CRITICAL THINKING PROJECT SYNOPSIS ON

"Bitcoin Price Predictor"

A critical thinking project synopsis submitted in partial fulfillment of the requirement for the award of Degree of

Bachelor of Engineering
(Sixth Semester)
In
COMPUTER SCIENCE & ENGINEERING
Session 2022-2023

Prescribed By
Dr. Babasaheb Ambedkar Technological University, Lonere (DBATU)

Guided By

Prof. Madhvi Sadu

Prof. Sachin Dhawas

Submitted By(Group 5)

Pratiksha Rajpurohit(B604)

Vanshita Meshram(B649)

Sanika chaudhari(B620)

Sneha Sarkar(B629)

Tejal Bobade(B637)

Rimpa Singh(B612)

Prerna Pal(B606)

ABSTRACT

After the boom and bust of cryptocurrencies' prices in recent years, Bitcoin has been increasingly regarded as an investment asset. Because of its highly volatile nature, there is a need for good predictions on which to base investment decisions. Although existing studies have leveraged machine learning for more accurate Bitcoin price prediction, few have focused on the feasibility of applying different modelling techniques to samples with different data structures and dimensional features. To predict Bitcoin price at different frequencies using machine learning techniques, we first classify Bitcoin price by daily price and high-frequency price. we attempt to predict the Bitcoin price accurately taking into consideration various parameters that affect the Bitcoin value. Using the available data we will predict the sign of the daily price change with highest possible accuracy. We have used Random Forest Classifier and compared with benchmark results for daily price prediction, we achieve a better performance, with the highest accuracies of the statistical methods and machine learning algorithms of 99%. Our investigation of Bitcoin price prediction can be considered a pilot study of the importance of the sample dimension in machine learning techniques.

Keyword: Bitcoin, Machine Learning, Random Forest Classifier, Cryptocurrencies

CONTENT

INTRODUCTION

Bitcoin is a decentralized digital currency that uses cryptography for security and is not controlled by any government or financial institution. It was created in 2008 by an individual or group of individuals using the pseudonym Satoshi Nakamoto (2008) with a paper titled “Bitcoin: A Peer-to-Peer (P2P) Electronic Cash System”. Transactions with bitcoin are recorded on a public ledger called the blockchain, which allows anyone to view the history of a specific Bitcoin. The decentralized nature of Bitcoin allows it to operate independently of central banks and can be transferred instantly across the globe. It has gained popularity as a means of exchange and a store of value (Baur and Dimpfl 2021). In the past 10 years, after experiencing several ups and downs, it broke through USD 68,000 per coin in November 2021, and the total current price once exceeded USD 1.2 trillion.

However, as a commodity, Bitcoin has the problem of high volatility. During the seven years from April 2015 to April 2022, the standard deviation of Bitcoin’s daily return rate was 3.85%, which was 2.68 times the standard deviation of gold’s return rate during the same period and 3.36 times that of the S&P500. Due to the large price fluctuations, the function of Bitcoin as a store of value as a commodity and as a transaction payment function as a currency has been questioned.

While enjoying the advantages of Bitcoin’s security and decentralization, how to grasp the trend of Bitcoin to minimize the risk of Bitcoin floating has become a difficult problem. Many researchers try to grasp the trend of Bitcoin through the correlation between the price of Bitcoin and the price of other commodities. But whether it is gold (Baur and Hoang 2021; Kim et al. 2020b; Blake 2019), which is often used for comparison, stock market index (Erdas and Caglar 2018), or crude oil price (Selmi et al. 2018), past studies have shown that the correlation between Bitcoin and them is weak.

In past studies, another type of research direction to grasp the price trend of Bitcoin is to predict the price of Bitcoin in the future through AI algorithms and powerful computing power of computers. With the improvement of hardware performance in the 21st century, machine learning technology which has become a hot field of research. Primarily, machine learning has been used across a variety of areas such as that of stock markets (Huang and Liu 2020; Philip 2020); crude oil markets (Fan et al. 2016); gold markets (Chen et al. 2020b); and futures markets (Kim et al. 2020a).

Prediction of Bitcoin by AI is mainly divided into two categories. The first category is the classification research of predicting the rise or fall of Bitcoin in the future. The error standard is DA and F1. The other category is regression research on predicting Bitcoin prices, while the corresponding errors are RMSE and MAPE. Due to the sharp fluctuations in the price of Bitcoin, only grasping the rise or fall of the price of Bitcoin in the future cannot help investors avoid risks. In contrast, getting the specific bitcoin price as a reference price is more useful.

LITERATURE SURVEY

1) "Predicting the price of the Bitcoin Using Machine Learning" –

Sean McNally, Jason Roche, Simon Caton – 2018 IEEE 26th Euromicro International Conference on Parallel, Distributed, Network - Based Processing. The goal of this paper is to ascertain with what accuracy the direction of Bitcoin price in USD can be predicted. The price data is sourced from the Bitcoin Price Index. The task is achieved with varying degrees of success through the implementation of a Bayesian optimized recurrent neural network (RNN) and a Long Short-Term Memory (LSTM) network. The LSTM achieves the highest classification accuracy of 52% and a RMSE of 8%. The popular ARIMA model for time series forecasting is implemented as a comparison to the deep learning models. As expected, the non-linear deep learning methods outperform the ARIMA forecast which performs poorly. Finally, both deep learning models are benchmarked on both a GPU and a CPU with the training time on the GPU outperforming the CPU implementation by 67.7%.

2) "A New Forecasting Framework for Bitcoin price with LSTM" –

Wu Chih – Hung, Ma Yu – Feng, Lu Chih – Chiang – 2018 IEEE International Conference on Data Mining Workshops (ICDMW). Long short-term memory (LSTM) networks are a state-of-the-art sequence learning in deep learning for time series forecasting. However, less study applied to financial time series forecasting especially in cryptocurrency prediction. Therefore, we propose a new forecasting framework with LSTM model to forecasting bitcoin daily price with two various LSTM models (conventional LSTM model and LSTM with ARIMA model. The performance of the proposed models are evaluated using daily bitcoin price data during 2018/1/1 to 2018/7/28 in total 208 records. The results confirmed the excellent forecasting accuracy of the proposed model with ARIMA. The test mean squared error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) for bitcoin price prediction, respectively. The our proposed LSTM with AR(2) model outperformed than conventional LSTM model. The contribution of this study is providing a new forecasting framework for bitcoin price prediction can overcome and improve the problem of input variables selection in LSTM without strict assumptions of data assumption. The results revealed its possible applicability in various cryptocurrencies prediction, industry instances such as medical data or financial time-series data.

3) "A Study of Opinion Mining and Data Mining Techniques to Analyze the Cryptocurrency Market –

Akhilesh P. Patil, T.S. Akarsh, A. Parkavi – 2018 3rd International Conference on Computational Systems and Information Technology of Sustainable Solutions (CSITSS). The value of various Cryptocurrencies such as Bitcoin, Litecoin, Ethereum are always elusive. Hence, it would be a great value addition to investors if a model is able to predict what would be the nature of the crypto market for the next day. Through this paper, a time-series model using Long Short-Term Memory Networks is built to determine the value of cryptocurrency in the future. As a study, three cryptocurrencies - Bitcoin, Litecoin and Ethereum has been taken into consideration. A comparison of the results by using opinion mining to interpret the mood of the market on the current day for different currencies has been done. The sentiment scores got from natural language processing of textual data are used as features to the model used for predictions. The time-series charts are plotted using Plotly - python library for graphing plots. The Mean Absolute Error calculated between the actual and predicted values is used as the uncertainty quantification method. These uncertainty quantification methods are compared to analyze the present-day scenario of the market using opinion mining

RELATED WORK

Aggarwal et al. (2019) studied whether gold price can predict Bitcoin price through three deep learning algorithms of CNN, LSTM, and GRU. The conclusion is that the predicted price of the model which only uses gold price deviates from the true Bitcoin price, and the prediction accuracy of the LSTM model is the best of three. Liu et al. (2021) expanded the range of explanatory variables, based on the cryptocurrency market and macro market index (stock market index, crude oil price, exchange rate, etc.) and search index, a total of 40 explanatory variables for Bitcoin price prediction. SDAE algorism shows better prediction performance than BPNN, PCA-SVR, and SVR.

Regarding the prediction research of Bitcoin price, the methods are divided into time series and machine learning. Multiple studies have concluded that the prediction accuracy of ARIMA is not as good as that of machine learning (McNally et al. 2018; Shin et al. 2021; Chen et al. 2020a; Akyildirim et al. 2021).

LSTM, as a controlled study of random forest regression in this study, has been studied as a target model many times in the past literature (Shin et al. 2021; Jagannath et al. 2021; Rizwan et al. 2019). Phaladisailoed and Numnonda (2018) used four deep learning algorithms (Theil–Sen regression, Huber regression, LSTM, and GRU) to predict the price of Bitcoin. The 52.78% accuracy of the LSTM algorithm is the highest. Based on the same explanatory variables, Tandon et al. (2019) found that adding 10-fold cross-validation to the LSTM training process can increase the accuracy of LSTM by 14.7%. However, the selection of explanatory variables in Phaladisailoed's and Tandon's studies is limited to OHLC, volume from top exchange and market cap. In the research done by Aggarwal et al. (2019), in addition to the price of Bitcoin itself, gold price was added to explanatory variables. The experimental results show that the RMSE of the LSTM algorithm is 47.91, which is better than CNN and GRU. McNally et al. (2018) added the variables difficulty and hash rate related to Bitcoin attributes in his research, the 52.78% prediction accuracy of LSTM is also better than the accuracy of RNN and ARIMA. Chen et al. (2020a) used LSTM, SVR, ANFIS, and ARIMA, four algorithms to predict the Bitcoin price. While Chen added eight kinds of Bitcoin attribute variables, public attention variables (Google Trends and Twitter data) and economic category variables. In the four subsample periods, LSTM all showed better prediction accuracy than the other three. Livieris et al. (2020) introduced a novel framework by preprocessing, which performed a series of transformations based on first differences or returns, to make data "suitable" for fitting a deep learning model based on the stationarity property.

In addition to predicting the price of Bitcoin, there are many studies using LSTM to predict other digital currencies (Sebastião and Godinho 2021; Saadah and Whafa 2020; Derbentsev et al. 2020). Politis et al. (2021) used LSTM to predict the price of Ether with an accuracy of 84.2%. Livieris et al. (2021) used hybrid CNN-LSTM to conduct prediction experiments on Bitcoin (BTC), Ethereum (ETH), and Ripple (XRP) with the highest market value at the time and obtained BTC The prediction accuracy of 55.03% is higher than ETH's 51.51% and XRP's 49.61%.

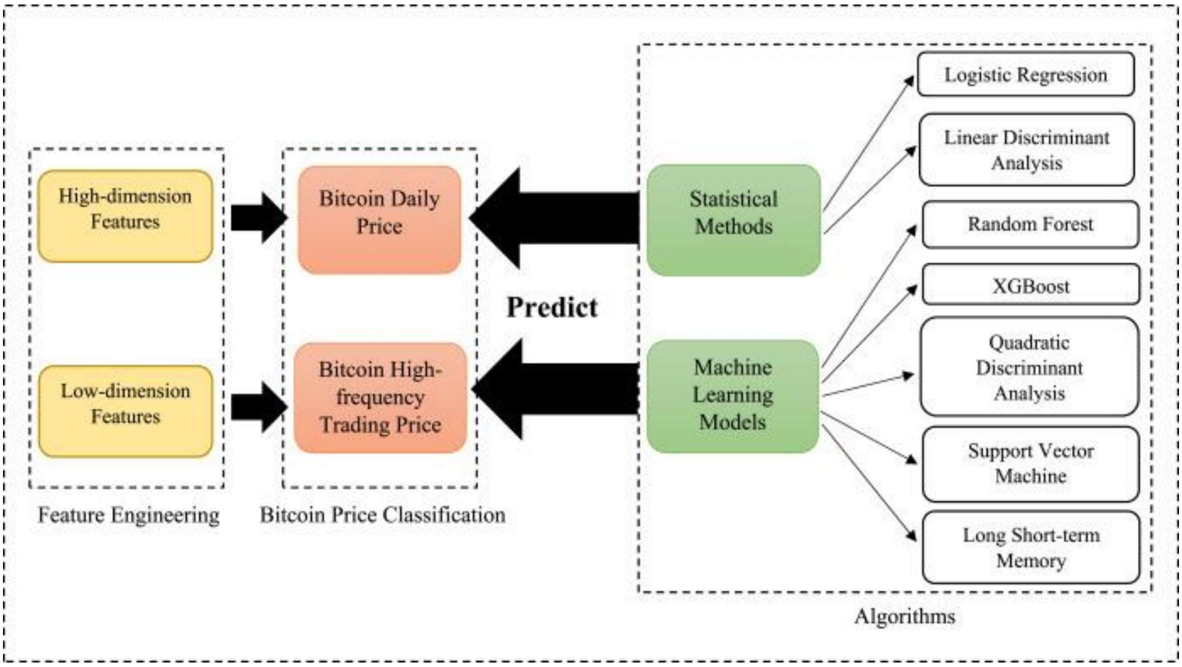
In McNally et al.'s (2018), García-Medina and Duc Huynh's (2021), and Chen et al.'s (2020a) studies, it is mentioned that adding Dropout layers between each layer of LSTM can reduce the effect of overlearning. But there are differences in the choice of dropout coefficients (0.1, 0.3, 0.5) among the three works of literature above.

Regarding the selection of explanatory variables, in addition to the macroeconomic variables used in many works of literature, Jagannath et al.'s (2021) research focuses on the core variables of the Bitcoin blockchain, including users, miners, and exchanges. Technical indicators have proven useful for predicting Bitcoin prices (Jaquart et al. 2021; Mudassir et al. 2020). The LSTM based on the self-adaptive technique also gets good prediction performance, but the article lacks a comparative experiment with the model added macroeconomic variables. Regarding the explanatory power of variables on Bitcoin price, García-Medina and Duc Huynh (2021) innovatively studied variables such as social media (E. Musk and D. Trump's remarks) and Tesla stock price. During the ups and downs in the second half of 2020, the conclusion was that the explanatory power of these variables that were of great interest at the time was not found. Carbó and Gorjón (2022), in their appendix, compare the effect of adding the previous period's Bitcoin price to the explanatory variables based on the LSTM algorithm. The RMSE accuracy of the model that added the previous Bitcoin price as an explanatory variable improved significantly from the original 21% to 11%.

The selection of time unit prices is also a point that has been analyzed by many researchers. Most research use days or minutes as the sample unit. In the quarterly research of DSVR, DNDT, and DRCNN conducted by Lamothe-Fernández et al. (2020), each model obtained more than 60% prediction accuracy, but this high accuracy may be related to Bitcoin's general uptrend between 2011 and 2019 in the sample, as well as the long quarterly units. The work of Shin et al. (2021) is based on the LSTM model, with sample units in a minute, hour, and day. The results show that the prediction accuracy of the day model and minute model is similar, and both better than the model with an hour unit.

Bitcoin has a history of 15 years since its birth in 2008, although it is not long compared to other assets. In previous studies, researchers are more willing to subdivide data samples into small samples before conducting prediction research (Shin et al. 2021; Chen et al. 2020a; Carbó and Gorjón 2022). In Jagannath et al.'s (2021) and Awoke et al.'s (2021) experiments, the longest period of a single sample does not exceed 4 years.

TOPIC OF DISCUSSION



Our project addresses leveraging appropriate machine learning techniques to engineer sample dimensions for Bitcoin price prediction. Inspired by the principle of Occam's razor and the characteristics of our datasets, we tackle the problem as follows. First, the prediction sample is divided into daily intervals with small sample size and 5-minute intervals with a big sample size. Second, we conduct the features engineering: select high-dimension features for daily price and few features for 5-minute interval trading data respectively. Third, we conduct simple statistical models including Logistic Regression and Linear Discriminant Analysis and the more complicated machine learning models including Random Forest, XGBoost, Quadratic Discriminant Analysis, Support Vector Machine and Long Short-term Memory. Fourth, we adopt the simple statistical methods to predicting Bitcoin daily price with high- dimensional features to avoid overfitting. Meanwhile, the machine learning models are leveraged in high-frequency price few features. Above figure shows the overview of our research framework.

Our project makes observations in two ways. One is to extend the feature dimensions, and the other is to evaluate different machine learning techniques for solving problems of multiple frequency Bitcoin prices. The study makes the following contributions. (1) To the best of our knowledge, we are at the forefront of establishing higher dimensional features for problems of Bitcoin daily price prediction by integrating investor attention, media hype and XAU gold spot features with common and traditional features such as network and market. (2) We address the importance of the sample dimension by classifying Bitcoin price data by interval. The real-time 5-minute interval trading data acquired from the top cryptocurrency exchange is high-frequency and large scale, and the aggregated Bitcoin daily price obtained from CoinMarketCap is low-frequency and small scale. Hence, the problem of Bitcoin price prediction is addressed from a broad perspective. (3) To find appropriately complex models and meet the requirement of accuracy, we evaluate different machine learning techniques using problems of multiple frequency Bitcoin price. Specifically, we lower the complexity of algorithms for low-frequency daily price prediction with higher-dimension features and apply more complicated models for high-frequency price prediction with a few features. The results show that simple statistical methods outperform machine learning models for daily Bitcoin price prediction while more complicated models should be adopted for high frequency Bitcoin price prediction. We envision this study as a pilot for dealing with datasets with different scales and intervals, which can shed light on other industrial prediction problems in the context of machine learning.

CONCLUSION AND FUTURE WORK

Deep learning models such as the RNN and LSTM are evidently effective for Bitcoin prediction with the LSTM more capable for recognizing longer-term dependencies. However, a high variance task of this nature makes it difficult to transpire this into impressive validation results. As a result, it remains a difficult task. There is a fine line between overfitting a model and preventing it from learning sufficiently. Dropout is a valuable feature to assist in improving this. However, despite using Bayesian optimization to optimize the selection of dropout it still couldn't guarantee good validation results. Despite the metrics of sensitivity, specificity and precision indicating good performance, the actual performance of the ARIMA forecast based on error was significantly worse than the neural network models

REFERENCE

- https://jpinfotech.org/bitcoin-price-prediction-using-machine-learning/
- https://www.mdpi.com/1911-8074/16/1/51
- https://www.granthaalayahpublication.org/ijetmr-ojms/index.php/ijetmr/article/view/IJETMR21_A05_2586/771