Comprehensive Literature Review on Machine Learning Structures that familiarise with Market cap prices of various Crypto-Currencies

M Srikanth, Ananya K S, M Srivathsan, Naga Shreya C

[1]Information Science Undergraduate Student, [2]Information Science Undergraduate Student, [3]Information Science Undergraduate Student, [4]Information Science Undergraduate Student
[1]Department of Information Science and Engineering,
[1]RV Institute of Technology and Management, Bengaluru, India

Abstract - With the advent of blockchain technology, the use of cryptocurrencies has increased tremendously.However, because of the volatile markets and large price swings, cryptocurrencies are not seen as a good investment. Motivated by the above considerations, let's look at different methods to effectively measure and capture price movements in the market. Some of the methods analyzed are a stochastic neural network model for predicting cryptocurrency prices, Bitcoin and Ethereum prices using Twitter data, and Google Trends data. Twitter is used as a source of information that influences purchasing decisions: it informs the user about the currency and its growing popularity, and the theoretical framework of time series analysis is based on fairly general properties of time series. To forecast the price and movement of the three most well-known cryptocurrencies, we also assess cutting-edge deep learning methods.

Index Terms - Machine Learning, Crypto-currency, Market cap, High price, Time series, Sentiment Analysis, Trends, LTSM, BiLTSM, CNN, ARIMA, Regression.

I. INTRODUCTION

There has been a paradigm shift in how transactions are carried out, moving away from tangible payments like cash and cheques and towards digital ones in an environment where technology is constantly evolving. The value of digital currencies is based on the reliability and security of the platform used to distribute them.

Using blockchain-based cryptocurrencies is a reliable answer to the aforementioned issues. A currency that functions effectively needs security, decentralisation, and transparency, and blockchain is a new technology that consistently saves information across the network. (2018) Gandomi, A. H., Haider, A.

The first and most well-known cryptocurrency, Bitcoin (BTC), was created in 2009 by an unidentified organisation or person. Since then, 4,000 alternative cryptocurrencies have been developed, including Ethereum (ETH) and Ripple (XRP), demonstrating the emergence of the cryptocurrency industry in finance. The positives are that BTC, ETH, and XRP are well-known cryptocurrencies and account for about 79.5% of the market capitalization of all cryptocurrencies

Cryptocurrencies can be utilised as a store of value as well as a medium of exchange due to the advantages described above and their widespread availability.

Machine learning algorithms can forecast bitcoin prices thanks to the abundance of cryptocurrency data that is now available and knowledge of social trends. Without explicitly programming the computer to carry out a particular task, these algorithms are a collection of techniques for discovering mathematical patterns from data. Yet, alternative models are required to capture more sophisticated representations of data as the complexity of data in the cryptocurrency market increases. The time series problem of bitcoin price prediction can be resolved using deep learning models, notably recursive neural networks. In the past, multiple research employing machine learning and deep learning algorithms to forecast the value of stocks and securities have been examined by a variety of authors. Unfortunately, there hasn't been a lot of study done on forecasting bitcoin prices.

II. RELATED WORK

Beyond Twitter and social media, internet data has also been a fruitful area of research. As far as we know, Ettredge et al. discovered a connection between job-related searches and unemployment rates. According to Bordino et al., there was a correlation between the number of enquiries and the volume of NASDAQ stock trading. After a thorough analysis of Google Trends data, Hyunyoung Choi and Hal Varian came to the conclusion that simple seasonal autoregressive models that used Google Trends data as input outperformed models that did not use Google Trends data by 5 to 20 percent. (J. Abraham, D. W. Higdon, H. Choi, and J. Ibarra, 2018)

According to research by Azure et al., the quantity of tweets about recently released films was a reliable indicator of box office earnings. In their analysis of current stock prices and twitter data for 30 stocks in the Dow Jones Industrial Average, Peter de Jonget et al. discovered that tweets were responsible for 87% of market returns. They also searched for data supporting the reverse hypothesis, that stock prices had an impact on tweets, but came up empty-handed.

Over the past few decades, extensive research has been conducted to forecast the stability and price of stocks and other market assets. Yet, due of it's modernity, unrelated study has been done to estimate the worth of cryptocurrencies. The amount of research being done to forecast cryptocurrency prices is, however, on the rise. We highlight crucial measures in this area in this section. Here, we expose the reader to a range of machine-learning techniques that have been applied to forecast how the prices of different currencies will change.

III. EXISTING METHODOLOGY

Based on the random walk theory, which is used to model stock prices in the financial sector. In order to replicate market volatility, it introduces multi-layered randomization into the activation of observable neural network functions. Long Short Term Memory (LSTM) and Multilayer Perceptron (MLP) models for Bitcoin, Ethereum, and Litecoin. Deep neural network stochastic layers that describe the chaotic behaviour of the financial system have been theoretically developed. It establishes the market's response to any updated information.

The tweets were typically promotional, generic, or created by a bot. Only 50% of the tweets gathered on any given day got an objective VADER score, according to Bitcoin and Ethereum. Most tweets lacked any sort of bias. Moreover, tweets that indicated whether a VADER score was objectively positive or bad scored well below the 0.5 cutoff. For all tweets with a non-neutral score larger than 0.0, the gamma kernels for objective and neutral VADER emotion scores reveal little overlap in the distributions. Although both positive and negative sentiment was gathered, as can be seen from the distribution charts for Bitcoin and Ethereum, VADER's sentiment analysis tended to classify tweets as more neutral than objective.

Similar to predicting stock prices, cryptocurrency price prediction is a typical time series problem. Conventional time series techniques have been used to predict the prices and movements of cryptocurrencies, including the well-known AutoRegressive Integrated Moving Average (ARIMA) model. Contrary to deep learning algorithms, which are more effective at predicting time-series issues, these models are unable to capture the nonlinear models of extremely complicated prediction problems.

Recall that the primary rationale for using LSTM and BiLSTM in cryptocurrency price prediction problems is that, due to their unique architectural designs, they capture potentially useful pattern dependencies of long or short sequences and contribute to the prediction performance, while the convolutional layers of the CNN model can remove noise from the raw input and extract valuable features, creating a less complex dataset that would be more useful in the model's financial analysis.

The effectiveness of the DL models was also contrasted with that of three conventional, cutting-edge ML models: the Support Vector Regressor (SVR), 3-Nearest Neighbors (3NN), and Decision Tree Regressor (DTR)

IV. DATA SETS , PERFORMANCE EVALUATION AND PARAMETER SETTINGS

Google has been offering trend data since 2014, which is an impartial sampling of search data. The information provided by Google is the Search Volume Index, not the total number of search inquiries (SVI). Each data point is divided by the total number of searches in a particular area and time period to get the search volume index.We were able to gather tweets thanks to the Twitter API. Moreover, Tweepy only allows 1500 tweets to be gathered each instance. While this generated a sizable dataset of tweets that was a representative sample of all those sent out while the software was running, it prevented us from calculating the overall number of bitcoin tweets sent out each day. Yet, www.bitinfocharts.com provides a hazy count of the daily tweets made in April 2014 regarding these two cryptocurrencies. We obtained the total number of tweets mentioning Bitcoin and Ethereum from this website. (J. Abraham, D. W. Higdon, H. Choi, and J. Ibarra, 2018)

A training set of data from January 2018 to February 2019 (10,176 values) and a test set of data from March 2019 to August 2019 were created using data from January 2018 to August 2019 at hourly rates in USD (4416 values). This information was obtained from the bitcoin exchange transaction site www.kraken.com. Four prediction horizons, 4, 9, 12, and 16 hours, were also maintained. Nevertheless, in this study, only the findings for the 4 and 9 h horizon are presented because higher horizon values were shown to be less effective.

Our initial dataset contains solely Bitcoin trade data and spans the time frame from April 1, 2013, to June 23, 2018. Seventy-seven cryptocurrencies trade against the US dollar at this time, but there aren't many with data that goes back more than a year. 2 Other cryptocurrency data's earliest start date is August 10, 2017.

V. RESULTS AND DISCUSSIONS

The most efficient and reliable models for predicting next-day lead time are support vector machines, with more than 50% consistent fit, low variability between products and across time scales, and a good ability to generalize consistently between sub-periods.

In comparison, ARIMA-based predictions usually performed poorly across all metrics. The standout conclusion is that the classification algorithms outperform both EC and ARIMA on all metrics.

VI. CONCLUSIONS

We have demonstrated that these findings are partially attributable to a research carried out while cryptocurrency values were still soaring. Also, regardless of the course of the price, Twitter users have a favourable outlook for cryptocurrencies. Furthermore, based on the various prediction horizons, the forecast accuracy fluctuates a bit. To produce reliable, even better predictive models across shorter time periods, more complicated models or more features may be adopted at a high throughput rate because to the higher sample size of minute-frequency observations. Many untapped opportunities exist in the multidisciplinary subject of stochastic processes and neural networks that may be taken advantage of in the cryptocurrency markets.

VII.REFERENCES

[1] Pintelas, E., Livieris, I. E., Stavroyiannis, S., Kotsilieris, T., & Pintelas, P. (2020). Investigating the Problem of Cryptocurrency Price Prediction: A Deep Learning Approach. In International Conference on Engineering Applications of Neural Networks (pp. 372-382)

[2] Gandomi, A. H., & Haider, A. (2018). Cryptocurrency price prediction using deep learning algorithms with technical and market indicators. Journal of risk and financial management, 11(4), 70. doi: 10.3390/jrfm11040070

[3] Bonneau, J., Miller, A., Clark, J., & Narayanan, A. (2015). Cryptocurrencies: A Brief Review. IEEE Security & Privacy, 13(4), 18-28. doi: 10.1109/MSP.2015.61

[4] Akyildirim, E., Goncu, A., & Sensoy, A. (2018). Prediction of Cryptocurrency Returns using Machine Learning. arXiv preprint arXiv:1811.07751.

[5] Shah, D., & Amjad, M. (2016). Trading Bitcoins And Online Time Series Prediction. In NIPS 2016 Time Series Workshop (pp. 1-15).

[6] Abraham, J., Higdon, D.W., Choi, H. and Ibarra, J., 2018. Cryptocurrency price prediction using tweet volumes and sentiment analysis. Journal of big data, 5(1), p.28. doi: 10.1186/s40537-018-0135-0

[7] M. Andrychowicz, S. Dziembowski, D. Malinowski, and L. Mazurek. On the Malleability of Bitcoin Transactions. In Workshop on Bitcoin Research, 2015.

[8] A. Back, M. Corallo, L. Dashjr, M. Friedenbach, G. Maxwell, A. Miller, A. Poelstra, J. Timon, and P. Wuille. Enabling blockchain ´ innovations with pegged sidechains, 2014.

[9] J. Bonneau, A. Narayanan, A. Miller, J. Clark, J. A. Kroll, and E. W. Felten. Mixcoin: Anonymity for Bitcoin with accountable mixes. In Financial Cryptography, 2014.

[10] Aloysius, N., Geetha, M.: A review on deep convolutional neural networks. In: 2017 International Conference on Communication and Signal Processing (ICCSP), pp. 588–592. IEEE (2017)

[11] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.-A.: Deep learning for time series classification: a review. Data Min. Knowl. Discov. 33(4), 917–963 (2019).

[12] Huang, J.Z., Huang, W., Ni, J.: Predicting Bitcoin returns using high-dimensional technical indicators. J. Finan. Data Sci. 5(3), 140–155 (2019)

[13] Ji, S., Kim, J., Im, H.: A comparative study of Bitcoin price prediction using deep learning. Mathematics 7(10), 898 (2019)

[14] Montgomery, D.C., Jennings, C.L., Kulahci, M.: Introduction to Time Series Analysis and Forecasting. Wiley, Hoboken (2015)

[15] Olvera-Juarez, D., Huerta-Manzanilla, E.: Forecasting Bitcoin pricing with hybrid models: a review of the literature. Int. J. Adv. Eng. Res. Sci. 6(9), 161–164 (2019)