

The price prediction for cryptocurrency based on the state-of-art machine learning approaches

Ying Guan

Smeal College of Business, Pennsylvania State University, State College, United States

Corresponding Author: ymg5129@psu.edu

Abstract. Cryptocurrency has evolved from a fringe phenomenon to a far more popular method of investing and financing. For investors and traders, predicting the price of bitcoin is critical. Several machine learning algorithms are utilized to anticipate the price of digital money in this research paper. The analysis employed Decision Trees, Light Gradient Boosting Machines, and Neural Networks. The purpose of this study is to look at the predicted accuracy of each machine learning method. According to the analysis, decision tree, lightGBM, and neural networks have a very high accuracy rate when it comes to forecasting cryptocurrencies. These results shed light on guiding further exploration to help investors in building an appropriate digital currency portfolio and reducing risks.

Keywords: Cryptocurrency, machine learning, decision tree, light gradient boosting machine, neural network.

1. Introduction

The concept of cryptocurrency was initially established in 2009, when Satoshi Nakamoto established the world's first digital currency, Bitcoin, which is still widely traded today [1]. Every cryptocurrency in the market is a tradable digital asset which builds on blockchain technology that only exists online or people usually call it virtual currency. Contemporarily, cryptocurrency has piqued the curiosity of investors, experts, and the general public. A fundamental change in the model of transactions from physical payments (e.g., cash) and checks to digital transactions has occurred in a constantly expanding technology context. The cryptocurrency market has evolved irregularly and at an unparalleled rate over its short lifespan. In 2021, the cryptocurrency industry is valued more than \$3 trillions. By 2026, the cryptocurrency market is expected to have grown from \$1.6 billion to \$2.2 billion USD [2]. Along with the most well-known cryptocurrency, Bitcoin, there are over 7,000 altcoins on the market. By November 2021, the total cryptocurrency market capitalization will exceed three trillion dollars. Cryptocurrency is a digital currency that functions as a medium of trade through computer networks that are not maintained or supported by central authorities such as governments or banks. Cryptocurrency is now a topic of debate in a wide variety of daily financial talks, ranging from investments and portfolio management to incomes and purchases. According to the data from Statista, there are currently more than 9,900 cryptocurrencies in the market around the world. The top three cryptocurrencies with the highest market capture are BTC Bitcoin, ETH Ethereum, and USDT Tether.

The development of theoretical models of cryptocurrencies has received considerable scholarly interest. Predicting the anticipated value of money, whether as a means of exchange or an asset, is a crucial aspect of its use. Using vast quantities of publicly accessible data on the cryptocurrency market and social trends, machine learning might be used to predict the price of cryptocurrencies. In recent years, a number of researchers have examined the use of machine learning and deep learning algorithms to predict the value of stocks and cryptocurrencies. Patel et al. presented a study on the prediction of cryptocurrency prices, in which they mostly used stochastic neural networks, Multilayer Pervrptron, and Long Short-Term Memory models [3]. The Bitcoin market forecast is another intriguing issue that people are eager to investigate. Patrick et al. authored a study on the forecast of the short-term bitcoin market using a logistic regression model as a benchmark [4].

Ever since their emergence, cryptocurrency market has been regarded as one of the most volatile and fluctuating markets in the world. Bitcoin, one of the most prominent examples of cryptocurrency

in the current market, is a digital asset that lacks a physical embodiment and a global product that is regulated worldwide. Nevertheless, a huge number of individual investors nowadays are readily and blindly manipulated by “false” trends, leading to irrational investment actions. Additionally, the Bitcoin market is unregulated by the government. Regulators that oversee conventional financial markets are mostly absent from the world of cryptocurrencies. All these factors contribute to Bitcoin’s significant price fluctuation. To explain the explanation for bitcoin price shift, bitcoin price forecasting models often require well-designed features, which is a difficult process. Previous research has been proven to be successful on forecasting the price of Bitcoin using two methods: empirical analysis and analysis of robust machine learning techniques. By analyzing the specifics of previous occurrences, machine learning systems and models that generate predictions based on training data could be developed. Such algorithms can not only be implemented for the Bitcoin market, but also for the entire cryptocurrency market. This paper will investigate cryptocurrency price prediction using the state-of-art form of machine learning models including Decision Tree, Light Gradient Boosting Machine, and Neutral Networks, this paper examines and analyzes the future price prediction, as well as assessing the accuracy of each model. As many investors are concerned with whether the quick sugar or fall is worth following, my study would provide investors with insights and options to make the best decision.

2. Status of Cryptocurrency

In the 1980s, the developer of cryptocurrency, Satoshi Nakamoto, sought to create a currency that did not need centralized organizations like banks and was also untraceable. In October 2008, Satoshi Nakamoto, a pseudonym used in the released whitepaper, established the first blockchain database [1]. The Genesis Block was the first block in the Bitcoin network and it contained fifty bitcoins. During the first few months of its existence, Bitcoin was almost worthless. Six months after its debut to the market, in April 2010, a single Bitcoin was worth less than 14 cents. It increased to 36 cents between May and early November before stabilizing at 29 cents [5]. From 2010 to 2014, the price of Bitcoin skyrocketed, notably in 2013 when it surged from \$13.30 in January to \$946.92 in December [6]. Bitcoin's price increased yearly at a rapid pace. Bitcoin reached an all-time high of just less than \$20,000 at the end of 2017, and 2017 will be known as the Bitcoin year. As shown in Figure. 1, the price for Bitcoin is about \$20,000 in the beginning of 2021, and increased to its highest point at \$65,466.84 in November.



Figure 1. Bitcoin price candlestick collected from Yahoo Finance.



Figure 2. Dogecoin price candlestick collected from Yahoo Finance

In the era of Bitcoin's grandeur, a new blockchain project named Ethereum arrives on the market as the number two cryptocurrency. Vitalik Buterin initially defined Ethereum in a 2013 whitepaper. Ethereum is a decentralized, open source blockchain system with its own cryptocurrency, Ether [7]. Dogecoin is another well-known digital currency; it is a Bitcoin-like cryptocurrency made famous by the Shibainu dog meme; it is one of the most bizarre and charming efforts to create money. It was said that its community was friendlier than those of other cryptocurrencies. In many ways, the history of Dogecoin is a microcosm of the history of cryptocurrencies. Dogecoin is identical to Bitcoin, and mining is used to protect Ethereum. Thousands of computers compete throughout the globe to mine Dogecoin in order to distribute the millions of additional coins added to the supply daily and to verify and execute network transactions. Dogecoin is far more susceptible to social media sentiment than the other two currencies. Elon Musk, co-founder and chief executive officer of Tesla, is also a renowned and successful investor. His investment affects not just individual investors but also the market environment as a whole. One of his Dogecoin-related tweets, "Doge barking at the Moon," received around 20,800 comments, 52,300 retweets, and 314,100 likes. Contemporarily, anyone with a good graphics processor can mine Ethereum on their computer, but it is unlikely to be profitable. Small-scale miners compete with larger companies utilizing sophisticated software, much as they do with Bitcoin. In 2022, Ethereum mining will be phased out in favor of a method known as proof of stake. Figure 2 depicts the price fluctuation of Dogecoin as a result of Elon Musk's actions.

The global cryptocurrency market was estimated at \$1,782 billion in 2021. The market is anticipated to reach \$32,420 billion by 2027, expanding at a CAGR of 58.4% between 2022 and 2027. Increased sector-wide digitization is one of the primary drivers driving market growth. Consequently, simple access to and increasing penetration of high-speed internet connection in everyday activities contribute to the market's optimistic outlook [8].

3. Machine Learning Approaches

There are various of machine learning algorithms models that could be used to forecast cryptocurrency prices. This research paper will employ decision trees, LightGBM, and neural networks as my major algorithm models. These are three well-known and well-developed models. Decision tree is one of the simplest and often used classification techniques [9]. The most appealing decision tree technique attributes are its efficiency and robustness, as well as its simple form. The most valuable advantage of a decision tree, according to Quinlan, is that it can be simply comprehended after generating a prediction [9]. Figure 3 shows what the decision tree looks like, it looks like this flowchart. In the figure below, the major factor which affects the decision is whether to go to the cinema or not. Therefore, if the answer is yes and we choose to go to the Cinema. If the answer is no, this opens up an array of other conditions. If the Weather is sunny, one will play tennis. If it is windy, one is going to the shopping mall, and one will stay at home if it is rainy.

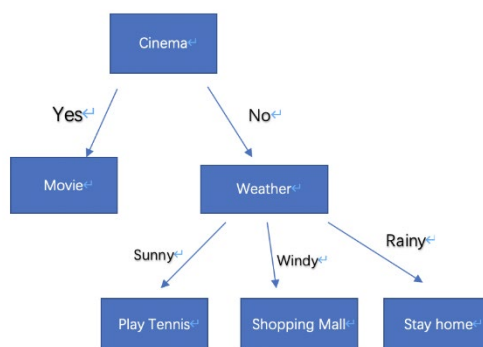


Figure 3. A sketch of Decision Tree.

Light Gradient Boosting Machine (LightGBM) is a gradient boosting framework that is very useful when used for computation. It uses a tree-based learning method, which is thought to be a quick-processing algorithm. Nowadays, the size of dataset has increased steadily, it has become more difficult for traditional algorithms to calculate the result very fast. As LightGBM has very strong computation power, it can provide quicker results compared to traditional algorithms. LightGBM uses less memory to operate the data and can handle massive volumes of data at the same time. LightGBM is very sensitive, it can easily overfit small amounts of data. It is very useful when it comes with a large number of datasets, it can be used with more than 10,000 rows of datasets [10]. There is no predetermined threshold for determining whether or not to use LightGBM, it can be used for small and large amounts of data, especially when great accuracy is required.

Neural network is a machine learning approach that is inspired by the human brain, it is a computer system made up of linked nodes that function similarly to nerves in the brain. Neural networks are very useful for stock price prediction. There are several types of neural networks such as RNN, CNN, and LSTM. RNN stands for recurrent neural networks, it is a type of artificial network that analyzes data sequences. These algorithms are made to handle a large number of inputs with no size limit [11]. RNN has multiple series predictions. Time series prediction is a type of application for RNN. Any time series forecasting problem can be solved using an RNN, such as prediction the prices of stocks or cryptocurrencies in a particular month or year. CNN stands for Convolutional Neural Network, the higher performance of convolutional neural networks with image, speech, or audio signal inputs sets them apart from conventional neural networks. Long Short-Term Memory (LSTM) is a type of RNN that can deal with long-term dependencies. The LSTM is a type of sophisticated RNN that permits information to be retained. It can solve the vanishing gradient problem that RNNs have.

4. Application

Mounika et al. collected Bitcoin data from 2011 to 2021 to make a 5-day price estimate [12]. For training and testing their model, they used a CNN and a LSTM. The objective was to determine which of these two models was more successful. The deep learning algorithms adopt a neural network structure with three layers, similar to that of the human brain: input, hidden, and output. Using Keras and Tensor flow, the author created a deep learning model using Sequential class model. Keras is a software program used to provide an interface for constructing artificial neural networks. Approximately 80% of the dataset is utilized for training, while the remaining 20% is used for testing. The Train set is used to train the neural network model, while the test set is utilized to produce predictions that are then compared to the actual data to evaluate accuracy and error metrics. This model operates so as to generate a time- or attribute-dependent behavior model. Back propagation is accomplished by sending back the neural network's findings at time t and the input network's results at time $t+1$. The best network for predicting stock prices is a recurrent neural network. Price prediction results by using CNN are shown in Figure 4 and Bitcoin price forecast through CNN is shown in Figure 5. According to the analysis, the CNN surpasses the other in price prediction.

Swamy et al. evaluated the future Bitcoin price using Linear Regression, Lasso Regression, and Decision Tree [13]. The accuracy of the decision tree among these models is 98.3895 percent. Using an algorithm that discovers several methods to partition a data set based on certain variables, decision trees are constructed. Besides, LightGBM was described in detail and this model was used to anticipate the price of digital currency [14], where 42 varieties of cryptocurrencies were used. And gathered information from January 1 to June 30, 2018. In the LightGBM algorithm, gradient-based one-side sampling and exclusive feature bundling are other unique strategies. Before employing LightGBM, the author of this work employed hyper-parameters to identify the quantity and range of variation, since hyper-parameters play a vital role in accuracy prediction. This study found that the usage of LightGBM is optimal for medium-term forecasting, such as 2 weeks, and that the greater the cryptocurrency's overall strength, the better its predictive performance. In terms of durability, the LightGBM is an exceptional forecasting model when dealing with a large number of data instances and attributes simultaneously.

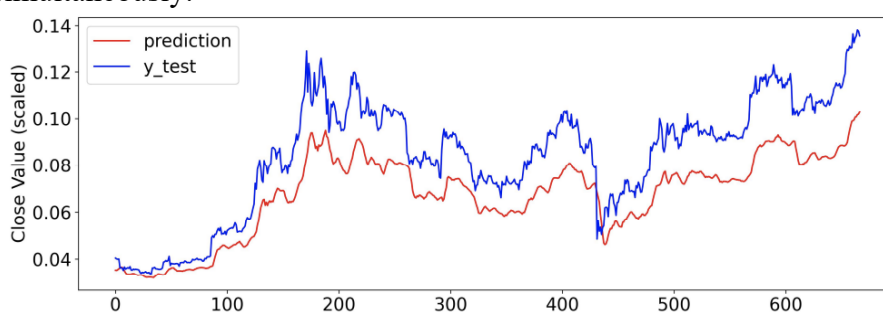


Figure 4. Bitcoin price prediction through CNN.

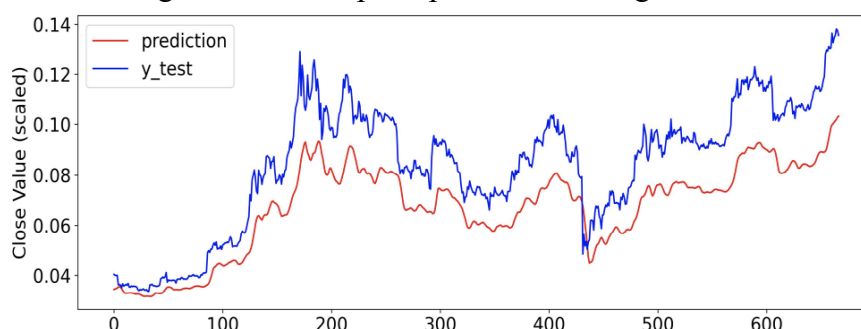


Figure 5. Bitcoin price prediction through LSTM model.

5. Limitation and Future Prospect

In principle, it is the intrinsic nature of cryptocurrency that it is not issued by any central authority, renders it immune to government intervention or manipulation. In this case, it is undoubtedly a unique characteristic that gives cryptocurrencies their most distinctive appeal. Meanwhile, it also creates enormous limitations of price prediction in the cryptocurrency market. Although cryptocurrency processes the potential of becoming a dominant force in the global system of payments, it is required to provide distinctive added value while also addressing and surmounting a number of formidable obstacles. The future of cryptocurrency is not quite foreseeable due to various factors including large data pools, taxability of transactions, external influences, liquidity risk, limitation on trading activity, as well as formal regulatory problems. The tedious amount of unfiltered data could potentially increase error rate and misdirect the research conclusion. As a “dependent” currency, it cannot escape being interacted with and affected by external influences such as the news and social media. There is currently no explicit legislation or guidance from governments regarding the taxability of transactions using cryptocurrencies. In the United States, many exchanges implement restrictions to limit the daily trading activity of their customers, therefore diminishing the liquidity pool against the cryptocurrency trading market. Future law regulation and newly-implemented policy may have an impact on the nature of cryptocurrency itself and thus impact the price prediction accuracy.

In future work, scholars ought to investigate more comprehensively on other factors that might affect the price fluctuation of the cryptocurrency market. Besides, researchers should intend to focus more on the effect of news and social media in general can have on the price and trading volume of cryptocurrencies by analyzing social media outputs based on sentiment analysis techniques. In addition, researchers should work on the construction of a new algorithm that has better efficiency on processing high-frequency data, as well as possessing the capability of filtering out unnecessary data and has better efficiency on processing high-frequency data sets directly target on valid data to generate more precise results.

6. Conclusion

In summary, this paper investigates the price prediction on cryptocurrency based on Decision Tree, Neural Networking, and LightGBM. The forecasting using the Decision Tree is about 98%, which is a high predicting accuracy rate among other models. On the five days price prediction through CNN and LSTM models, the experimental results show that convolutional neural networks outperform the other on price prediction. LightGBM is a model that is suitable for medium term forecasting, it shows a higher accuracy rate when forecasting the 2 weeks' price. Overall, these results offer a guideline for future investors and researchers, these machine learning approaches will help everyone when they are trading, and the accuracy rate will be increased in the future.

References

- [1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system[J]. Decentralized Business Review, 2008: 21260.
- [2] Coleman, Sara. "Buying a Car with Cryptocurrency 2022." Bankrate, Information on: <https://www.bankrate.com/insurance/car/buying-car-with-crypto/>.
- [3] Jay P, Kalariya V, Parmar P, et al. Stochastic neural networks for cryptocurrency price prediction[J]. Ieee access, 2020, 8: 82804-82818.
- [4] Jaquart P, Dann D, Weinhardt C. Short-term bitcoin market prediction via machine learning[J]. The journal of finance and data science, 2021, 7: 45-66.
- [5] Mukhopadhyay U, Skjellum A, Hambolu O, et al. A brief survey of cryptocurrency systems[C]. 2016 14th annual conference on privacy, security and trust (PST). IEEE, 2016: 745-752.
- [6] "Bitcoin Historical Prices." Bitcoin Price Chart and Tables | Finance Reference, Information on: <https://www.in2013dollars.com/bitcoin-price>.
- [7] "Ethereum Price Today, ETH to USD Live, Marketcap and Chart." CoinMarketCap, Information on: <https://coinmarketcap.com/currencies/ethereum/>.
- [8] "Global Cryptocurrency Market Report 2022-2027 - Industry to Cross a Staggering \$32.4 Trillion by 2027, Exploding with a CAGR of 58.4%." Yahoo! Finance, Yahoo!, Information on: <https://finance.yahoo.com/news/global-cryptocurrency-market-report-2022-120800380.html>.
- [9] Quinlan J R. Induction of decision trees[J]. Machine learning, 1986, 1(1): 81-106.
- [10] Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree[J]. Advances in neural information processing systems, 2017, 30.
- [11] Dhruv P, Naskar S. Image classification using convolutional neural network (CNN) and recurrent neural network (RNN): A review[J]. Machine Learning and Information Processing, 2020: 367-381.
- [12] Mounika S, Yadav P A, Yashaswi T, et al. Crypto-Currency Price Prediction using CNN and LSTM Models[J]. Preprint academia.edu, 2021.
- [13] U Raghavendra Swamy, D Rupini, G Shravya, V Akhila. Crypto-Currency Price Prediction Using Machine Learning Techniques [J]. JETIR 2021, 8(6): 2349-5162.
- [14] Sun X, Liu M, Sima Z. A novel cryptocurrency price trend forecasting model based on LightGBM[J]. Finance Research Letters, 2020, 32: 101084.