

Multi-factor Stock Forecasting Model Based on Independence Criterion Weighted Bagging Algorithm

Xinyi Sun ^{1, #}, Wendi Cai ^{2, *, #}, Mengyu Li ^{3, #}

¹ School of Statistics, Renmin University of China, Beijing, China

² School of Mathematics and Physics, Xi'an Jiaotong Liverpool University, Suzhou, China

³ School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, China

* Corresponding author: moyuhan4077@163.com

#These authors contributed equally.

Abstract. Stock price prediction and the selection of important factors are a hot issue for statisticians. Machine learning methods need to ensure the prediction accuracy of the model in high-dimensional data scenes on the one hand, and the interpretability of the output of the prediction model on the other. If a single model is used for forecasting, it is easy to cause the problems of deviation and variance imbalance of the forecasting model (over-fitting and under-fitting). Based on this, we propose an interpretable forecasting model weighted bagging based on independence criterion. On the one hand, the proposed method solves the problem that a single model is easy to fall into over-fitting and under-fitting by using the idea of model averaging. On the other hand, when the sub-model is an interpretable machine learning method in Bagging framework, the proposed method can also output the weighted feature importance of each factor. In addition, the choice of independence criterion and sub-model is free. The actual data analysis shows that the proposed method has smaller prediction error compared with the comparison methods such as LASSO, ridge regression, random forest and XGBoost. The factor selection model under the minimum prediction error criterion will be more interpretable.

Keywords: Weighted Bagging Algorithm, Feature Selection, Mutual Information, Stock Price Forecast s.

1. Introduction

The research on stock market forecast is of great significance to investors and enterprises, and also plays an important role in promoting China's economic development. However, there are many factors in stock price data, such as volatility, nonlinearity and random changes influenced by the market, which makes it difficult to simulate and model the stock price and predict the trend. The fluctuation trend charts of the following stocks over a period of time illustrate this problem, as shown in Figure 1:

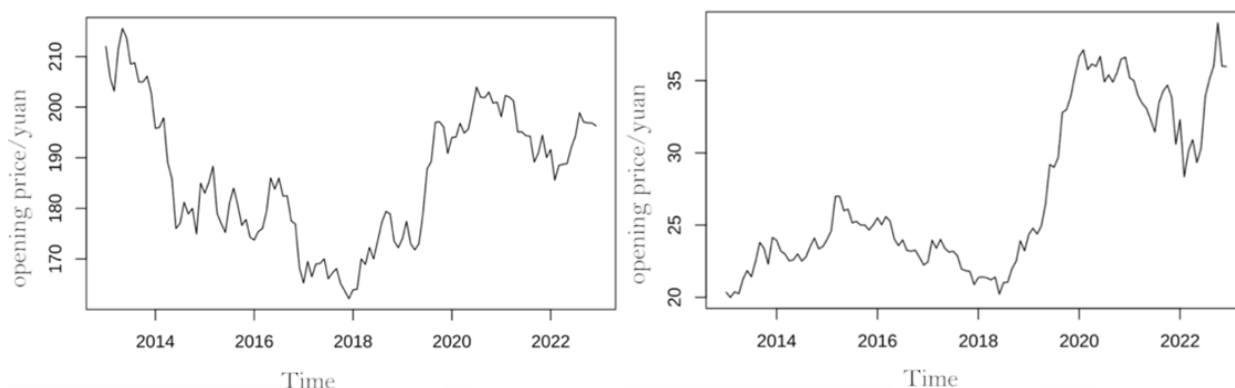


Figure 1. Time series of daily opening prices of Guizhou Maotai (left) and BYD (right) stocks

The stock price forecasting problem is essentially a high-dimensional data regression problem, and the commonly used forecasting models mainly include lasso, ridge regression, and other regularized multiple regression methods. Specifically, He Ping states that the combined LASSO-logistic algorithm can effectively identify the complex relationship between stock characteristic variables and expected returns [1]; According to Zhang Yiran invokes the analysis method of ridge regression to solve the problem of multicollinearity [2]. However, there may be complex nonlinear relationships between stock prices and multiple factors, when forecasting methods based on generalized linear models are no longer applicable, and decision tree models based on logistic rules and forecasting methods based on this technique will be more applicable. Specifically, Chen Quan et. al argue that combining financial knowledge using a single-factor strategy to filter factors and training with machine learning algorithms to predict stock returns shows that decision trees are excellent in stock forecasting [3]; Wen Guangjun states a genetic algorithm to iteratively calculate model parameters of decision trees to evaluate candidate stocks based on their fundamentals and technical characteristics [4]. Nevertheless, although the above methods achieve better learning results, they suffer from the problem that a single model tends to cause overfitting and underfitting. Neural network models based on deep learning can often deal with this problem, but their poor interpretation is not suitable for multi-factor selection in stock price prediction tasks. According to Ma Tian et. al used 148 feature variables of micro firm stocks to construct a factorial large dataset and used generative adversarial network (GAN) method to construct a deep learning model for data downscaling and feature extraction to solve the high-dimensional trap problem [5]. Another model is random forest, which can give factorial interpretability and is an integrated learning model. For example, according to Liang Zhou used 23 common factors to build a random forest model to predict the returns of CSI 500 index components and construct a portfolio, and found that the random forest model can fit and predict the relative returns of individual stocks better [6]. The random forest can deal with the problem of prediction model bias and variance imbalance. However, the model sets the prior probability distribution of each sub-model to be uniform, which is obviously unreasonable.

Based on this, a weighted Bagging algorithm based on independence criterion is proposed in this paper. The proposed method sets data-adaptive prior probability distributions for sub-models by using independence criteria (e.g., mutual information, distance correlation coefficients, and correlation measures based on normal distributions), which makes the prediction and multi-factor selection mechanisms of the models more reasonable. In addition, the choice of the base model in the proposed method is free, so it can be equipped with linear and nonlinear interpretable machine learning methods, which makes the proposed framework adaptable to more complex data scenarios. Practical data analysis shows that the proposed method is competitive.

2. Theory and Methodology

2.1. Weighted Bagging algorithm based on independence criterion

The regression model in this paper is trained from perspective of Bagging in the ensemble learning, and the flowchart is shown in Figure 2. The classical Bagging algorithm trains sub-models and then equally weights them to form a final model[7]. The weights of the sub-models are obtained using the weighted-averaging Bagging algorithm in this study, and they are then averaged and weighted in accordance with a variety of independence criterion, including mutual information and Pearson coefficients under the presumption of a normal distribution. The following are the steps involved in implementing the enhanced algorithm utilized in this paper:

- (1) Bootstrap method is used to randomly sub-sample from the original data set with put-back n samples;
- (2) To train all subsamples separately for m sub-models and calculate the weights of sub-model under a variety of independence criterion
- (3) To calculate the corresponding weights of feature factor by using the weights of sub-models;
- (4) Sub-models are used to fit the test datasets in the subsamples, and the mean square error is

then calculated to compare our improved technique to other conventional algorithms.

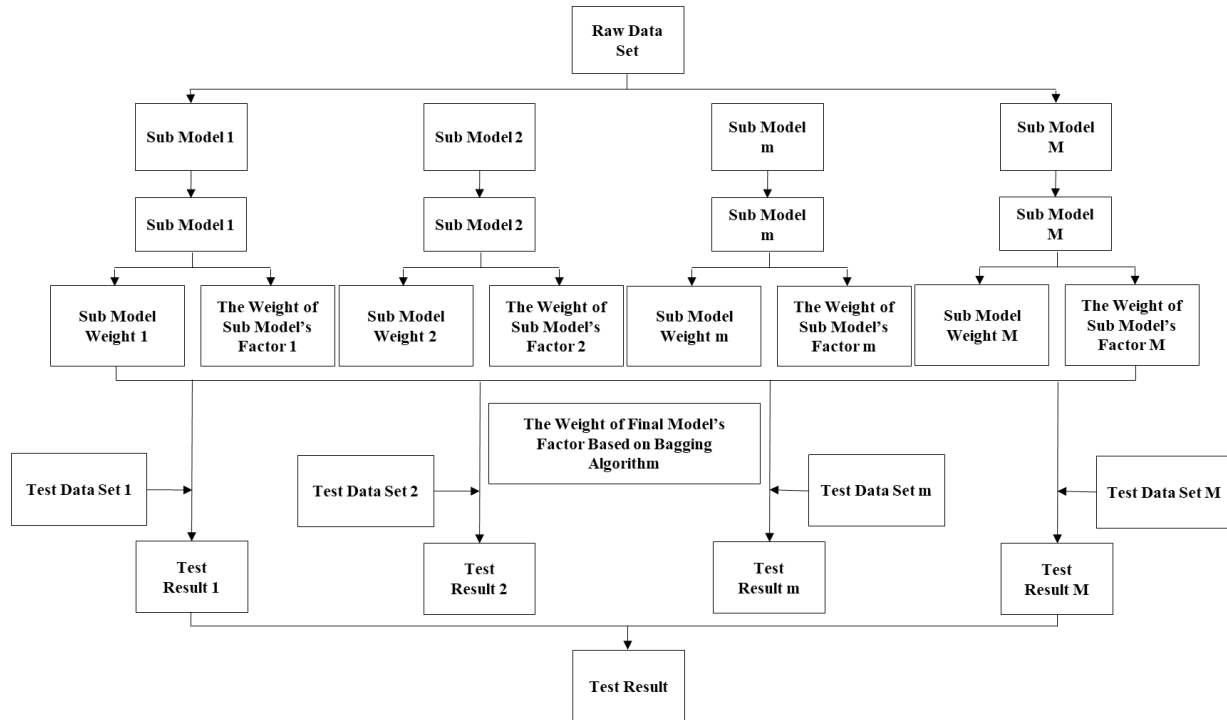


Figure 2. Research flowchart

2.2. Independence criterion and selection of base model

In terms of the base model, Bagging methodology does not limit the form of the base model, and it can be chosen based on the nature of the model. As for linear regression problem in particular, classical multiple regression linear regression method, Lasso [8], ridge regression [9], and decision tree regression algorithm [10] can be considered; for nonlinear regression, Support Vector Machine method [11] can be chosen. The Lasso method and decision tree algorithm are selected in this research and compared to determine the ideal base model considering taking into account the model complexity and accuracy.

Regarding the choice of independence criterion, we could consider the Pearson correlation coefficient, the mutual information and the distance correlation coefficient [12]. First of all, we consider the Pearson correlation coefficient which measures the correlation of random variables X, Y , defined as follows.

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \quad (1)$$

The Pearson correlation coefficient independently measures the relationship between each feature and the response variable Y relationship between them when it comes to feature selection.

Secondly, the mutual information can be used to measure the degree of independence of two random variables. For two random variables X and Y , their joint density function is $p(X,Y)$, the marginal probability functions are $p(X)$ and $p(Y)$. The mutual information $MI(X;Y)$ is defined as follows[13]:

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (2)$$

The standardized mutual information is normalized to the mutual information as follows for its equation.

$$NMI_m(X; Y) = \frac{MI_m(X; Y)}{\sum_m MI_m(X; Y)} \quad (3)$$

Where m is the first m sub-models.

In this paper, we choose to use all the factors in each subsample training set as X random variables and all fitted values in each subsample training set as Y random variables, aiming to explore the amount of reduction in the uncertainty of Y given X . This amount is used as the weight of the sub-model.

Finally, if the two variables X, Y are not correlated or independent of one another or cannot be judged, the distance correlation coefficient can be chosen for feature selection. The formula is as follow [14]:

$$dcorr(X, Y) = \frac{dcov(X, Y)}{\sqrt{dcov(X, X)dcov(Y, Y)}} \quad (4)$$

$$dcov^2(X, Y) = S_1 + S_2 - 2S_3 \quad (5)$$

Where

$$S_1 = \frac{1}{n^2} \sum_{i=1}^I \sum_{j=1}^J ||X_i - X_j|| ||Y_i - Y_j|| \quad (6)$$

$$S_2 = \frac{1}{n^2} \sum_{i=1}^J \sum_{j=1}^J ||X_i - X_j|| \left| \frac{1}{n^2} \sum_{i=1}^J \sum_{j=1}^J ||Y_i - Y_j|| \right| \quad (7)$$

$$S_3 = \frac{1}{n^3} \sum_{i=1}^J \sum_{j=1}^J \sum_{u=1}^J ||X_i - X_u|| ||Y_j - Y_u|| \quad (8)$$

2.3. Feature weights of the ensemble model under the weighted Bagging algorithm

Multiplying the feature weights of the sub-models by their weights, the feature weights of the ensemble model produced by the Bagging method are then normalized to produce the feature weights of the final model. The feature weights of the sub-models can be obtained in two ways. In general, the characteristic weights of the sub-models can be understood as the regression coefficients of the characteristic factors. We use the general classical linear regression model as an example, and its formula is as follows[15]:

$$Y_m = X_m \beta_m + \varepsilon_m \quad (9)$$

Where coefficient m represents the m -th sub-model, then X_m represents all factor values of the m -th sub-model, and the β_m represents factor regression coefficients of factor regression coefficients of the m -th sub-model, also called factor exposure, and we consider factor exposure as the weights of features in sub-models. If the sub-model has a total number of p factors, then we use

the factor exposure as the feature weight of the sub-model which is $\beta_m = \begin{bmatrix} \beta_{m1} \\ \beta_{m2} \\ \vdots \\ \beta_{mp} \end{bmatrix}$ If there are M

sub-models also M subsamples, then $\beta = \begin{bmatrix} \beta_{11} & \cdots & \beta_{M1} \\ \vdots & \ddots & \vdots \\ \beta_{1p} & \cdots & \beta_{Mp} \end{bmatrix}$. If the weight vector of sub-model is $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix}$, the weights of features under Bagging algorithm are $\pi = w\beta^T$ in final model.

When it comes to the decision tree model, the feature weights of sub-models can be transformed to the importance of feature factor. The methodology of Gini index weighting is generally utilized to calculate the feature factor importance, which represents the average amount of change in the impurity of node when splitting certain node of the tree model[16]. Suppose the factor variables are $X_1, X_2, \dots, X_j, \dots, X_J$, we need to calculate the weights of each feature in the aggregated model under the Bagging algorithm. The Gini index of the n -th node in the m -th Bootstrap self-help sub-sample or the m -th sub-tree is as followed [17]:

$$GI_{mn} = \sum_{k=1}^K p_{mnk}(1 - p_{mnk}) \quad (10)$$

Where K is the number of categories in the m -th Bootstrap self-help sub-sample or the m -th sub-tree and p_{mnk} represents the probability of the node n in the m -th Bootstrap self-help sub-sample or the m -th sub-tree which belongs to the k -th category. The importance of the factor variable X_j at the node n , i.e. the amount of change in Gini coefficient before and after branching at the node n is

$$V_{mjn} = GI_{mn} - GI_l - GI_r \quad (11)$$

At this time GI_l and GI_r are the Gini coefficients of the two new nodes split by node n Gini factors of the two new nodes that split. If the factor X_j appears for N occurrences in the m -th sub-model, then the weight or importance of variable X_j in the m -th sub-model is:

$$V_{mj} = \sum_{n=1}^N V_{mjn} \quad (12)$$

3. Data Analysis

3.1. Data sources and experimental description

The Tushare Big Data platform (<https://tushare.pro>) provided the stock data used for this article, and the Python SDK was used to retrieve daily and monthly data for the chosen Shanghai and Shenzhen stocks. These stocks were chosen for their excellent qualities of having a high market capitalization, high liquidity, and no material issues with financial reporting. The data omit the non-trading periods and cover the period from January 1, 2013, through December 31, 2022. Two sets of stock data, Guizhou Maotai (600519.SH) and BYD (002594.SZ), are chosen, and their daily and monthly data is evaluated and forecasted in both high and low frequency dimensions, respectively, to test the resilience of the model. The following variables were determined from the original data: the opening and closing prices for today, the high and low prices for today, the closing price for yesterday, the difference between the closing prices for the two days, the daily increase, volume, turnover, turnover ratio, P/E ratio, and the average P/N ratio. The remaining characteristic variables in Table 1, whose names and calculation techniques are presented in the following table, are constructed by this study by reviewing the literature based on the a forementioned original characteristic variables.

Table 1. Factor Description Table

Category	Name	Calculation method
Moving Averages	MA1	Sum of 5 days' closing prices / 5
	MA2	Sum of 10-day closing prices /10
	MA3	Sum of 20-day closing prices / 20
Difference Moving Averages	MACD	12-day EMA— 26-Day EMA
	MACDsignal	DIFF's M-day average of dissimilar moving averages
	MACDhist	DIFF— DEA
Energy Tide Indicator	obv	Yesterday obv+ sign× Today's volume
Bollinger Bands	upper	20-day SMA +2SD
	BOLL	20-day SMA
	lower	20-day SMA— 2SD
Relative Strength Indicators	RSI1	$A/(A+B) \times 100$ A=Sum of N intra-day closing gains B=Sum of N intra-day closing declines N=6
	RSI2	N=12
	RSI3	N=24
William Indicator	WR1	$(C - H_n) \div (H_n - L_n) \times 100$ C = latest closing price on day n Hn = highest price in the last n days Ln=Lowest price in the last n days n=6
	WR2	n=10
Stochastic Indicators	kdj-k	$K_i = \frac{2}{3} K_{i-1} + \frac{1}{3} RSV_i$
	kdj-d	$D_i = \frac{2}{3} D_{i-1} + \frac{1}{3} K_i$
	kdj-j	$J_i = 3K_{i-1} - 2D_i$
Deviation rate	bias_6	BIAS=(Closing price - N-day simple average of closing prices)/N-day simple average of closing prices*100 N=6
	bias_12	N=12
	bias_24	N=24
Rate of change indicators	ROC	(Today's closing price - 12 days ago closing price)/N days ago closing price

The prediction effect in this paper is also assessed using the additional assessment metrics listed below.

(1) Mean Square Error

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (13)$$

(2) Average relative error

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (14)$$

The experimental machine learning framework for this study is built using sklearn, numpy, and scipy in conjunction with the Anaconda-Python 3.8 experimental environment. The first n rows of data are chosen as the training set (the division of n is between 70% and 80%), and the data after that are used as the test set for the experiments because the data are time series data and the traditional

random division of training and test sets based on a certain ratio is not applicable. Training sets of various sizes are divided according to the sturdiness of the model, and the training sets are trained by training them at various sizes. The model's robustness is evaluated.

3.2. Comparison results

3.2.1. Stock price time series data forecasting

In this study, we aim to train models to correctly forecast the opening prices of daily and last day of the month using high and low frequency datasets made up of daily and daily data of Guizhou Maotai and BYD stocks. The novel model in this study makes use of the concept of a combination model based on bagging, which can employ various weighting schemes with various base models. In the experimental process, two core functions named result and DW Bagging model are set. Logical values are set by conditional statements in the result function, and when the value is true, it is possible to train on multiple training sets of varying sizes simultaneously; when the value is false, feature variables are chosen. While calling the DW Bagging model function in the result function to predict the data by the weighted bagging method, the weighting method can be artificially specified, and in this experiment, Pearson correlation coefficient and mutual information quantity two methods were used. By setting variables such as base model and number of submodels in the result function, one can achieve control of parameters such as base model and number of submodels of the innovation model. After training and prediction are complete, the MSE and MRE are determined. Many prediction models are chosen in order to examine the impact of the novel model in this paper. Multiple linear regression, Lasso regression, ridge regression, decision tree, random forest, and XGBoost were the models chosen. For analysis, the predicted outcomes under several models are collected and contrasted with the creative model. The comparison's findings are listed below.

Table 2. Experimental results display table for Guizhou Maotai's high frequency data

	SRE_Mean	SME_Mean	SRE_STD	SME_STD
Lasso returns	10.114545	13.535069	0.238463	0.337492
Ridge Return	8.895451	11.561160	0.194488	0.235008
Decision Trees	277.924171	324.581242	216.405258	219.645550
Random Forest	286.593031	331.403215	213.746893	219.337434
XGBoost	276.185259	321.439907	218.710225	224.236564
Mutual Information and Lasso	7.843642	11.330446	0.729830	0.790393
Mutual Information and Decision Trees	286.822384	331.257029	0.729830	219.859419
Pearson and Lasso	11.049864	14.923005	0.852913	1.106437
Pearson and Decision Tree	287.602348	331.977683	213.782918	219.564076

Table 3. Experimental results display table for BYD's high frequency data

	SRE_Mean	SME_Mean	SRE_STD	SME_STD
Lasso returns	3.507212	4.846982	0.473357	0.943073
Ridge Return	1.785272	2.449303	0.137726	0.124623
Decision Trees	124.652432	144.670454	21.689522	22.662242
Random Forest	126.962436	146.937638	20.664659	21.579726
XGBoost	122.742103	142.819516	21.662616	22.494551
Mutual Information + Lasso	1.749028	2.531105	0.065725	0.068694
Mutual Information + Decision Trees	129.137754	153.924986	10.421664	6.256422
Pearson + Lasso	1.886402	2.601444	0.163307	0.129292
Pearson + Decision Tree	129.092403	153.877785	10.423817	6.253027

Table 4. Experimental results display table for Guizhou Maotai's low-frequency data

	SRE_Mean	SME_Mean	SRE_STD	SME_STD
Lasso returns	59.030333	68.522158	5.521555	8.571648
Ridge Return	60.323046	69.756608	6.250029	9.512525
Decision Trees	77.440599	83.152896	75.889095	76.850098
Random Forest	93.750910	98.233131	90.847790	92.268670
XGBoost	66.932650	72.193422	80.456132	81.579844
Mutual Information + Lasso	24.493242	31.226056	3.956555	5.741272
Mutual Information and Decision Trees	94.181882	98.997216	91.101341	92.278196
Pearson and Lasso	27.991612	35.327365	3.945107	6.916891
Pearson and Decision Tree	93.565108	97.957327	90.852780	92.100539

Table 5. Experimental results display table for BYD's low frequency data

	SRE_Mean	SME_Mean	SRE_STD	SME_STD
Lasso returns	2.690528	3.169293	0.448524	0.489545
Ridge Return	3.595791	4.242371	0.783366	0.919204
Decision Trees	11.384209	14.470707	1.774626	1.671875
Random Forest	12.256557	15.327733	1.773176	1.179623
XGBoost	10.802823	13.781264	1.384095	1.405557
Mutual Information and Lasso	1.632432	2.075663	0.505785	0.606347
Mutual Information and Decision Trees	13.068416	16.152920	1.921118	1.389614
Pearson and Lasso	1.737759	2.205734	0.736747	0.886263
Pearson and Decision Tree	13.046002	16.130056	1.767249	1.296152

The base model chosen for the innovation model based on the above results is Lasso regression, weighted by the mutual information quantity. In the high frequency data set of Guizhou Maotai presented in Table 2, the MRE of this innovative model is 7.84 and MSE is 11.33; in the high frequency data set of BYD shown in Table 3, the MRE is 1.75 and MSE is 2.53; in the low frequency data set of Guizhou Maotai presented in Table 4, the MRE is 24.5 and MSE is 31.2; as shown in Table 5 following the low frequency data set of BYD, the MRE is 1.63 and MSE is 0.51. The model showed good results in all four data sets.

Since the innovative model base model and weighting approach in this paper are free, different methods can be selected for different data for prediction and the optimal combination model is found, which improves the accuracy of the submodel. Meanwhile, the use of weighting solves the problem of model overfitting and underfitting, which makes the variance and deviation of the prediction model trade-off. Therefore, it can be considered that the innovative model in this paper is universal and generalizable.

3.2.2. Multi-factor selection results

In this paper, the model is applied to different data backgrounds for variable selection, which expands the application scenario of this innovative model. The results of feature selection in the above four data contexts are as follows:

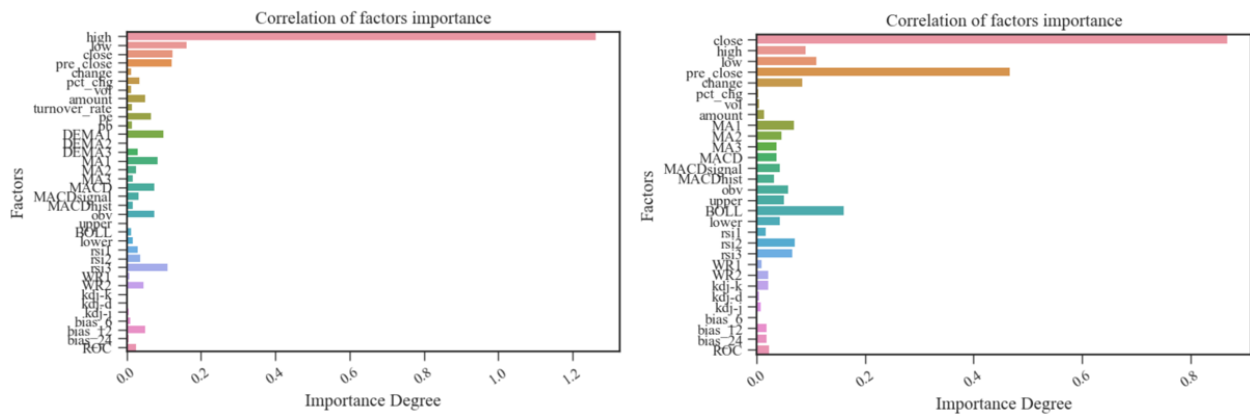


Figure 3. Guizhou Maotai stock high (left) and low (right) frequency data feature selection results

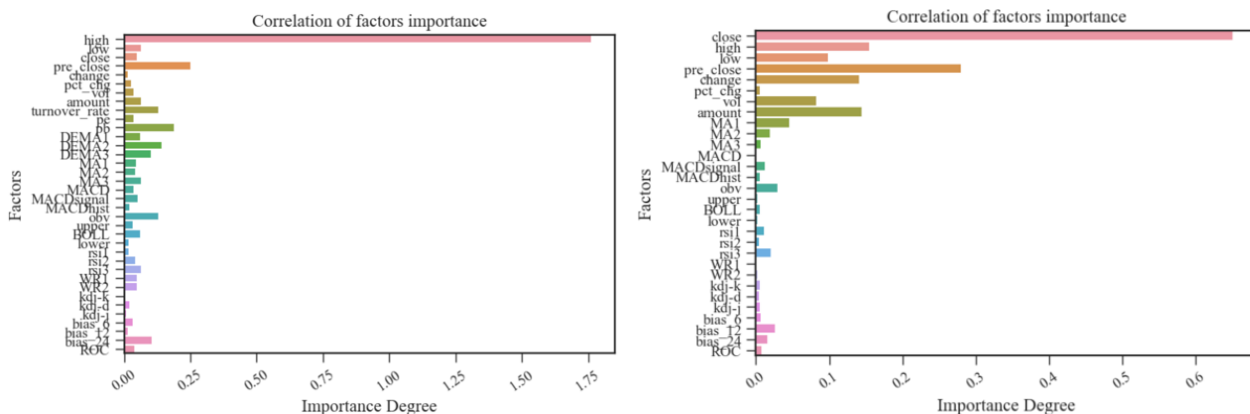


Figure 4. BYD stock high (left) and low (right) frequency data feature selection results

The results of feature variable selection under different data sets can be derived from the above graph. In Guizhou Maotai's high-frequency data set, Figure 3-left shows the top five feature importance from high to low are high price, low price, closing price, 24-day relative strength indicator, and previous day's closing price; in Guizhou Maotai's low-frequency data set, the right one in Figure 3 presents the top five feature importance from high to low are closing price, previous period's closing price, Bollinger band, high price, and volume; in BYD's high-frequency data set, the top five feature importance from high to low are The top five highest price, previous day's closing price, average P/E ratio, energy wave indicator, and 24-day deviation, which is shown in the left one in Figure 4; the top five low-frequency data set features in BYD presented in the left one in Figure 4 from highest to lowest importance are closing price, previous period's closing price, highest price, volume, and lowest price.

4. Conclusions and Recommendations

This paper compares our enhanced model of this paper with the traditional models such as Lasso regression, ridge regression, decision tree, random forest, and Extreme Gradient Boosting (XGBoost) through empirical analysis using both mean square error and mean relative error criterion, and concludes that the weighted Bagging algorithm with the base model of Lasso regression and the independence criterion of mutual information in the data set selected in this paper all present better results. For the feature weights of the final model derived from the improved weighted Bagging algorithm, the feature factor weights are different for different frequencies of data namely low frequency(monthly) and high-frequency(daily). Based on the above conclusions, the model in this paper can be considered as an innovative and constructive one. To begin with, we could flexibly select the base model and independence criterion without restriction to ensure the accuracy of the ensemble model while solving the problem of overfitting or under-fitting of a single model.

Furthermore, the Bagging algorithm, as an ensemble learning algorithm, incorporates multiple seed models to improve the interpretation of the ensemble model.

The innovative approach in this paper could also be applied to spectral data, and the independent criteria can also be used in methods such as maximum information coefficient and method selection method as well. In summary, the heuristic innovation method employed in this study adaptively incorporates a number of base models by various independent criteria to generate the ensemble model feature factor weights, which can be further discussed in terms of feature importance as well as feature selection.

References

- [1] He, P., Lan, W., Ding, Yue. Is China's stock market predictable? -- A perspective based on a combined LASSO-logistic approach [J]. Statistical Research, 2021, 38 (05): 82 - 96. DOI: 10.19343/j.cnki.11-1302/c.2021.05.007.
- [2] Zhang Yiran. Linear regression analysis of stock prices--based on matlab ridge regression analysis [J]. Time Finance, 2013 (03): 198.
- [3] Chen, Quan, Gong, Xuantao. Research on multi-factor optimization model strategy for stocks based on decision tree [J]. Information Technology and Informatization, 2020, No.238 (01): 209 - 211.
- [4] Wen GJ, Yuan LX, Ma XG, et al. A genetic algorithm-based method for CART tree stock selection [J]. Investment and Entrepreneurship, 2022, 33 (15): 31 - 33+50.
- [5] Ma, T., Jiang, F. W., Tang, G. H. Deep learning and factor investment in Chinese stock market-based on generative adversarial network approach [J]. Economics (Quarterly), 2022, 22 (03): 819 - 842. DOI: 10.13821/j.cnki.ceq.2022.03.05.
- [6] Zhou, Liang. A study on multi-factor investment in stocks based on random forest model. Financial Theory and Practice, 2021, 7: 97 - 103 [2023-02-27].
- [7] Meng, S. Y. Research on Bagging regression algorithm based on attribute weights. Modern Electronics Technology, 2017, 40 (1) [2023-02-27]. DOI: 10.16652/j.issn.1004-373x.2017.01.027.<http://kns.cnki.net/kcms/detail/11.2127.TP.20191101.1122.002.html>.
- [8] Tibshirani R. Regression Shrinkage and Selection via the Lasso [J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58 (1): 267 - 288.
- [9] Yao, H. L., Ma, X. Q., Wang H., Li, J. Z. Stock Market Trend Prediction Algorithm Based on Morphological Characteristics and Causal Ridge Regression, Computer Engineering, 2016, 42 (2): 175 - 183,
- [10] Random Forest Model Stock Price Prediction Based on Pearson Feature Selection, Yan Z. X., Qin C., Song G. Computer Engineering and Applications, 2021, 57 (15): 286 - 296,
- [11] Saunders C., Stitson M. O., Weston J. Support Vector Machine Reference Matual, Technical Report, Departm, ent of Computer Science, Royal Holloway, CSD-TR-98-03, 1998 [2023-02-27].
- [12] Li Zhiqin, Du Jianqiang, Nie Bin, Xiong Wangping, Huang Changyi, Li Huan. A review of feature selection methods. Computer Engineering and Applications, 2019, 55 (24) [2023-02-27].
- [13] Yong Ju-Ya, Zhou Zhong-Mei. Multilevel feature selection algorithm based on mutual information. Computer Applications, 2020, 40 (12): 3478 - 3484 [2023-02-27].
- [14] Thomas M. C., Joy A. T., Ruan Jishou, Zhang H. Translation. Fundamentals of Information Theory (original book, second edition). Machine Industry Press, 2007, ISBN 978 - 7 - 111 - 22040 - 4.
- [15] Tao-Yun Cao. A Study of Variable Importance Based on Random Forest. Statistics and Decision Making. 2022, 4: 60 - 63 [2023-02-27]. DOI: 10.13546/j.cnki.tjyjc.2022.04.011.
- [16] <https://kns.cnki.net/kcms/detail/42.1009.C.20220223.1117.011.html>.
- [17] Janitza S., Strobl C., Boulesteix A., An AUC-based Permutation Variable Importance Measure for Random Forests, BMC Bioinformatics 2013, 14-119 [2023-02-27], <http://www.biomedcentral.com/1471-2105/14/119>.
- [18] Yang K, Hou Y, Li K. Variable importance Measure of Random Forest and Its Progress, Public Health School, Harbin Medical University, Harbin 150081. [2023-02-27].