Research on Price Prediction of Digital Currency Based on Machine Learning

Lirui Liu[1,a,*]

[1]Upper School, Ranney School, New Jersey, NJ 07724, USA
a. 2024liul@myranney.org
corresponding author

Abstract: Since the invention of digital currency, the social recognition and demand for special currency and similar cryptocurrencies have increased significantly with the development of digital currency and blockchain technology. The rapid rise in the price of digital currency and its significant volatility attracts a large number of users to invest in it as a digital asset. Before the formation of a regulatory strategy with a standardized system, the development of digital currency will undoubtedly have an increasing impact on society, and its price fluctuation will become an unstable factor in society by increasing the risk to users. Therefore, finding out the factors that affect the price of digital currency and forecasting its price has become the focus of research on Bitcoin in recent years. This will not only help investors and relevant institutions understand Bitcoin and the digital currency market but also help improve the financial market and its policies. In this paper, I use a machine learning model to predict the price of digital currency from both numerical and trend aspects. The main contents include the following elements.First, in terms of data characteristics, I comprehensively summarized and referred to the previous research on the price of special currency at home and abroad. Second, in terms of model prediction, I designed a two-stage feature processing method, used three kinds of recurrent neural network models to compare and predict the price of digital currency, and used recursive feature elimination (RFE) and logical regression (LR), random forest (RF), linear discriminant analysis (LDA) Support Vector Machine (SVM) and Naive Bayes (NB), five commonly used machine learning models, are combined to predict the price trend of Bitcoin.

Keywords: digital currency, price forecasting, machine learning, artificial neural network

1. Introduction

Since the birth of Bitcoin, a digital currency in 2007, it has attracted more and more attention due to its characteristics of decentralization, strong privacy, and constant total amount. As the first decentralized digital currency, Bitcoin has the characteristics of low transaction fees, anonymity, and high autonomy. With the increase in the number of digital currency users, the digital currency network can only handle up to seven transactions per second, making the digital currency network more congested. Slow and impractical transfer: the design of Bitcoin, such as intelligent contract expansion, is becoming increasingly prominent. This makes the competitive currency of digital currency develop rapidly. Since 2017, with the decentralized financial ecology of Ethereum smart contract technology, the number of tokens in the digital currency market has increased rapidly. As of August 2021, the

number of digital currencies in BINANCE, the digital currency exchange with the highest turnover in the world, has reached 320. In addition to numerous investment targets, the digital money market has now derived applications with mortgage lending, exchanges, insurance, games, and other scenarios. The rapid development of the digital currency market has attracted many institutional investors. The scale of digital currency currently managed by the digital asset management company Grayscale is worth 45 billion dollars. Huo Coin Technology obtained the No. 9 license issued by the Hong Kong Securities Regulatory Commission, becoming the first listed company in Hong Kong to be approved to issue digital currency funds. In April 2021, Huo Coin Technology issued three digital currency investment funds, and the asset scale under its management has exceeded 1 billion dollars. In addition to professional asset management companies, the digital currency has also begun to become part of the asset allocation of listed companies. According to incomplete statistics, more than 20 listed companies currently hold digital currency, with a total market value of more than 9.8 billion dollars. Although institutional investors poured into the digital currency market, the sharp fluctuations in the price of digital currency brought considerable challenges to investors. From April 2021 to May 2021, Bitcoin fell from a high point of $65000 to a low point near $30000, with a price drop of 53% in just 40 trading days. Based on this, it is of great practical significance to study the factors affecting digital currency's price.

2. Literature Review

2.1. Research on the Price Forecast of Digital Currency

Foreign research on digital currency price prediction: Çayir et al. used PROPHET and ARIMA methods to predict bitcoin prices [1]. Based on the selected bitcoin price series from May 2016 to March 2018, they compared the prediction effects of the two models, and the final results showed that PROPHET was better than ARIMA. Saad et al. used a multiple regression model to predict the price of Bitcoin using data before 2017 [2]. The author also uses the price data of other cryptocurrencies (Ripple, Litecoin, Dash, Ethereum) and provides the relevant matrix of their prices through the Pearson method.

Zhengyang et al. analyzed and predicted Bitcoin price data from August 2015 to June 2019 through the ANN-LSTM model. At the same time, the author used several experiments to conduct joint experiments on multiple cryptocurrencies [3]. The author believes that the model error is related to the scale of prediction time. Livieris et al. used a hybrid model of convolutional neural network (CNN) and LSTM to predict the prices of the three cryptocurrencies with the highest market capitalization: Bitcoin (BTC), Ethereum (ETH), and Ripple (XRP). The MICDL model consists of 3 convolution layers and 16 filters. The experimental results show that MICDL is a performance prediction aspect [4].

The most appropriate model. Jaquart et al. focused on short-term trend prediction of the cryptocurrency market and tried to solve this problem using LSTM and GRU (Gated Cyclic Unit) methods [5]. The model realizes feedforward, LSTM, and GRU, artificial neural network technology based on tree and random forest, and finally believes that RNN and GBC models are most suitable for forecasting the short-term bitcoin market. Biswas et al. proposed a new method to predict the value of a digital currency based on the LSTM model by considering several variables, such as stock market value, quantity, distribution, and high-end delivery [6]. The actual prediction results are effective. Li proposed a new support vector regression (SVR) based on data segmentation and compared it with the prediction effect of some commonly used machine learning algorithms [7]. The results show that the new model can significantly improve the accuracy and effectiveness of virtual currency price prediction compared with the standard models. Jiang predicted cryptocurrency prices through the KNN algorithm and improved the KNN model according to the short-term price trend

before the trading day [8]. The experimental results show that the improved KNN model has more accurate prediction results than the logistic regression model and the traditional KNN model.

2.2. Domestic Research on Digital Currency Price Prediction

Li Jing predicted the price of special currency through the BPNN model. Divide the price series data from August 2010 to March 2016 into multiple periods, and build models for each period to forecast [9]. The research results show that the closer the data is, the more accurate the prediction results will be, and the short-term prediction effect will be better than the long-term prediction effect. Li Yinglu passes through BPNN [10].

The model predicts the cryptocurrency index. By building a three-layer improved BPNN network to analyze and forecast the daily dimension data from March 2018 to March 2019, the final results show that the improved model prediction results are good at most locations except for individual points. He Xiongwei et al. predicted this by comparing the LSTM model with the particular currency price series [11]. Through comparison of the forecast data in the past 30 days, it is found that LSTM has a certain value in forecasting the trend of Bitcoin, but it needs to be more accurate in forecasting the price. Zhao Lei et al. The PEC coefficient, VAR, and LPPL models are used to predict the price foam of Bitcoin [12]. The final experimental results show that the PEC coefficient can reflect the size of the foam to some extent, And the results predicted by the LPPL model are very similar to the actual price foam. Zhang Ning et al. The results show that the model is better than the single model [13].

3. Price Forecast Model and Evaluation Index

3.1. Price Forecast Model

The deep learning model is a kind of artificial neural network model (ANN), which emphasizes the number of hidden layers of the model and is a complex algorithm. With the continuous development of artificial intelligence technology, deep learning is widely used in speech recognition, automatic driving, image recognition, etc. Among them, the financial field is also a significant application scenario of deep learning.

In the traditional stock market, the deep learning model represented by the recurrent neural network is often used. Bai Wankuan used three recurrent neural network models (RNN) to predict the Hong Kong Hang Seng Index [14]. As a new variant of the RNN model, the CW-RNN model achieved the best effect among the three models. Long-term and short-term memory model (LSTM) is an essential variant of recurrent neural networks. Bao Zhenshan et al. Convolutional neural network (CNN) is a commonly used model for deep learning [15]. Gunduz et al. used the CNN model to predict stock prices [16]. Geng Jingjing et al. combined the CNN model with LSTM to indicate the three stock indexes [17]. When comparing the other models, LSTM-CNN achieved the best results. In the price study of Bitcoin, LSTM has also been used to predict Bitcoin prices many times [18]. McNally et al., Fischer and Krauss, Mallqui and Fernandes also used Simple RNN to forecast the price of the particular currency [19-21].

Inspired by the above research, this paper uses three kinds of recurrent neural network models (RNN) to forecast the price of digital currency.
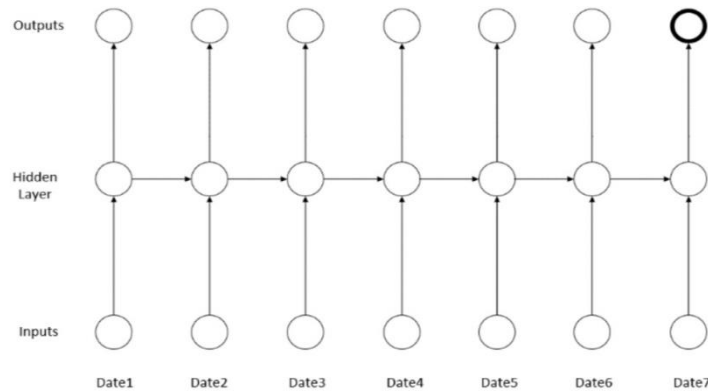
Figure 1: Basic structure of RNN.

3.1.1. Cyclic Neural Network

Cyclic neural network (RNN) is a deep learning model with memory function, which is suitable for processing serial data and is often used to model natural language processing (NLP) and time series data. It is a chain connected directed recurrent neural network, and the loop unit is a hidden layer it contains. The output of the hidden layer uses the following input data as the network input for the next training. At the same time, RNN updates the parameters through backpropagation. It uses Gradient Descent to optimize the model to reduce the error between the model output value and the actual value.
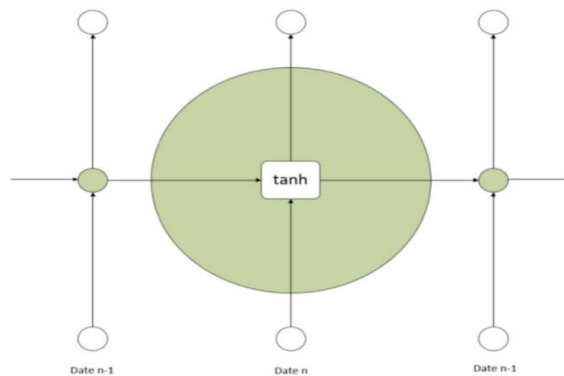

Figure 2: Basic structure of simple RNN.

3.2. Bitcoin Price Forecast Results

The super parameter settings are common to all models when forecasting Bitcoin prices. The model unit will undergo nonlinear conversion. That is, the internal weight matrix will convert the dimensions of the input data into a specific size output. This study uses a single hidden layer with an output dimension of 20, and this study uses Linear as the hidden layer activation function, which means that the input tensor is unchanged. The MAE is used as a loss function using the Adam optimizer. Adam uses gradient first-moment estimation and second-moment estimation to adjust the learning rate of each parameter dynamically. It combines the advantages of Adagrad's sound processing of sparse gradient and RMSprop's good processing of non-stationary targets. After the parameter adjustment

test of this study, when the window length is 5-7 days, the model can obtain the best results. Therefore, the window length of this study is seven days.

3.3. Price Trend Prediction Model

Bitcoin price trend prediction is the prediction of price rise and fall, which belongs to classified prediction. The commonly classified prediction models include support vector machine (SVM), random forest (RF), logical regression (LR), k-nearest neighbor algorithm (KNN), naive Bayes (NB), etc. In the early Bitcoin price trend prediction research, most studies mainly adopted machine learning methods. Mallqui and Fernandes found that the direction of using SVM to predict Bitcoin prices can reach 59.45% [21]. Valencia et al. successfully predicted the price of Bitcoin using random forest (RF) [22]. In the stock market, Li Zhongchen used a variety of traditional machine learning models and deep learning models to predict the rise and fall of 100 A-share stocks by category [23]. Zhang Guisheng and Zhang Xindong combined SVM and GARCH models and compared them with the traditional ARMA-GARCH model, achieving better stock forecasting results [24].

Drawing on previous research results, this paper uses several commonly used machine learning models and recursive feature elimination methods to predict the price trend of special currency land. These machine learning methods include support vector machine (SVM), logical regression (LR), linear discriminant analysis (LDA), random forest (RF), and naive Bayes (NB).

3.3.1. Support Vector Machines

Support Vector Machine (SVM) is a classical machine learning classification algorithm that appeared earlier and was widely used. The basic idea of SVM is to find a classification algorithm that can correctly partition data and achieve the maximum separation hyperplane on the training set. SVM can be divided into linear SVM and nonlinear SVM.

The linear SVM model is a linear classifier with the most considerable interval in the feature space. For linear indivisible data, nonlinear SVM is required. Nonlinear SVM maps data to high-dimensional space through kernel function and then uses linear SVM to classify the mapped data. Different kernel functions are used for different data. The commonly used kernel functions are Polynomial kernel functions and radial basis functions (RBF).

3.3.2. Logical Regression

Logistic regression (LR), also known as logistic regression, is also a widely used classic classification model. Logical regression model is a kind of generalized linear regression model, which has the same form as the traditional multiple linear regression model, as shown in the following formula:

$Z = \omega T x$ where ω is a parameter and x is a variable. The dependent variables of the logical regression model differ from those of the traditional multiple regression linear model. In order to convert the linear regression to the binary classification algorithm, the logical regression model uses the Sigmoid function to map the predictive value obtained by the linear regression to the range of $0 - 1$ and then sets the classification threshold, usually 0.5. Sigmoid functions have

It has strong robustness. The specific formula is as follows:

$$g(z) \; = \; 1$$

$1 + exp\,(-\,z)$ logistic regression uses the maximum likelihood method to estimate parameters. The model finds a set of ω, which makes the function.

$G\,(z)$ has the highest probability of correct classification.

3.3.3. Linear Discriminant Analysis

Linear discriminant analysis (LDA), also known as Fisher linear discriminant analysis (FLD), can be used to reduce the dimension of data and be used for classification prediction. Its basic idea is to project high-dimensional samples onto the best discriminant vector space. The optimization goal of LDA is to obtain the maximum inter-class and minimum intra-class distances.

3.3.4. Random Forest

Random Forest (RF) is a classical integration algorithm. The integrated algorithm combines multiple single algorithms for learning. Compared with a single model, it improves the learning performance and reduces the generalization error. The integrated models of classical machine learning classification algorithms are mainly divided into Bagging and Boosting algorithms. The bagging algorithm trains multiple models and then combines the results of each model to vote to determine the final model result. The random forest belongs to the Bagging algorithm composed of numerous decision trees.

The decision tree is a tree structure, with each leaf node representing the classification category and each internal node representing the division of each attribute. The decision tree can be divided into ID3, C4.5, CART, and other algorithms according to the decision tree's different rules of attribute division. Each decision tree is a classifier, and these decision trees vote for the final classification result of the random forest. The basic idea of the Bagging algorithm is to extract n times of data from the data set as the training set and train n models. For a Bagging-based random forest, the data will be randomly sampled n times from where they are put back, and each sample data will train the decision tree. This sampling method can help the random forest algorithm effectively avoid overfitting and has robustness.

4. Conclusion

In the era of diversified investment, digital currency represented by Bitcoin has received more and more attention, and more and more research has been done on its price. According to the current research situation at home and abroad, the price prediction of digital currency is mainly based on econometric models and machine learning models. Artificial neural network (ANN) is the most widely used among machine learning models. Although existing studies have drawn some conclusions, what needs to be more comprehensive is that these studies did not comprehensively consider the factors that affect the price of digital currency and did not compare the factors that affect the price of bitcoin in different periods.

In view of this situation, the main work of this paper is to use machine learning methods to compare the price and price trends of digital currency. The main conclusions of this paper are as follows.

Deep learning can better predict a special currency's price than this typical time series. In the aspect of the digital currency price trend prediction model, this paper uses the method of combining recursive feature elimination (RFE) with a machine learning model. As an integrated model, the prediction results show that random forest (RF) shows better stability in digital currency price trend prediction. With regard to the influencing factors of digital currency price, this paper draws the following two conclusions through the above empirical research. First, many factors affect the price of digital currency. Among them, the market transaction information and blockchain information of digital currency plays a crucial role in model prediction. Second, affected by the social acceptance of special currency, the price fluctuation of digital currency varies significantly in different development stages. In addition to the market transaction information and blockchain information of Bitcoin, other major influencing factors are also various. In this paper, I use the machine learning method to predict the price of digital currency and the price trend of digital currency. Although the prediction of digital

currency's price can be comparable to the previous research because the research is to help investors make effective investments, the price prediction only serves as a reference. If it is only for digital currency investors, it may be more intuitive to predict the return rate of Bitcoin investment. The following are the prospects for future research:

First, it focuses on the rate of return, including the prediction of the rate of return of digital currency investment or the research of optimization algorithms to maximize the rate of return or control the investment risk.

Second, many researchers now do quantitative investment research in the stock market. They study multiple stocks for portfolio investment to maximize returns while controlling risks. Referring to its research, my idea is that portfolio investment can also be made in the digital currency market, which is still rare. Because there are thousands of digital currencies, their fluctuations could be more consistent, which can ultimately achieve the purpose of hedging risks.

References

[1] Yenidoğan I, Çayir A, Kozan O, et al. Bitcoin forecasting using ARIMA and PROPHET[C]//2018 3rd international conference on computer science and engineering (UBMK). IEEE, 2018: 621-624.

[2] Saad M, Choi J, Nyang D H, et al. Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions[J]. IEEE Systems Journal, 2019, 14(1): 321-332.

[3] Zhengyang W, Xingzhou L, Jinjin R, et al. Prediction of cryptocurrency price dynamics with multiple machine learning techniques[C]//Proceedings of the 2019 4th International Conference on Machine Learning Technologies. 2019: 15-19.

[4] Livieris I E, Kiriakidou N, Stavroyiannis S, et al. An advanced CNN-LSTM model for cryptocurrency forecasting[J]. Electronics, 2021, 10(3): 287.

[5] 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.

[6] Biswas S, Pawar M, Badole S, et al. Cryptocurrency price prediction using neural networks and deep learning[C]//2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2021, 1: 408-413.

[7] Li Y. The price prediction of virtual currency base on improved support vector regression[C]//2021 4th International Conference on Information Systems and Computer Aided Education. 2021: 2587-2591.

[8] Jiang H. Cryptocurrency price forecasting based on shortterm trend KNN model[C]//2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT). IEEE, 2021: 1165-1169.

[9] Li Jing. Building a Bitcoin Market Forecast Model Using BP Neural Network [J]. Monthly Journal of Finance and Accounting, 2016 (21): 33-36.

[10] Li Yinglu. Prediction of Cryptocurrency Index Based on BP Neural Network [J]. Market Weekly, 2019 (08): 104-105.

[11] He Xiongwei, Lin Hai. Empirical Analysis of Time Series Prediction of Bitcoin Based on LSTM [J]. Modern Computer, 2020 (36): 40-46.

[12] Zhao Lei, Liu Qing. Risk Identification of Bitcoin Price foam Based on LPPL Model [J]. Statistics and Decision, 2020,36 (18): 128-131.

[13] Zhang Ning, Fang Jingwen, Zhao Yuxuan. Bitcoin price prediction based on LSTM hybrid model [J]. Computer Science, 2021,48 (S2): 39-45.

[14] Bai Wankuan. Research and Application of RNN Neural Network in Stock Index Price Forecasting Model [D]. Chongqing University, 2018.

[15] Bao Zhenshan, Guo Junnan, Xie Yuan, Zhang Wenbo. Prediction Model of Stock Price Rise and Fall Based on LSTM-GA [J]. Computer Science, 2020, 47 (S1): 467-473.

[16] Gunduz, H., Yaslan, Y., & Cataltepe, Z. Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations[J]. Knowledge Based Systems, 2017, 137:138–148.

[17] Geng Jingjing, Liu Yumin, Li Yang, Zhao Zheyun. Prediction Model of Stock Index Based on CNN-LSTM [J]. Statistics and Decision Making, 2021,37 (05): 134-138.

[18] Chen, Z., Li, C., & Sun, W. Bitcoin price prediction using machine learning: An approach to sample dimension engineering[J]. Journal of Computational and Applied Mathematics, 2020, 365.

[19] Mcnally S, Roche J, Caton S . Predicting the Price of Bitcoin Using Machine Learning. 2018:339-343.

[20] Fischer, T., & Krauss, C. Deep learning with long short-term memory networks for financial market predictions[J]. European Journal of Operational Research, 2018, 270(2):654–669.

[21] Mallqui, D. C. A., & Fernandes, R. A. S. Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques[J]. Applied Soft Computing Journal, 2019, 75:596–606.

[22] Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. Price movement prediction of cryptocurrencies using sentiment analysis and machine learning[J]. Entropy, 2019, 21(6).

[23] Li Zhongchen. Research on Stock Trend Prediction Based on Machine Learning [D]. University of Electronic Science and Technology of China, 2020.

[24] Zhang Guisheng, Zhang Xindong. Research on SVM-GARCH Stock Price Forecasting Model Based on Neighborhood Mutual Information[J]. China Management Science, 2016,24 (09): 11-20.