INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 109-112

e-ISSN : 2583-1062

Impact Factor: 5.725

STOCK PRICE PREDICTION USING MACHINE LEARNING

Dr. Antony Cynthia[1], Rithankumar S[2], kiruthic M[3]

[1,2,3]Department of Computer Applications Sri Krishna Arts and Science College Coimbatore,641008, India.

DOI: https://www.doi.org/10.58257/IJPREMS32829

ABSTRACT

At this stage, China's economic development continues to progress, and the emergence of various emerging industries has caused the stock market to show strong volatility. However, people's research and prediction on the stock market have never stopped. In order to improve the accuracy of stock prediction, this paper studies the network model based on the improved support vector machine (SVM) algorithm to realize the correct judgment of the stock price trend, so as to achieve the purpose of accurate stock price prediction, and improve the accuracyof the prediction while ensuring the speed of the model. Experiments show that the proposed prediction model can approximate the short-term price trend of thestock market, and provide a more reliable data basis for the accurate prediction of stock prices, which benefit the high-tech development and progress of the stockmarket

Keywords- Regression, Decision Tree, Random Forest, Support Vector Machine

1. INTRODUCTION

In the past hundred years, the world economy has developed rapidly, which has brought great development opportunities to the financial industry. At the same time, the impact of capital market on the quality and speed of economic developmentis also expanding, so the role of the stock market is becoming more and more important. Stock market is not only one of the important ways of financing for listed companies, but also with the improvement of people's living standards, buying stocks has become a very common way of investment in people's life Investors earn the price difference by buying and selling stocks. Predicting the rise and fall of stock price is an importantbasis for predicting the stock market and making buying and selling operations. The rise and fall of stock pricedetermines the loss and profit, so there is a certain risk in investing in stocks. People expect to reduce the risk and increase the return in the stock investment.If people only rely on subjective judgment and estimation to predict the stock priceand operate, it will not onlyincrease the risk, but also greatly reduce the return.Stock selection, timing and trading strategy are the three main research directionsto reduce the risk and increase the return of stock investment. Among them, quantitative investment and portfolio investment strategy are the two most commonly used methods. The formulation of these investment strategies generally needsto be combined with the prediction of the trend of stock price rise and fall. Therefore, how to improve the accuracy of stock price prediction has always been a hot issue.

2. METHODOLOGY

LOGISTIC REGRESSION

Logistic regression is a statistical method used for binary classification tasks, where the goal is to predict the probability of an observation belonging to one oftwo possible classes. It's called "logistic" because it models the probability using the logistic function. Here's a brief explanation:

Model Representation: Logistic regression models the relationship between the independent variables and the categorical dependent variable by estimating probabilities using a logistic function

Probability Estimation: The logistic function (also known as the sigmoidfunction) transforms any real-valued number into a value between 0 and 1,which can be interpreted as a probability. This function maps any input toa value between 0 and 1, making it suitable for binary classification problems.

Decision Boundary: Logistic regression calculates a decision boundary that separates the classes. This boundary is linear in the feature space, dividing it into two regions corresponding to the two classes.

Parameter Estimation: The model's parameters (coefficients) are estimated using optimization techniques like maximum likelihoodestimation. The objective is to find the parameters that maximize the likelihood of the observed data given the model.

Predictions: Once the model is trained, it can predict the probability of anobservation belonging to a particular class. Typically, a threshold (e.g., 0.5)is applied to these probabilities to make binary predictions.

Evaluation: Logistic regression models are evaluated using metrics like accuracy, precision, recall, F1-score, or area under the ROC curve (AUC-ROC) depending on the specific requirements of the classification task

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 109-112

e-ISSN : 2583-1062

Impact Factor: 5.725

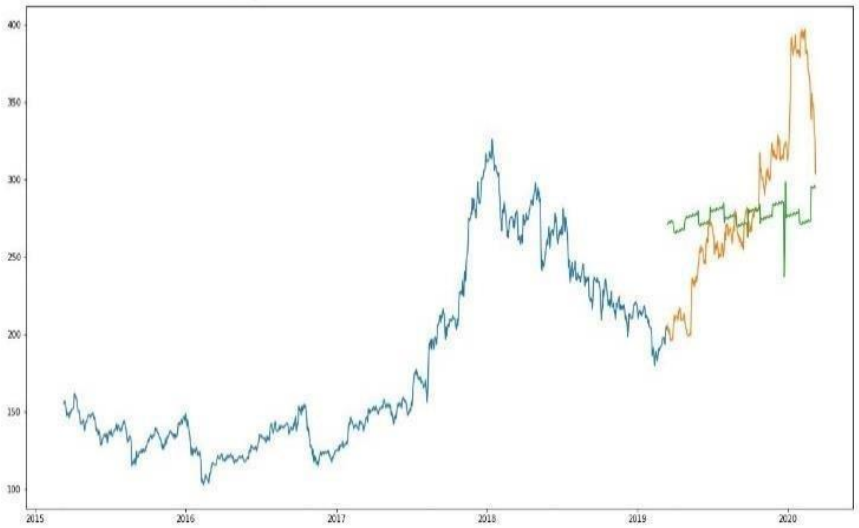

Fig. 1. Logestic regression

3. SUPPORT VECTOR CLASSIFIER

Support Vector Machine (SVM) is a supervised machine learning algorithm thatcan be used for stock price prediction. Here's a brief explanation of how SVM can be applied

Model Representation: In SVM, the goal is to find the hyperplane that best separates the data into different classes. For stock price prediction, theclasses could be "buy" and "sell" or "increase" and "decrease" in price.

Feature Selection: Features used for prediction could include historical stock prices, trading volume, technical indicators (such as moving averages, relative strength index), and fundamental factors (like earnings reports, economic indicators).

Kernel Trick: SVM can use different kernel functions to map data into higher-dimensional space where it might be more separable. Common kernels used include linear, polynomial, radial basis function (RBF), and sigmoid.

Training: The SVM algorithm aims to find the hyperplane that maximizesthe margin between the classes. This margin represents the separation between different classes, providing robustness to new data points.

Prediction: Once trained, the SVM model can predict the movement of stock prices based on new input features. For example, if the model predicts an increase in stock price, it suggests buying, whereas if it predictsa decrease, it suggests selling.

Evaluation: The performance of the SVM model for stock price predictioncan be evaluated using metrics such as accuracy, precision, recall, or the profitability of the trading strategy based on the model's predictions.

It's important to note that while SVM can be applied to stock price prediction, it'sjust one of many possible approaches, and its effectiveness can vary depending on factors such as the choice of features, the quality of data, and market dynamics. Additionally, predicting stock prices accurately is challenging due to the inherentvolatility and complexity of financial markets. Therefore, it's often recommendedto use SVM in combination with other techniques and to consider expert advice when making investment

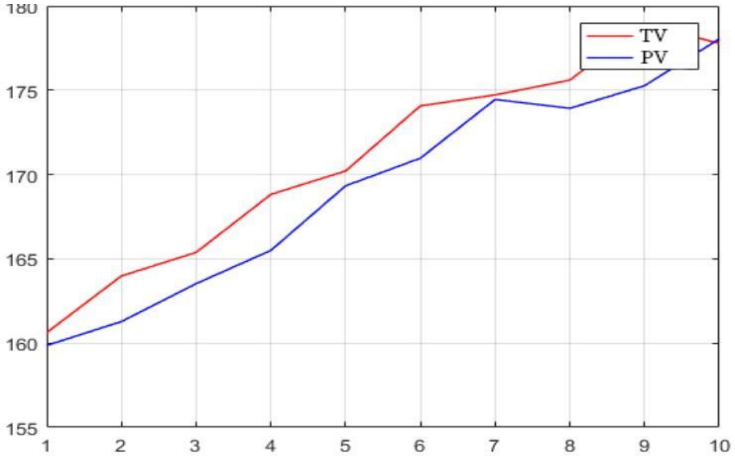

Fig. 2. Logestic support vector classifier

.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 109-112

e-ISSN : 2583-1062

Impact Factor: 5.725

4. RANDOM FOREST CLASSIFIER

A Random Forest Classifier is a machine learning algorithm that utilizes an ensemble of decision trees to make predictions. Here's a brief explanation of howit can be used in stock price prediction:

Ensemble of Decision Trees: A Random Forest consists of a collection of decision trees. Each tree is trained independently on a subset of the data and makes its own prediction.

Feature Selection: At each node of a decision tree, a subset of features isconsidered for splitting. This randomness helps in reducing correlation between trees and prevents overfitting.

Voting Mechanism: Once all trees are trained, predictions are made by each tree. In classification tasks, the class predicted by the majority of trees(mode) is chosen as the final prediction. In regression tasks like stock priceprediction, the average of predictions from all trees is often taken.

Robustness and Generalization: Random Forests are robust against overfitting and noise in the data due to the randomness in feature selectionand the aggregation of predictions from multiple trees. They also generalize well to unseen data.

Feature Importance: Random Forests can provide insights into feature importance, indicating which features (such as historical stock prices, trading volumes, economic indicators, etc.) are most influential in predicting stock prices.

A Random Forest Classifier algorithm can be effective in stock price prediction by leveraging the power of ensemble learning, robustness against overfitting, andproviding insights into feature importance.

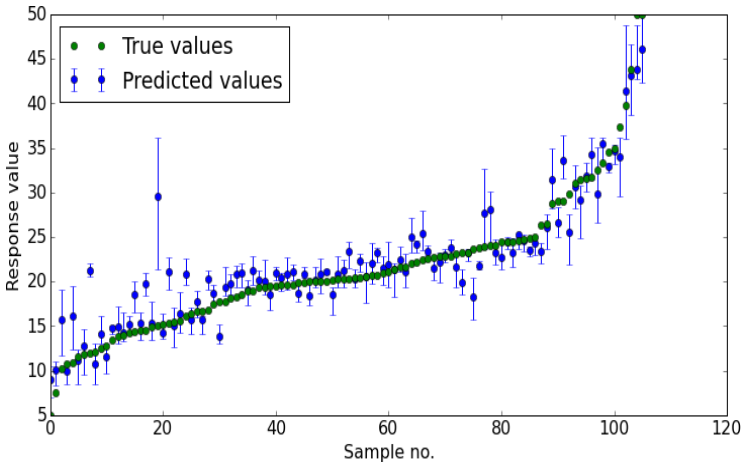

Fig. 3. Random Forest classifier

5. EXTRA TREE CLASSIFIER

Extra Trees Classifier (Extremely Randomized Trees) is a variation of theRandom Forest algorithm that introduces additional randomness in the tree- building process. Here's a brief explanation of how it can be applied to stock price prediction:

Random Feature Selection: Similar to Random Forests, Extra Trees randomly select a subset of features at each node for splitting. However, instead of searching for the best possible split, Extra Trees select random thresholds for each feature, leading to even more randomness in the model.

Randomly Sampled Data: Like Random Forests, Extra Trees also build multiple decision trees on randomly sampled subsets of the training data. This further increases the diversity among the trees in the ensemble.

Voting Mechanism: Predictions are made by aggregating the predictions from all the trees in the ensemble. In classification tasks, the majority classpredicted by the trees is chosen as the final prediction. For regression taskslike stock price prediction, the average of predictions from all trees is oftentaken.

Robustness and Generalization: Extra Trees, by introducing more randomness in the tree-building process, are robust against overfitting andnoise in the data. They can generalize well to unseen data, similar to Random Forests.

Computational Efficiency: Extra Trees can be computationally more efficient compared to Random Forests since they do not require searchingfor the best split at each node. This can be beneficial when dealing with large datasets or when computational resources are limited.

Extra Trees Classifier offers a variation of the Random Forest algorithm with additional randomness, which can be useful in stock price prediction tasks by providing robustness against overfitting and noise, along with potential computational efficiency benefits.

.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 109-112

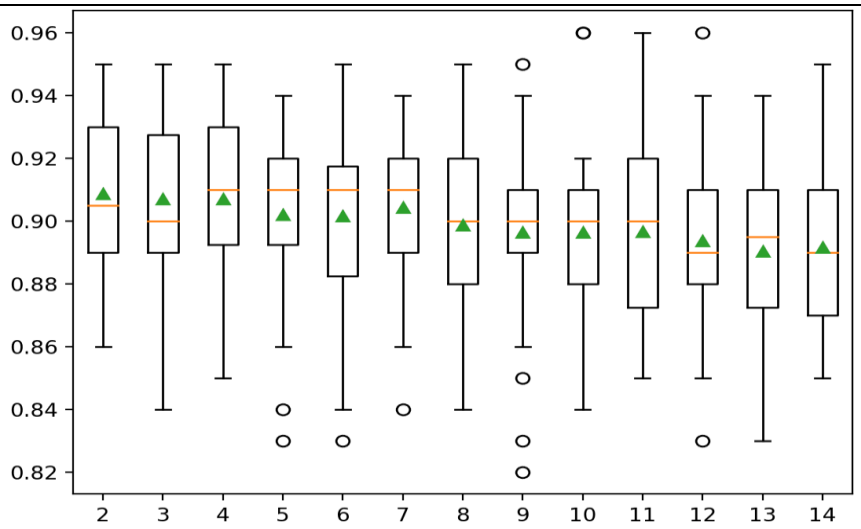

Fig. 2. Extra Tree classifier

6. CONCLUSION

The current stock market prices in China are showing relatively obvious volatility, which poses a greater challenge to the prediction of stock prices. In view of the difficulties in the current stage of stock market price prediction, this paper opens up a new research idea. Based on the traditional SVM algorithm, thispaper optimizes and improves the model according to the characteristics of the overall trend of stock prices, realizes a stock price prediction model with higher accuracy, and provides a reliable basis for applications in stock market investmentand stock analysis. However, in order to realize the practical value of prediction models in quantitative investment, the stability and applicability of prediction need to be further strengthened, and the accuracy of forecasting results also needto be improved to a certain extent.

7. REFERENCES

[1] A. Kong, H. Zhu, "Predicting Trend of High Frequency CSI 300 Index Using Adaptive Input Selection and Machine Learning Techniques," Journal of Systems Science and Information, vol. 6, pp. 120–133, 2018

[2] A. K. Jain, J. Mao, K. M. Mohiuddin, "Artificial Neural Networks: A Tutorial," Computer, vol. 29, pp. 31–44, 2015

[3] M. Santoso, R. Sutjiadi, R. Lim, et al, "Indonesian Stock Prediction Using Support Vector Machine (SVM)," Matec Web of Conferences, pp. 164, 2018.

[4] L. I. Hong, Y. W. Hong, "Prediction of Stock Index Based on Fractal Interpolation and SVM," Journal of Jishou University (Natural Sciences Edition), vol. 3, pp. 19–24, 2018.

[5] J. Y. Heo, J. Y. Yang, et al, "Stock Price Prediction Based on Financial Statements Using SVM," International Journal of Hybrid Information Technology, vol. 9, pp. 57–66, 2016.

[6] H. Li, Y. H. Zhao, "Stock Trend Predicting Method Based on DFS-BPSO-SVM Predicting Model," Software Guide, vol. 16, pp. 147–151, 2017.

[7] J. H. Xiao, X. H. Zhu, C. X. Huang, et al, "A New Approach for Stock Price Analysis and Prediction Based on SSA and SVM," International Journal of Information Technology & Decision Making, pp. 1–17, 2018.

[8] L. Tang, S. Zhang, L. He, et al, "Research on Stock Prediction in China based on Social Network and SVM Algorithm," 2nd International Conference on Economic Development and Education Management (ICEDEM 2018), pp. 435-438, 2018.

[9] H. Y. Wang, H. Li, J. Y. Shen, et al, "A Novel Hybrid Fractal Interpolation-SVM Model for Forecasting Stock Price Indexes," Fractals, pp. 1-20, 2019.

[10] C. Xiao, W. Xia, J. Jiang, et al, "Stock Price Forecast Based on CombinedModel of ARI-MA-LS-SVM," Neural Computing and Applications, pp. 1-10, 2020.

[11] Y. Zhang, B. Li, H. Lu, et al, "Sample-Specific SVM Learning for Person Re-identification," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[12] S. M. Erfani, S. Rajasegarar, S. Karunasekera, et al, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," Pattern Recognition, pp. 121-134, 2016.