Research on the Stock Price Prediction Model of Banks Based on SVM and BP Neural Network

Junbin Zhang[1], Peiying Zhang[1,2], Jinsui Huang[1,3](✉), Wennan Wang[4], and Tiancheng Wang[5]

[1] Faculty of Finance, City University of Macau, Macao, China
F20092100174@cityu.mo, huangjscn@outlook.com
[2] Zhongkai University of Agriculture and Engineering, Guangzhou, China
[3] Lingnan Normal University, Zhanjiang, China
[4] Academy of Management, Guangdong University of Science and Technology, Dongguan, China
[5] Alibaba Cloud Big Data Application College, Zhuhai College of Science and Technology, Zhuhai, China

Abstract. The classification and prediction of stock price fluctuation pattern is a very important issue in stock market research in recent years. The rise and fall of stock price directly reflects the change of investors' assets, and if the rise of stock price can be predicted more accurately, it can provide investors with a certain degree of assisted decision-making.

The banking industry is an important part of the stock market, and the use of models to accurately predict stock price movements is of great practical significance to investors in formulating investment strategies and avoiding market risks, thus more and more scholars are focusing on stock price related research in the banking industry. As stock price movements become more and more complex, simple time series models cannot solve the non-stationarity and non-linearity of stock data, and the shortcomings of traditional linear models are gradually exposed. Historical studies more often use the historical information of individual stocks to predict the future trend of stock prices, and rarely consider the linkage between stocks in the same market. In this paper, we use the information of linked banking stocks for forecasting, and use a combination of neural network and support vector machine to predict stock price patterns. Stock data from January 2020 to June 2020 for the two major indices of the Chinese stock market are used to predict stock trends and stock closing price regressions using two algorithms, support vector machine classification and BP neural network, respectively, and the results show that the prediction results of the support vector machine model after data preprocessing have better prediction results than the original data to build the support vector machine model. This is because they have an average accuracy of 79.13% for support vector machines and 78.7% for BP neural networks. To compare the model accuracy, convolutional neural network and traditional ARIMA model are also built in this paper, and the results show that the support vector machine has the highest prediction accuracy.

Keywords: Support Vector Machine · Machine learning · Stock price model prediction · BP neural network

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 1273–1280, 2023.
https://doi.org/10.2991/978-94-6463-198-2_132

1 Introduction

It is essential that investors on the stock market make the right decisions by having access to fast and accurate information. Due to the fact that a large number of stocks trade on stock exchanges, there are a number of factors that influence stock prices. Economic, social, and political factors may also have an impact on stock prices. In recent years, more and more financial anomalies appear in the stock market, accompanied by plate rotation, concept speculation and other phenomena, which are mostly caused by the complex and changeable factors of the market.[1] Furthermore, it is difficult to predict stock prices since they are uncertain. Research into machine learning techniques and algorithms has progressed substantially in the last few years. This has led to improvements in the accuracy of stock price prediction that can be derived from a large amount of related literature [2]. As far back as a few decades ago, scholars only treated stock data as a traditional time series when studying it. Because stock price forecasting is similar to time series analysis, this method of research was the most common method of study used by previous researchers. However, the volatility that stock prices suffer from is unsteady and noisy, so the traditional time series approach has its limitations due to the volatility of stock prices. Stock prices are often non-stationary because they are affected by many factors, and as a consequence, prediction is also limited by the fact that they are not always predictable. As we move into the era of colossal data, machine learning methods are gradually gaining traction among scholars. They are being used widely in the field of finance as they become more widely available. In order to form an effective machine learning model, historical data is usually used, in order to form a model that can be applied to the future. The process consists mainly of input samples, continuous training, adjusting parameters, and finally displaying the final prediction result as a result of inputting updated data. This is in order to demonstrate the accuracy of the prediction.

The support vector machine regression model (SVR) has been used for predicting stock prices in three different markets of large and small capitalization, and it has been compared with the random wandering model in previous studies. In these studies, it has been demonstrated that the SVR model has high prediction accuracy even during periods of low volatility [3]. As part of the analysis of the short-term stock price increase and decrease prediction model, a system of characteristic variables constructed by selecting six types of financial indicators was selected for building mathematical models for comparative analysis based on three machine learning algorithms: random forest, support vector machine, and ridge regression. The system of characteristic variables was constructed by selecting six types of financial indicators. As a result of the results of the analysis, it was found that the highest performance was achieved by using a combination of principal component analysis and support vector machines [4]. As a machine learning method, the support vector machine and BP neural network methods are used in this paper. This machine learning method has the advantage of providing better performance in both theoretical and practical simulations than their more popular counterparts.

2 Research Methodology

Stock historical data is a kind of time series with noise, how can we extract effective information from it is very important, this paper cleans the data by principal component analysis, extracts the maximum possible effective information, and then selects the support vector machine model with optimal parameters to predict the stock closing price, comparing the traditional time series model and BP neural network which is widely used in financial research. A good development of predictive analysis. Similarly, in practical applications, the model is trained to predict stock prices through historical data. If the model predicts that the closing price on the next day is higher than that of the same day, then the stock price will rise in the future, and stockholders can continue to hold their positions to gain income. Similarly, the stock price prediction can also enable the country to have a deeper understanding of the current economic situation and industry development, so as to make timely macro-control.

2.1 SVM

In addition to statistical learning theory and supervised learning, support vector machines are founded on the theory of supervised learning. They are machine learning methods based on the theory of statistical learning, which is a branch of machine learning. As shown in Fig. 1. The "VC Dimension Theory" is based on the concept of structural risk minimization, which is also rooted in the principles of probability theory, which is a theory that is founded on the concept of probability theory. Despite the method's strengths, it remains limited because of its ability to generalize. This means a small sample size can still produce reliable results even when there is a large number of participants involved [5]. In the field of classification and regression, Support vector machines are commonly used for approximation problems, nonlinear problems, and dimensional catastrophes. This is due to their power to solve these types of problems. The reason for this is that global optimal solutions and fast convergence are the two advantages of global optimal solutions [6].

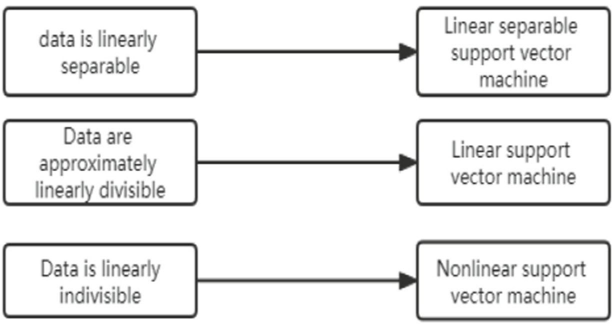

Fig. 1. Support vector machine model type

The idea of the support vector machine method is to find a suitable separation hyperplane to separate the sample data set, and then maximize the distance of each classification data to the separation hyperplane. In the case that the sample data is linearly separable, a linear function is constructed to separate the sample data according to the current spatial characteristics, and in the case that the sample data is nonlinear, a kernel function technique is used to map the low-dimensional data into the space with higher dimensions, and then a linear classification method is used to construct a separation hyperplane to separate the data, which successfully transforms the nonlinear problem into a linear one. Given a linearly separable training data set, the separated hyperplane obtained by solving the corresponding convex quadratic programming problem by interval maximization or equivalently learning is:

$$w^* \cdot x + b^* = 0 \tag{1}$$

Corresponding classification decision function:

$$f(x) = sing\left(w^* \cdot x + b^*\right) \tag{2}$$

Let the sample training set be:

$$\mathrm{F} = \left\{\left(\mathrm{x_i}\,,\ \mathrm{y_i}\right) \mathrm{i}\ \in\ \mathrm{N}^*, \mathrm{x}\ \in\ \mathrm{R}^*,\ \mathrm{y}\ \in (-1, 1)\right\} \tag{3}$$

2.2 BP Neural Network

In terms of structure and implementation mechanism, the artificial neural network is considered to be a simulation of a biological neural network. It is composed of a large number of neurons connected by a complete and rich network of links. There has been a great deal of development of neural networks in various fields such as data mining, economics, healthcare, transportation, etc. It is widely accepted that error back propagation neural networks, also known as BP networks, are used widely in the financial industry. As part of reducing the dimensionality of data, multi-layer neural networks can be applied to provide a more essential picture of the data and allow for better classifications and visualizations of the data as well as reducing the dimensionality of data. It is a process called "layer-by-layer initialization" that is used to set up deep neural network algorithms layer-by-layer in order to overcome the problem of deep neural networks. A linear fit is more likely to provide an explanation when analyzing small samples (30), and in particular when analyzing small samples. In this case, the linear analysis results are not satisfactory when analyzing large samples, and the problem of poor adaptability appears when analyzing large samples. The nonlinear characteristics of BP neural networks are therefore used to handle large samples, and to find more relevant financial indicators by using their nonlinear characteristics. In order to measure stock price fluctuations, it is necessary to fit the correlation between different financial indicators and stock price fluctuations into the model. The second type of model relies on computer simulations, which can be trained to adapt to various situations as well as be self-trained to adapt to different situations with the help of these simulations [7]. As multilayer neural networks can provide a more fundamental depiction of data, it is considered beneficial to reduce the dimensionality of the data represented by multilayer neural networks. This is because they can provide a more essential representation of data. By doing this, it is possible to classify and visualize data in a more effective manner. As a result of a related study, neural networks were used to create an accurate forecast of the London stock market over the past twenty-five years with an accuracy rate of more than 60% [8].

Therefore, the purpose of this paper is to use support vector machines and BP neural networks to predict stock prices. In order to achieve the most effective result, both advantages of these methods will be combined after comparison. This will allow us to utilize more comprehensive stock market information in order to improve the accuracy of our stock price model prediction in the future.

3 Research Data

It is critical to choose different data indicators for different research purposes. It is pertinent to note that in cluster analysis of stocks, data indicators relating to the number of daily trades on stocks as well as financial indicators including profitability, growth capability, and solvency are the primary considerations. It is usually the case in forecasting models that trend-oriented and momentum-oriented indicators, such as moving averages and their combinations, or price volatility, are chosen for their construction in order to enhance the forecasting effect of the models, which are more sensitive to changes in stock prices, as well as time-sensitive, as well. Data selected for this study are real-time data for the stocks of 10 Chinese banks from January 1, 2020 to June 30, 2020 for the period from January 1, 2020 to June 30, 2020. For the purpose of the analysis, more recent dates are considered to have a higher practical significance compared with older dates. Throughout this study, the neural network training model for the BP neural network was developed using Python version 3.6 as the primary computing environment. It was also developed using Windows 10 x64 as a secondary computing environment. It was decided to apply the standard BP neural network parameters for writing the BP neural network based on the comprehensive characteristics of financial indicators and stock price fluctuations. A generalized model with basic parameters has the advantage of being malleable on the one hand, and robust on the other. A generalized model can be used for a variety of purposes due to the flexibility of the language structure used in the program. A brand-new function can be easily added to the program in order to enhance its functionality. In order to get a basic idea of what it's like to trade on a particular day, you can use the daily stock quotes indicator. As measured by the indicators in Table 1, the purpose of this document is to enable the researcher to be able to identify the general movements and quotes of the stock as they are reported in Table 1. The information in this document is based on studies related to this topic [9]. In the dataset, there are 936 valid pieces of information. The first 70% of these data are extracted to form the training dataset, while the last 30% are used to construct the test dataset.

4 Experimental Results

As shown in Fig. 2. Based on the results of the study, such prediction models are more accurate in predicting the price increase classification of bank sector stocks in general. This is because they have an average accuracy of 79.13% for support vector machines and

Table 1. The trading day market indicators used in this study

Act Pre Close Price	Actual yesterday closing price
Open Price	The opening price of the stock
Highest Price	The stock's high for the day
Lowest Price	Lowest price of the day
Close Price	Closing price
Turnover Vol	Trading volume
Turnover Value	Daily turnover of shares (Yuan)
Turnover Rate	Daily turnover rate = (Turnover/number of unlimited outstanding shares)
Chg Pct	Increase or decrease = (Closing price/Previous closing price -1)
PE	Forward P/E = (total market value/net profit attributable to parent owner)
PB	Price-to-book ratio = (Total market value/total equity attributable to parent company)
Market Open Day	Whether the stock is open for the day
vwap	Transaction amount/volume

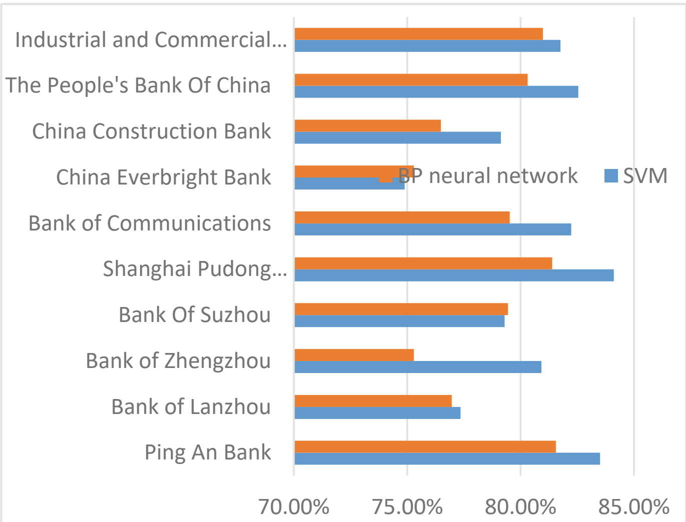

Fig. 2. Different methods of investment success rate comparison table

78.7% for BP neural networks, in predicting the stock prices of 10 banks. A helpful way to discover why bank stocks have increased in price is to look at the specific performance of the stock. This can be used to assist investors in making better investment decisions.

In order to increase the reliability of data, the commonly used autoregressive moving average model (ARMA model) is added in this paper, which is a random time series analysis model proposed by Box and Jenkins in the 1970s. Its basic idea is that some time series are a family of random variables that depend on time. Although the single column value of the time series is uncertain, the change of the whole series has certain regularity, which can be approximated by the corresponding mathematical model. AEMA model is used to forecast the stock price of ICBC 20 days after January 31, 2020 in the data adopted in this paper. The predicted results were compared with the actual stock prices. In fact, the predicted results within 5 days were significant, and the similarity between them and the results of support vector machine algorithm reached 80%.

5 Conclusion

In this study, a support vector machine model as well as a BP neural network have been employed. They were used. They were used in order to forecast price fluctuations associated with the stocks of ten banks, using a support vector machine model. Moreover, it is pertinent to note that the support vector machine model has better results than the BP neural network when it comes to predicting human behavior. In other words, it has a lower RMSE and MAE than the BP neural network, which is due to the lower RMSE and MAE. Thus, the support vector machine is capable of predicting the future with significantly higher accuracy than the regression machine when it comes to predicting the future. As can be seen from the table, when comparing the two models, there does not appear to be a significant difference between them when compared side by side. Despite the fact that the closing price of the stock of the company is more volatile, it is more likely that the BP neural network will be less effective if the stock price is more volatile. The reason for this is that it runs for a longer period of time than other programs. The results of an in-depth comparison of the support vector machine model with the rule-based model have revealed that the support vector machine model is more effective when it comes to analyzing stock market data than the rule-based model and is, therefore, a better tool for predicting the Chinese stock market than the rule-based model. It is also possible to use the preprocessed data as a reference for future studies of the stock market based on the preprocessed data. This can be used in conjunction with the preprocessed data itself.

References

1. Dai Shugeng, Lu Bin. An empirical analysis of the changing trend of incomplete information trading and herd effect in stock market [J]. Broad Virtual Economy Research,2016,7(04):33-41.
2. Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction: Algorithms and techniques.
3. Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. Neural Computing & Applications, 10, 184-192.
4. Liu Xinyue, & Cheng Ximing. (2022). Stock Price fluctuation prediction model based on financial indicators. Journal of Information Science and Technology University of Beijing (Natural Science Edition).

5. Yang, R., Yu, L., Zhao, Y., Yu, H., Xu, G., Wu, Y., & Liu, Z. (2020). Big data analytics for financial Market volatility forecast based on support vector machine. International Journal of Information Management, 50, 452-462.
6. Ding, Y., Song, X., & Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. Expert Systems with Applications, 34(4), 3081-3089.
7. Jin, W., Li, Z. J., Wei, L. S., & Zhen, H. (2000, August). The improvements of BP neural network learning algorithm. In WCC 2000-ICSP 2000. 2000 5th international conference on signal processing proceedings. 16th world computer congress 2000 (Vol. 3, pp. 1647–1649). IEEE.
8. Zhang, Z., Zohren, S., & Roberts, S. (2019). Deeplob: Deep convolutional neural networks for limit order books. IEEE Transactions on Signal Processing, 67(11), 3001-3012.
9. Yang Jia. (2020). Based on principal component analysis and support vector machine (SVM) classification of share price increase prediction research (a master's degree thesis, relearn). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202202&filename=1020355973.nh