

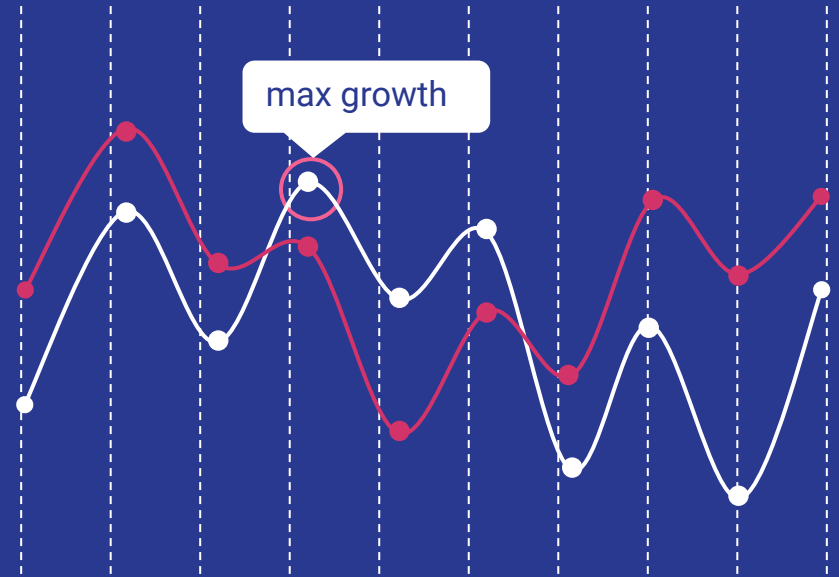
# Selección de cripto activos para una cartera de inversión

2da presentación

**Nicola Mori**  
**Ezequiel Kinigsberg**

*Mentora - María de los Ángeles Martínez*

*Diplomatura en Ciencia de Datos, Aprendizaje Automático  
y sus Aplicaciones*



# Un poco de contexto y objetivo

- La pandemia del COVID trajo aparejado un ascenso en la popularidad de los cripto activos y un cuestionamiento a las monedas tradicionales
- El aumento de la demanda impulsó los precios de los cripto activos, que tocaron valores récord a mediados de mayo y se mantienen muy por encima de los valores pre pandemia.
- Es un mercado muy volátil y la diversificación es clave para reducir el riesgo y definir perfiles de inversión.
- Nos propusimos investigar cómo elaborar una cartera de monedas para reducir el riesgo y maximizar los rendimientos, con rebalanceos semanales o mensuales.

# Exploración

# Dataset utilizado

- 16 criptoactivos.
- Se cuenta con las variables *open*, *high*, *low*, *close* y *volumen* por hora.
- El período analizado va desde el *2018-01-01 12am* hasta el *2021-05-16 15pm*.
- No todos los activos están presentes desde el inicio.
- Hay pocos datos faltantes (menos del 0,4% de los datos): Se les aplicó backwards fill.

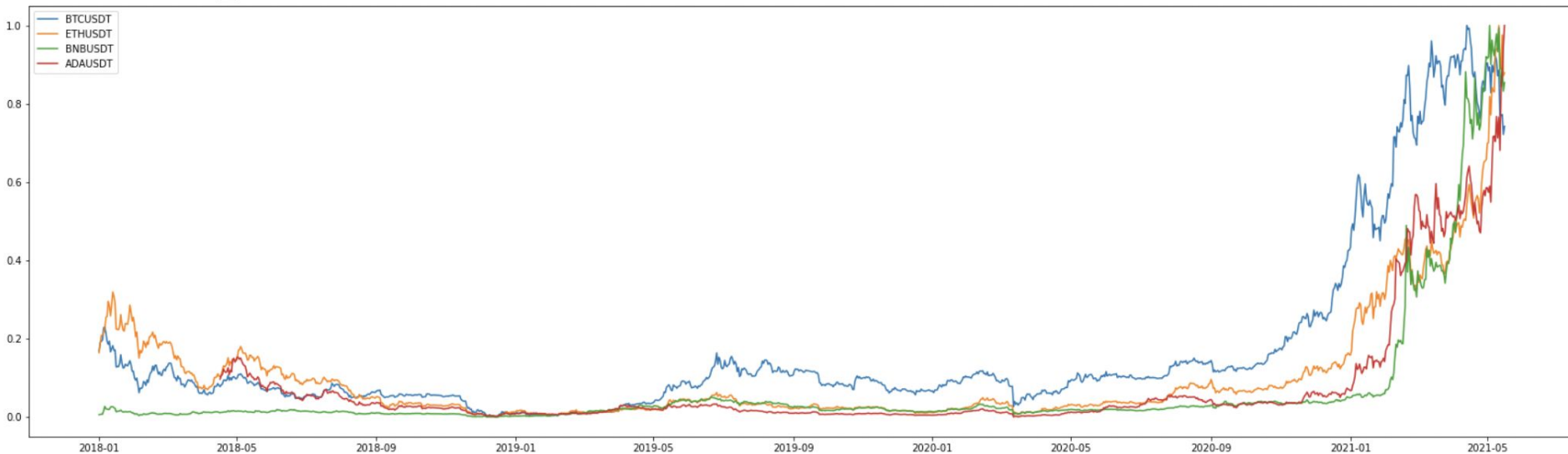
YMD	2018-01	2018-02	2018-06	2018-07	2018-10	2018-11	2019-03	2019-05	2019-06	2019-08	2019-11	2020-02	2020-03	2020-04	2020-06	2020-11	2020-12	2021-02	2021-03	2021-04
ETHUSDT_NaN	1	32	11	7	3	7	6	10	1	8	4	6	1	2	3	1	5	2	1	5

# Bitcoin como lider

AAVEUSD	1	0.9	0.76	0.76	0.9	0.94	0.84	0.92	0.7	0.54	0.88	0.92	0.87	0.86	0.74	0.76
ADAUSD	0.9	1	0.89	0.93	0.92	0.93	0.96	0.97	0.86	0.7	0.94	0.93	0.94	0.91	0.89	0.91
BCHUSD	0.76	0.89	1	0.92	0.82	0.91	0.95	0.81	0.94	0.88	0.96	0.92	0.95	0.96	0.95	0.95
BNBUSD	0.76	0.93	0.92	1	0.83	0.91	0.95	0.87	0.91	0.78	0.92	0.89	0.91	0.93	0.97	0.96
BTCUSD	0.9	0.92	0.82	0.83	1	0.85	0.9	0.96	0.71	0.51	0.91	0.92	0.95	0.83	0.79	0.88
COMPUSD	0.94	0.93	0.91	0.91	0.85	1	0.93	0.91	0.85	0.75	0.95	0.93	0.91	0.97	0.9	0.9
DASHUSD	0.84	0.96	0.95	0.95	0.9	0.93	1	0.93	0.91	0.77	0.95	0.94	0.96	0.93	0.94	0.95
DOTUSD	0.92	0.97	0.81	0.87	0.96	0.91	0.93	1	0.76	0.57	0.9	0.94	0.91	0.86	0.81	0.85
EOSUSD	0.7	0.86	0.94	0.91	0.71	0.85	0.91	0.76	1	0.9	0.89	0.85	0.86	0.87	0.92	0.88
ETCUSD	0.54	0.7	0.88	0.78	0.51	0.75	0.77	0.57	0.9	1	0.79	0.7	0.73	0.82	0.83	0.77
ETHUSD	0.88	0.94	0.96	0.92	0.91	0.95	0.95	0.9	0.89	0.79	1	0.96	0.98	0.97	0.93	0.96
LINKUSD	0.92	0.93	0.92	0.89	0.92	0.93	0.94	0.94	0.85	0.7	0.96	1	0.96	0.95	0.91	0.94
LTCUSD	0.87	0.94	0.95	0.91	0.95	0.91	0.96	0.91	0.86	0.73	0.98	0.96	1	0.93	0.91	0.95
MKRUSD	0.86	0.91	0.96	0.93	0.83	0.97	0.93	0.86	0.87	0.82	0.97	0.95	0.93	1	0.95	0.95
NEOUSD	0.74	0.89	0.95	0.97	0.79	0.9	0.94	0.81	0.92	0.83	0.93	0.91	0.91	0.95	1	0.96
XMRUSD	0.76	0.91	0.95	0.96	0.88	0.9	0.95	0.85	0.88	0.77	0.96	0.94	0.95	0.95	0.96	1
AAVEUSD																
ADAUSD																
BCHUSD																
BNBUSD																
BTCUSD																
COMPUSD																
DASHUSD																
DOTUSD																
EOSUSD																
ETCUSD																
ETHUSD																
LINKUSD																
LTCUSD																
MKRUSD																
NEOUSD																
XMRUSD																

- Existe en general una alta correlación de precios entre los distintos cripto activos
- En la mayoría de los casos las monedas copian la tendencia del Bitcoin, la moneda con mayor capitalización y popularidad
- Esto representa una dificultad cuando hablamos de diversificar

# Bitcoin como lider

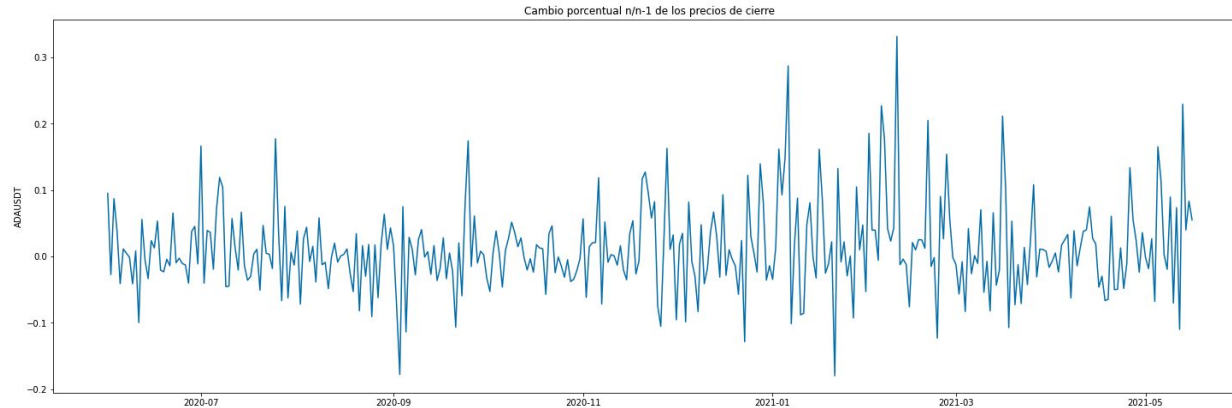
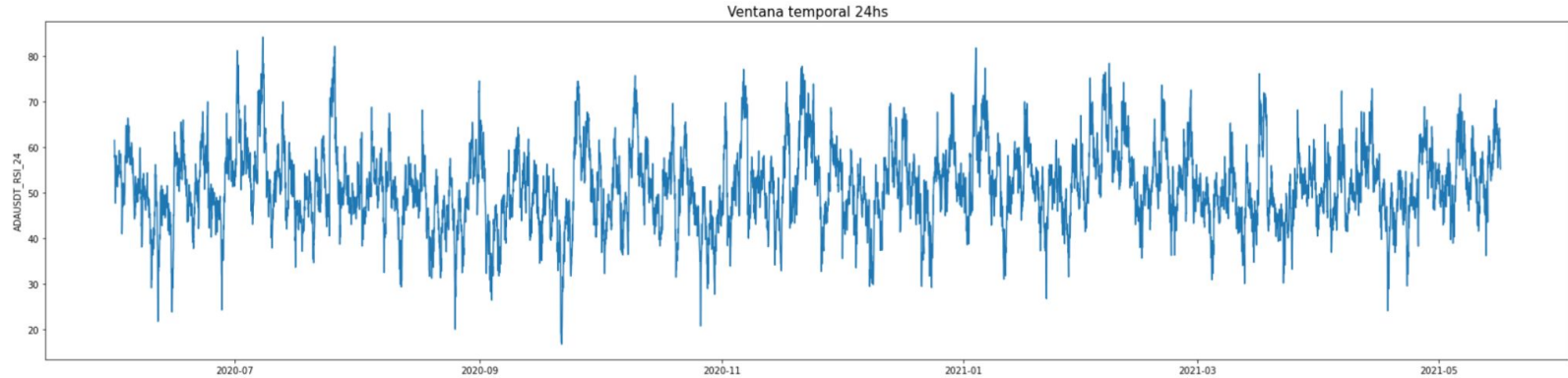


# Indicadores financieros y features

A partir de los datos de precio y volumen se elaboraron una serie de indicadores financieros y features:

- Media móvil: mide el precio promedio en una ventana determinada de tiempo.
- Media móvil exponencial (EMA): mide el precio promedio en una ventana determinada de tiempo, dándole más peso a los datos de precios más actuales.
- Índice direccional promedio (ADX): mide la fuerza de una tendencia, pero no su dirección real.
- Índice de fuerza relativa (RSI): mide la fuerza y dirección de la tendencia del mercado con valores entre 0 y 100.
- Cambio porcentual: cambio del precio de cierre respecto a la hora anterior. Se usó como punto de comparación del rendimiento de los distintos activos.

# Indicadores financieros y features



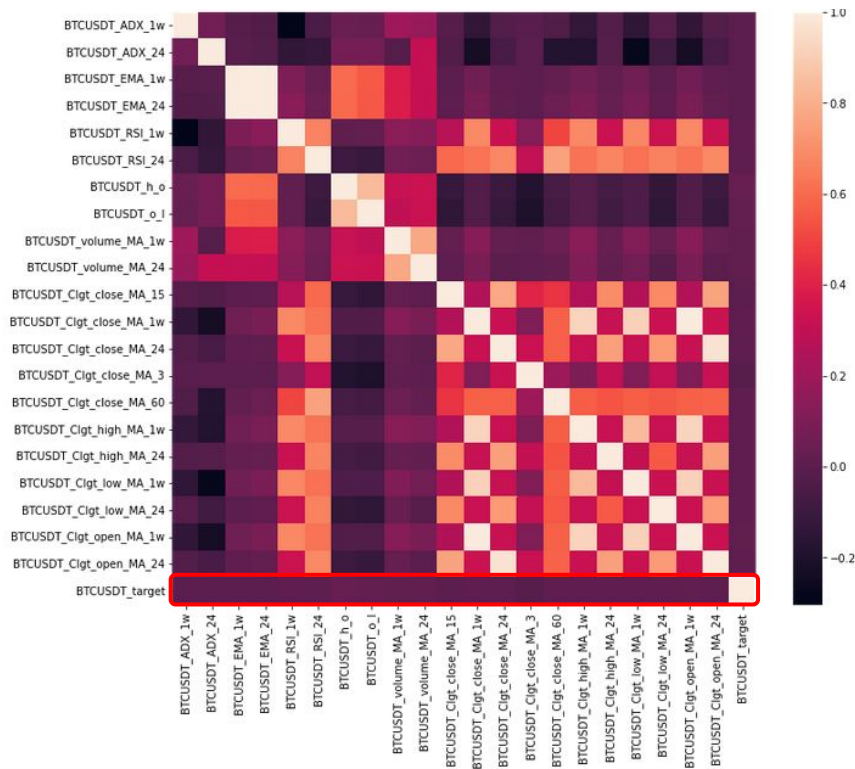


# Modelado

# Primer aproximación al modelado del problema

- Se modeló al problema como una regresión, tratando de predecir el cambio porcentual del precio de cierre de cada crypto activo, a la hora siguiente.
- Se creó un baseline, una regresión lineal y un árbol de decisión.
- Se utilizó CV para la búsqueda de hiperparámetros.
- Se calculó el error cuadrático medio, el error absoluto mediano y error absoluto medio, para los conjuntos de train y test.

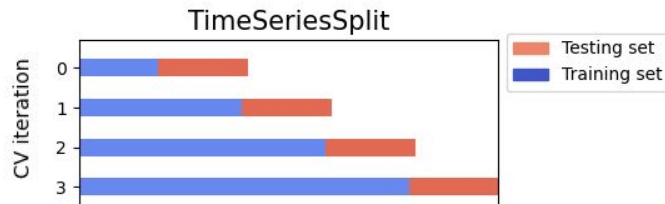
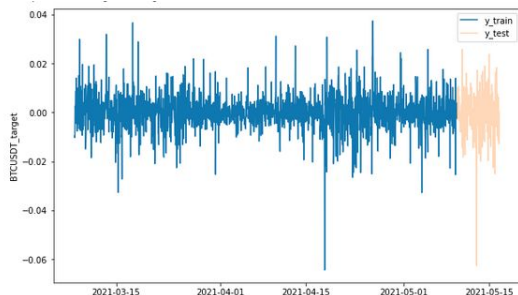
# Correlación entre features y target



- Se puede observar una muy baja correlación entre la variable objetivo y el resto de los features por lo que intuimos que una aproximación lineal no arrojará buenos resultados.

# Train-Test

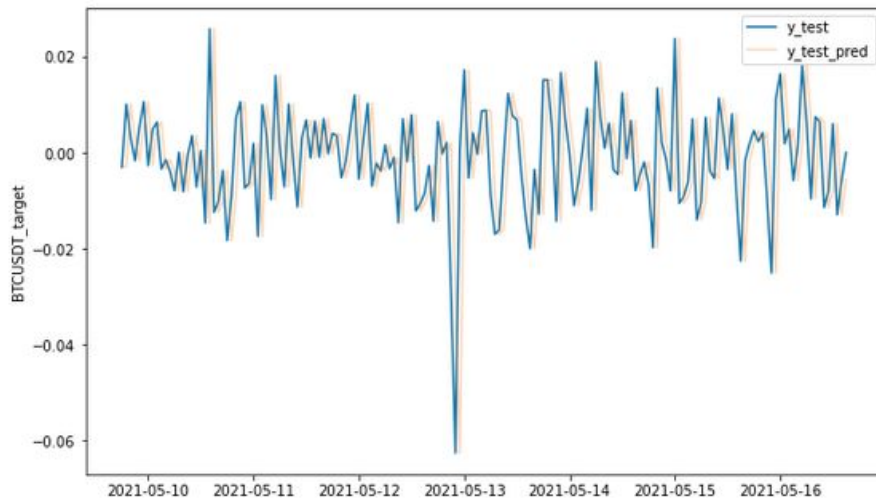
- Dado que se está trabajando con series de tiempo, es necesario tener especial cuidado al generar los conjuntos de train y test, ya que se debe mantener la coherencia temporal.
- Existen distintas estrategias para el armado donde se tenga a su vez algún tipo de validación.
  - Partición estandar: Se define una partición de train que abarca  $t_k:t_n$  y otra de test que comienza en  $t_{n+1}$ .
  - Walk forward validation: Se entrena con  $t_k:t_n$  particiones y se evalúa con una sola instancia  $t_{n+1}$ . Luego el conjunto de train se agranda para incluir  $t_{n+1}$  y test pasa a ser  $t_{n+2}$ , etc.
  - Símil K-Fold (*TimeSeriesSplit*): El conjunto de train son las primeras  $K$  particiones y el conjunto de test la partición  $K+1$ . Con cada nuevo entrenamiento el conjunto de train incrementa su tamaño en  $K$  y el conjunto de test se mantiene de tamaño fijo.



Fuente: [https://scikit-learn.org/stable/auto\\_examples/model\\_selection/plot\\_cv\\_indices.html](https://scikit-learn.org/stable/auto_examples/model_selection/plot_cv_indices.html)

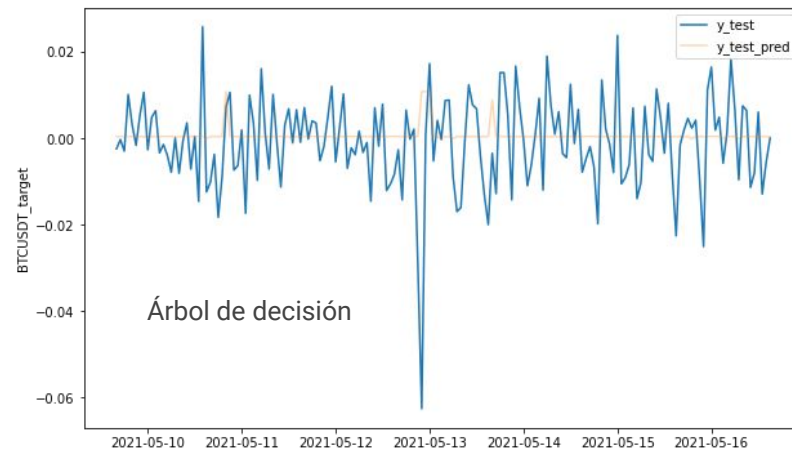
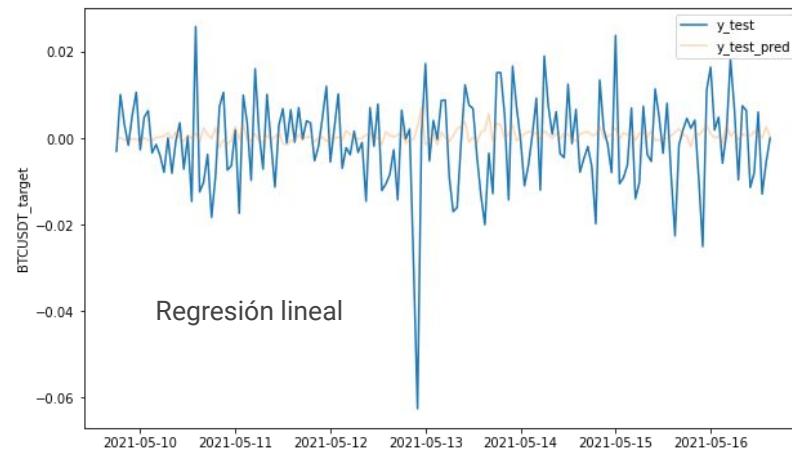
# Baseline

- Consiste en un modelo naive (*algoritmo persistente*) en donde se utiliza el valor en el intervalo previo  $t-1$  para predecir el output en el intervalo  $t$ .



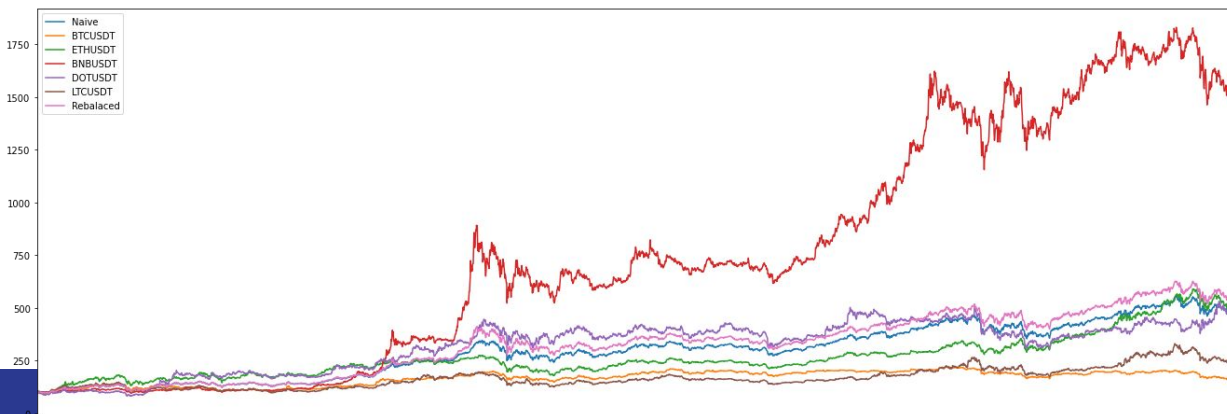
# Resultados: Bitcoin

	baseline	linreg	tree
train_mean_squared_error	0.000155	0.000074	0.000072
train_median_absolute_error	0.007506	0.004911	0.004888
train_mean_absolute_error	0.004488	0.002727	0.002740
test_mean_squared_error	0.000214	0.000116	0.000122
test_median_absolute_error	0.011250	0.008002	0.007988
test_mean_absolute_error	0.008553	0.006574	0.006564



# Segunda aproximación al modelado del problema (WIP)

- Construcción de algoritmo para la compra-venta de una selección de activos cada cierto tiempo  $T \rightarrow$  cartera de inversión.
- Se cambió el enfoque a un problema de clasificación donde se intentará predecir el retorno.
  - Se define un retorno  $R$  (5%), una ventana  $T$  (7 días) y un target binario (1|0).
    - Si el retorno en la ventana  $T$  es  $\geq R$ , se mantiene el activo.
    - Si el retorno en la ventana  $T$  es  $\leq -R$ , se vende el activo.
    - Si el retorno en la ventana  $T$  se encuentra entre  $\pm R$ , se mantiene el activo.
- Decision tree, Random Forest y XGboost.
- Periodo  $\geq 2020-01-01$



# Segunda aproximación al modelado del problema (WIP)

- Este proceso se itera, moviendo la ventana T para entrenar y evaluar/predecir cada crypto-activo e ir recomponiendo la cartera semana a semana en función de las predicciones.
- Dado que la predicción se puede realizar a nivel hora (mínima unidad con la que contamos información) o a nivel día, establecemos la clasificación definitiva para el periodo deseado (1 semana) por voto mayoritario de las clases predecidas.
- Finalmente una vez recorrido todo el espacio temporal con los rebalanceos correspondientes, utilizaremos distintas métricas de performance para evaluar y comparar los rendimientos de los portafolios.

	Diciembre				Enero			
	Semana1	Semana2	Semana3	Semana4	Semana1	Semana2	Semana3	Semana4
Iteración1	Train	Train	Train	Train	test/predict: idxmax clase mayoritaria			
Iteración2		Train	Train	Train	Train	test/predict: idxmax clase mayoritaria		
Iteración3			Train	Train	Train	Train	test/predict: idxmax clase mayoritaria	
Iteración4				Train	Train	Train	Train	test/predict: idxmax clase mayoritaria



# Conclusiones

# Dificultades

- La gran volatilidad de muchos de los cripto-activos, puede dificultar encontrar patrones subyacentes.
- Hay bastante correlación entre las distintas monedas (siguiente al bitcoin), lo cual puede presentar un desafío a la hora de buscar diversificar las opciones de inversión.
- Existen muchos indicadores de análisis técnico posibles a la hora de generar nuevos features para entrenar el modelo y no está claro en este momento cuáles deberíamos usar o si algunos presentan ventajas en este caso respecto a otros.
- Es difícil determinar cuál es la ventana temporal adecuada para el conjunto de entrenamiento.

# Trabajo a futuro

- Contamos con información a nivel hora, pero se podría realizar un remuestreo para obtener información diaria.
- Se puede considerar rebalancear la cartera en otro intervalo de tiempo diferente a una semana (15 días, 30 días, etc.).
- Considerar las comisiones de las transacciones a la hora de decidir si rebalancear la cartera.
- Se podría probar utilizar librerías específicas de machine learning para series de tiempo multivariadas (sktime por ejemplo).



# Gracias!

Nicola Mori  
Ezequiel Kinigsberg