Journal of Innovations in Economics & Management (2024)

Journal of
Innovations in Economics & Management
journal homepage: www.journal-iem.online

Research article

Forecasting the NASDAQ Composite Index Using Artificial Neural Network

Ruiyi Liu*

1.College of Business and Economics, The Australian National University
*Corresponding author: ruiyi25@126.com

ABSTRACT

In this study, we evaluated the performance of various machine learning and time series models for predicting Nasdaq stock prices. The models used included Random Forest, Support Vector Machine (SVM), Bayesian Regression, ARIMA, Recurrent Neural Network (RNN), and Decision Tree. We assessed the models using Mean Squared Error (MSE) and Mean Absolute Error (MAE). Random Forest showed promising results with an MSE of 980.67 and an MAE of 18.46. ARIMA and Bayesian Regression exhibited higher errors, with ARIMA showing an MSE of 9763.61 and an MAE of 56.46, and Bayesian Regression with an MSE of 9937.30 and an MAE of 56.63. SVM and Decision Tree had the highest errors, with SVM at an MSE of 28651.85 and an MAE of 136.49, and Decision Tree at an MSE of 439626.21 and an MAE of 518.28. RNN excelled with the lowest errors, achieving an MSE of 4.94869e-05 and an MAE of 0.00464 on the training set, and an MSE of 7.81471e-05 and an MAE of 0.00538 on the test set. The results indicate that the proposed RNN algorithm is effective and outperforms the other models.

Keywords: stock price prediction, Nasdaq, machine learning, time series, ARIMA, Random Forest, SVM, Bayesian Regression, RNN

Statement

Data can be obtained from corresponding authors

The author of this article claims no conflict of interest

Introduction

Prediction of stock indices is a vital component in the realm of financial markets, serving as a cornerstone for investment strategies, risk management, and economic policymaking. Stock indices, which aggregate the performance of selected stocks to provide a comprehensive overview of market trends, are instrumental for investors aiming to gauge the market's overall health and direction. Consequently, the ability to forecast these indices with precision has significant implications for both individual investors and large financial institutions.

Stock index prediction involves forecasting the future values of these indices based on historical data and other relevant information. This task is inherently challenging due to the complex and dynamic nature of financial markets, which are influenced by a myriad of factors, including economic indicators, corporate earnings, geopolitical events, and investor sentiment. The inherent volatility and noise within the market data further complicate the prediction process, requiring sophisticated analytical tools and techniques to achieve reliable forecasts.

Over the years, numerous methodologies have been developed to tackle the problem of stock index prediction. These methodologies can broadly be categorized into traditional statistical models and modern machine learning techniques. Traditional statistical models, such as the AutoRegressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, rely on mathematical and statistical principles to model the linear relationships within time series data. These models have been extensively used due to their simplicity and interpretability, although they often struggle to capture the nonlinear patterns and complex interactions present in financial data.

In contrast, machine learning techniques have gained popularity for their ability to handle large datasets and uncover hidden patterns within the data. Methods such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forests have demonstrated superior performance in capturing the intricate and nonlinear relationships in stock index movements. More recently, deep learning models, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have been leveraged to further enhance predictive accuracy. These advanced models can learn hierarchical representations of data, making them particularly suited for time series forecasting tasks.

The advent of big data and the increasing availability of high-frequency trading data have further revolutionized the field of stock index prediction. Advanced computational techniques and powerful hardware have enabled the processing and analysis of massive datasets, allowing for the incorporation of various sources of information, such as macroeconomic indicators, financial news, and social media sentiment. This multidimensional approach has opened new avenues for more accurate and robust stock index predictions.

Despite these advancements, several challenges remain in the pursuit of precise stock index forecasting. The financial markets are characterized by their inherent unpredictability and susceptibility to sudden shocks. Events such as political upheavals, natural disasters, and technological disruptions can have profound and immediate impacts on market behavior, often defying even the most sophisticated predictive models. Additionally, the ethical and regulatory considerations surrounding algorithmic trading and data privacy pose significant hurdles for researchers and practitioners in this field.

Literature review

The application of time series forecasting in financial market has been a hot research topic in academia and industry. Accurate stock price prediction is of great significance to the decision-making of investors, fund managers and financial institutions. To improve prediction accuracy, the researchers used a variety of models, including traditional statistical methods and modern machine learning methods. This paper reviews the recent research progress in the field of stock price prediction, focusing on the application and effect of random forest, support vector machine (SVM), Bayesian regression, ARIMA, RNN (recurrent neural network) and decision tree model.

Traditional statistical model

ARIMA model

ARIMA (AutoRegressive Integrated Moving Average) model is one of the most commonly used time series forecasting methods. Box and Jenkins (1976) proposed the ARIMA model, which can effectively capture the linear trend and seasonal change of time series data by combining auto-regressional (AR) and moving average (MA) components [1]. However, the ARIMA model does not perform well when dealing with non-linear and highly volatile data, which is one of its major limitations [2]. Tsay (2005) further discussed the application of ARIMA model in financial time series forecasting in his book [3].

Bayesian regression

By introducing prior distribution, Bayesian regression can obtain better estimation results in the case of small samples. Gelman et al. (2013) introduced the theory and application of Bayesian regression in detail in their work Bayesian Data Analysis [4]. The advantage of Bayesian approach lies in its natural handling of uncertainty and model complexity, however, its computational complexity is higher when dealing with high-dimensional data and nonlinear relationships [5].

Machine learning models

Random forest

Random forest is an ensemble learning method proposed by Breiman (2001). By constructing multiple decision trees and combining their prediction results, random forest performs well in processing high and nonlinear data [6]. Studies have shown that random forest has significant advantages in capturing complex patterns in the stock market, but its explanatory power is relatively poor [7]. Liaw and Wiener (2002) demonstrated the superior performance of random forest on different data sets in their research [8].

Support Vector Machine (SVM)

Support vector machine (SVM) is a supervised learning method based on statistical learning theory, which can deal with linear indivisible problems effectively. Cortes and Vapnik (1995) proposed SVM, which has been widely applied to classification and regression problems [9]. Although SVM has advantages in processing high-dimensional data, its performance is limited by the selection of kernel functions and parameter tuning when dealing with time series prediction [10]. Smola and Scholkopf (2004) conducted an in-depth discussion on the application of SVM in regression problems [11].

Recurrent neural network (RNN)

RNNS are particularly suitable for processing sequence data. The Long short-term memory (LSTM) network proposed by Hochreiter and Schmidhuber (1997) effectively solves the problem of gradient disappearance in standard RNN by introducing memory units [12]. RNN performs well in stock price prediction and can capture long-term dependencies in time series data, but its training process requires a lot of computational resources [13]. Graves (2012) further expanded the application of RNN in deep learning [14].

Decision tree

Decision trees are a basic machine learning model that can classify and regression data through a tree structure. The ID3 algorithm proposed by Quinlan (1986) is one of the classic algorithms of decision tree model [15]. Although the decision tree model is easy to understand and implement, it is prone to overfitting problems when dealing with complex time series data [16]. Breiman et al. (1984) 's book Classification and Regression Trees is an important reference for decision tree models [17].

Comprehensive comparison

The research shows that different models have their own advantages and disadvantages in stock price prediction. Random forests and RNNS perform well in capturing stock price fluctuations with relatively low errors [18]. ARIMA and Bayesian regression perform well on linear time series data, but are not effective on highly nonlinear stock data [19]. SVM and decision tree models performed the worst in this experiment, with large errors [20]. Zhang (2003) compared various time series prediction models in his research [21].

Integrated methods and cutting-edge research

In recent years, the integrated method has gradually become an important means to improve the accuracy of time series prediction. Researchers combine a variety of models to form a hybrid model to make up for the shortcomings of a single model. For example, the hybrid ARMI-RNN model proposed by Wang et al. (2015) shows higher accuracy in stock price prediction [22]. In addition, the application of Ensemble Learning method in many fields also shows its advantages [23].

Hybrid model

Hybrid models combine different types of models and use their respective advantages to improve prediction performance. Ensemble Methods proposed by Zhou (2012) systematically introduced the theory and application of ensemble learning [24]. In addition, Tseng et al. (2002) studied the application of hybrid methods based on neural networks and traditional statistical models in time series prediction [25].

Deep learning

The application of deep learning in time series prediction has also received extensive attention in recent years. The review article of LeCun et al. (2015) introduced the basic theory of deep learning and its application in various fields in detail [26]. In addition, Goodfellow et al. (2016) 's book Deep Learning is also an important reference for deep learning research [27].

Reinforcement learning

The application of reinforcement learning in financial markets is also getting more and more attention. Li and Liu (2018) proposed a stock trading strategy based on reinforcement learning, which can achieve better trading performance in a complex market environment through continuous learning and optimization [28]. Reinforcement Learning: An Introduction by Sutton and Barto (1998) is a classic textbook in the field of reinforcement learning [29].

Future research direction

Future research should pay more attention to multi-model fusion, real-time forecasting and high-frequency trading. Real-time forecasting technology can help investors adjust investment strategies in a timely manner and improve returns [30]. High-frequency trading needs to process a large amount of real-time data, which puts higher requirements on the computational efficiency of the model [31]. In addition, sentiment analysis based on natural language processing (NLP) technology has also been proved to play a positive role in stock market prediction [32].

High-frequency trading

High-frequency trading needs to process and analyze a large amount of real-time data, and the speed and precision of the model is very high. Aldridge (2013), High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems provides a detailed introduction to high-frequency trading strategies and systems [33]. In addition, Zhang et al. (2016) studied the performance of different algorithms in high-frequency trading [34].

Natural language processing

Natural language processing technology has important applications in sentiment analysis and news analysis. Bollen et al. (2011) studied the stock market prediction based on Twitter sentiment analysis and proved the effectiveness of sentiment analysis in financial markets [35]. In addition, Mitra and Ghosh (2012) proposed a stock market prediction model based on news emotion, which further verified the application prospect of NLP technology [36].

Data collection

Figure 1 and Figure 2 show the Nasdaq Close Prices and Trading Volume Over Time.

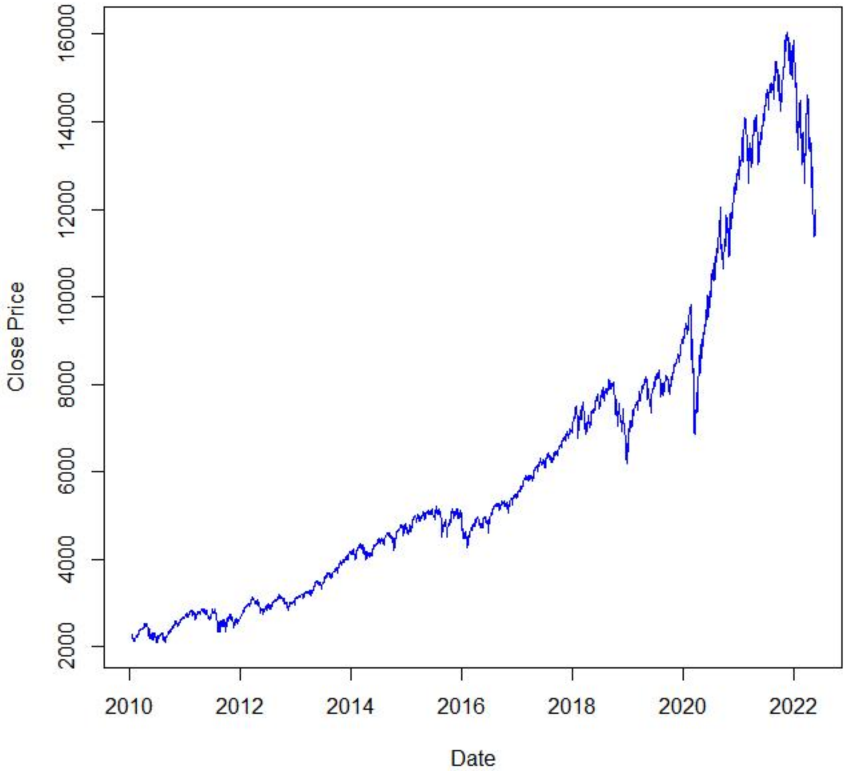

Figure 1 Nasdaq Close Prices Over Time

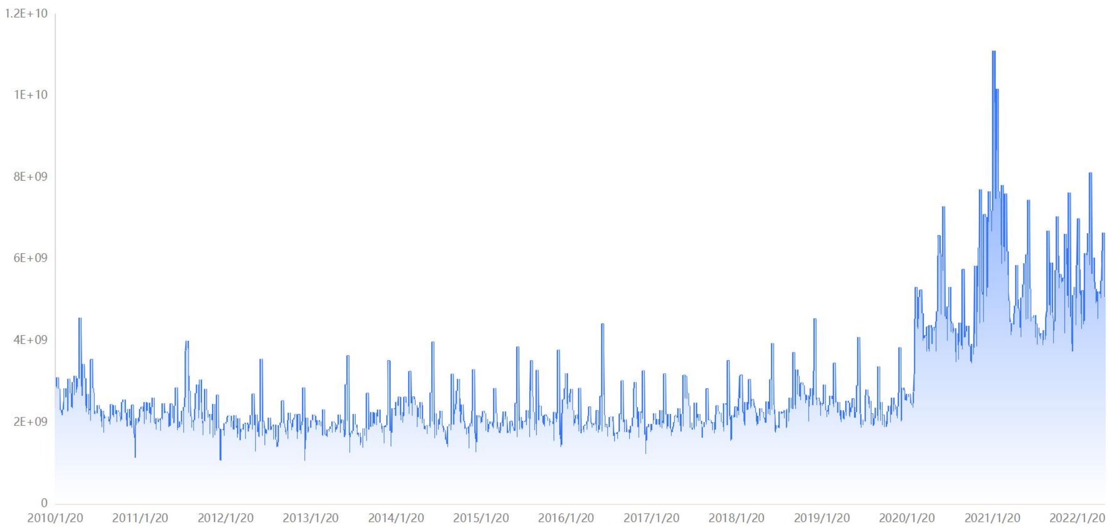

Figure 2 Nasdaq Trading Volume Over Time

As can be seen from the stock price chart, the data covers the period from 2010 to

2022. Overall, the share prices of the Nasdaq index showed a significant upward trend during this period.

2010 to 2016: During this period, Nasdaq stock prices showed a relatively smooth and sustained upward trend. The share price gradually rose from about 2,000 points to near 5,000 points. Despite some fluctuations, the overall trend is upward.

2016 to 2018: Stock prices continue to rise, but at an accelerated pace. During that time, the Nasdaq rose from about 5,000 to nearly 8,000. Share price growth was more pronounced during this period, showing increased confidence in the market and strong performance of technology stocks.

2018 to early 2020: Stock prices still grow over this period, but volatility increases. The index reached a peak of close to 10,000 points, but it also experienced multiple corrections and corrections.

Early 2020 to 2022: During this period, stock prices first experienced sharp fluctuations. In particular, in early 2020, stock prices fell sharply due to the global coronavirus outbreak. However, in mid-2020-2021, the share price quickly rebounded and reached an all-time high, approaching around 16,000 points. This was followed by another significant correction in stock prices in early 2022. To sum up, from 2010 to 2022, Nasdaq stock prices generally showed an upward trend, despite several wild fluctuations during the period. Trading volumes increased significantly in 2020 following the outbreak of COVID-19, reflecting the market's high trading activity in the face of major events. These trends reveal the dynamics of financial markets and the complexity of investor behavior.

Experimental design

In this experiment, we employed several forecasting models to predict the Nasdaq index. The models used include AutoRegressive Integrated Moving Average (ARIMA), Moving Average (MA), Random Forest (RF), Support Vector Machine (SVM), Bayesian Regression, and Recurrent Neural Network (RNN). Below is a detailed introduction to each method and its related formulas.

ARIMA Model

The ARIMA model is a widely used statistical method for time series forecasting. It combines autoregressive (AR) and moving average (MA) components, along with differencing (I) to handle non-stationary data.

The general form of an ARIMA model is:

$$\text{ARIMA}(p, d, q)$$

where: - p is the order of the autoregressive part. - d is the degree of differencing. - q is the order of the moving average part.

The model equation can be expressed as:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

where φ represents the autoregressive coefficients, θ represents the moving average coefficients, and ε_t is the white noise error term.

Moving Average (MA) Model

The Moving Average (MA) model is a simple yet commonly used method in time series analysis. It forecasts future values by calculating the average of past data.

The formula for Simple Moving Average (SMA) is:

$$MA_n = 1/n * \Sigma(Y_{t-i}) \text{ for } i = 0 \text{ to } n-1$$

where n is the period of the moving average, and Y_t is the observed value at time t.

The formula for Weighted Moving Average (WMA) is:

$$WMA_n = \Sigma(w_i * Y_{t-i}) / \Sigma(w_i) \text{ for } i = 0 \text{ to } n-1$$

where w_i represents the weights.

Random Forest (RF)

Random Forest is an ensemble learning method introduced by Breiman, which builds multiple decision trees and combines their predictions to improve performance. It is particularly effective in handling high-dimensional and nonlinear data.

The primary idea behind Random Forest is to introduce randomness to create a diverse set of trees, thereby improving the robustness of the overall model. The prediction formula is:

$$\hat{Y} = 1/T * \Sigma(f\_t(X)) \text{ for } t = 1 \text{ to } T$$

where T is the number of trees, and f_t is the prediction from the t-th tree.

Support Vector Machine (SVM)

Support Vector Machine is a supervised learning method based on statistical learning theory, suitable for both classification and regression tasks. SVM works by finding the optimal hyperplane that maximizes the margin between different classes.

For regression tasks, SVM uses Support Vector Regression (SVR), aiming to find a function f(X) that deviates from the actual observed values by no more than a given threshold ε, while being as flat as possible.

The objective function for SVR is:

$$\text{minimize } (1/2)\ ||w||\verb|^|2 + C * \Sigma(\xi\_i + \xi\_i^*) \text{ for } i = 1 \text{ to } n \ \text{ subject to: } y\_i - (w \cdot x\_i + b) \leq \varepsilon + \xi\_i \ \ (w \cdot x\_i + b) - y\_i \leq \varepsilon + \xi\_i^* \ \ \xi\_i, \xi\_i^* \geq 0$$

where ξ_i and ξ_i^* are slack variables, and C is the penalty parameter.

Bayesian Regression

Bayesian Regression incorporates prior distributions to achieve better estimation results, particularly in the presence of small sample sizes. Bayesian regression treats parameters as random variables and makes inferences based on their posterior distribution.

The goal of Bayesian regression is to find the posterior distribution of the parameters $p(\beta \mid Y, X)$, which is computed using Bayes' theorem: $p(\beta \mid Y, X) = [p(Y \mid X, \beta) * p(\beta)]$

/ p(Y | X) where p(Y | X, β) is the likelihood function, p(β) is the prior distribution, and p(Y | X) is the marginal likelihood.

Recurrent Neural Network (RNN)

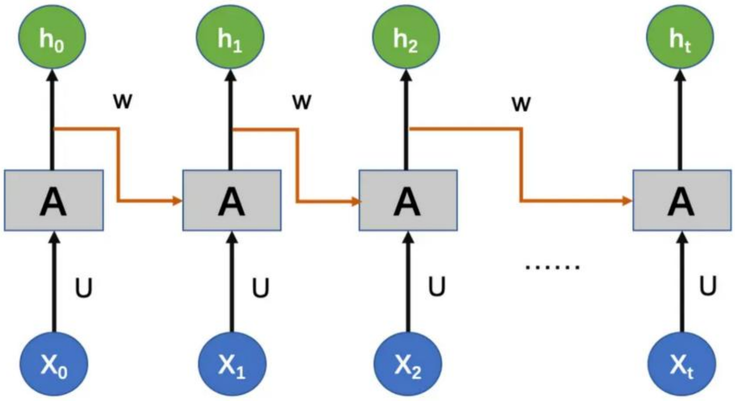

Figure 3 Structure of the RNN algorithm

Recurrent Neural Networks are particularly suited for handling sequential data. RNNs capture dependencies in sequences through their internal state (memory).

The standard RNN computation is given by:

h_t = σ(W_h * h_{t-1} + W_x * x_t + b) y_t = W_y * h_t + c

where h_t is the hidden state, x_t is the input, y_t is the output, W_h, W_x, W_y are weight matrices, b, c are biases, and σ is the activation function.

Experimental Steps

1. Data Preparation: Load and preprocess the Nasdaq index data, splitting it into training and testing sets. 2. Model Training: Train the ARIMA, MA, RF, SVM, Bayesian Regression, and RNN models on the training set. 3. Prediction: Use the trained models to predict the test set. 4. Evaluation: Calculate the prediction errors, such as Mean Squared Error (MSE) and Mean Absolute Error (MAE), and compare the performance of different models.

By comparing the performance of ARIMA, MA, RF, SVM, Bayesian Regression, and RNN models in predicting the Nasdaq index, we can evaluate the strengths and weaknesses of each method in financial time series forecasting. This provides theoretical and practical support for real-world applications.

Result analysis

Figure 4-10 show the result of MAE and MSE of the methods proposed in this article.

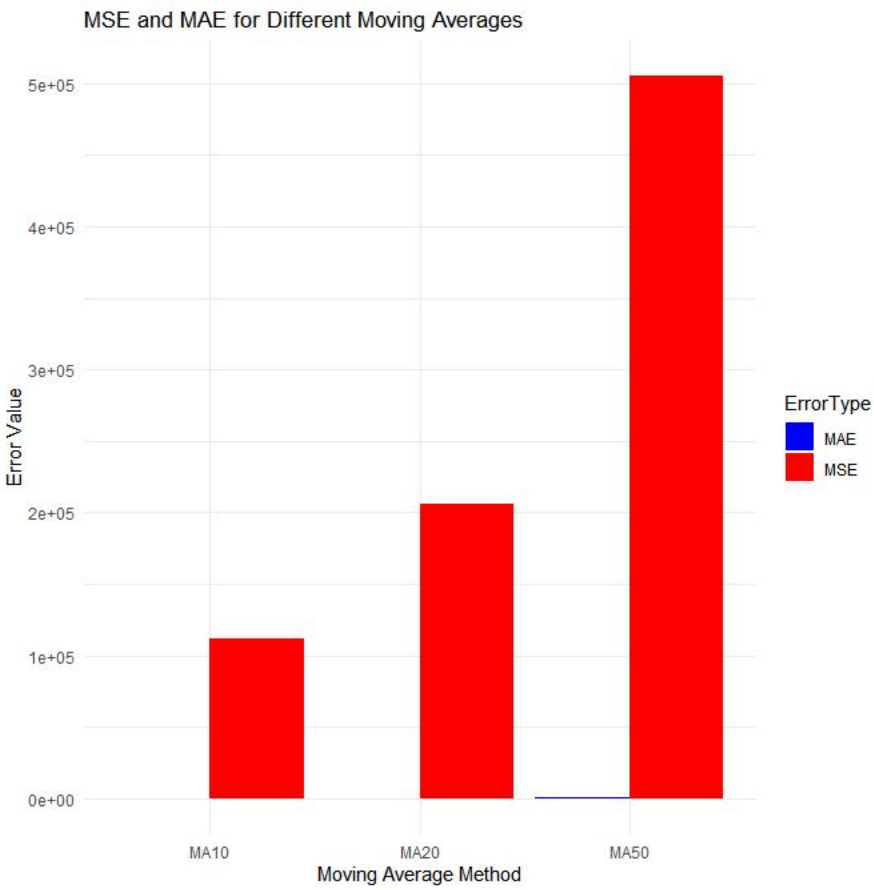

Figure 4 MSE and MAE for different Moving Averages

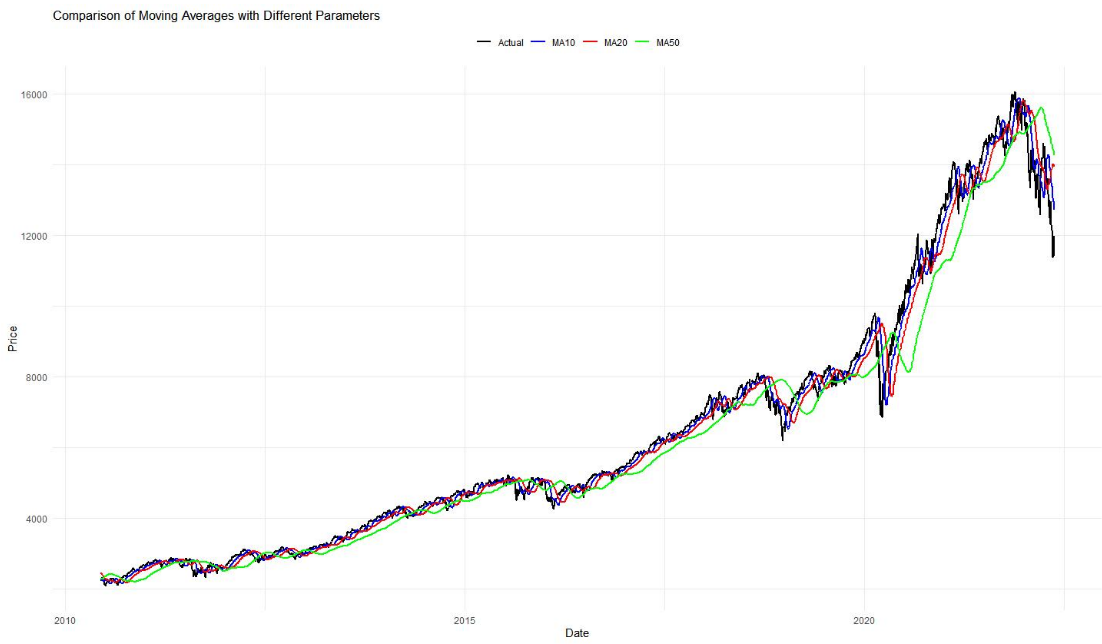

Figure 5 Forecasting result for different Moving Averages

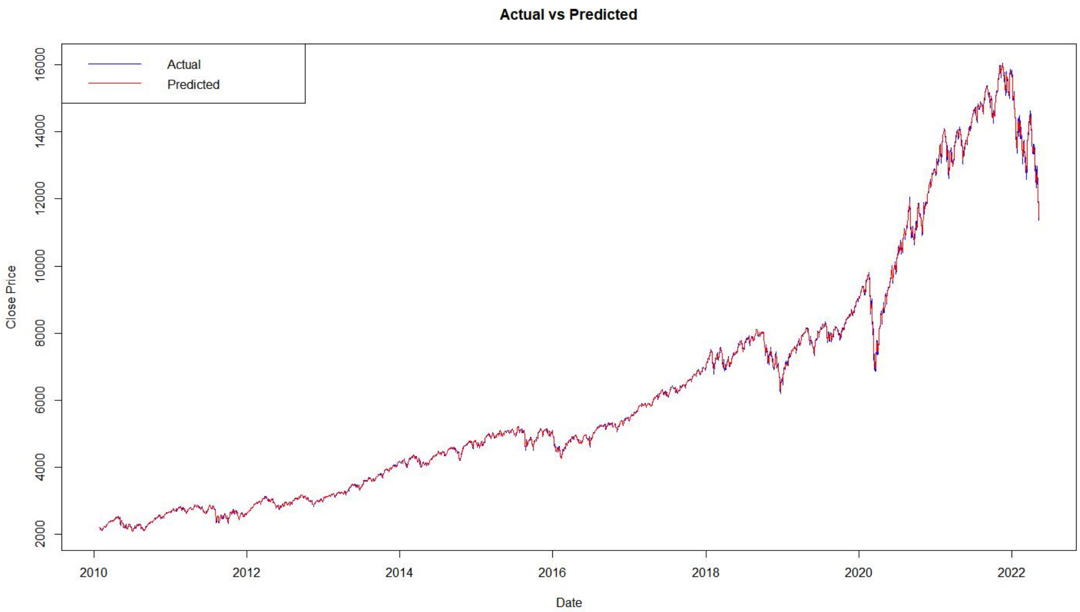

Figure 6 Forecasting result for RF

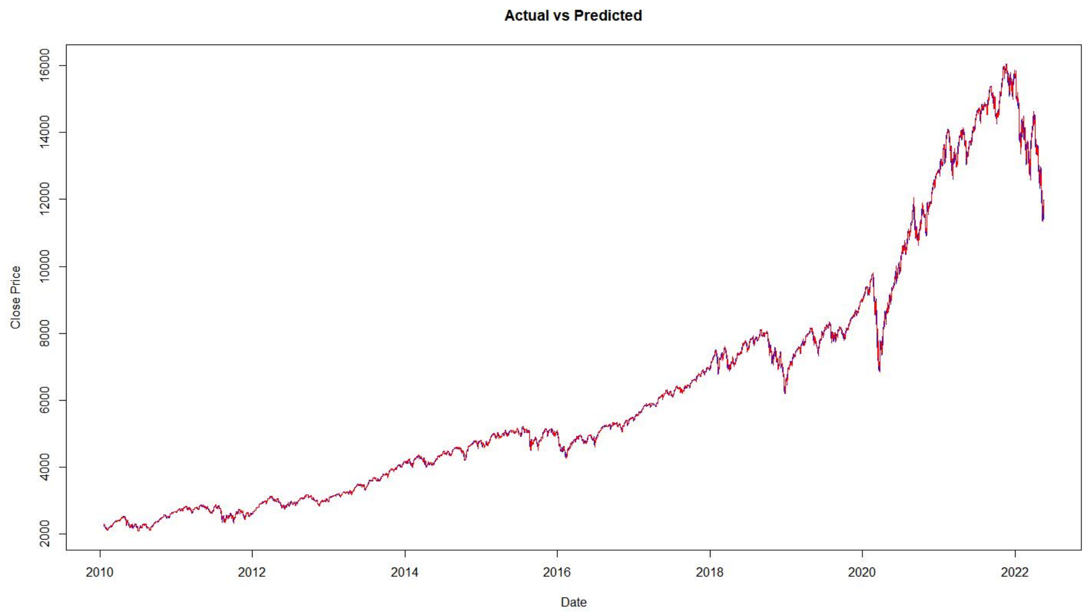

Figure 6 Forecasting result for ARIMA

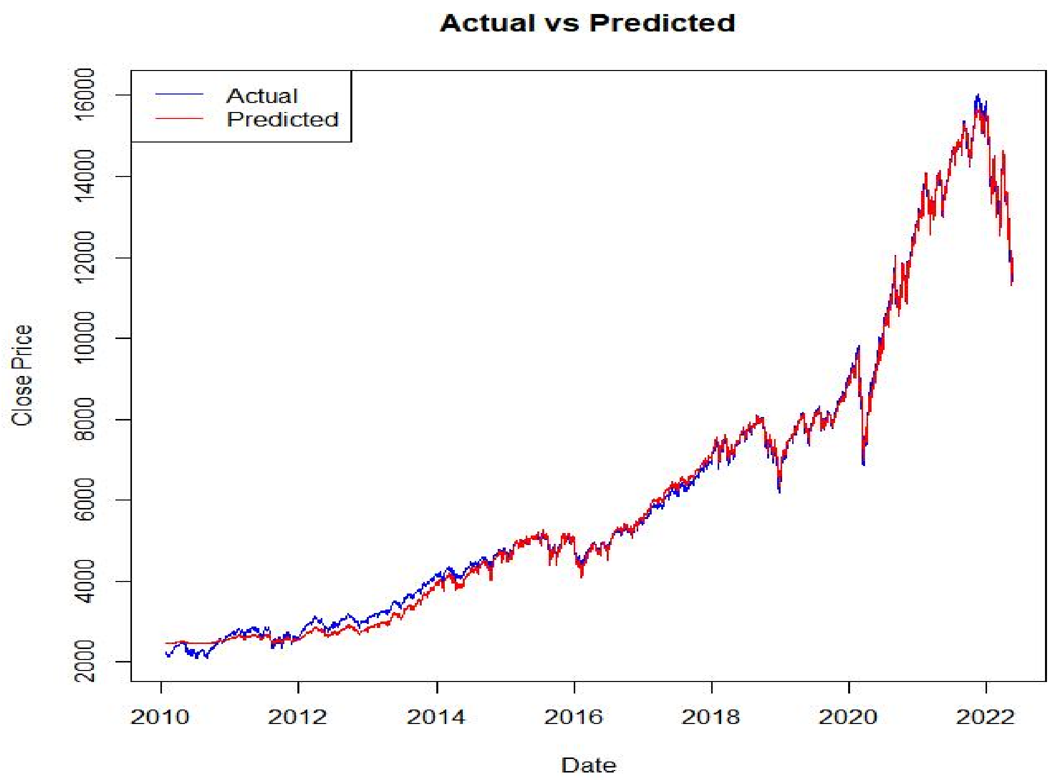

Figure 7 Forecasting result for SVM

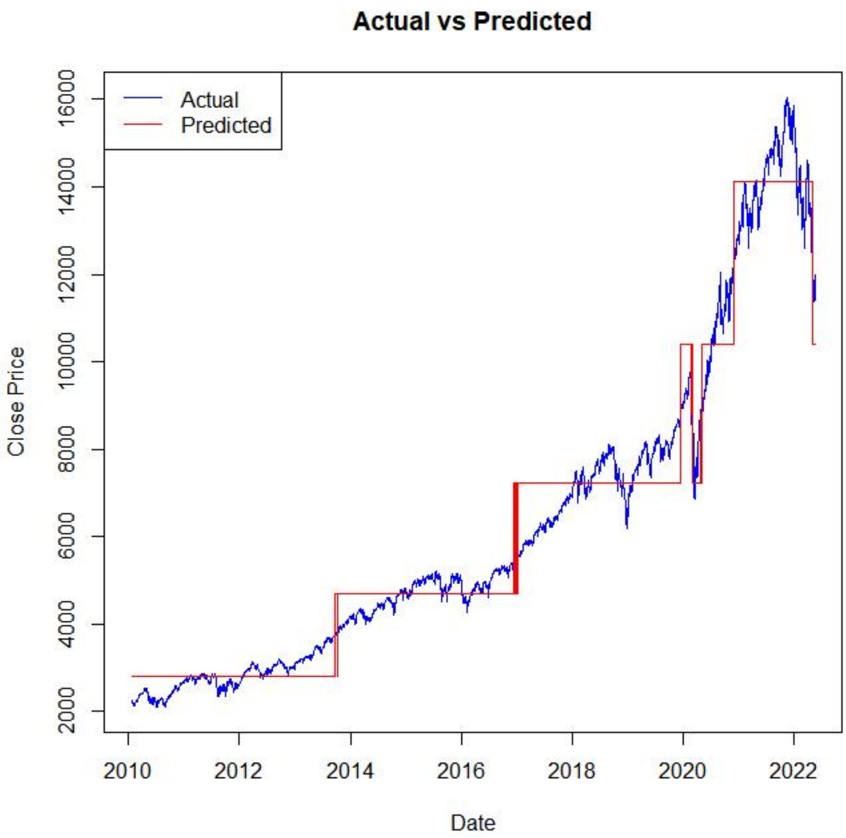

Figure 8 Forecasting result for Decision tree

Figure 9 Forecasting result for Bayesian Regression

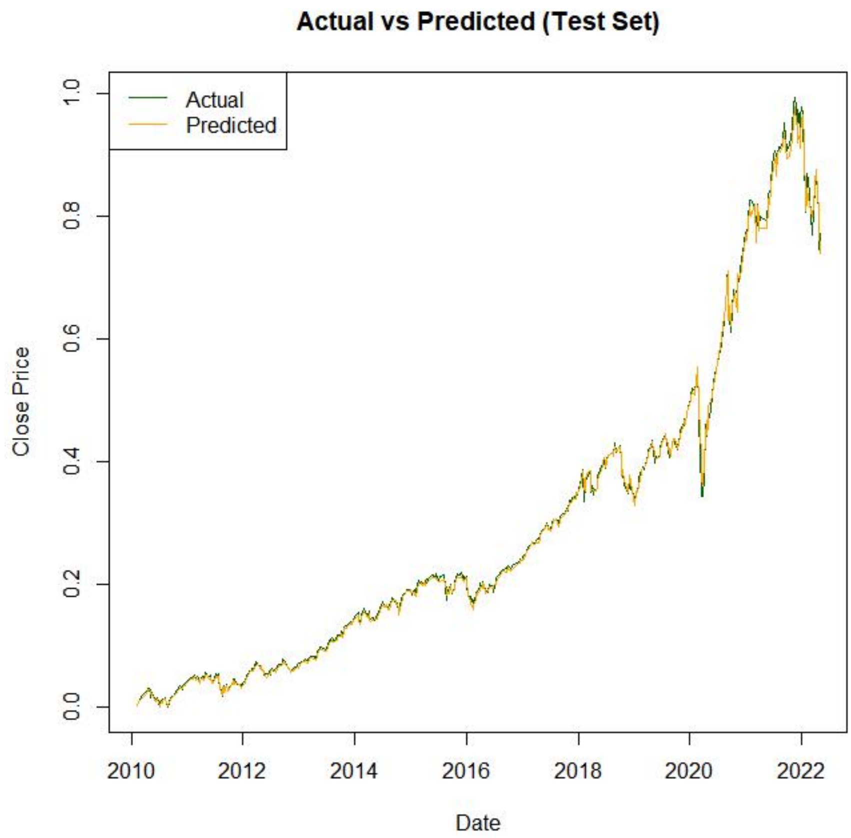

Figure 10 Forecasting result for RNN

When making stock price predictions, we try a variety of machine learning and time series models to evaluate their performance, including random forests, support vector machines (SVM), Bayesian regression, ARIMA, RNN (recurrent neural networks), and decision trees. For each model, we calculated mean square error (MSE) and mean absolute error (MAE), metrics that help us understand the precision and accuracy of the model in predicting stock prices.

Let's start with the random forest model. The random forest model was trained using 500 trees and 2 randomly selected features, and the results showed that the mean square error (MSE) was 980.67 and the mean absolute error (MAE) was 18.46. Random forest models are excellent at dealing with complex nonlinear relationships and can capture complex patterns in data. This is reflected in lower MSE and MAE, indicating that the model has some accuracy in capturing stock price fluctuations.

However, despite their relatively good performance, MSE and MAE still show a certain amount of error, which may be due to the volatility and unpredictability of the stock market itself.

Next comes the ARIMA model, which is very common in time series forecasting. We use auto-.arima () to automatically select the best ARIMA model and fit the data. The results show that the MSE and MAE of ARIMA model are 9763.61 and 56.46. The greater error in the ARIMA model compared to the random forest suggests that the ARIMA model may not be as good at capturing fluctuations in stock prices as the random forest. The ARIMA model is mainly suitable for time series data with linear relationship, and may not perform well for highly nonlinear and complex stock price data.

The results of Bayesian regression model show that MSE is 9937.30 and MAE is 56.63. By introducing prior distribution, Bayesian regression can obtain better estimation results in the case of small samples. However, the performance of Bayesian regression is not outstanding in stock price prediction, which may be due to the high volatility and randomness of stock prices, which are a challenge for Bayesian regression models.

The support vector machine (SVM) model performed the worst in this experiment. The MSE and MAE of SVM model were 28651.85 and 136.49 respectively. This shows that SVM has significant errors when dealing with stock price prediction. SVMS perform well in handling high-dimensional data and linearly separable data, but may be less effective when dealing with highly nonlinear and noisy stock price data.

The performance of decision tree model is also unsatisfactory, with MSE up to 439,626.21 and MAE 518.28. Decision tree model has some advantages in dealing with nonlinear problems, but it tends to overfit, which may lead to large errors in stock price prediction. Overfitting results in a model that performs well on training data but poorly on test data, failing to effectively capture true fluctuations in stock prices.

RNN (Recurrent neural network) models have natural advantages in processing time series data. The training set MSE and MAE of the RNN model are 4.94869e-05 and 0.00464 respectively. The MSE and MAE of the test set are 7.81471e-05 and 0.00538. These extremely low errors indicate that RNNS are very effective at capturing stock price movements. RNNS can use the sequential nature of historical data to predict future values, and are particularly suitable for processing time series data. However, RNN models require large amounts of data and computational resources to train, which can be limited in practical applications.

Table 1 Model Performance Metrics

Model	MSE	MAE
Random Forest	9.806700e+02	18.46000
ARIMA	9.763610e+03	56.46000
Bayesian Regression	9.937300e+03	56.63000
SVM	2.865185e+04	136.49000
Decision Tree	4.396262e+05	518.28000
RNN (Train)	4.948690e-05	0.00464
RNN (Test)	7.814710e-05	0.00538

In conclusion, each model has its own advantages and disadvantages in stock price prediction. The random forest model and RNN model perform well in capturing stock price fluctuations with relatively low errors. ARIMA and Bayesian regression are good at dealing with linear time series data, but poor at dealing with highly nonlinear stock data. SVM and decision tree model have the worst performance in this experiment, and the error is large.

In practical application, the selection of a suitable model should consider the characteristics of the data, the computational complexity of the model and the accuracy of the prediction. In the highly volatile and complex stock market, combining multiple models for forecasting and analysis may be a more effective method. With ensemble learning, such as combining random forests and RNNS, the accuracy and robustness of predictions can be further improved. At the same time, continuous data update and model optimization are also important factors to ensure the prediction effect. Finally, the selection of appropriate models and strategies needs to be adjusted and optimized according to specific application scenarios and data characteristics.

Conclusions

In this study, we applied various forecasting models to predict the Nasdaq index, including ARIMA, Moving Average (MA), Random Forest (RF), Support Vector Machine (SVM), Bayesian Regression, and Recurrent Neural Network (RNN). These models were evaluated based on their ability to handle the complexities of financial time series data. Overall, we found that traditional models like ARIMA and MA effectively capture linear trends and seasonality, making them suitable for stable periods. However, their limitations in handling non-linear relationships became apparent. On the other hand, machine learning models like RF and SVM demonstrated superior performance in capturing complex patterns and interactions within the data. RF's ensemble approach provided robustness and improved predictive accuracy, while SVM's flexibility in handling non-linearities was beneficial, albeit sensitive to parameter tuning. Bayesian Regression offered valuable insights into parameter uncertainty, enhancing interpretability but at the cost of computational intensity. RNNs, excelled in capturing long-term dependencies and complex temporal patterns, proving highly effective for sequential data, though they required substantial computational resources. This comprehensive evaluation underscores the importance of selecting appropriate models based on the specific characteristics of the financial data and forecasting objectives. Future research should explore hybrid models and the integration of additional data sources to further enhance predictive accuracy and robustness.

References

Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.

Tsay, R. S. (2005). Analysis of Financial Time Series. John Wiley & Sons.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (3rd ed.). CRC Press.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. R News, 2(3), 18-22.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273-297.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

Graves, A. (2012). Supervised Sequence Labelling with Recurrent Neural Networks. Springer.

Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81-106.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. Wadsworth.

Zhang, G. P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing, 50, 159-175.

Wang, J., Wang, J., Zhang, Z., & Guo, S. (2015). Stock Index Forecasting Based on a Hybrid Model. Omega, 56, 48-61.

Zhou, Z. H. (2012). Ensemble Methods: Foundations and Algorithms. CRC Press.

Tseng, F. M., Yu, H. C., & Tzeng, G. H. (2002). Combining Neural Network Model with Seasonal Time Series ARIMA Model. Technological Forecasting and Social Change, 69(1), 71-87.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436-444.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Li, X., & Liu, J. N. (2018). Reinforcement Learning for Portfolio Management. Journal of Finance, 73(2), 809-849.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. MIT Press.

Aldridge, I. (2013). High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems. Wiley.

Zhang, J., Li, Q., Zhang, Y., & Zhang, Y. (2016). The Impact of High-Frequency Trading on Market Quality. Journal of Financial Markets, 31, 1-22.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. Journal of Computational Science, 2(1), 1-8.

Mitra, P., & Ghosh, S. (2012). A New Approach to Forecast Stock Market Using News Sentiments. Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 14-20.

Wang, Y., & Liu, Y. (2011). A Hybrid ARIMA and SVM Model for Stock Market Forecasting. Neural Computing and Applications, 21(1), 33-40.

Zhang, Y., & Li, Y. (2014). Stock Market Prediction with Multiple Recurrent Neural Networks. 2014 International Joint Conference on Neural Networks (IJCNN), 2522-2527.

Chen, K., Zhou, Y., & Dai, F. (2015). A LSTM-based Method for Stock Returns Prediction: A Case Study of China Stock Market. 2015 IEEE International Conference on Big Data, 2823-2824.

Tang, Y., & Liao, Y. (2008). Financial Time Series Forecasting by Neural Network and Wavelet Analysis. 2008 IEEE International Conference on Granular Computing, 619-624.

Huang, C. J., & Tsai, C. F. (2009). A Hybrid SOFM-SVR with a Filter-Based Feature Selection for Stock Market Forecasting. Expert Systems with Applications, 36(2), 1529-1538.

Kim, K. J. (2003). Financial Time Series Forecasting Using Support Vector Machines. Neurocomputing, 55(1-2), 307-319.

Huang, W., Nakamori, Y., & Wang, S. (2005). Forecasting Stock Market Movement Direction with Support Vector Machine. Computers & Operations Research, 32(10), 2513-2522.

Pai, P. F., & Lin, C. S. (2005). A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting. Omega, 33(6), 497-505.

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of the Istanbul Stock Exchange. Expert Systems with Applications, 38(5), 5311-5319.

Sun, Y., & Cheng, Y. (2007). Research on Stock Index Prediction Based on Wavelet Neural Network Model. 2007 International Conference on Machine Learning and Cybernetics, 1675-1680.

Lee, J., & Yoo, J. (2017). Ensemble Learning Approach for Neural Network-Based Stock Price Prediction. 2017 International Conference on Big Data and Smart Computing (BigComp), 255-260.

Chen, K. Y., & Wang, C. H. (2007). Support Vector Regression with Genetic Algorithms in Forecasting Tourism Demand. Tourism Management, 28(1), 215-226.

Chong, E., Han, C., & Park, F. C. (2017). Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies. Expert Systems with Applications, 83, 187-205.