CSE523 Machine Learning

Project: Crypto Forecasting

Weekly Project Report: February 16, 2022

Team: CryptoPolice

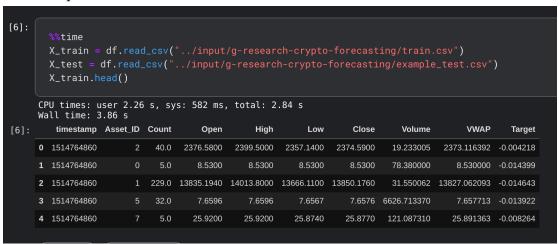| Sameep Vani	AU1940049
Kashvi Gandhi	AU1940175
Khushi Shah	AU1920171
Kavya Patel	AU1940144

Tasks performed this week:

1) Used multiple frameworks and libraries to read the dataset.

Outcomes of the task performed:

1) Time taken by multiple frameworks and libraries to read the dataset:
- Pandas

```
[5]:   %%time
       X_train_pandas = pd.read_csv("../input/g-research-crypto-forecasting/train.csv")

       CPU times: user 30.1 s, sys: 4.38 s, total: 34.5 s
       Wall time: 34.7 s
```

- Apache Spark :

```
[8]:   %%time
       X_train_spark = spark.read.csv(path="../input/g-research-crypto-forecasting/train.csv", header=True, inferSchema=True

       [Stage 1:=====================================================>  (21 + 1) / 22]
       CPU times: user 79.8 ms, sys: 25.8 ms, total: 106 ms
       Wall time: 1min 50s
```

- Rapids :

```
[6]:    %%time
        X_train = df.read_csv("../input/g-research-crypto-forecasting/train.csv")
        X_test = df.read_csv("../input/g-research-crypto-forecasting/example_test.csv")
        X_train.head()
```

CPU times: user 2.26 s, sys: 582 ms, total: 2.84 s
Wall time: 3.86 s

[6]:

	timestamp	Asset_ID	Count	Open	High	Low	Close	Volume	VWAP	Target
0	1514764860	2	40.0	2376.5800	2399.5000	2357.1400	2374.5900	19.233005	2373.116392	-0.004218
1	1514764860	0	5.0	8.5300	8.5300	8.5300	8.5300	78.380000	8.530000	-0.014399
2	1514764860	1	229.0	13835.1940	14013.8000	13666.1100	13850.1760	31.550062	13827.062093	-0.014643
3	1514764860	5	32.0	7.6596	7.6596	7.6567	7.6576	6626.713370	7.657713	-0.013922
4	1514764860	7	5.0	25.9200	25.9200	25.8740	25.8770	121.087310	25.891363	-0.008264

Tasks to be performed next week:

1) Finalizing framework
2) Understanding time series API
3) Figuring out the relationship and linkages between different csv files of our dataset.