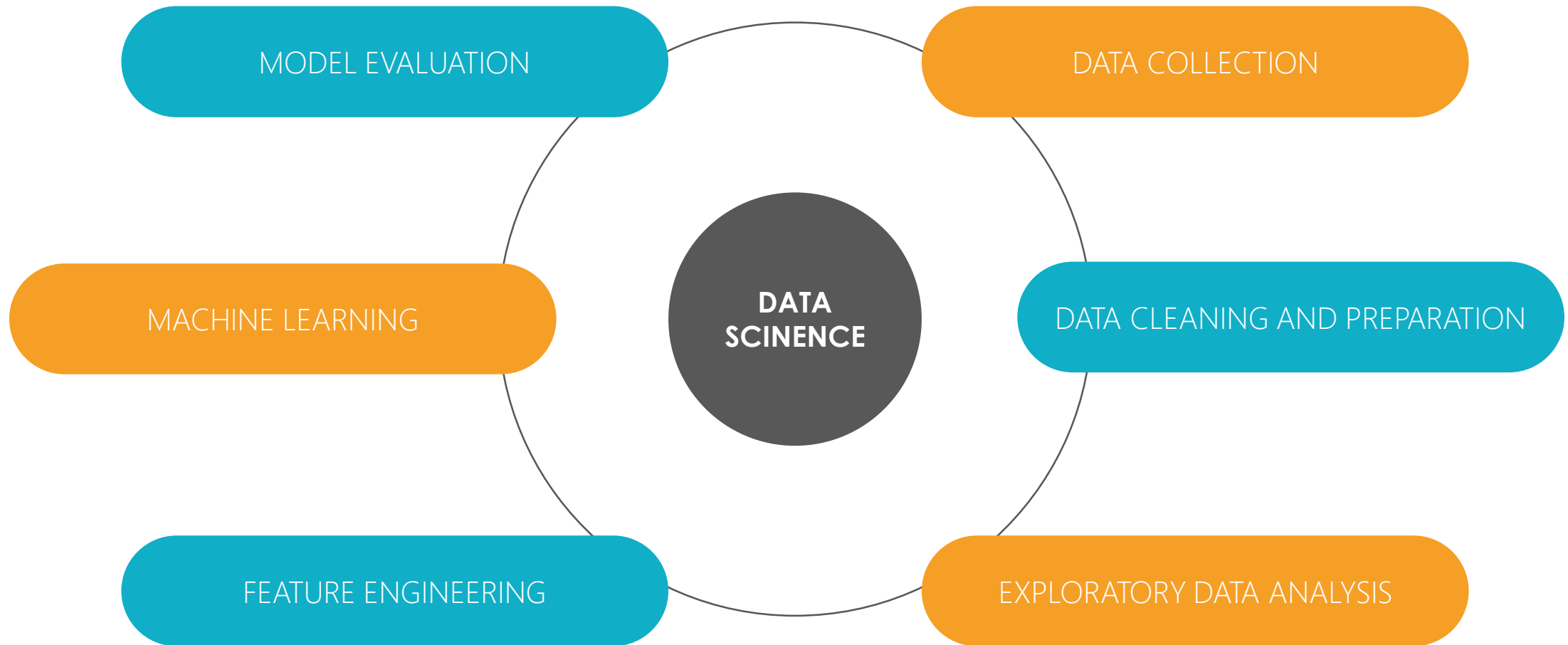




Predicción de Bitcoin

Presentación

• Análisis del proyecto •



Data Collection

```
symbol = 'BTCUSDT'  
tick_interval = '2h'
```

```
def get_candles(start='', symbol='BTCUSDT', tick_interval='2h', limit=1000):  
  
    base_url = 'https://api.binance.com/'  
    endpoint = 'api/v3/klines?'  
  
    if start:  
        query = 'symbol=' + symbol + '&interval=' + tick_interval + '&startTime=' + str(start) + '&limit=' + str(limit)  
    else:  
        query = 'symbol=' + symbol + '&interval=' + tick_interval + '&limit=' + str(limit)  
  
    candles = requests.get(base_url + endpoint + query).json()  
  
    return candles, candles[-1][6] # return candles and last close time in a tuple  
  
def get_all_candles_from_start(symbol, tick_interval): # devuelve una lista de velas, cada vela es una lista tb  
    start=1502942400000 # 17 de agosto de 2017  
    _, last_time = get_candles(start='', symbol=symbol, tick_interval=tick_interval, limit=1)  
    candles = []  
    while start < last_time:  
        i_candles, next_hop = get_candles(start, symbol, tick_interval)  
        candles = candles + i_candles  
        start = next_hop  
    return candles
```

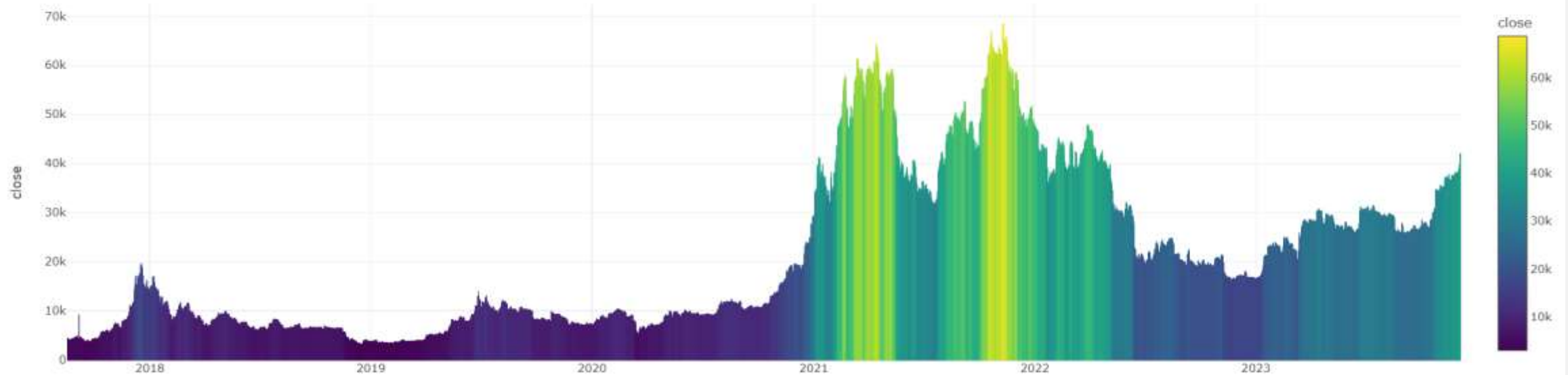
Data Cleaning

- Crear data set, con sus columnas
- Pasar horas a datetime y hora de Madrid
 - Hacer la columna target
 - Quitar los nulos
- Mirar duplicados y outliers
 - Guardamos el dataset

open_time	open	high	low	close	volume	close_time	quote	trades	takers_buy_base	takers_buy_quote	ignore	target
2023-12-05 05:00:00+01:00	41829.98000000	41918.00000000	41724.75000000	41911.09000000	1670.53622000	2023-12-05 06:59:59.999000+01:00	69863035.35224590	92522	877.45543000	36695932.50422930	0	41528.00000000
2023-12-05 07:00:00+01:00	41911.10000000	41919.08000000	41414.00000000	41528.00000000	4831.57684000	2023-12-05 08:59:59.999000+01:00	200961967.19007840	164668	1921.19232000	79943375.99872620	0	41705.54000000
2023-12-05 09:00:00+01:00	41528.00000000	41728.00000000	41420.00000000	41705.54000000	2719.06619000	2023-12-05 10:59:59.999000+01:00	113102276.96793730	124334	1342.83865000	55860335.67331080	0	41726.09000000
2023-12-05 11:00:00+01:00	41705.55000000	41758.77000000	41565.13000000	41726.09000000	2563.16304000	2023-12-05 12:59:59.999000+01:00	106830010.77133510	98961	1214.70509000	50631074.56260810	0	41786.10000000

Exploratory Data Analysis

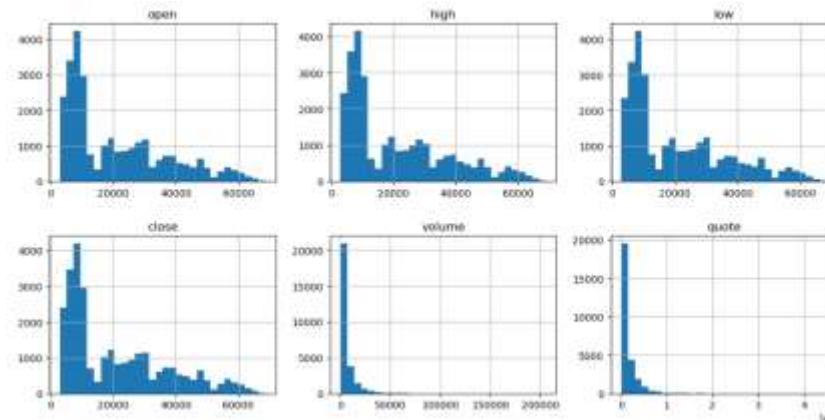
Evolution Bitcoin cada 2h



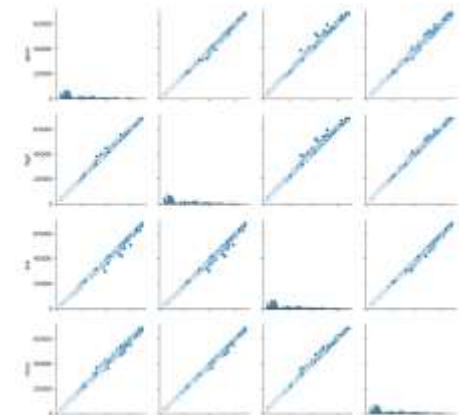
Correlation

open	1.00	1.00	1.00	1.00	0.97	0.40	0.21	0.06	0.39
high	1.00	1.00	1.00	1.00	0.97	0.40	0.24	0.06	0.40
low	1.00	1.00	1.00	1.00	0.96	0.39	0.22	0.06	0.39
close	1.00	1.00	1.00	1.00	0.97	0.40	0.23	0.06	0.39
volume	0.57	0.57	0.06	0.57	1.00	0.39	0.91	1.00	0.09
quote	0.40	0.40	0.39	0.40	0.39	1.00	0.91	0.09	1.00
trades	0.13	0.19	0.23	0.13	0.31	0.91	1.00	0.91	0.91
takers_buy_base	0.06	0.06	0.06	0.06	1.00	0.39	0.91	1.00	0.09
takers_buy_quote	0.39	0.40	0.39	0.39	0.39	1.00	0.91	0.09	1.00

Numerical features distribution



Pair Plot



Feature Engineering

- Pasar todo a numérico y a datetime si es necesario
 - Borrar atributos innecesarios
- One hot encoding y Ordinal encoding
 - Mirar correlaciones
- Crear nuevos atributos
 - Guardar dataset

price_range	fib_61.8	fib_50	halving	day_of_week	buy_sell
67.37	4287.05534	4295.005	0	3	1.0
62.58	4311.31556	4318.700	0	3	1.0
112.46	4376.27972	4389.550	0	3	1.0
85.58	4432.50156	4442.600	0	3	0.0
48.00	4429.33600	4435.000	0	3	1.0

Machine Learning

- Train / Test
- Cross-validation, custom
- Baseline 155,23 de media / diferencia entre close to close
- LR, KNN, DT, RF, GBDT, SVR, EN
 - Randomsearch
- RNNs / Simple, LSTM, GRU

Mean train MAE: 154.88 \$
Mean validation MAE: 125.45 \$

Mean train MSE: 92891.17 \$
Mean validation MSE: 25939.21 \$

Mean train MedAE: 54.64 \$
Mean validation MedAE: 111.11 \$

Mean validation Bacc: 0.51 %

Mean validation Prec: 0.51 %

Mean validation Rec: 0.59 %

- Tests

Test Bacc: 0.5463318884636483

Test Prec: 0.5653043269615292

Test Rec: 0.49862433862433864



Gracias