Forecasting in Cryptocurrencies

Simran Singh Gulati (31125301)
Supervisor – Dr. Christoph Bergmeir

MONASH University

Abstract

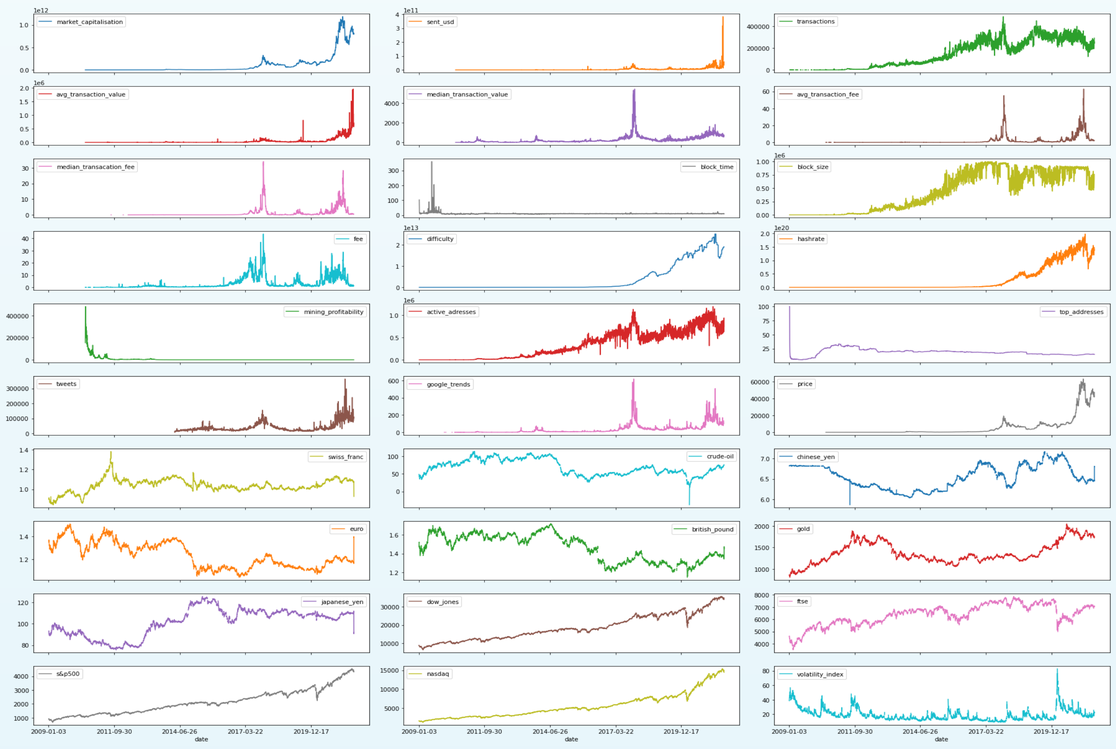
In the last one decade there has been a tremendous increase in Bitcoin's market value and the public attention around Cryptocurrencies. Its highly volatile nature has ignited a great buzz among machine learning practitioners to predict its prices in real time. In this study we have scraped over 30 time series encompassing global *macroeconomic factors*, *blockchain attributes* and *public opinion* to forecast the **Bitcoin** price. The economic features include some influential stocks, fiat currencies, gold and crude oil prices. The technological features include hash rate, mining difficulty, block size, etc. Lastly, the public sentiment was drawn using number of Tweets and Google Trends. We deployed a Random Forest to draw non-linear relationships among the mentioned feature set. The model predicted log-returns which were further used to compute the actual price of the Bitcoin for the upcoming day. The study focussed on last 7 years of data which had the most fluctuating trends. Thereby the dataset was split into different periods to account for potential market cycles. Finally, the model predictions were compared against the naive forecasts. Although the algorithm had at par performance with the naive forecasts, a Diebold-Mariano test revealed that the forecasts were insignificantly less accurate.

Introduction

- Cryptocurrency refers to a virtual asset functioning as a medium of exchange over the internet.
- Bitcoin, the first cryptocurrency was founded in 2009 during The Great Recession.
- Devised to develop distributed applications and promote a peer-to-peer transactional network.
- Among 6000+ different cryptocurrencies, Bitcoin holds the largest market capitalization.
- Bitcoin was valued at $0 at its launch in 2009 (U.S. dollar).
- While the global economy crashed during the pandemic, Bitcoin proved to be virus-immune and hit an exorbitantly high price of $66,000 in 2021.
- Growing awareness and high volatility motivates us to forecast bitcoin price in real time.
- While most of studies stick to historic pricing, some of them have also tried to unveil the latent factors affecting bitcoin price.
- The potential determinants include –
  - Blockchain Attributes: hash rate, block size, mining difficulty, etc.
  - Media Sentiment
    - Tweets
    - Google
  - Macroeconomic Factors:
    - Stock Market (S&P500, DOW30, etc.)
    - Global Commodities (Gold, Crude Oil, etc.)
- Presuming cryptocurrency behaves like stock market, the general agreement says that returns are not predictable but the volatility might be. We check the same for bitcoin and aim to find its predictors.

Methodology

Data Collection : We scraped the following 30 daily time-series from the internet:



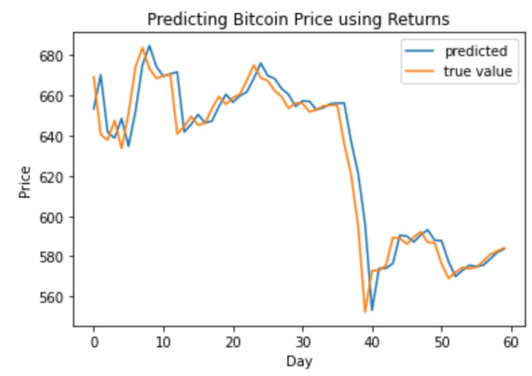
Data Imputation : While blockchain values were readily available, stock market only works during the business hours. So we used a forward fill for the economic features.

Data Exploration : We observed notable periods of price gain and split the timeline into 4 datasets:

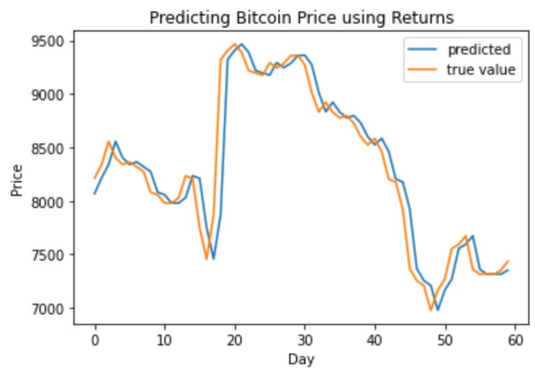
I. Early 2015 to mid 2016
II. Late 2016 to early 2018
III. Mid 2018 to late 2019
IV. Early 2020 onwards

Data Modelling : A matrix with 2400 rows and a 100 columns was computed and fed into a Random Forest. The algorithm trained on 90% data using a 1000 trees and made one-step ahead predictions. Three different techniques were used for forecasting wherein the final model was targeted to predict the log-returns. These were later transformed and added to the current day price. A Naïve Forecast was used as a benchmark for our model.

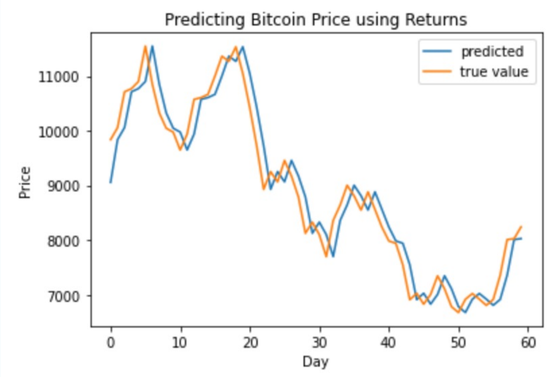
```
Test Set : 1 of 4

Model MAE:  7.299444919001799
Naive MAE:  7.3395666666666655
Model RMSE: 11.511005963291662
Naive RMSE: 11.361866291826068
```



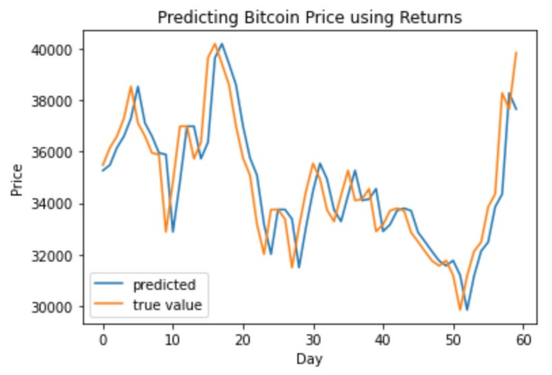

```
Test Set : 3 of 4

Model MAE:  144.0980894499704
Naive MAE:  143.93333333333334
Model RMSE: 250.64169913842812
Naive RMSE: 250.59874966434555
```




```
Test Set : 2 of 4

Model MAE:  328.734238393509
Naive MAE:  328.73333333333335
Model RMSE: 390.18100202253174
Naive RMSE: 390.1454002462842
```



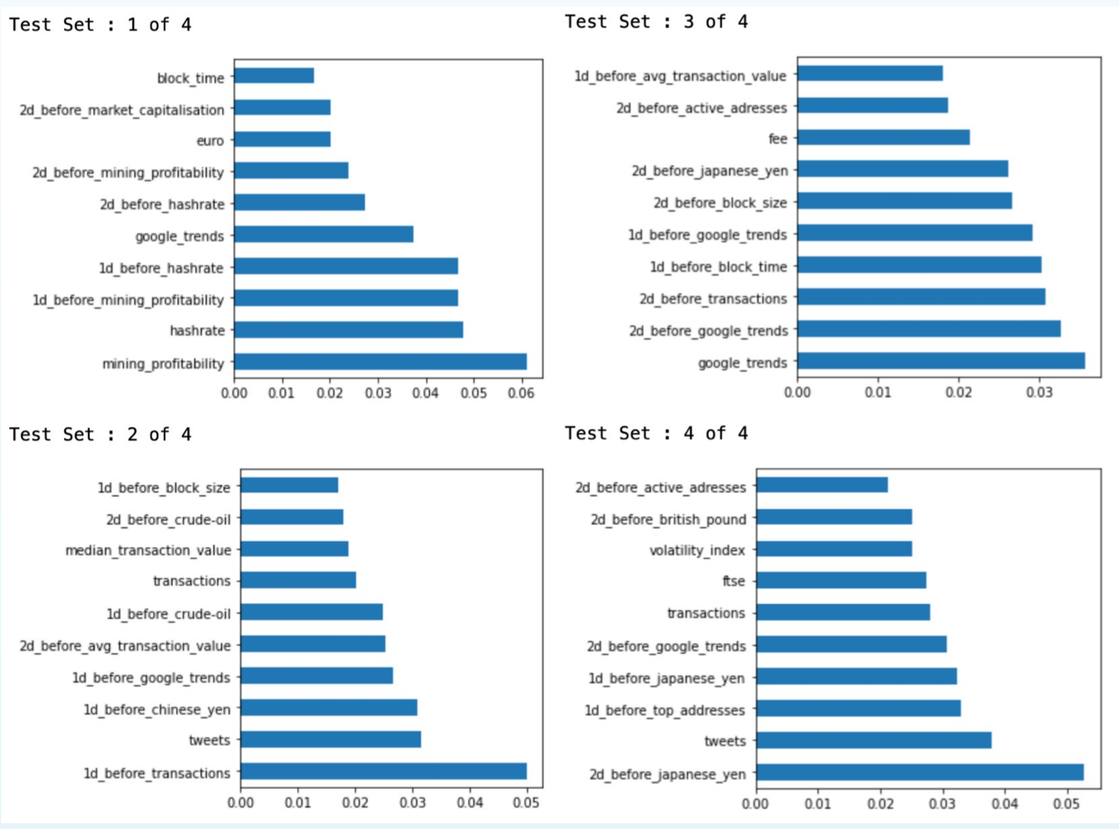

```
Test Set : 4 of 4

Model MAE:  984.9998224105944
Naive MAE:  985.0
Model RMSE: 1271.2235962055329
Naive RMSE: 1271.2830395575436
```



Experimental Results: The model traced the overall trend really well. The high noise in our series motivated us to do an error comparison using RMSE (lower is better).

Dataset	Forecast 1	Forecast 2	Forecast 3	Naïve Forecast
1	36.99	26.07	11.50	**11.36**
2	487.66	404.67	390.18	**390.14**
3	276.62	286.86	250.64	**250.59**
4	1936.08	1498.63	**1271.22**	1271.28

Feature Importance: The error-rate continuously dropped with each iteration which led to feature-extraction for all the datasets.



Statistical Testing : Diebold-Mariano test was used to compare the forecast accuracies for our model against the naïve.

Conclusion

- The potential influence on bitcoin price transitioned from technological features (in the early years) towards the economic ones (current day).
- The random forest and naïve forecast have different forecast-accuracies and the results were insignificantly less accurate.
- It is a non-stationary time series with stochastic trends and thus could not be predicted.
- The dynamics for Bitcoin change drastically and determining their returns is not plausible.