (BTC\_metals\_fv).ipynb

Se llevaron adelante diversos modelos supervisado de ML de tipo boosting.

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV () en un algoritmo de HistGradientBoostingClassifier()):
  - % de aciertos sobre el set de evaluación para test: 0. 0.6686542891990784
  - Accuracy of the classifier is: 0.6706281833616299
  - Precision Score of the classifier is: 0.6656441717791411
  - Recall Score of the classifier is: 0.7185430463576159
  - F1 Score of the classifier is: 0.6910828025477708
  - AUC for our classifier is: 0.7368588042550246

*Notebook:* Algoritmos de boosting (BTC\_metals\_fv).ipynb

[https://drive.google.com/file/d/13uUfBMDIQiWvFIH\\_jvzZzX3LYwXzMKKP/view?usp=sharing](https://drive.google.com/file/d/13uUfBMDIQiWvFIH_jvzZzX3LYwXzMKKP/view?usp=sharing)

#### 8.3.2. Algoritmos de boosting (BTC\_ind\_trend\_ET\_fv).ipynb

Se llevaron adelante diversos modelos supervisado de ML de tipo boosting.

- El mejor modelo encontrado fue el siguiente (utilizando LeaveOneOut() en un algoritmo de HistGradientBoostingClassifier()):
  - % de aciertos sobre el set de evaluación para test: 0.680

*Notebook:* Algoritmos de boosting (BTC\_ind\_trend\_ET\_fv).ipynb

<https://drive.google.com/file/d/1oBQCm4clft2NnR7fzLyz19a9b5n1jqzW/view?usp=sharing>

### 8.3.3. *Algoritmos de boosting (BTC\_various\_indicators\_fv).ipynb*

Se llevó adelante diversos modelos supervisado de ML de tipo boosting.

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV () en un algoritmo de GradientBoostingClassifier ()):
  - Accuracy of the classifier is: 0.6494565217391305
  - Precision Score of the classifier is: 0.6798029556650246
  - Recall Score of the classifier is: 0.6831683168316832
  - F1 Score of the classifier is: 0.6814814814814815
  - AUC for our classifier is: 0.7015872897530717

*Notebook:* Algoritmos de boosting (BTC\_various\_indicators\_fv).ipynb

<https://drive.google.com/file/d/1hhh40WhJq9rupF2JQrzR13XzTZXRoezk/view?usp=sharing>

## 8.4. Mejor modelo

Se buscó hypertunear el mejor modelo de ML.

- El mejor modelo encontrado fue el siguiente (utilizando RandomizedSearchCV () en un algoritmo de HistGradientBoostingClassifier ()):
  - Accuracy of the classifier is: 0.729483282674772
  - Precision Score of the classifier is: 0.7115072933549432
  - Recall Score of the classifier is: 0.7115072933549432
  - F1 Score of the classifier is: 0.7115072933549431
  - AUC for our classifier is: 0.8009775483847034

*Notebook:* Hypertunning del mejor modelo (BTC\_multiple\_indicators).ipynb

<https://drive.google.com/file/d/1-xpmxsBEuQhFRxMC3eQUX8tIgyDvhZjS/view?usp=sharing>

## 9. Metadata

La metadata del proyecto se puede encontrar en el archivo: “Metadata.docx” (incluido versiones de librerías utilizadas). Link al repositorio de Google Drive:

[https://docs.google.com/document/d/12yfdnVTwmsX9Y6gXGe3\\_Ew-xuha35iBk/edit?usp=sharing&ouid=101791130377378188501&rtfpof=true&sd=true](https://docs.google.com/document/d/12yfdnVTwmsX9Y6gXGe3_Ew-xuha35iBk/edit?usp=sharing&ouid=101791130377378188501&rtfpof=true&sd=true)