International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-10, Issue-3, May 2022
https://doi.org/10.55524/ijircst.2022.10.3.63
Article ID IRPP1212, Pages 398- 402
www.ijircst.org

Crypto Currency Price Prediction with Machine Learning Using Python

Balisetty Naga Nikitha[1], Addanki Sudha Maheswari[2], Dudekula Shameena[3], Bandaru Poojasri[4], Heena Kauser[5], and G. Sambasiva Rao[6]

[1,2,3,4,5,6]Department of Computer Science and Engineering, PACE Institute of Technology & Sciences, Vallur, Ongole, Andhra Pradesh, India

Correspondence should be addressed to Balisetty Naga Nikitha; 18kq1a0505@pace.ac.in

ABSTRACT- We use and study a wide range of machine learning methods to predict and trade in the daily crypto currency market. We teach the algorithms to make daily market predictions based on how the 100 cryptocurrencies with the most market value change in price. Based on our research, all of the used models are able to make estimates that are statistically sound, with the average accuracy of all crypto currencies falling between 52.9% and 54.1%. When these accurate numbers are based on the 10% most confident expectations for each class and day, they go up to somewhere between 57.5% and 59.5%. A well-known case study in the field of data science looks at how people try to figure out how much different digital currencies are worth. Stock prices and the prices of cryptocurrencies are based on more than just the amount of buy and sell orders. At the moment, the government's financial policies about digital currencies affect how the prices of these things change. People's views about a crypto currency or a star who directly or indirectly backs a crypto currency can also cause a big rise in buying and selling of that currency. This study looks at the trustworthiness of the three most famous coins on the market today: bitcoin, how well buying strategies for ethereum and litecoin that are based on machine learning work. The models are checked and tested with both good and bad market situations. This lets us figure out how accurate the forecasts are in light of any changes in how the market feels between the proof and test times.

I. INTRODUCTION

Since its launch, which coincided with the global financial crisis of 2008 and the ensuing lack of trust in the financial system, bitcoin has taken centre stage in the global financial landscape, gaining the attention of regulators, governmental organisations, institutional and individual investors, academics, and the general public.

For instance, in the United States and the United Kingdom, "What is bitcoin?" was the most frequently searched Google query in 2018. (Marsh 2018). Futures options for bitcoin were added to the Chicago Board Options Exchange (CBOE) and the Chicago Mercantile Exchange (CME) in December 2017. This is another example of how the market has changed. This is a good example of how the traditional banking business tries to keep up with the market trend.

As a response to Bitcoin's quick rise in popularity and value, many new digital currencies, also called "altcoins," have been made. Most of them are very similar to Bitcoin, with only small differences. The rate at which Bitcoin's market value and price are going up is a good sign of success. (e.g., block time, currency supply, and issuance scheme). There is a big market for digital money that is not regulated. (Foley et al. 2019). At the moment, there are more than 5,700 different currencies, more than 23,000 different websites, and a market worth more than $270 billion USD.

Even though Bitcoin and other crypto currencies were created to be independent, peer-to-peer electronic payment systems, they quickly got a bad name as risky investments soon after they were created. (Nakamoto 2008). Even though their values have very little to do with the major types of financial assets and are mostly based on human behavior, the usefulness of their data is still debatable. As a consequence of this, a significant number of hedge funds and asset managers started including crypto currencies in their portfolios, while the scholarly community invested a substantial amount of effort into investigating crypto currencies trading, with a concentration on machine learning (ML) algorithms. (Fang et al. 2020). The focus of this study is on the three most famous cryptocurrencies: Bitcoin, Ethereum, and Litecoin. ML methods are used to look into how well they work and how reliable they are. So, it is an important addition to the growing number of books and articles about cryptocurrency. These three crypto assets were picked because of their maturity, shared characteristics, and prominence in terms of media attention, trading activity, and market capitalization (according to CoinMarketCap, these three assets currently account for about 75% of the total market capitalization of all crypto assets).

Because it is based on cryptography, Bitcoin can work as a peer-to-peer (P2P) virtual currency without a trusted third party, which is something that is required for the majority of other virtual currencies. This indicates that there will be no further expenditure of funds on items that aren't absolutely necessary. The technology known as blockchain is what enables bitcoin to function. It is a digital database that is distributed rather than centralized, and it records transactions that take place between users without the requirement for a third party to check or confirm these transactions. Because there is no centralized authority, it is possible to make copies of this paper without the assistance of any other users on the network. These users collaborate to ensure that the information is current by employing specialist software and working together. (Yaga et al. 2019). It is immaterial (being an electronic system based on cryptographic entities without any physical representation or intrinsic value), decentralized (does not need a trusted third-party intermediary), accessible and consensual (is open source, with the network managing the balances and transfers of bitcoins), integer (solves the double-spending problem), and transparent (information on all transactions is public). These components make up what is known as the "ecosystem" of bitcoin. i.e. ten eight hundredths of a bitcoin), resilient (the network has been shown to be resilient to attacks), pseudonymous (the system does not expose the identity of users but discloses the addresses of their wallets), and the quantity of bitcoins is capped at 21 million units.

Litecoin and Ethereum were the second and third major cryptocurrencies to make an appearance on the market when they were introduced to the general public for the first time in October 2011 and August 2015, respectively. The same mathematical theory that underpins bitcoin also underpins Litecoin, another kind of digital currency that, like bitcoin, can only ever be produced in a finite number (84 million in total). It was developed to lessen the amount of central processing unit (CPU) power required for mining in order to increase the processing times and work speeds that could be achieved. There will be occasions when you absolutely need to be able to do this. Ether, Ethereum's native currency, has an endless supply despite the fact that Ethereum, like Bitcoin and Litecoin, is a stateless network. (or ether for short, as it is more generally known in the financial writing community). In addition to this, the Ethereum system offers a method for programs to function on its public network. This indicates that anyone can use the blockchain to store and organize their own information if they choose to do so. To be more specific, it makes it simpler to create online contracts, often known as "smart contracts," that are more resistant to things like censorship, theft, and intervention from third parties. Because of this, Ethereum has risen to become the second most important kind of currency, and it also explains why it has been so popular ever since it was first introduced.

The primary purpose of this research is not to develop a superior way of machine learning, analyze the benefits and drawbacks of existing ML approaches, or determine how machine learning might be improved. It is possible to accurately forecast output based on input group factors. Instead, the primary objective is to determine whether or not the profitability of machine learning-based trading strategies, which is frequently supported by empirical research, holds true not only for bitcoin but also for ethereum and litecoin, even when market conditions change, and within a more realistic framework in which trading costs are taken into account and short selling is prohibited. In spite of the fact that some of these identical topics have been investigated in other studies, what makes our paper stand out is the depth to which we investigated these subjects. To put it another way, what makes our work unique is that we conduct it using an approach to study that considers the complete picture. An analysis of the purchase processes from a mathematical and financial perspective also provides us with evidence from the real world to support our conclusions. As a direct consequence of this, more individuals are likely to believe what we have discovered.

To make sure there is no misunderstanding, when we refer to "market conditions," we are referring to the same thing as Fang et al. (2020). According to Fang et al., it is especially likely for odd market occurrences such as booms, crashes, and other anomalies to take place with digital currencies. (2020). Alternating between periods characterized by strong optimistic markets, in which the majority of returns are in the upper-tail of the distribution, and periods characterized by strong pessimistic markets, in which the majority of returns are in the lower-tail of the distribution, is what it means for market conditions to change. When market conditions change, they alternate between these two types of periods.

II. REVIEW LITERATURE

In the early stages of bitcoin research, there was significant discussion regarding the nature of bitcoin and whether or not it should be considered a currency or merely a speculative asset. According to the authors, the general acceptance of the latter position can be attributed to a number of factors, including high volatility, disproportionate short-run gains, and bubble-like market activity. This claim is being extended to include further integrated crypto-currencies such as Ethereum, Litecoin, and Ripple at this time. Researchers were motivated to investigate the probable linkages between crypto currencies and macroeconomic and financial factors, as well as other price drivers in the investing behavior sector, due to the widespread misconception that crypto currencies are worthless speculative assets. Even in more traditional marketplaces, it has been demonstrated that these criteria are of the utmost importance. Wen et al. (2019) highlight the fact that Chinese companies that receive more notice from private purchasers tend to have a decreased chance of their stock values plummeting. This is one of the findings that they highlight in their study.

According to Kristoufek (2013), the price of bitcoin has a significant relationship with the number of page views that are received on Wikipedia and Google Trends. The findings that Kristoufek (2015) obtained corroborate the findings of earlier studies and demonstrate that fundamental parameters like the Financial Stress Index and the price of gold in Swiss francs do not show a statistically significant association with one another. According to the findings of Bouoiyour and Selmi's (2015) research on the relationship between bitcoin values and a number of different variables, including the current price of gold on the market, the volatility of bitcoin, and Google searches, the only factor that has a substantial effect at the 1% level is lagging Google searches. This was found to be the case when the researchers examined the relationship between bitcoin values and a number of different variables. Researchers Polasik et al. (2015) came to the conclusion that the trajectory of the price of bitcoin was mostly determined by three factors: the number of news, the tone of the news, and the volume of bitcoin exchanged. The authors of Panagiotidis et al. (2018) evaluate twenty-one different variables that could influence bitcoin yields, and they come to the conclusion that search traffic, as assessed by Google Trends, is one of the most important variables. New research conducted by Panagiotidis and colleagues (2019) has found that fluctuations in the price of gold have a much more positive effect on the price of bitcoin than does the volume of online inquiries. According to Ciaian et al. (2016), the factors that ultimately determine the price of bitcoin are market pressures and the attractiveness of bitcoin as an investment. In addition, it does not appear that macro-financial issues have any kind of influence over the long term. Zhu et al. (2017) demonstrate that the monthly price of bitcoin is influenced by a wide variety of economic data. Some of these indicators include the Consumer Price Index (CPI), the Dow Jones Industrial Average (DJIA), the federal funds rate (FFR), the price of gold, and the value of the US Dollar Index. According to the findings of Li and Wang (2017), the early market price of bitcoin was mostly driven by speculative investment and significantly strayed from economic principles. The passage of time caused swings in market prices to become an increasingly accurate reflection of changes in economic parameters such as GDP, inflation, interest rates, and money supply. According to Dastgir et al. (2019), the level of attention paid to bitcoin (as measured by Google Trends) is a direct cause of the cryptocurrency's extraordinary profits. When we examine the extremes of the distribution, we see that there is a correlation between these two things. According to the findings of Baur et al. (2018), bitcoin's lack of correlation with traditional asset classes such as equities, bonds, exchange rates, and commodities was observed during both times of relative financial stability and volatility. Bouri et al. (2017) observed only a weak association between bitcoin and other fundamental financial parameters, such as other major global market indexes, stocks, energy, gold, the general commodities index, and the U.S. dollar index. In addition, the researchers found no link between bitcoin and the U.S. dollar index. These authors further demonstrate that cryptocurrency is not related to other important economic indicators in their research. Pyo and Lee (2019) came to the conclusion that the value of bitcoin does not appear to be tied to the U.S. job rate, PPI, or CPI statements; rather, the value does appear to react to comments made by the FOMC regarding monetary policy. Please provide citations for the following paraphrase from: Please provide citations for the following paraphrase from:

Li and Wang (2017) investigated the theory that bitcoin's value is mostly driven by its popularity, and they discovered that this was true of other digital currencies as well. Their findings support the concept. This was measured by looking at comments from online groups and news outlets, as well as from inquiries submitted to Google and Wikipedia, postings made on Twitter and Facebook, and discussions in speciality forums. For example, Kim et al. (2016) analyzed user comments and replies in online crypto currency groups to anticipate changes in the daily values and trades of bitcoin, ethereum, and ripple. They found favorable findings, particularly for bitcoin. We did this by reading comments from our customers and responding to questions expressed in internet discussion boards. Phillips and Gorse (2017) develop profitable trading techniques for a variety of cryptocurrencies by employing hidden Markov models that are constructed from data taken from online social media platforms. Corbet et al. (2018b) discovered that there were no correlations in either the time or frequency domains between bitcoin, ripple, and litecoin and a wide range of economic and financial issues. According to the findings of Sovbetov (2018), the market beta, trading volume, volatility, and perceived appeal are some of the elements that influence the monthly prices of bitcoin, ethereum, dash, litecoin, and monero. Other aspects include supply and demand. Phillips and Gorse (2018) investigate the ways in which internet and social media factors are related to the prices of bitcoin, ethereum, litecoin, and monero. Specifically, the authors focus on the linkages between these aspects. They make the interesting discovery that short-term associations appear to be influenced by particular market events, such as hacks or security breaches, whereas medium-term positive correlations rise dramatically during bubble-like regimes.

The potential for irrational behavior in the bitcoin market, such as swarming, has generated a number of studies, like the one that was conducted by Stavroyiannis and Babalos (2019).

Gurdgiev and O'Loughlin (2020) investigate the relationship between the price fluctuations of 10 cryptocurrencies and proxies for fear (the VIX index), uncertainty, investors' sentiment toward cryptocurrencies (measured using opinions posted by investors in a bitcoin forum), and investors' perceptions of bullishness or bearishness in the overall financial markets (measured by the CBOE put-call ratio). They emphasize that cryptocurrencies can be used as a hedge against ambiguity, but not against fear, and that investor opinion is a powerful indicator of the direction that the cryptocurrency market will take. A number of the compositions concentrate on just this subject. The findings imply that buyers of crypto assets are subject to grouping biases, and they show that anchoring and recency biases, to the extent that they do exist, are non-linear and reliant on the context in which they occur. Chen et al. (2020a) evaluate the impact of dread mood on Bitcoin prices and show that increased fears about the spread of the coronavirus have resulted in negative returns and large trading volume. This research follows a similar line of thinking as the previous two studies. This one also appears in the aforementioned paper, and it is comparable to the research that was cited earlier. The authors reach the conclusion that bitcoin behaves in a manner that is more comparable to that of other financial tools during times of market distress (such as during the coronavirus pandemic), and that it does not function as a safe haven during these times.

III. METHODOLOGY

In this study, both the rates of main cryptocurrencies and trading approaches based on machine learning are analyzed. Cambridge studied. The system is comprised of linear models, RFs, and SVMs respectively. (SVMs).

These models are used for regression as well as trading suggestions for binary options. The income from trading cryptocurrency is dependent. (classification models).

Regression forests (RFs) are a hybrid of regression and classification trees. The regression RFs make predictions about the yield, whereas the classification RFs provide predictions about the price. The regression RFs make predictions about the yield, whereas the classification RFs provide predictions about the price. Profit may be predicted using regression RFs, whereas price change can be predicted using classification RFs. RFs are built upon the base of regression and categorization trees. The space occupied by the independent variables can be divided into subspaces using simple regression or classification trees. Both options are valid. The structure of the tree is reconstructed via projection, starting at the root node. Following that, the tries and branch choices continue until they reach the leaf node. It provides a definition of the outcome-dependent projection of the variable. (the forecast for the next return or the binary signal that predicts whether the price is likely to climb or fall the following day). RF trees vary. The branch test for each tree node is determined by a random split of the independent factors and the observations from the training dataset. to select the optimal course of action. Following this stage, the RF forecasts are either the binary signal that was selected by the majority of trees (referred to as the regression RF), or the average of the trees' guesses (referred to as the classified RF). (in the event that a categorization RF is involved).

Classification and error correction are also performed using support vector networks. (SVMs). The binary classification hyperplanes are located by SVMs. This gap, which is defined as the sum of the shortest lengths to the nearest data point for both categories, widens as the precision of the categories improves. (Yu and Kim 2012). There is room for classification errors thanks to slack parameters that gauge perplexity as well as a margin size-error measure. Possible misclassification. Misclassifications are sometimes made.

The data for time series used in machine learning are typically divided up into training, validation, and test sets. The test set is responsible for the analysis of the outcomes, whereas the training set is responsible for making predictions about the models. Models can be predicted using training, validation, and test sets. Training data estimates models. The primary concerns that need to be addressed when it comes to identifying the various data segments for this task are as follows: on the one hand, to eliminate any and all dangers associated with data surveillance; and on the other hand, to ensure that the findings acquired in the test set can be considered representative. Both of these concerns need to be addressed in order to successfully complete this task. The dataset is cut in half to create two subsamples of equal length using this procedure. The main purpose of the "training" subsample is to develop early models by "fitting" model parameters to data as part of the process. Its one and only purpose gave rise to its eponymous moniker. After then, the remaining fifty percent is divided into two subsamples: one for validation (twenty-five percent), and one for testing (twenty-five percent). The data for these subsamples are comparable. The validation sub-sample is used to select the model that works best for each group, and the test sub-sample is used to evaluate income and estimates. The best model is determined by both of the subsamples.

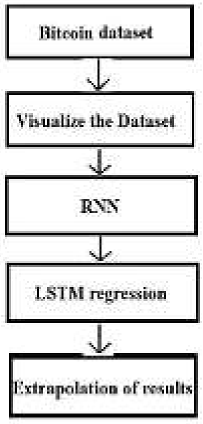

Figure 1: Work flow of the proposed model

In order to forecast the price of Bitcoin in the future using data mining and deep learning methods, we apply a Neural Network (RNN) model trained with the Long Short-Term Memory (LSTM) regression method to the collected collection of digital currencies. This allows us to make accurate price predictions. Predictions of the price of Bitcoin based on a Neural Network (RNN) model that was trained using the Long Short-Term Memory (LSTM) regression method using an existing database of cryptocurrency values. In order to compile the data that was used in this study, numerous characteristics that are linked to Bitcoins' data values were utilised. The overarching purpose of this research is to devise a formula for forecasting the price of bitcoin in the future with a level of precision that is superior to that of existing methods over an extended period of time. Making precise forecasts about the market is an endeavor that is fraught with difficulty. As a result of this, we have made the decision to simplify matters by basing our price projection on only the following three possible outcomes: an increase, a reduction, or a continuance of the level that is now being seen. The analysis of forecasts will be based, in part, on the future numbers that will be produced by the various programs. The development of a system that is able to make accurate forecasts regarding the value of bitcoin in the future is the primary objective of the suggestion to incorporate RNN components into the model.In order to forecast the price of Bitcoin in the future using data mining and deep learning methods, we apply a Neural Network (RNN) model trained with the Long Short-Term Memory (LSTM) regression method to the collected collection of digital currencies. This allows us to make accurate price predictions. Predictions of the price of Bitcoin based on a Neural Network (RNN) model that was trained using the Long Short-Term Memory (LSTM) regression method using an existing database of cryptocurrency values. In order to compile the data that was used in this study, numerous characteristics that are linked to Bitcoins' data values were utilised. The overarching purpose of this research is to devise a formula for forecasting the price of bitcoin in the future with a level of precision that is superior to that of existing methods over an extended period of time. Making precise forecasts about the market is an endeavor that is fraught with difficulty. As a result of this, we have made the decision to simplify matters by basing our price projection on only the following three possible outcomes: an increase, a reduction, or a continuance of the level that is now being seen. The analysis of forecasts will be based, in part, on the future numbers that will be produced by the various programs. The development of a system that is able to make accurate forecasts regarding the value of bitcoin in the future is the primary objective of the suggestion to incorporate RNN components into the model.

IV. LOGISTIC REGRESSION

The logistic regression model is frequently utilized as a standard because of the ease with which it can be trained and because of its overall simplicity. Both LR and basic linear regression make predictions about the possibility of a binary event happening based on a linear mixture of a collection of factors. For categorization problems that involve binary response variables, LR is comparable to basic linear regression because both techniques make these predictions. In contrast to traditional linear regression, which has a closed-form solution available, the convexity of the loss function makes it possible for the global optimal to be located quickly and easily through the use of computational techniques. Due to the fact that it has a unique response and is not improved using a haphazard process, the LR model is the only model that is used for inference that does not require initially constructing a collection of individual models obtained with various seeds. This is because the LR model is the only model that has a distinct response.

When binary cross-entropy is used as the loss function, the LR model can also be represented by a straightforward one-layer neural network consisting of a single neuron and a sigmoid activation function. This is possible because the LR model is so straightforward. In order to answer the optimization problem, Scikit-learn's LR is applied; the Newton-CG learning algorithm is utilized, and the maximum number of repetitions that are permitted is 1000. We do not make any changes to any of the other hyperparameters beyond the baseline settings.

V. CONCLUSION

To determine the best parameters for each model type, we take the average deal yield from our validation sample. These profits are the result of a day-trading strategy that takes into account the direction of the return forecast (in the case of regression models) or the binary prediction of an increase or decrease in price (in the case of classification models), both of which are obtained in a rolling-window framework. This study contributes to the growing but still modest body of literature on machine learning in bitcoin, which may prove useful to researchers and practitioners. In particular, it examines the market turmoil that has occurred since the middle of 2017 and the bear market that followed, it employs trading variables as well as network variables as important inputs to the data set, and it offers a comprehensive statistical and economic analysis of the examined trading strategies in the market for crypto currencies. Even though values in the evidence phase rise sharply before falling suddenly and sustainably, the average return remains positive.

The survey results are more reliable, notwithstanding the negative average yield.

You can see how well your trading strategy holds up under stress by putting it through a series of tests. There is no consistent trend during proof and test durations to identify the best model or currency, and the reliability of projections varies widely depending on the model and crypto currencies used. When compared to other research, the models' predictive accuracy is poor. The best program in its category maximizes average earnings from one step forward rather than minimizing the number of mistakes it makes. The most noticeable pattern is the decline in forecast accuracy between test and validation. This may be due to the drastically divergent price trends of the two eras.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] A Prototype for Grievance Redressal System (2018): Shaligram Prajapat, Vaibhav Sabarwal and Varun Wadhwani.

[2] Online Grievance Redressal System (2019): Mukesh Buldak, Shrikant Pandhekar, Afzal Gigani, Amreshsinh Kachwah.

[3] Complaints and Grievance Management System: prof. H. N. Renushe, prof. Mrs S. V. Deshmukh, Prof. Mrs. D. Y. Jadhav

[4] Online Complaint Management System (2015): Osmannasr, enayatalkhider.

[5] Devika radha krishnan, Nisarggandhewar

[6] Subhash, C., Ashwani K.: Assessing grievances redressing mechanism in India. Int. J. Comput. Appl. 52(5), 12–19 (2012)

[7] Dipankar, M.: Solution towards effective complaint registration system in Indian scenario. In: IJCA Proceedings on National Conference on Advancement of Technologies—Information Systems & Computer Networks (ISCON—2012), vol. 1, pp. 1–2 (2012)

[8] Prasadu Peddi (2019), Data Pull out and facts unearthing in biological Databases, International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.

[9] Nripendra, P.R., Yogesh, K.D., Michael, D.W., Vishanth, W.: Investigating success of an e-government initiative: validation of an integrated IS success model. Inf. Syst. Front. 17(1), 127–142 (2015)

[10] Rajesh, K.Y., Sarvesh, M.: Role of insurance ombudsman and grievance management in life insurance services in Indian perspective. Int. Lett. Soc. Humanist. Sci. 31, 9–13 (2014).

[11] Varun, G.: Redressing grievances and complaints regarding basic service delivery. World Dev. 41, 109–119 (2013)