Stock Price Prediction and Portfolio Optimization Based on Mean Variance Model and Random Forest Model

Rundong Chen

Finance, Jinan University, Guangzhou, 511436, China
wangyuping@tzc.edu.cn

Abstract. In order to show the applicability and accuracy of different models for predicting stock prices, it is necessary to select classical models for effective comparison. In this paper, the termination model and random forest model are used to analyze and compare five representative American stocks in detail. This paper constructed two different portfolios, each with a unique investment strategy, ranging from maximizing the Sharpe ratio to minimizing risk. To predict future returns and optimize these portfolios, this study utilized the Random Forest method, a robust machine learning algorithm known for its versatility in handling various types of data and its ability to model complex interactions. The analysis of the actual stock market data indicates that the random forest algorithm has a better prediction effect in the stock market quantification. The algorithm can accurately predict the rise and fall trend of stock prices, and can provide the probability of each stock's rise or fall. In a word, it provides more decision basis for investors.

Keywords: Mean variance model, Random forest model, Portfolio.

1 Introduction

For people in modern society, stocks are an essential way to invest and finance. Since the birth of stocks in the world, people have thought of different ways to predict the price of stocks. In the past, people usually predicted the stock market based on their own experience or intuition. These methods are often not very effective. However, these methods cannot scientifically analyze the entire historical data and market conditions and have considerable personal subjectivity. With the rapid development of data science and artificial intelligence, many new models have emerged that can effectively deal with large-scale data and more complex nonlinear relationships. These models are suitable for helping investors make predictions.

In the past, many studies focused on one single model, such as studying time series analysis, machine learning, or natural language processing. An article from 2014, for example, mentioned the use of trading networks to predict stock prices. They characterize the daily trading relationship among its investors using a trading network [1]. The other article is to study the machine learning algorithm to predict the stock price trend of the stock market, mainly studying the algorithm of Long Short-Term

J. Lu (ed.), *Proceedings of the International Workshop on Navigating the Digital Business Frontier for Sustainable Financial Innovation (ICDEBA 2024)*, Advances in Economics, Business and Management Research 315,
https://doi.org/10.2991/978-94-6463-652-9_67

Memory (LSTM) [2-4]. The applicability and accuracy of the algorithm are tested [5]. Specifically, it explores the effect of the method on the stock price.

At the same time, many papers study the performance and effect of different models in stock price prediction. Different models' characteristics, advantages, and disadvantages are analyzed by comparing the stability, prediction accuracy, or complexity of different models. It is feasible to analyze how accurately different models and combinations match specific prediction tasks. For example, in a 2020 article, the paper compared nine different machine learning approaches to two powerful deep learning approaches [6].

Much of the research has focused on the role of individual models in predicting stock prices. And in fact, different models' characteristics, advantages, and disadvantages are analyzed by comparing the stability, prediction accuracy, or complexity of different models. It is feasible to analyze how accurately different models and combinations match specific prediction tasks.

Therefore, this paper focuses on selecting two important representative models from the two key types of mean-variance model and machine learning model respectively to analyze the rate of return problem based on stock selection optimization: mean-variance model and random forest model. These two representative models will significantly differ in the final analysis and results. As a result, different models can be analyzed relatively objectively.

In order to show the accuracy and timeliness of stock data, this article collects stock information (open, high, low, close/adjacent close, and volume price) from these companies from Yahoo Finance: the stocks of Apple, Microsoft, Amazon, Tesla, Nvidia. And use data such as “Moving averages, Bollinger Bands, and RSI” to analyze their performance from 2018 to 2023. After data collection, the above two representative models will be used to make predictions. Based on the research results, further analysis and comparison will be conducted.

2 Method

2.1 Data Preparation

This article focuses on evaluating the performance and predictability of a diverse portfolio consisting of 5 major stocks from various industries: the stocks of Apple, Microsoft, Amazon, Tesla, Nvidia, including technology, finance, consumer goods, and energy. They have an important position in the market. The data spans from July 1, 2018, to July 1, 2023, a period that encompasses significant global events such as the COVID-19 pandemic, supply chain disruptions, and geopolitical tensions. These events have considerably impacted impacts on market dynamics, making this period a rich source for analysis.

In order to show the accuracy and timeliness of stock data, this article collects stock information (open, high, low, close/adjacent close, and volume price) from these companies from Yahoo Finance: the stocks of Apple, Microsoft, Amazon, Tesla, Nvidia. Reviewing the history of each stock, analyzing their historical trend, analyzing volatility, historical return and other data, it helps to get a preliminary understanding of the history of these stocks. And use data such as “Moving averages, Bollinger Bands, and RSI” to analyze their performance from 2018 to 2023. Crunching the data upfront allows them to assess the market effect of these five stocks from 2018 to 2023. The historical data for some sample stocks are provided in Fig. 1.

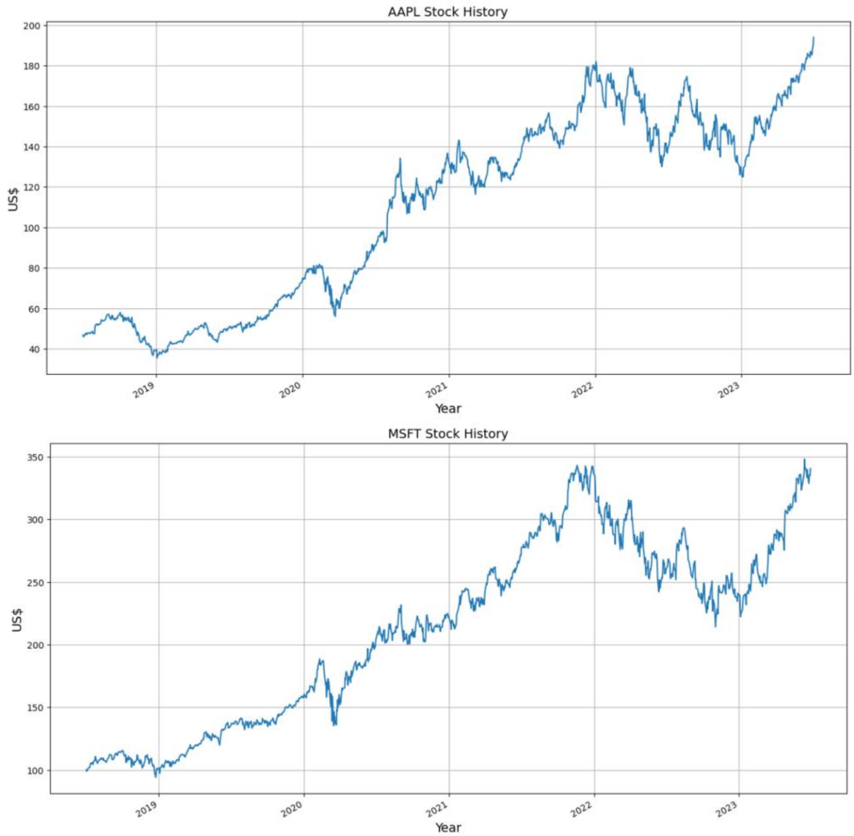

Fig. 1. The historical data for some sample stocks (Photo/Picture credit: Original).

2.2 Stock Price Prediction

In order to compare and reflect the differences between different model methods in predicting stocks, it is important to select some representative models. So, this paper focuses on selecting two representative models: mean-variance model and random forest model. It is easier to focus through the contrast of the two models, which can be detailed for more in-depth discussion and intuitive presentation.

Mean Variance Model. Mean-Variance Model is the classic mean-variance model of portfolio Model, which is a portfolio selection based on portfolio returns and risks [7, 8]. The model is designed to help investors make trade-offs between risk and return and optimize portfolio choices. The objective of portfolio allocation is to build portfolios with varying investment preferences, aided by the Mean-variance Model. The primary aim of the Mean-variance Model is to maximize the Sharpe ratio or minimize risk. This paper typically constructs several portfolios, taking into account additional factors such as Sharpe ratio, volatility, momentum, and others, to create personalized portfolios.

Random Forest Model. Random Forest is an ensemble learning method that improves the overall prediction accuracy by building multiple decision trees and summarizing their predictions [9, 10]. This method is mainly used for classification and regression tasks, and is widely used because of its ease of implementation, fast running speed, high accuracy, and strong robustness to noise and outliers in data sets.

This article trains a Random Forest model for each stock, which refers to a classifier that uses multiple trees to train and predict samples. Random forest algorithm, as an excellent machine learning algorithm, has been widely used in stock market quantification. The random forest model has many advantages. It can produce highly accurate classifiers while also handling a large number of input variables that can assess the importance of each feature to the model. In the stock market, it has high accuracy and real-time performance and can be well predicted.

2.3 Portfolio Optimization

The purpose of the portfolio allocation is to construct portfolios with different investment preferences with the aid of Mean-variance Model. The primary goal of Mean-variance Model is to maximize the Sharpe ratio or minimize risk. This paper typically constructs some portfolios with additional consideration of Sharpe ratio, Volatility, Momentum and so on to make personalized portfolios.

The first one is Maximization of Sharpe Ratio. The objective of this portfolio is to maximize the Sharpe ratio, which is the ratio of excess return over the risk-free rate to the portfolio's standard deviation. It is the ideal model of risk-return trade-off, aiming to achieve the highest risk-adjusted performance.

This article also has Minimize Risk. This portfolio's goal is to minimize risk, measured by the standard deviation of returns. The optimization process allocates higher weights to less volatile stocks. The primary advantage of this portfolio is its risk minimization, which is consistent with the preference for risk-averse investors, but the low expected return might not meet the objectives of more aggressive investors.

3 Results and Discussion

Table 1. Experimental result

Heading level	Apple	Microsoft	Amazon	Tesla	Nvidia
Maximum Sharpe (Mean Variance)	0.3931	0	0.0284	0.2302	0.3483

Minimize Risk (Mean variance)	0.3057	0.1705	0.5238	0	0

Table 1 provides related experimental results. The primary goal of Mean-variance Model is to maximize the Sharpe ratio or minimize risk. For the first portfolio of maximization of Sharpe ratio, it has a Sharpe ratio of 1.168, a return of 51.84% and a risk of 41.02%. This goal of minimizing risk is measured by the standard deviation of returns. It has a Sharpe ratio of 0.849, a return of 29.27% and a risk of 29.86%.

Then given different portfolio from part2 and apply matching learning method to predict the returns, make investments in Jan 2018. For the first portfolio of maximization of Sharpe ratio, predicted return of portfolio 1 for July 2023 is 0.01380 and actual return of portfolio 1 for July 2023: 0.01380. For the minimize risk, predicted return of portfolio 2 for July 2023: 0.00855. actual return of portfolio 2 for July 2023: 0.00855.

From the results, the first portfolio has a Sharpe ratio of 1.168, a return of 51.84% and a risk of 41.02%, indicating strong risk-adjusted returns. In the case of high returns, there is also a high risk. The concentration of assets (39.31% of weight is assigned to Apple and 34.83% in Nvidia) may expose the portfolio to idiosyncratic risks, which can be a drawback.

This minimize risk portfolio's goal is to minimize risk, measured by the standard deviation of returns. The resulting portfolio has the lowest standard deviation of 29.86%, with a Sharpe ratio of 0.849, reflecting lower expected returns in exchange for reduced risk. The primary advantage of this portfolio is its risk minimization, which is consistent with the preference for risk-averse investors, but the low expected return might not meet the objectives of more aggressive investors.

Through the analysis of the actual stock market data, it can be found that the random forest algorithm has a better prediction effect in the stock market quantification. After proper parameter adjustment and model optimization, the algorithm can accurately predict the rise and fall trend of stock prices and can provide the probability of each stock's rise or fall, so as to provide more decision basis for investors.

But at the same time, the results of integrating them may require more computing resources, and appropriate parameter tuning is required to obtain the best performance. The output can be difficult to interpret, especially when multiple features and complex interactions are involved. These problems have been solved to some extent, and the application prospect of random forest model in the field of stock price prediction is very broad. Through the analysis of the actual stock market data, it can be found that the random forest algorithm has a good forecasting effect in the stock market quantification, but any forecasting model has some errors, investors need to combine other factors for comprehensive analysis, in order to develop a more scientific and reasonable investment strategy.

The mean-variance model offers several advantages. Rooted in modern portfolio theory, it is both intuitive and straightforward to implement. Additionally, it aids investors in making more informed decisions by optimizing the balance between risk and expected return. However, it needs a strong assumption that the model assumes that the market is efficient and that asset returns follow a normal distribution, which is usually not true in reality. It is also sensitive to extreme values. Extreme market conditions, such as financial crises, can cause models to produce inaccurate predictions.

4 Conclusion

This paper focuses on mean-variance model and random forest model to forecast five typical stocks. The advantages and disadvantages of the two important models are discussed by comparing and analyzing the experimental results. First of all, through data collection and processing, the data analysis is done, and then the selected stocks are allocated with different models, and experimental conclusions are drawn through comparative analysis. Portfolios focusing on maximizing the Sharpe ratio or leveraging momentum strategies demonstrated strong returns, albeit with varying levels of risk. Additionally, the application of machine learning, specifically the Random Forest method, shows promise in predicting stock returns and aiding in portfolio optimization. For further study, classical models of different aspects can be used. Forecast and optimize the investment portfolio of important stocks, including new energy, finance, automobiles and other fields. It is possible to compare not only the applicability of the model but also the differences between different types of stocks. Similar control variables can be used to better determine the similarities and differences between different models or different stocks.

References

1. Sun, X. Q., Shen, H. W., Cheng, X. Q.: Trading network predicts stock price. Scientific Reports 4(1), 3711 (2014).
2. Yu, Y., Si, X., Hu, C., Zhang, J.: A review of recurrent neural networks: LSTM cells and network architectures. Neural Computation 31(7), 1235-1270 (2019).
3. Staudemeyer, R. C., Morris, E. R.: Understanding LSTM: A tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586 (2019).
4. Sherstinsky, A.: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena 404, 132306 (2020).
5. Phuoc, T., Anh, P. T. K., Tam, P. H., Nguyen, C. V.: Applying machine learning algorithms to predict the stock price trend in the stock market: The case of Vietnam. Humanities and Social Sciences Communications 11(1), 1-18 (2024).
6. Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., Mosavi, A.: Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data: A comparative analysis. IEEE Access 8, 150199-150212 (2020).
7. Steinbach, M. C.: Markowitz revisited: Mean-variance models in financial portfolio analysis. SIAM Review 43(1), 31-85 (2001).
8. Chiu, C. H., Choi, T. M.: Supply chain risk analysis with mean-variance models: A technical review. Annals of Operations Research 240(2), 489-507 (2016).
9. Rigatti, S. J.: Random forest. Journal of Insurance Medicine 47(1), 31-39 (2017).

10. Biau, G.: Analysis of a random forests model. The Journal of Machine Learning Research 13(1), 1063-1095 (2012).