Zero Theorem Literature Review

"Bitcoin Price Prediction Using Machine Learning: An Approach to Sample Dimension Engineering, Z. Chen, C. Li, W. Sun, 2019"

By: SoReZ

https://www.zerotheorem.com

Highlights

- Developed an approach that utilizes statistical and machine learning approaches to accurately forecast Bitcoin price.
- Selecting high-dimension features for daily pricing and a few features for 5-minute interval trading data as part of the feature engineering process.

Background

With the boom and bust in value of cryptocurrency, BitCoin has gained a lot of importance. Hence, due to its highly volatile nature, value of BitCoin reflects the confidence of investors in cryptocurrency. However, existing studies have leveraged machine learning for more accurate Bitcoin price prediction, but still there remain some limitations that can predict BitCoin price at different frequencies. Therefore, machine learning techniques that focused on the practicality of applying alternative modelling techniques to samples with diverse data structures and dimensional attributes are needed to predict Bitcoin price at different frequencies.

Introduction

To overcome the limitations, there is a need to predict Bitcoin price changes using robust machine learning approaches. In this context, Chen et al. (2020) predicted Bitcoin price at different frequencies using statistical methods such as Linear Discriminant Analysis and Logistic Regression and machine learning algorithms including Random Forest, Quadratic Discriminant Analysis, XGBoost, Long Short-term Memory (LSTM) and Support Vector Machine (SVM). For this purpose, Chen et al. (2020) leverage appropriate machine learning techniques to engineer sample dimensions for Bitcoin price prediction.

Proposed Methodology

To do this first the prediction sample is divided into daily intervals with small sample size and 5-minute intervals with a big sample size. After dividing the prediction sample, Chen et al. (2020) conduct the feature engineering process by selecting high-dimension features for daily price and few features for 5-minute interval trading data respectively. After applying feature engineering Chen et al. (2020) conduct simple statistical models and complicated ML models that predict Bitcoin daily price with high-dimensional features to avoid over-fitting. To have a better understanding, let's have a look at the Figure 1

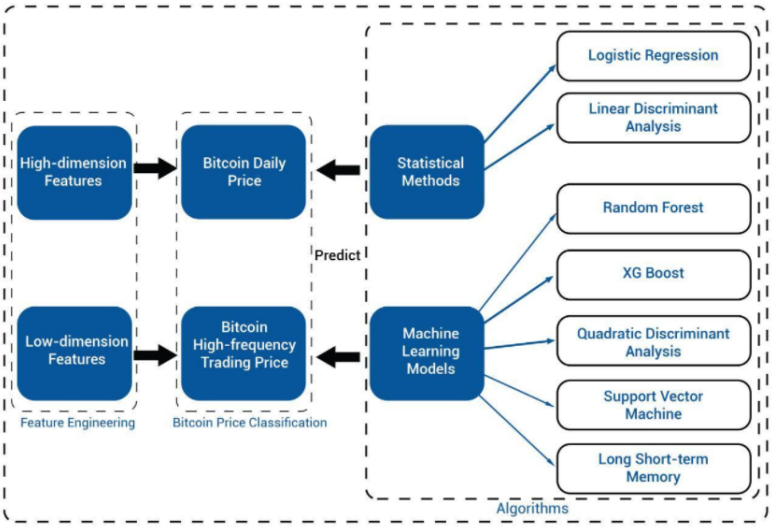

Figure 1: An Overview of the Framework by Chen et al. (2020)

Hence, multiple features are selected where first feature type to select for daily Bitcoin price prediction relates to the Bitcoin market and trading including independent features like transaction value and market capitalization and second feature is based on the significance of the price prediction to problem observed. For this purpose, data for these features were obtained from Blockchain Explorer and Bitcoinity.org. Another feature considered is attention from investors and media. Finally, Chen et al. (2020)) included 12 major features in the prediction models such as Block Size, Number of Transactions, Hash Rate, Confirmed Transactions per Day, Mining Difficulty, Mempool Transaction Count, Mempool Size, Estimated Transaction Value, Market Capitalization, Gold Spot Price, Google Trend Search Volume Index, and Total Transaction Fees.

Details of Proposed Methodology

Chen et al. (2020) employed two datasets for the study. The first includes aggregated Bitcoin daily price from CoinMarketCap.com. It also includes trading and market data, property and network data, gold spot price, and media and investor attention. The second dataset was obtained from Binance and it includes 5-minute interval Bitcoin trading price data in real-time. Two statistical methods, logistic regression (LR) and linear discriminant analysis (LDA) were implemented for Bitcoin daily price with higher dimensional features. However, for the 5-minute interval price, machine learning models were used. The models include random forest (RF), quadratic discriminant analysis (QDA), support vector machine (SVM), long short-term memory (LSTM), and XGBoost (XGB).

Results and Discussion

For Bitcoin daily price prediction, the statistical methods have an average accuracy of 65% while the machine learning models have an average accuracy of 55.3%. The LR model gave the best results with an accuracy of 66%. SVM gave the best performance among the machine learning models with an accuracy of 65.3% while XGB gave the worst performance with an accuracy of 48.3%. For Bitcoin 5-minute interval price prediction, the average accuracy of the machine learning models was 62.2% (with LSTM having the best performance, 67.2%) while that of statistical methods was 53.0% (LR gave 54.5% while LDA gave 51.5%).

Conclusion

This led to the conclusion that statistical methods are better for Bitcoin daily price prediction while machine learning models are better for Bitcoin 5-minute interval price prediction. From results it is clear that there exist several limitations in data analysis and sources. For a more comprehensive study in the future, price data with different features, granularities, and more dimensions should be collected. Other methods like statistical method, ARIMA, and machine learning model, RNN, should be considered in future.

References

Chen, Z., Li, C., and Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, 365:112395.