ISSN: 2664-9640

The Best Investment Strategy for Gold and Bitcoin Combinations

Guochi Zhang, Zihao He

School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, Shaanxi, 710048, China

Abstract

In this paper, we model an optimal trading strategy to improve the return on investment. For the optimal trading strategy model based on supervised learning, two sets of officially provided data are adjusted and supplemented, and the data is visualized to verify that prices and yields have transaction stability. Select 8 eight factors to filter the data. Introduce 5 classic algorithms in supervised machine learning to determine the optimal LGB classification model. For the trading strategy backtesting model, by using a new evaluation indicator for backtesting to compare five machine learning models, the results show that the LGB quantitative trading strategy model has the highest annual rate of return at 32.80307, and a Sharpe ratio of 5.213687, which was finally determined. Finally, robustness and sensitivity analyses are performed on the considered models.

Keywords

KNN, SVC, gold and bitcoin, trading backtest, plain bayesian, light GBM model.

1. Introduction

Market traders often buy and sell volatile assets with the goal to maximize their total returns. To determine whether traders should buy, hold, or sell assets in their portfolio daily, use computer programs to design reasonable trading strategies to minimize transaction costs and increase investment returns [1]. Traders usually have a commission at every sale. Two of these assets are gold and bitcoin.

This work requires us to propose the best trading strategy for traders to maximize their returns, and to prove the reliability of the trading strategy based on the amount changes and returns derived from different models. Our work mainly includes: Modeling optimal trading strategies based on historical trading dates and prices for gold and bitcoin, giving daily whether to buy, hold or sell assets in their portfolio, and considering whether to buy. The proportion of assets sold, and the optimality of the trading strategy is proved from the evaluation indicators such as annualized rate of return and Sharpe ratio.

Based on the 5-year trading dates and prices from 2016 to 2021, this paper uses the trained machine learning method to obtain the optimal trading strategy under the combination of cash, gold and bitcoin [C, G, B] that the trader will have, And the returns of various trading strategies are evaluated from multiple evaluation indicators, which verifies the optimality of the model. First, model the optimal trading strategy. We extract data features and financial labels from the dataset given in the question to formulate optimal trading strategies, and use five supervised learning machine learning methods (KNN, SVC, logistic regression, plain Bayes, and Light GBW) to make predictions [2]. We use new evaluation metrics to compare the effectiveness of each classification model and its performance in the market to simulate whether a trading strategy is optimal [3]. Finally, a sensitivity analysis of the trading strategy is carried out. We compare cross-asset trading results at different transaction costs and discuss whether the model is reliable under changing transaction costs.

ISSN: 2664-9640

2. Supervised Learning based Model for Optimal Trading Strategies

In order to more intuitively and clearly represent the best transaction measurement model based on supervised learning established in this paper, the basic flow chart shown in the Fig.1.

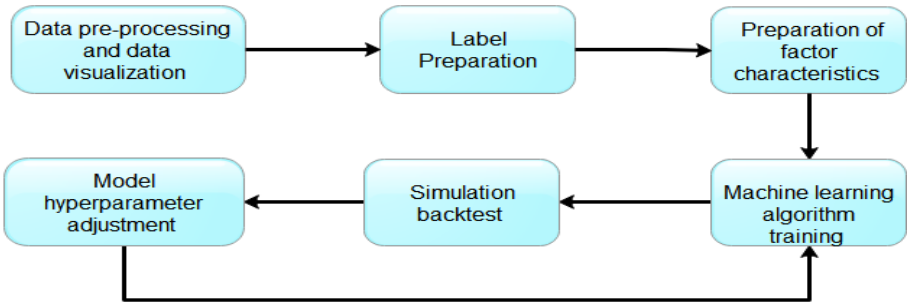


Figure 1 Basic flowchart of a supervised learning-based trading strategy

We adapt the data format to a year/month/day format for the time data from 2016 to 2021 so that they are arranged according to a timeline, which is shown in table 1.

Table 1 Partially adjusted data for bitcoin time and price

Data	Value	Data	Value
2016-9-11	621.65	2016-9-16	609.11
2016-9-12	609.67	2016-9-17	607.04
2016-9-13	610.92	2016-9-18	611.58
2016-9-14	608.82	2016-9-19	610.19
2016-9-15	610.38	2016-9-20	608.66

By analyzing and comparing the given dataset, we found that the data for bitcoin is completely lined up according to the timeline of each day, while the data for gold has the problem of missing data for non-trading days. Since the price of gold is unchanged on non-trading days, here we use the price of the trading day before the non-trading day as the price on the non-trading day, thus filling in the missing data along the time axis. Finally, we supplement the time of non-trading gold and price in the dataset and applied it to later models.

To better visualize the price fluctuation trends of bitcoin and gold over the five-year period of 2016-2021, this paper uses the processed data to make a line chart of price fluctuations reflecting the dynamics of bitcoin and gold prices, which is shown in Fig.2.

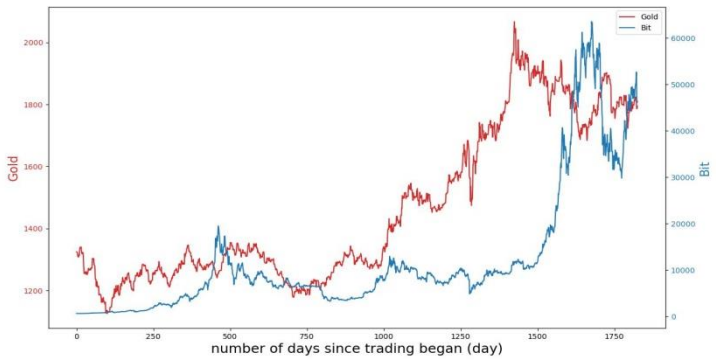


Figure 2 Changes in bitcoin and gold price volatility

To verify that bitcoin and gold price changes and yield changes have trading smoothness, this paper calculates five-day rolling average returns for both, i.e., creates an average of every five data in the entire data set to analyze the data points and plots the time trend of the returns as shown in Fig.3 and Fig.4.

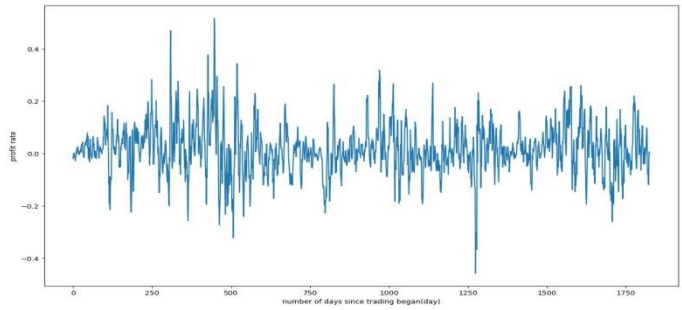


Figure 3 Time trends in average gold returns

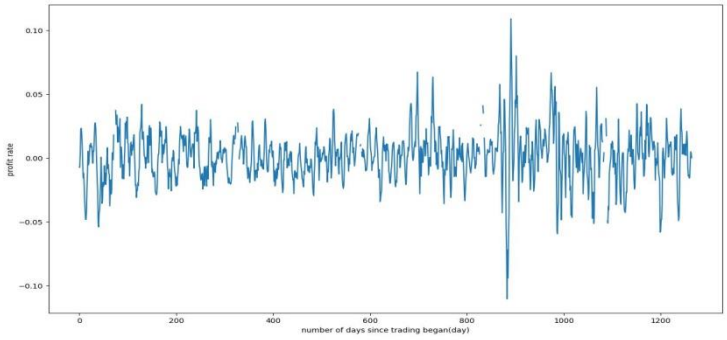


Figure 4 Time trend of average bitcoin returns

We need to prepare the list of features to be used in the training process to provide data for the strategy. The table2 is the feature data commonly used in this study, with a total of nine feature factors, and each row represents one data feature.

Table2 Definition of feature factors

Feature variable name	Feature Definition
CLRCLE	Week and month information
BIAS	(Price - Average price of the previous 10 days)/Average price of the previous 10 days
VOL	Standard deviation of the previous three days' prices
RANGE	The difference between the maximum and minimum value of the asset price in the previous 3 days
MA	Average price of asset price in the previous 5 days
PSY	Number of up days in 5 days/5
TRIX	(3-day average price - previous day's 3-day average price)/3-day average price
DM	5-daySMA - 10-daySMA

The KNN algorithm, also known as the K-nearest neighbor algorithm, is a simple and effective algorithm for classification and logistic regression in machine learning, which can be expressed as:

$$c = \arg\max_{v \in C} \sum_{i=1}^{k} \Upsilon(v, class(y_i)) \tag{1}$$

The transformation of sample features by the kernel function allows the samples to be classified linearly in a high-dimensional space, and the Gaussian kernel is the most frequently used kernel function in support vector machine algorithms due to its good different ability [4]. In this paper, the kernel functions are selected in turn for prediction, and it is found that the Gaussian kernel function has the best classification effect, so the Gaussian kernel function is adopted, and its general form is:

$$\kappa(x,x)=e^{\frac{\left\|x-x^T\right\|^2}{2\sigma^2}} \tag{2}$$

Logistic regression algorithms can predict the likelihood of an event occurring under the action of many different input variables [5], which can be expressed as:

$$\begin{aligned} l(w) &= \sum(c_i \times \ln f(x_i)+(1-c_i)\times \ln(1-f(x_i))) \\ &= \sum c_i(\frac{\ln f(x_i)}{1-f(x_i)}-\ln(1-f(x_i)+\ln(1-f(x_i)) \\ &= \sum(c_i \times (w^T x_i)-\ln(1+e^{w^T x_i})) \end{aligned} \tag{3}$$

Parsimonious bayes was proposed based on bayesian classification, and the algorithm satisfies a simple assumption that the attribute values are conditionally independent of each other at a given target value, which can be expressed as:

$$C(X)=\arg\max_{1\le i\le n} P(C_i)\bullet\prod_{i=1}^{n} P(x_i \mid C_j) \tag{4}$$

Light GBM is an improved model for gradient boosting decision tree GBDT, which can be expressed as:

$$L_t=\sum_{j=1}^{J}[(\sum_{i\in I_j} g_i)\varpi_j+\frac{1}{2}(\sum_{i\in I_j} h_i+\lambda)\varpi_j^2] \tag{5}$$

This paper also uses the "bitcoin-gold balanced strategy" to control the position of the corresponding assets by referring to the equity-bond balanced strategy proposed by Graham. Specifically, the investor is assumed to have a risk-neutral preference and set the gold-bitcoin ratio to 1:1 of the total position. That is, when the price of asset rises and its position exceeds the target position (50%), sell the excess of asset and buy asset. Conversely, when the price of the asset falls and the position of asset falls below the target position (50%), sell part of asset and buy asset [7]. With such dynamic position adjustments, we maintain a 1:1 target gold - bitcoin ration.

The objective of this question is to construct the optimal classification prediction model for bitcoin and gold, and to predict by the five algorithms KNN, SVC, LGB, Logistic, and NB mentioned above, and three accuracy indicators Accuracy, recall, and F1 are selected to evaluate the classification results and illustrate the feasibility of the model. Where recall is the future price of each asset class and F1 is the nine characteristic variables developed in the previous section. A comparison of the accuracy indicators of the prediction results of the five models was made, as shown in the table 3.

Table 3 Comparison of prediction accuracy of 5 models

	KNN	SVC	Logistic	LGB	NB
Accuracy	0.682	0.789	0.485	0.623	0.485
Optimum accuracy	0.71	0.70	0.69	0.73	0.70
recall	0.708	0.702	0.694	0.683	0.697
F1	0.686	0.585	0.571	0.621	0.572

This paper also visualizes the accuracy metrics of the above models, as shown in Fig. 5, and the results also support this conclusion.

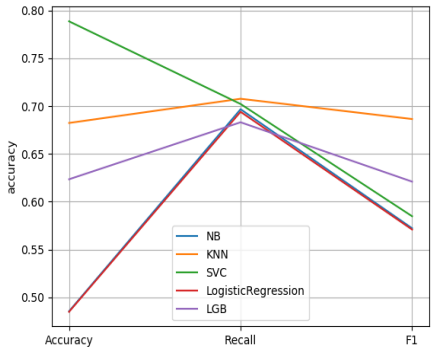


Figure 5 Comparison of prediction accuracy of 5 classification models

To ensure the robustness of the model, this paper also calculates the total asset value at September 10, 2021 predicted by the five trading strategy models mentioned above, as shown in table 4.

Table 4 Final predicted value of the 5 trading strategy models (in USD)

	KNN	SVM	Logistic	NB	LGB
Final Value	57885.34	38098.38	34604.31	36810.18	164105.21

We plot the trend of total asset value calculated based on this transaction strategy, as shown in Fig. 6.

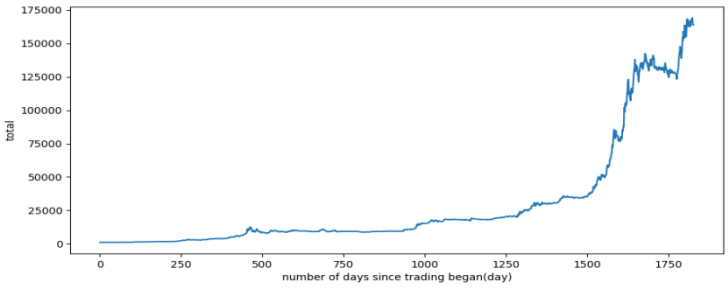


Figure 6 Daily asset returns based on optimal trading strategies

From the Fig.6, it can be observed that the total return obtained from the Light GBM quantitative timing model established in this paper is in a stable upward trend, which can well select the trading time and the number of trades with the highest return and the maximum total asset value, therefore, the model is reasonable and effective.

3. Trading Strategy Backtesting Model

We apply supervised machine learning methods to quantitative timing to select the best trade timing and predict the probability of future rises, falls and flatness of gold and bitcoin to construct a gold-bitcoin high frequency quantitative trading strategy based on the LGB model.

This section will verify the feasibility of the best trading model through the back test model and compare the effectiveness of the each trading model and its performance in the market. The asset pool used in this retest model is bitcoin and gold, and the factor pool used is the feature pool of the previous part of the trading strategy [8].

The annualized rate of return is the rate of return within one year based on the current rate of return and is an important indicator for evaluating quantitative strategies [6], which is calculated as follows:

$$R_p = (\frac{P_e}{P_s})^{(250/n)} - 1 \tag{6}$$

where P_e denotes the strategy's final total assets, P_s denotes the strategy's initial total assets, and n denotes the number of backtest trading days, here we take 5.

The Sharpe Ratio, also known as the Sharpe Index, is an evaluation indicator of the profitability of a quantitative model, with the objective of calculating how much excess reward a portfolio will generate for each unit of risk taken. The Sharpe ratio is proportional to profitability and is calculated as:

$$SR = \frac{E(R_p) - R_f}{\sigma_p} \tag{7}$$

where $\mathrm{E}(R_p)$ denotes the expected portfolio payoff, R_p denotes the annualized rate of return, R_f denotes the risk-free rate , and σ_p denotes the standard deviation of the portfolio.

Through machine learning methods to predict the future probability of gold and bitcoin rise, fall and flat, according to the predicted results of the purchase and purchase operations, in order to verify the rationality of the trading strategy, the trading behavior of gold and bitcoin during the backtest is plotted, as shown in Fig. 7.

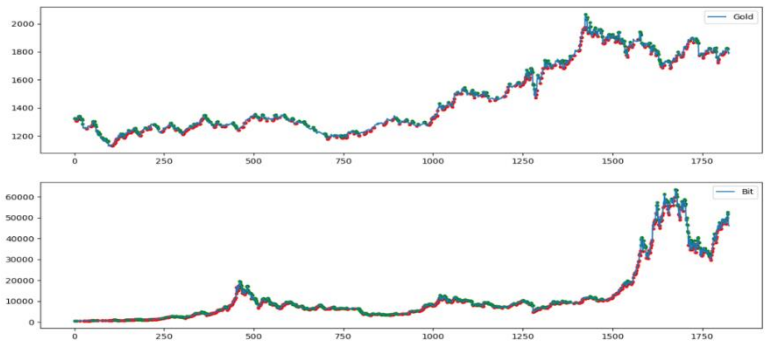


Figure 7 Gold, bitcoin daily purchase and purchase distribution chart

Also, this section back-tests the total asset value predicted by each specific quantitative timing model at different trading times based on the five different machine learning quantitative timing models in the previous section to obtain the daily total asset value for each model at the time of quantitative trading, as shown in Fig. 8.

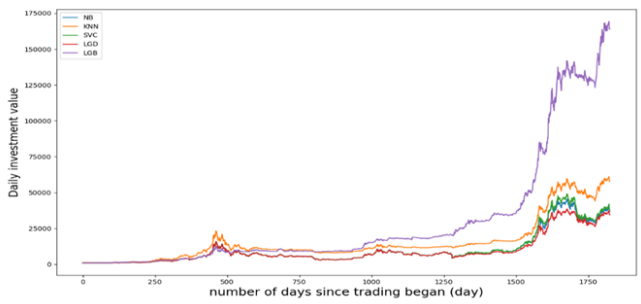


Figure 8 Total asset value of 5 different models in quantitative trading

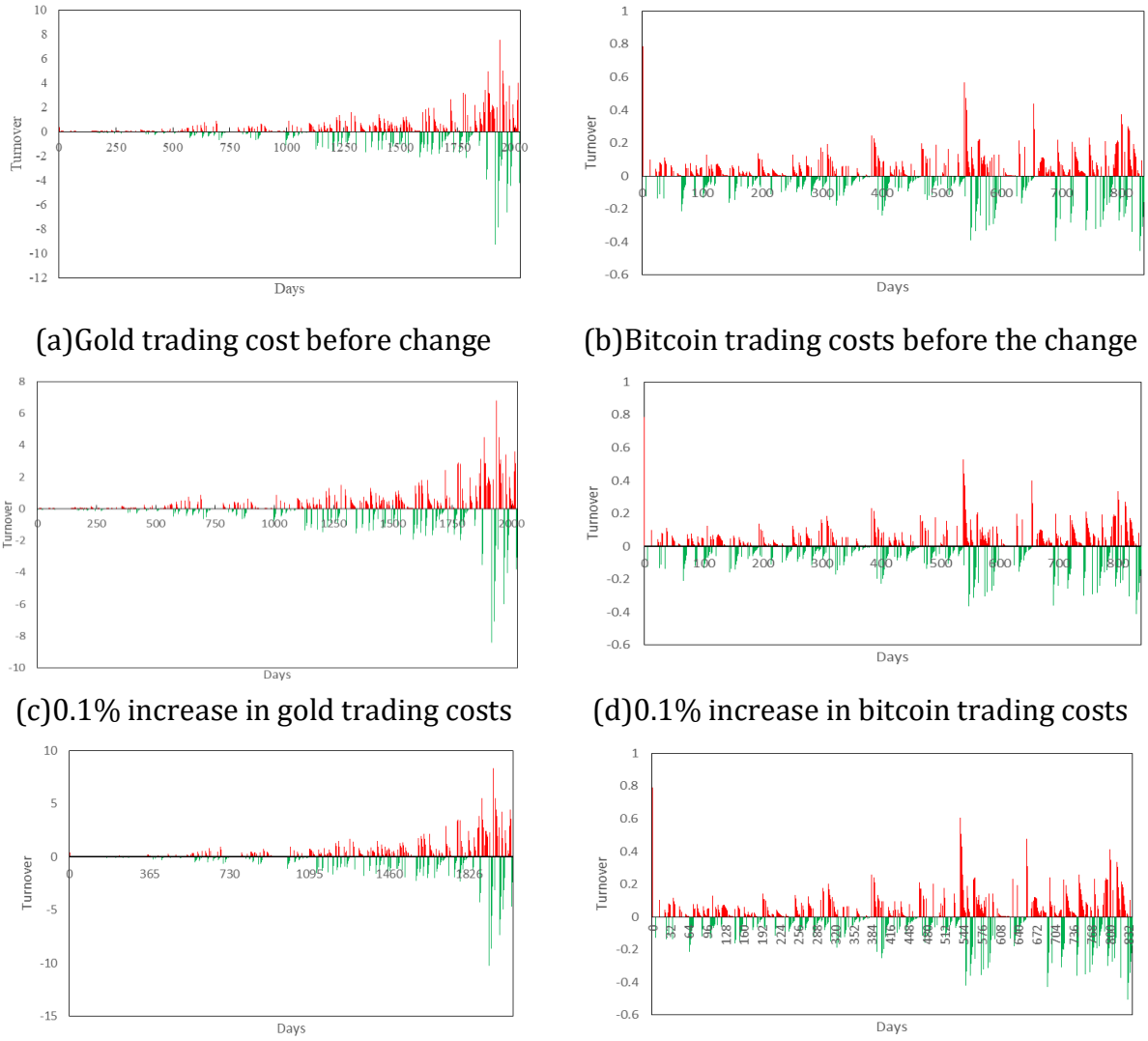
4. Sensitivity Analysis

This paper needs to analyze the sensitivity of the total value of the transaction strategy to the transaction cost. We compare the impact of the transaction cost of gold and bitcoin by 0.1%

reduction and 0.1% increase on the total value of assets, respectively, and compare the results before and after the change, as shown in Fig.9.

Figure 9 Sensitivity analysis of total asset value to transaction costs

This paper also maps the impact of transaction costs on transaction volumes of various assets after increasing and reducing perturbations by 0.1%, as shown in Fig.10.



(a)Gold trading cost before change (b)Bitcoin trading costs before the change

(c)0.1% increase in gold trading costs (d)0.1% increase in bitcoin trading costs

(e)0.1% reduction in gold trading costs (f)0.1% reduction in bitcoin trading costs

Figure 10 Changes in trading volume before and after changes in gold and bitcoin trading costs

5. Conclusion

In this paper, five machine learning algorithms, KNN, SVC, LGB, Logistic, and NB, are used for the study. First, data features and financial labels are extracted from the official dataset. Based on this, the optimal classification prediction model for bitcoin and gold, the LGB-based trading strategy model, is constructed by combining the five machine learning models in the comparison of return and evaluation indexes. Based on this model, this paper predicts the future probability of up, down and flat for gold and bitcoin. Finally, each asset class is discounted accordingly, and its total asset value is obtained as $164,105.21.

References

[1]Kearns, Kulasza, Nevmyvaka. Empirical Limitations on High Frequency profit ableity [D]. Working Paper.2010.
[2] Palace Xiujun. The Bayesian Theory of Learning and its Research [D]. The year is in 2002.
[3] Xu Weihong, a stock selection strategy based on the classic four-factor model, Southwest Securities, 2013,1~14.
[4] Li GZ, Wang M, Zeng HJ. Introduction to support vector machines [M]. Beijing: Electronic Industry Press, 2004: 13-16.
[5] Liu Dawen, Fan Wise. Support vector machine based stock price index modeling and forecasting [J]. Statistics and Decision Making, 2013(02):76-78.
[6] Ding Peng. Quantitative investment: strategies and techniques [M]. Electronic Industry Press, 2012.
[7] Li B, Lin Y, Tang WX. ML-TEA:A set of quantitative investment algorithms based on machine learning and technical analysis[J]. Systems Engineering Theory and Practice, 2017, 37(05):1089-1100.
[8] Ma R. Research on stock trading model based on deep reinforcement learning [D]. Shanghai International Studies University, 2021.