Research on the Price Prediction of Bitcoin and Gold Based on Random Forest Model

Jingben Lu[1, a], Yawei Song[2, b], Qianhui Li[3, c], Junrong Tang[4, d], Yuke Hou[5, e*]
[a]e-mail: 250762821@qq.com, [b]e-mail: rain2896056605@gmail.com, [c] e-mail: liqianhui1001@163.com,
[d]e-mail: 2710194728@qq.com, [e]e-mail: hyk20000827@qq.com

[1]Guangxi Normal University, School of Mathematical and Statistics, Department of Statistics, Guilin, Guangxi, China

[2] Guangxi Normal University, School of Mathematical and Statistics, Department of Statistics, Guilin, Guangxi, China

[3] North China University of Technology, School of Economics and Management, Accounting, Beijing, China

[4] Guangxi Normal University, School of Economics and Management, Department of Economics, Guilin, Guangxi, China

[5] Guangxi Normal University, Mathematical and Statistics, Department of Statistics, Guilin, Guangxi, China

Abstract—In recent years, machine learning has achieved good results in the field of asset prices. Compared with traditional data analysis and technical analysis, using machine learning methods can show unique advantages in various aspects. In this paper, we combine the correlation between bull and bear markets and bitcoin and gold prices in the market, and apply a random forest model to predict them. The results of the study show that the random forest has high explanations and the accuracy of the model predictions are above 0.9, indicating that the model is a good fit for bitcoin and gold price predictions; bitcoin price is volatile and not suitable as a long-term investment, and it is suitable for gold at the beginning of each year.

Keyword-Bitcoin; Gold Price; Random Forest; Investment

1 Introduction

In modern society, market traders frequently buy and sell volatile assets with the goal of maximizing total returns.

However, individual investors account for a large proportion, and the irrational trading situation is serious, resulting in a sharp rise or fall in asset prices, which makes market transactions complicated and unstable, and it is difficult to maximize total returns by human subjective judgment alone. Advanced mathematical models predict the trend of asset prices, which can effectively avoid irrational investment strategies in dynamic and unstable market transactions.

An important part of asset prices is the price of gold, which is a reliable means of preserving value because it can be used as a hedge against the risk of fiat currency devaluation. In the

ICICA 2022, December 02-04, Chongqing, People's Republic of China
Copyright © 2023 EAI
DOI 10.4108/eai.2-12-2022.2327931

general perception of society, gold is indestructible, constant in value, segmentable as needed, and to some extent portable. But most importantly, gold is scarce in nature and in limited supply. In recent years, a "new member" of the traditional wealth consensus has begun to emerge. It is true that fiat money and gold are still symbols of wealth, but with the advent of the digital age, more and more people are embracing the new idea of "digital gold". The term "digital gold" refers primarily to Bitcoin, which has been given value by a consensus that Bitcoin is as "valuable" as gold. As a result, Bitcoin has become what people call an asset, as has gold.

As a time series of time series, asset price has always attracted much attention. Grey forecasting, RF (Random Forest), ANN (Artificial Neural Network), SVM (Support Vector Machine) and other methods have all been used for asset price forecasting. Experts such as He Yu investigated the intrinsic connection between the changes in the financial ratio indicators of the company and the trend of stock price changes through decision tree, logistic regression model, and neural network model mining, and finally combined model construction to confirm the feasibility and rationality of the technique for stock price prediction [1]; experts such as Zhao Wei used the reinforcement learning model to enhance the effect [2]. In terms of trading strategies, Wu Di and other experts use a variety of factor stock selection models to capture excess returns [3]; Neftci and other scholars use price moving average strategy to improve returns [4]; Jegadeesh and other scholars put forward momentum strategy [5]

Among them, the random forest algorithm has the characteristics of high accuracy, hard to over-fit and high tolerance, and the random forest model tends to be more accurate in short-term fund trend prediction. In this paper, we learn from previous research ideas on stock prices and financial assets, and select random forest to forecast bitcoin and gold prices, to provide theoretical basis and channels for investors and financial institutions to diversify risky investments, and government departments to strengthen market regulation.

Before making predictions on bitcoin and gold prices, we consider the impact of bear and bull markets on their changes, such as price fluctuations and investment risks. The so-called short market, also known as a bear market, refers to a security market in which prices have a long-term downward trend. The general trend of price changes is constantly going down, characterized by large drops and small increases. The overall trend of the short market is downward, although there is a rebound, but a wave of lower than a wave, the majority of people are losing money, although there are occasional opportunities but fleeting, not easy to capture, difficult to operate. While the long market, also known as bull market, refers to the long-term upward trend of the securities market. The general trend of price changes is constantly going higher, characterized by large increases and small decreases.

The overall trend of the long market is upward, although there are declines, but one wave is higher than the other. There are more buyers than sellers, demand is greater than supply, popularity is constantly gathered, investors are willing to chase high, the number of new accounts is increasing, and new funds are constantly pouring in.

2 Model Construction

Since the vast majority of the data format returned by NASDAQ and London Bullion Market is pandas DataFrame type, which is very convenient for data analysis and visualization and

reduces the workload, the data selected for this experiment comes from NASDAQ and London Bullion Market. 2017- 2021 data of bitcoin and gold prices are selected as the object of analysis.

3 Data pre-processing

(1). NA processing: there are missing gold price values in the dataset. Therefore, the gold price of the non-gold trading days will be added first, and its value will be taken as the gold price of the previous trading day.

(2). data transformation: the single series data with time series, is the population of the fit, in order to achieve the purpose of forecasting, need to transform the single series data into regression data.

Random forest is a classifier that uses multiple trees to train and predict the samples. The number of features is chosen randomly, the training data is chosen randomly, and the prediction label with the most occurrences for the same prediction data is taken as the final prediction label. Random forest is a decision tree model based on the bagging framework. Random forest contains many trees, each tree gives classification results, and the rules for generating each tree are as follows:

(1). If the training set size is N, for each tree, N training samples are randomly and with put-back taken from the training as the training set of the tree, and repeated K times to generate K sets of training samples.

(2). If the sample dimension of each feature is M, specify a constant m<<M and randomly select m features from the M features.

(3). Use m features to grow each tree to the maximum extent possible and without pruning process. The flowchart of the classification algorithm is as follows:

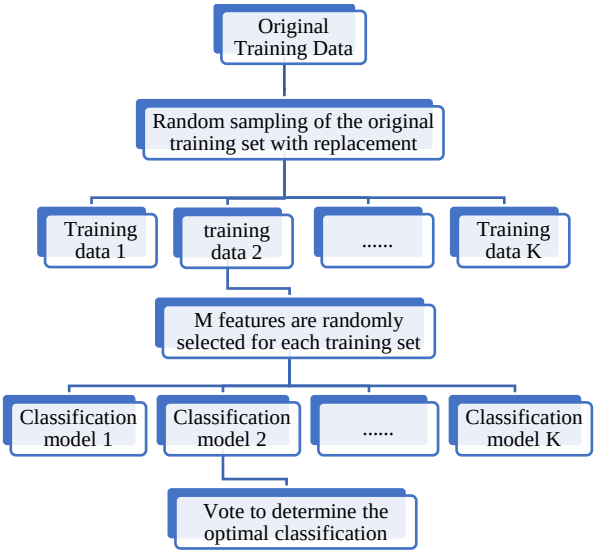

Figure 1. Modeling process

The characteristics of random forest:

(1). simple implementation, strong generalization ability, and can be implemented in parallel because the trees are independent of each other during training;

(2). compared with a single decision tree, it can learn the interaction between features and is less possible to overfit;

(3). it can directly feature a lot of high-dimensional data because some features are still randomly selected from these features during the training process to training; 4) the feature importance can be given after the training is completed. Of course, this advantage mainly comes from the decision tree. Because the decision tree calculates entropy or Gini coefficient during training, the further to the root of the tree, the more important the features are.

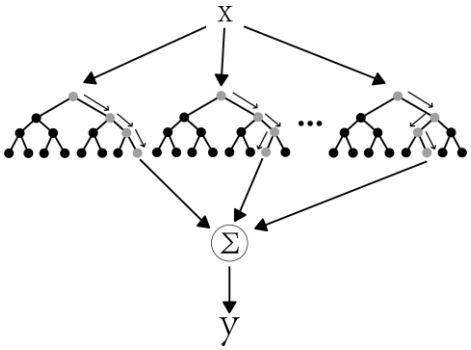

Figure 2. Bagging Tree Ensemble

3.1 Determining Forecast Outcome Metrics

In this paper, R^2(determination coefficent), D_{MSE}(mean square error)and D_{MAE}(mean absolute error)are uesd as the measurement indicators of prediction results.In

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

$$D_{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 \times 100\% \tag{2}$$

$$D_{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - f(x_i)| \times 100\% \tag{3}$$

y_i is the true value, $f(x_i)$ is the predicted value, $\bar{y}$ is the mean of the true value, and n is the number of samples. According to the above formula, we can get:

Table 1. Gold and Bitcoin Outcome Metrices Values

	R^2	D_{RMSE}	D_{MSE}	D_{MAE}
Gold	0.995	1076.978	1159881.637	497.394
Bitcoin	0.996	15.849	251.179	11.067

The closer the value of R^2 is to 1, the more accurate the model is. It can be seen from the above table that $R^2{}_{\text{Gold}} = 0.995$ and $R^2{}_{\text{Bitcoin}} = 0.996$ are both close to 1, which shows that the model has high accuracy.

Bull market and bear market: Bull market and bear market are two different trends expectations in the market. In this article, bull market refers to the rising price of Bitcoin/Gold, which is mainly characterized by big rises and small falls, that is, the overall market price of Bitcoin/Gold is running in an upward trend, while the bears are running in the opposite direction.

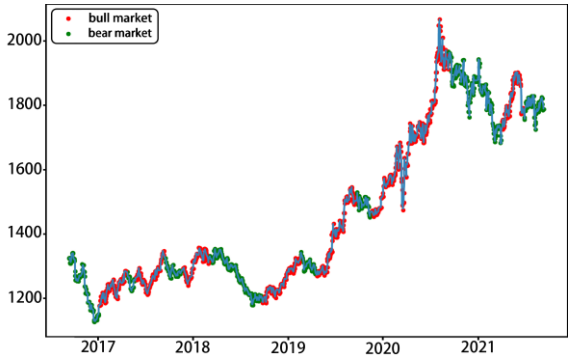

Figure 3. Gold bear market and bull market contrast

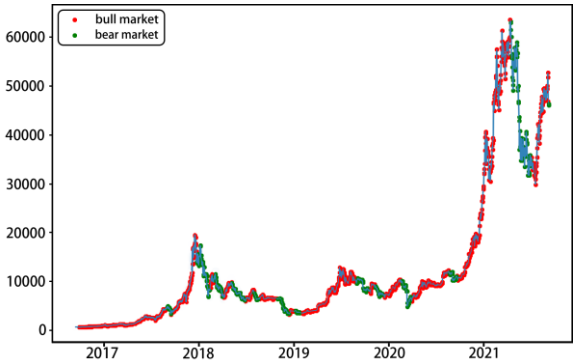

Figure 4. Bitcoin bear market and bull market contrast

As can be seen from the above graph(), the bitcoin bull market is more coherent and lasts longer, with peak points of returns around 2018 and 2021, and after 2021 the bitcoin price grows sharply and changes more frequently, with increased oscillations; the bitcoin bear market is more sparse and lasts shorter, but becomes more coherent and lasts longer after 2021. The gold bull market is more sparse and fragmented, becoming more coherent and longer in duration around 2020,

with peak points of gains before 2021; the gold bear market is more sparse before 2020, but becomes more coherent and longer in duration around 2021, with two sharp increases in gold prices in both mid-2019 and mid-2020, with some price increases after 2021 back down, but the price shocks are smaller and are more stable

Purchase risk: Generally, it refers to the risk that the purchaser may bear for the loss or bankruptcy of future operations and financial activities in order to realize its purchase purpose. This paper defines the expression of purchase risk as:

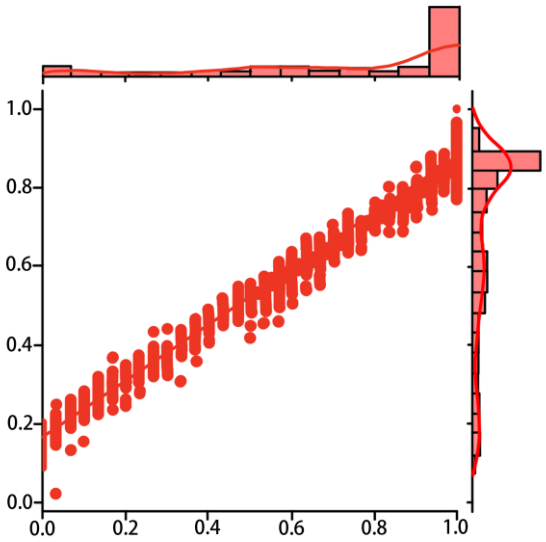

Figure 5. Bitcoin risk and bull market related chart

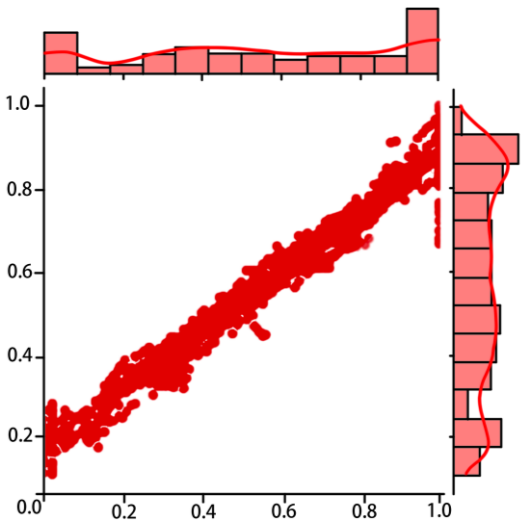

Figure 6. Gold risk and bull market related chart

As can be seen from the chart above, the risk of buying bitcoin is positively correlated with the bull market and the risk of buying gold is positively correlated with the bull market. The risk of buying bitcoin is more volatile, and the risk of buying bitcoin increases sharply when the bull market correlation coefficient is between 0.8 and 0.9; the risk of buying gold is more concentrated and less volatile, and the risk of buying gold is higher at the beginning and the end of the bull market.

3.2 Prediction results and analysis.

The results of calling sklearn in Python environment to predict the price of bitcoin and gold are shown below:

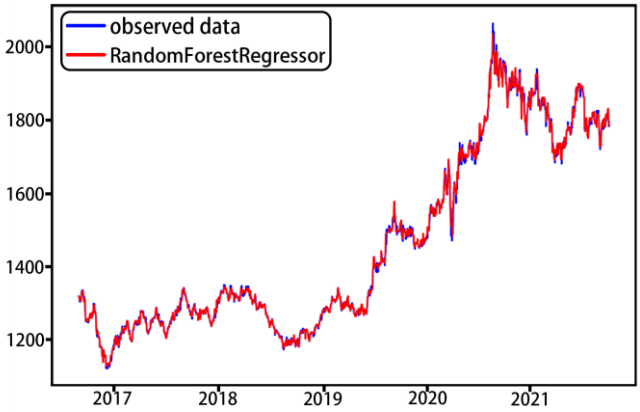

Figure 7. Gold Price Prediction

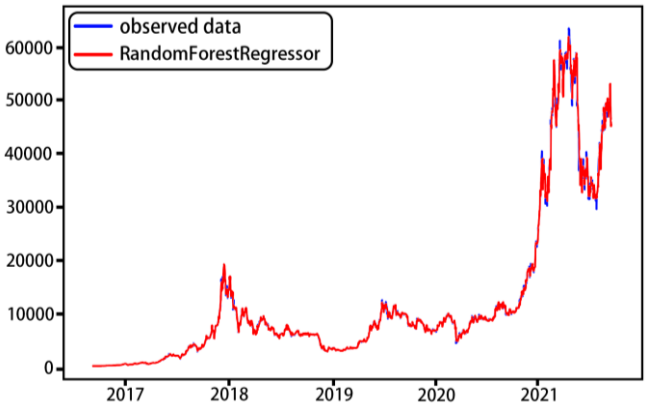

Figure 8. Bitcoin Price Prediction

The figure6,7 is a line chart between the predicted value of the model and the actual value, in which the blue discount is the actual value of bitcoin and gold prices, and the red line is the predicted value. It can be seen from the figure6,7 that the predicted value of the model is basically consistent with the actual value, and the degree of fit is as high as 0.995 and 0.996.

The stability of the prediction is strong, and the proposed model can predict the trend of bitcoin and gold prices more efficiently and accurately. For investors, the accurate prediction of bitcoin and gold price trend, the gain acquisition and risk prevention are inextricably linked; for the national economic development, the trend of bitcoin and gold price reflects and influences the national macroeconomic policy and the ability to participate in international economic activities. Therefore, the research on bitcoin and gold price prediction has important theoretical significance and application prospect.

4 Conclusions

In this paper, we apply machine learning theory and propose a bull market and bear market correlation and random forest model to predict bitcoin and gold prices based on the influence of market conditions on bitcoin and gold prices. From this experiment, it can be seen that bitcoin is more volatile in the short term and influenced by external factors is not suitable for long-term investment. During the peak of the bull market, bitcoin has huge returns, while the risk increases sharply and there is a possibility of a crash at any time, so it is also not suitable for sound investment. It is recommended that government agencies establish relevant firewall spot and derivatives markets, and pursue more transparent regulatory rules to reduce the volatility of bitcoin. The price of gold is less volatile and the risk fluctuations are also less regardless of whether it is a bull or bear market. It is also found that the returns on gold increase at the beginning of each year, making it suitable for investment in that period. It can be observed that the fast training speed of this model of random forest is because the feature subsets is randomly selected, so there is no need to do feature selection, and the prediction fit for both bitcoin and gold prices is higher than 0.9, which has extremely high data suitability, and there is high correctness and feasibility of using this model for short-term prediction of bitcoin and gold prices.

Acknowledgments. We thank our teachers in statistics for their great support and help in solving the problems that arose during our work, our classmates for their help in our research process and our friends for their help in our work during the final layout and correction of errors, due to our limited knowledge, and to our teachers for their non-negligible help in our field of expertise

References

[1]He Yu. Research on stock price prediction based on data mining combination model [D][D]. Southwest University of Finance and Economics, 2014:34-56

[2]Zhao Wei. Research on the Strategy of Diversion Selection Based on Machine Learning Met hod [J]. Economic Research Guide, 2022 (02): 106-108.

[3]Wu Di.A Empirical Analysis of China 's Stock Market Selection Model [D]. Fudan Un iversity, 2011.

[4]Lu Yinqiu,Neftci Salih N.. Financial Instruments to Hedge Commodity Price Risk for Developing Countries[J]. IMF Working Papers,2008,08(6).pp.1-1

[5]Narasimhan Jegadeesh, Sheridan Titman. Profitability of Momentum Strategies: An Evaluation of Alternative Explanations[J]. The Journal of Finance,2001,56(2).

[6]Omar Abdullah Bin, Huang Shuai, Salameh Anas A., Khurram Haris, Fareed Muhammad. Stock Market Forecasting Using the Random Forest and Deep Neural Network Models Before and During the COVID-19 Period[J]. Frontiers in Environmental Science,2022.
[7]Zhou J,Mao Q, Zhang J,et al. Selection of breast features for young women in northwestern China based on the random forest algorithm[J]. Textile Research Journal, 2022, 92(7-8): 957-973.