

Project Proposal: CSC240 Spring2022

Beilei Guo, Liwei Jiang

March 14, 2022

1. DATASET CHOICE

The proposal is to analyze the dataset: "Cryptocurrency Historical Prices" available via Kaggle. This dataset aligns with Beilei's interest in cryptocurrency. By doing a master's degree in data science, Beilei would like to pursue a career in quantitative research after graduation. Speculating on cryptocurrency's historical price could help her get familiar with digital assets and build a knowledge foundation for future investment. As a student pursuing a computer science degree, Liwei has long shown a keen interest in time-series data, especially for the forecasting field. An analysis of the historical prices of cryptocurrencies also satisfied his curiosity.

Cryptocurrencies themselves are complicated and unique existence. On the surface, they are worthless and contain significant credit risk, but their prices surpassed people's expectations again and again. Although paying prices to leave words on a large, open tally book itself is hard to be defined rational or not, the transactions in the market are obviously not always rational enough. But we could still separate some rules based on observation; for example, the demand for cryptocurrencies would increase whenever the economic conditions get difficult. Thus, a forecast on their prices is a worth exploring and exciting topic.

2. TEAM COMPOSITION

Liwei Jiang and Beilei Guo will be working together on this final project. Liwei is an undergraduate student pursuing a B.S. in Computer Science, and Beilei Guo is a part-time student pursuing a M.S. in Data Science.

3. GOAL OF THE ANALYSIS

This project aims to employ various techniques to develop a model for predicting the future price and transaction volume. The final way to evaluate the performance of the result is still under discussion.

4. PLANNED TECHNICAL APPROACH

4.1 Visualization

The analysis will start from basic visualization to achieve a deeper understanding of the dataset. Results will be displayed in various formats, including frequency table, histogram, color printing essay, and n-gram.

4.2 Data Preprocessing

A cursory review of the data reveals some unnecessary columns like the name and symbol. Furthermore, the value of the cryptocurrencies of each date is divided into four detailed values and may require to be processed into one value for further analysis. Although we didn't find any missing index yet, a preprocess is still needed.

4.3 Model Selection

We will start with pure linear regression, which employs time as a feature to forecast the future price. And the result would also perform as the baseline to compare with the output of more complex modeling followed by it.

The cryptocurrencies' prices do not always tend to be independent from the prices before or after that time point. Instead, they are linked in some sort of correlation, and therefore, an autoregressive (AR) model could be introduced here. Similarly, as another frequently used model, the moving average (MA) model is also a candidate for this project. Furthermore, although too many parameters may increase the risk of over-fitting, the auto-regressive moving average (ARMA) model, which combined the AR and MA model, is also worth trying in this project.

5. ROLE OF TEAM MEMBERS

Beilei will be responsible for future model selection, optimization, and result evaluation. Liwei will take charge of implementing and modifying the selected mining algorithms. Liwei and Beilei will work together to develop the model and deliver a final presentation in class.