

Hybrid Deep Learning for AI-Based Financial Time Series Prediction

ZHONG, Yuqiang^{1*} CHENG, Qishuo² QIN, Lichen³ XU, Jinxin⁴ WANG, Han⁵

¹ Henan Agricultural University, China

² University of Chicago, USA

³ University of Rochester, USA

⁴ Southern Methodist University, USA

⁵ University of Southern California, USA

* ZHONG, Yuqiang is the corresponding author, E-mail: markovmj7@gmail.com

Abstract: As the influencing factors of the development of things are not clear or the data collection is difficult, many forecasting problems have evolved into univariate time series forecasting problems, that is, mining its own inherent laws from the past time series data, so as to predict its future development trend. The purpose of this study is to explore the feasibility of using random forest classifier to predict the long-term trend of stocks. By analyzing data sets such as Apple, Samsung, and General Electric, we built a random forest model and found that its prediction accuracy was 85 to 95 percent. With the increase of the number of decision trees, the prediction results of the model tend to be stable, indicating that increasing the number of decision trees can improve the prediction accuracy. In addition, we point out that the method is also suitable for short-term trend forecasting, and suggest training with more fine-grained transaction data to improve forecasting accuracy. Finally, we look forward to the potential of the field of artificial intelligence in time series forecasting, emphasizing that the application of technologies such as deep learning will further improve forecasting accuracy and provide more reliable decision support for financial investors.

Keywords: Financial time series forecasting, Random forest model, Machine learning, Deep learning

DOI: <https://doi.org/10.5281/zenodo.10938270>

1 INTRODUCTION

Owing to its substantial return on investment, the financial market garners significant attention from investors and scholars alike, who seek to forecast the price trends of financial products. However, the financial market is affected by many complex factors such as market sentiment and policy, which make its trend contain a lot of noise. The earliest scholars used traditional statistical methods such as ARIMA and other models to study the influence of various factors on the price trend of financial products. However, financial time series are mostly non-linear and unstable, and the prediction performance of traditional models is not satisfactory. Therefore, constructing a stable and effective financial time series forecasting model is a challenging and practical task.

Currently, financial time series forecasting methods can be divided into measurement forecasting methods and machine learning methods. [1] On one hand, econometric forecasting methods include differential integrated moving average autoregressive model (ARIMA), dynamic model average (DMA), generalized autoregressive conditional heteroscedasticity model (GARCH), etc. Zhu, Mengran, et al. (2024) explore innovations in credit default prediction

through the utilization of ensemble methodology. Their study investigates the efficacy of ensemble techniques, specifically employing [2] LightGBM, XGBoost, and LocalEnsemble algorithms. By integrating these advanced machine learning models, the authors aim to enhance the accuracy and reliability of credit default prediction systems. Through their research, Zhu et al. contribute to the ongoing development of predictive analytics in financial risk management, offering valuable insights into the application of ensemble methodologies for credit risk assessment. However, econometric models have some conditional restrictions on time series, requiring the stationarity of time series, and their processing effect is poor for nonlinear and non-stationary data. On the other hand, common machine learning methods include support vector machine (SVM), BP neural network (BPNN) [3], recurrent neural network (RNN) [4], etc. Because these models do not need to impose specific conditions when processing complex nonlinear and non-stationary data, they offer more advantages and have been widely used. Although machine learning methods do not necessarily improve the prediction accuracy of complex dynamic systems, they can often extract detailed data information from nonlinear time series data related to targeted financial markets, thus improving prediction accuracy.

2 RELATED WORK

Time Series is a set of data sorted by time, which is equally spaced in time, such as every minute, every hour, every month or every quarter. Starting from this concept, we will find that the price data, trading volume data and yield data of all financial products are typical time series data, whether it is stocks, bonds, gold or bitcoin.

2.1 Time series forecasting

Time series prediction is the result of observing the underlying process of time series data. Since time series data is a variable with random process, the mining of time series data by time series prediction is more adequate than the regression analysis of random sampling of cross-section data. The main feature of time series data is time sequence, which is the natural carrier of various dynamic information. Li, Zhenglin et al. (2024) propose a novel approach to change detection utilizing Feature Manipulation for DDPM (Density-Dependent Probabilistic Model). Their work, outlined in the preprint "Feature Manipulation for DDPM based Change Detection," introduces a methodological framework that leverages the power of DDPMs to effectively detect changes in various datasets. By integrating feature manipulation techniques, the authors aim to enhance the sensitivity and accuracy of change detection algorithms. [5] This research contributes to advancing the field by introducing innovative methodologies that hold promise for improving the detection of changes in dynamic environments. It can reflect the trend, periodicity and hysteresis effect. From the perspective of probability statistics, a time series is a sample realization of a set of random variables at a series of moments.

According to whether the statistical characteristics of the series change with time, it can be divided into non-stationary series and stationary series. If the sequence is stationary, it means that the correlation between the two moments of the sequence remains stable. [6-7] As a result, the future can be well predicted based on statistical patterns presented by historical data. On the contrary, if the sequence is non-stationary, it means that the correlation between the two moments of the series is unstable, that is, the mining of the influencing factors of the data is incorrect or insufficient, and it is difficult to predict the future data.

One commonly used model for time series analysis is Autoregressive Process, which refers to the regression of variables to themselves. In a time series, this means that the current value is linearly dependent on its past value, which is similar to the logic of technical analysis. The p-order autoregressive process is denoted as AR(p), and AR(p) represents the current value as a linear combination of the drift term constant C, the current error term (ε_t) and the first p values of the sequence (y_{t-p}). The formula is expressed as follows:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

For the first order autoregressive process, that is, AR (1), the formula is:

$$y_t = C + \phi_1 y_{t-1} + \varepsilon_t \quad (2)$$

Where ϕ is the Root of the time series, its value is related to an important concept in the time series - Stationarity.

The properties of a stationary time series do not change with the passage of time, that is, its mean, variance, autocorrelation and other statistical values are not a function of time.

Stationarity is the fundamental assumption of many time series models, including AR and the more complex ARMA, ARIMA, SARIMA, etc. These models can only be used if the time series data passes the stationarity test, otherwise the predictions of the models will not be reliable.

However, time series forecasting has changed dramatically in the last two years, especially since the advent of MAE Aiming, which now allows unsupervised pre-training of models of time series using methods similar to MAE.

The Makridakis M-Competitions series (known as the M4 and M5 respectively) took place in 2018 and 2020 respectively [8] (the M6 also took place this year). The recent Ventilator Pressure Prediction competition demonstrated the importance of using deep learning methods to tackle real-time time series challenges. The goal of the competition is to predict the chronological order of pressure in the mechanical lung. Each training instance is its own time series, so the task is a multi-time series problem. The winning team submitted a multi-layer deep architecture that included an LSTM network and a Transformer block.

2.2 Financial time series

Due to the characteristics of noise and volatility in finance, the prediction of financial time series is generally considered to be one of the most challenging problems in time series prediction (Wang, 2012). For the economic and social organization of modern society, how to accurately predict stock movements is still an open question. In the past few decades, machine learning models such as artificial neural networks (ANNs) and support vector regression (SVR) (Cherkassky, 1997) have been widely used to predict financial time series and have achieved high prediction accuracy, (Huang et al. (2005)). However, in the literature, recent trends in the machine learning and pattern recognition communities argue that deep nonlinear topology should be applied to time series prediction. As an improvement over traditional machine learning models, new deep learning models can successfully model complex real-world data by extracting robust features of relevant information and achieve better performance than before. Considering the complexity of financial time series, combining deep learning with financial market forecasting is considered to be one of the most fascinating topics. However, this area remains

relatively unexplored.

Some current work on the application of deep learning to the financial field has introduced the first two methods into research. For example, Ding et al. (2015) combined neural tensor network and deep learning stochastic forest time series prediction to predict the short - and long-term impact of events on stock price trends. In addition, some studies have used random forest time forecasting in financial market forecasting. [9] However, there has been little in-depth research on whether the stackable autoencoder approach can be applied to financial market forecasting. RF Random forest algorithm is an ensemble learning algorithm that improves prediction accuracy by combining multiple decision trees. In RF random forest algorithms, the training data for each decision tree is randomly extracted from a portion of the original data, which reduces the risk of overfitting. The final prediction is the average of the predictions of all decision trees.

When making stock price forecast, we need to collect relevant data, including historical stock prices, company financial statements, macroeconomic indicators and so on. Therefore, this paper analyzes the time series prediction of financial markets through the application status of the stochastic forest time series prediction model based on deep learning in financial markets and case experiments.

2.3 Random forest model

Random forest is a powerful ensemble learning method for classification and regression tasks. [10] They belong to a family of decision tree-based algorithms known for their versatility, robustness, and ability to handle complex data. Random forests combine predictions from multiple independent decision trees to make more accurate and robust predictions. The idea behind ensemble learning is that by combining the results of multiple models, you can often achieve better overall performance than a single model.

The heart of a random forest is a decision tree. A decision tree is a supervised machine learning model that makes decisions by subsetting input data based on specific eigenvalues and creating a tree-like structure of decision nodes. Random Forest employs a technique called Bootstrap aggregating, or Bagging for short. Bagging involves creating multiple random subsets (replacing) the original data set.

Each subset is used to train a separate decision tree. [11] This process reduces variance and helps prevent overfitting because each tree sees a slightly different version of the data. In addition to using self-service samples, random forests also introduce randomness by selecting only a random subset of features on each node of the decision tree. This helps with de-correlation and further improves the robustness of the model.

Once all the individual decision trees have been trained, they are then combined to make predictions. For the classification task, each tree "votes" for one class, and the class with the most votes becomes the final prediction.

The following is the schematic diagram of the random forest principle:

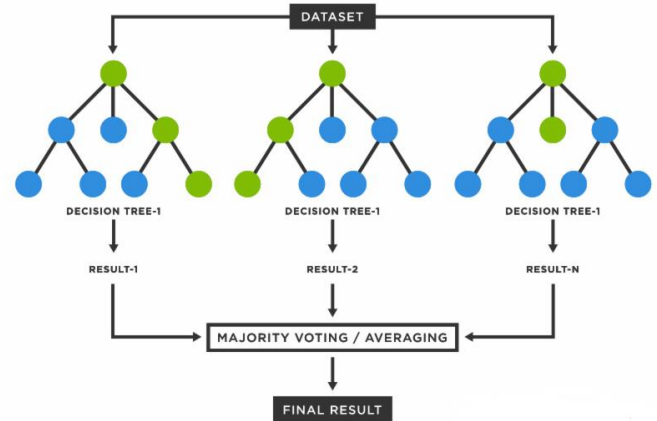


Figure 1. Schematic diagram of random forest principle

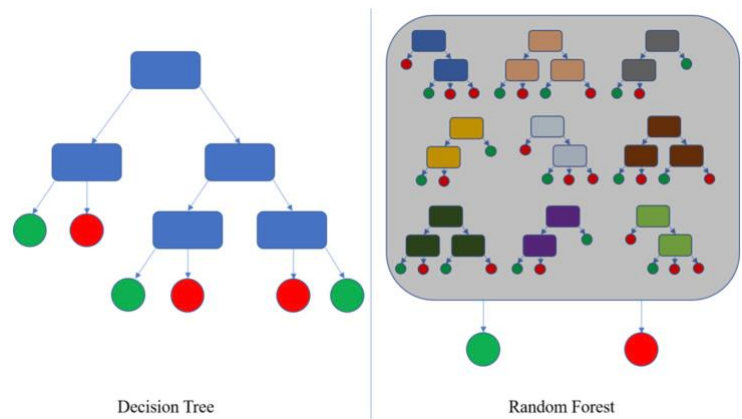


Figure 2. Random forest Decision tree architecture diagram

According to the random forest decision tree in FIG. 2, it can be seen that decision tree is a supervised learning method (both eigenvalues and label data are given during training, which is commonly understood as a given set of data with both x and y). It is a tree-shaped structure in terms of morphology, with each internal node representing a test on an attribute and each branch representing a test output. Each leaf node represents a category [12].

Random forest models have gained significant popularity in the realm of financial markets due to their robustness and effectiveness in handling complex data structures like time series data, which are prevalent in financial markets. These models offer several advantages when it comes to predicting stock prices or market movements.

Handling Non-linearity: Financial data often exhibit non-linear relationships between various factors influencing market movements. Random forest models are capable of capturing these non-linearities efficiently, making them suitable for predicting stock prices or market trends.

Feature Importance: Random forest models provide a mechanism to measure the importance of different features in predicting stock prices [13]. This feature importance analysis can offer valuable insights into which factors are most influential in driving market movements, aiding investors and analysts in making informed decisions.

Robustness to Overfitting: Overfitting is a common challenge in building predictive models for financial markets due to the noisy and high-dimensional nature of financial data. Random forest models mitigate the risk of overfitting by aggregating predictions from multiple decision trees trained on different subsets of the data, thus improving generalization performance.

Ensemble Learning: By combining predictions from multiple decision trees, random forest models harness the power of ensemble learning, which often leads to superior predictive performance compared to individual models. This ensemble approach helps in reducing variance and improving the overall stability of predictions [14].

Adaptability to Time Series Data: Financial data, such as stock prices, exhibit temporal dependencies, making them inherently sequential in nature. Random forest models can be adapted to handle time series data by incorporating lagged variables or by using rolling window approaches, enabling them to capture temporal patterns and dynamics in the data.

2.4 Importance of Random Forests in Financial Time Series

In financial time series analysis, random forest models serve a pivotal role due to their capability to enhance prediction accuracy. These models excel in capturing intricate patterns and relationships inherent in financial time series data. By leveraging historical data, random forest algorithms can effectively forecast stock prices, market indices, and other [15] financial metrics with greater precision compared to traditional methods. This enhanced accuracy empowers investors and financial institutions to make more informed decisions regarding their investment strategies, thus improving overall performance and profitability.

Furthermore, random forest models significantly contribute to risk management within the realm of finance. By providing accurate forecasts of future market movements, these models enable investors to anticipate potential risks and adjust their portfolios accordingly. This proactive approach to risk management allows investors to mitigate potential losses and optimize their investment strategies to align with their risk tolerance and financial goals. Ultimately, the use of random forest models enhances risk-adjusted returns and fosters greater stability within the financial markets.

Additionally, random forest models play a vital role in algorithmic trading systems, facilitating automated decision-

making processes based on predefined rules or strategies. [16] These models can efficiently analyze large volumes of historical data and swiftly make predictions, enabling algorithmic traders to execute trades more effectively and profitably. By leveraging the predictive power of random forest algorithms, algorithmic traders can capitalize on market opportunities and adapt to changing market conditions in real-time, thus gaining a competitive edge in the financial markets.

3 METHODOLOGY

Machine learning has become an important tool in asset market analysis. However, among the many machine learning algorithms, which algorithm can achieve better prediction results is still a key question. In a study published in the journal *Applied Mathematical Finance*, researchers used a random forest algorithm to predict where stock prices would rise or fall d days in the future.

Compared with support vector machine (SVM), linear discriminant analysis (LDA) [17] and other models, it is found that random forest can achieve better prediction results. Specifically, the random forest model was able to predict the direction of stock prices with 85 to 95 percent accuracy. This finding provides strong support for asset market analysis, indicating that random forest algorithm has significant advantages in the field of stock price prediction. Therefore, based on the research published in the journal *Applied Mathematical Finance*, this experiment uses the random forest algorithm to predict the direction of the rise and fall of stock prices in the next d days. The results show that compared with support vector machine [18-19] (SVM), linear discriminant analysis and other models, random forest can achieve better prediction results, with an accuracy of 85 to 95 percent.

3.1 Experimental model

In order to minimize the prediction error, the paper treats the prediction of the stock price trend as a binary classification problem (up or down), which is solved by integrated machine learning modeling. In this paper, RSI (Relative Strength index), KD random index, MACD and other 6 commonly used technical indicators are used as classification characteristics to train the random forest model [19-20]. Finally, it is found that with the increase of the number of decision trees, the accuracy of the model increases and tends to converge. Moreover, the longer the forecast time window, the more accurate the model.

Random forest is an algorithm that integrates multiple trees through the idea of Ensemble Learning. Its basic unit is decision tree, and its essence belongs to ensemble learning, a branch of machine learning. There are two key words in the name of random forest, one is "random" and the other is "forest". "Forest" we understand very well, one is called a tree, then hundreds of trees can be called a forest, this analogy is still very appropriate, in fact, this is the main idea

of random forest - the embodiment of the idea of integration. The meaning of "random" will be discussed in the following section.

From an intuitive point of view, every decision tree is a classifier (assuming we're dealing with a classification problem). We want to classify an input sample, and we need to classify the input sample into each tree. So for an input sample, N trees will have N classification results. To draw a vivid analogy, there is a meeting in the forest to discuss whether an animal is a mouse or a squirrel, and each tree has to express its own opinion on the question independently, that is, each tree has to vote. Whether the animal is a mouse or a squirrel depends on the number of votes cast, and the category that gets the most votes is the forest category. Each tree in the forest is independent, and 99.9 percent of unrelated trees make predictions that will cancel each other out. The predictions of a few good trees will rise above the noise and make a good prediction. Random forest integrates all the classification voting results, designating the category with the most votes as the final output, which is the Bagging idea of random forest.

3.2 Experimental Design

First, for each decision tree, the training set is randomly selected from the whole sample set. In this paper, the standard of decision tree classification is the technical index in the feature matrix X, and the technical index classification is used until the Gini impurity is small enough to meet the requirements.

These decision trees predict independently, and then vote on the outcome of each decision tree prediction, with the most votes becoming the prediction of the random forest. [21] This avoids overfitting of a single decision tree.

Due to random sampling, each decision tree does not use the full Sample (only about 2323 samples are drawn), and the samples that are not drawn are the non-sample set of this decision tree (Out of Bag Sample). For all out-of-pocket samples generated by decision trees, for each sample, the classification of the tree it is used as an oob sample (about 1313 trees) is calculated, and then a simple majority vote is used as the classification result of the sample, and finally the ratio of the number of mismarks to the total number of samples is used as the OOB mismarks rate of the random forest. Therefore, the smaller the OOB deviation, the lower the proportion of misclassification and the more accurate the random forest classification.

3.3 Data Preprocessing

The original experiment collected the stock price trend data of Apple, Samsung and GE for a total of about 7,000 trading days, and predicted the stock trend after 30 days, 60 days and 90 days, respectively.

In order to remove the noise in the historical data and show the actual law of the historical data, the author adopts the exponential smoothing method to preprocess the stock

price data:

$$S_0 = Y_0 \text{ for } t > 0 \quad (3)$$

$$S_t = \alpha * Y_t + 1 - (\alpha) * S_{t-1} > 0 \quad (4)$$

$$1 S_0 = Y_0 \text{ for } S_t = \alpha * Y_t + 1 - (\alpha) * S_{t-1} \quad (5)$$

Alpha is between 0 and 1 and is closer to 1, giving more weight to recent data. Because recent movements are somewhat more consistent, greater weight is placed on the most recent data.

$$\text{target}_i = \text{Sign}(\text{close}_{i+d} - \text{close}_i) \quad \text{target}_i = \text{Sign}(\text{close}_{i+d} - \text{close}_i) \quad (6)$$

d is the predicted time window and Sign is the symbolic function. When the value of target_i is 1, it means that at this moment i, the closing price after d is higher than the closing price today, that is, the stock will rise after d; Vice versa. target_i is also the target that the model needs to predict.

3.4 Build Random forest model

In this experiment, the stock price of Apple was modeled by random forest, and the stock price trend was predicted 30, 60 and 90 days later. The author finds that the accuracy of the model increases and tends to converge with the increase of the number of decision trees in the model. Moreover, the longer the forecast time window, the more accurate the model.

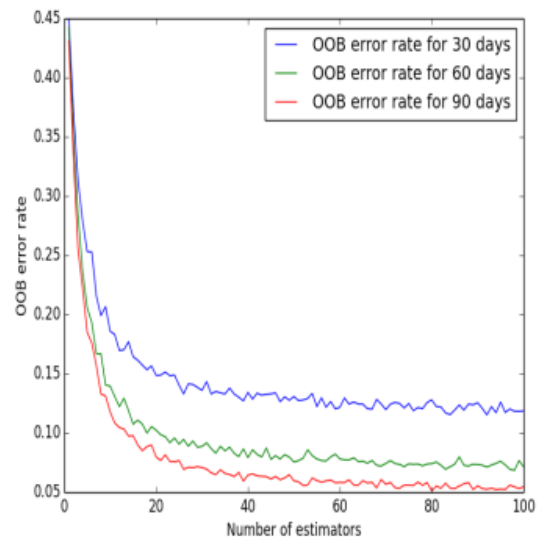


Figure 3. ROC curve of random forest model results

3.5 OOB error Specific result

Table 1. Error test results table

Trading Period (Days)	No. of Trees	Sample Size	OOB error
30	5	6590	0.241729893778
30	25	6590	0.149165402124
30	45	6590	0.127617602428
30	65	6590	0.123672230653
60	5	6545	0.198472116119
60	25	6545	0.0890756302521
60	45	6545	0.0786860198625
60	65	6545	0.0707410236822
90	5	6500	0.191384615385
90	25	6500	0.0741538461538
90	45	6500	0.0647692307692
90	65	6500	0.0555384615385

According to Table 1 above, OOB errors gradually decrease and converge, and reach a steady state when the number of decision trees is greater than 45. The larger the forecast time window, the smaller the OOB error, but this reduction is marginal.

3.6 Model result

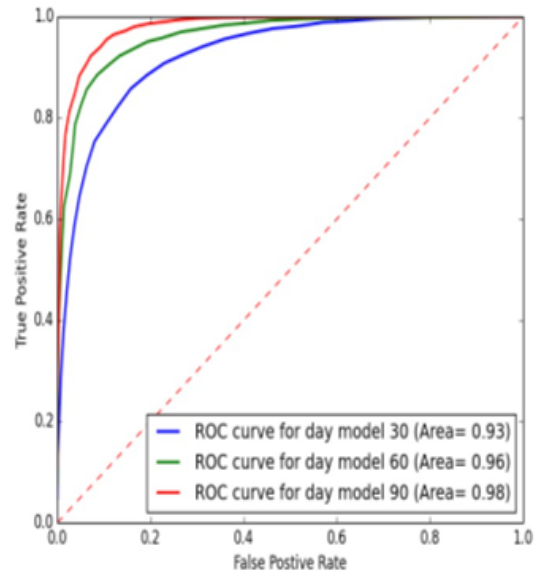
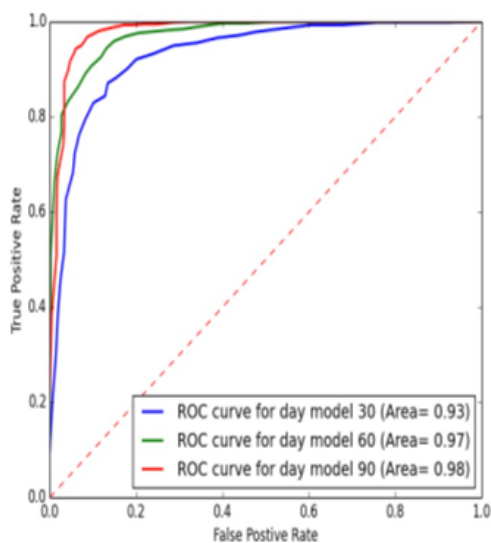
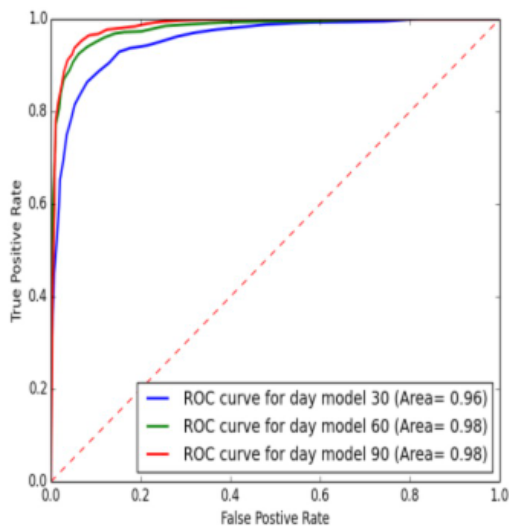


Figure 4. ROC curve of stock price prediction accuracy of Apple, Samsung and General Electric

The experimental results (Figure 4) show that using random forest classifier to predict the long-term trend of stocks has remarkable accuracy. Through the analysis of data sets such as Apple, Samsung, and General Electric, our model was able to achieve 85-95% accuracy, which demonstrates the potential and effectiveness of the method in predicting stock trends. [22-24] The advantage of the random forest model is that it can handle a large amount of feature data and can effectively deal with noise and complexity in the data. In addition, as the number of decision trees in the random forest increases, we observe that the model's predictions show a more stable trend. This shows that increasing the number of decision trees can improve the prediction accuracy of the model and reduce the volatility of the results, making the prediction more reliable.

Although this study focuses on the prediction of long-term stock movements, the same method can be applied to the prediction of shorter periods. By training with hour - or minute-level trading data, we can make more accurate predictions about short-term volatility, giving investors more flexibility and precision. This expansion can help investors better respond to market changes and adjust investment strategies in a timely manner, thus maximizing the return on investment. In summary, our research results provide a reliable method for investors to predict the stock trend by using random forest classifiers to guide their investment decisions, and with the refining of data granularity, our model can also adapt to the prediction needs of shorter cycles, so as to further improve investment benefits.

4 CONCLUSION

In this study, the random forest classifier is used to predict the long-term trend of stocks, and remarkable results have been achieved. [25] Analyzing data sets such as Apple, Samsung, and General Electric, our model shows 85-95% accuracy, demonstrating the potential and effectiveness of this approach in predicting stock trends. Random forest models are known for their ability to process large amounts of feature data and for their effective handling of data noise and complexity. In addition, as the number of decision trees in the random forest increases, we observe that the model's predictions show a more stable trend. This shows that increasing the number of decision trees can improve the prediction accuracy of the model and reduce the volatility of the results, making the prediction more reliable.

While this study focuses on the prediction of long-term stock movements, the same method can be applied to the prediction of shorter periods. By training with hour-level or minute-level trading data, we can more accurately predict short-term volatility, providing investors with more flexibility and precision. This expansion can help investors better respond to market changes and adjust investment strategies in a timely manner, thus maximizing the return on investment.

However, as the field of artificial intelligence continues to evolve, techniques such as deep learning show great potential in the field of time series prediction. In recent years, deep learning models have made significant progress in time series analysis, especially in the financial field. By combining deep learning models, we can better capture complex patterns and dynamic features in the data, thereby improving the accuracy and stability of predictions. In the future, with the continuous development and application of deep learning technology, we have reason to believe that artificial intelligence will play an increasingly important role in the field of time series forecasting, providing investors with more accurate and reliable forecast results, helping them to better formulate investment strategies and obtain higher returns.

Acknowledgments

We would like to express our sincere thanks to the first author of this article, Wu Binbin, and his team. In the article "Enterprise Digital Intelligent Remote Control System Based on Industrial Internet of Things", they conducted in-depth research and discussion on the key issues of enterprise digital transformation, providing us with valuable ideas and inspiration. Thanks to Wu Binbin and his team for their contributions in artificial intelligence, deep learning and other related fields. Their research has played a key role in integrating artificial intelligence and the Industrial Internet of Things, a concept that provides strong support for digital transformation across industries and provides an important

reference for our future research and practice. These achievements have made important contributions to promoting the development of financial intelligence and laid a solid foundation for our future research and practice.

Funding

Not applicable.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author Contributions

Not applicable.

About the Authors

ZHONG, Yuqiang

Department of Information and Computer Sciences, Henan Agricultural University, Shenzhen, Guangdong, China

CHENG, Qishuo

Department of Economics, University of Chicago, Chicago, IL, USA

QIN, Lichen

Department of Computer Science, University of Rochester, Rochester, NY, USA

XU, Jinxin

Department of Cox Business School, Southern Methodist University, Dallas, TX, USA

WANG, Han

Department of Mathematics, University of Southern California, Alhambra, CA, USA

References

- [1] Li, Zhenglin, et al. "Feature Manipulation for DDPM based Change Detection." arXiv preprint arXiv:2403.15943 (2024).
- [2] Xu, J., Wu, B., Huang, J., Gong, Y., Zhang, Y., & Liu, B. (2024). Practical Applications of Advanced Cloud Services and Generative AI Systems in Medical Image Analysis. arXiv preprint arXiv:2403.17549.
- [3] Zhu, Mengran, et al. "Ensemble Methodology: Innovations in Credit Default Prediction Using LightGBM, XGBoost, and LocalEnsemble." arXiv preprint arXiv:2402.17979 (2024).
- [4] Zhang, Y., Liu, B., Gong, Y., Huang, J., Xu, J., & Wan, W. (2024). Application of Machine Learning Optimization in Cloud Computing Resource Scheduling and Management. arXiv preprint arXiv:2402.17216.
- [5] Gong, Y., Huang, J., Liu, B., Xu, J., Wu, B., & Zhang, Y. (2024). Dynamic Resource Allocation for Virtual Machine Migration Optimization using Machine Learning. arXiv preprint arXiv:2403.13619.
- [6] Liu, B. (2023). Based on intelligent advertising recommendation and abnormal advertising monitoring system in the field of machine learning. *International Journal of Computer Science and Information Technology*, 1(1), 17-23.
- [7] Wu, Binbin, et al. "Enterprise Digital Intelligent Remote Control System Based on Industrial Internet of Things." (2024).
- [8] Pan, Y., Wu, B., Zheng, H., Zong, Y., & Wang, C. (2024, March). THE APPLICATION OF SOCIAL MEDIA SENTIMENT ANALYSIS BASED ON NATURAL LANGUAGE PROCESSING TO CHARITY. In The 11th International scientific and practical conference "Advanced technologies for the implementation of educational initiatives"(March 19–22, 2024) Boston, USA. International Science Group. 2024. 254 p. (p. 216).
- [9] K. Xu, X. Wang, Z. Hu and Z. Zhang, "3D Face Recognition Based on Twin Neural Network Combining Deep Map and Texture," 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, 2019, pp. 1665-1668, doi: 10.1109/ICCT46805.2019.8947113.
- [10] Shi, Peng, Yulin Cui, Kangming Xu, Mingmei Zhang, and Lianhong Ding. 2019. "Data Consistency Theory and Case Study for Scientific Big Data" *Information 10*, no. 4: 137. <https://doi.org/10.3390/info10040137>.
- [11] Zhenghua Hu, Xianmei Wang, Kangming Xu, and Pu Dong. 2020. Real-time Target Tracking Based on PCANet-CSK Algorithm. In *Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence (CSAI '19)*. Association for Computing Machinery, New York, NY, USA, 343–346. <https://doi.org/10.1145/3374587.3374607>.
- [12] Medication Recommendation System Based on Natural Language Processing for Patient Emotion Analysis. (2024). *Academic Journal of Science and Technology*, 10(1), 62-68. <https://doi.org/10.54097/v160aa61>
- [13] Li, X., Zheng, H., Chen, J., Zong, Y., & Yu, L. (2024). User Interaction Interface Design and Innovation Based on Artificial Intelligence Technology. *Journal of Theory and Practice of Engineering Science*, 4(03), 1-8.
- [14] Xu, K., Zhou, H., Zheng, H., Zhu, M., & Xin, Q. (2024). Intelligent Classification and Personalized Recommendation of E-commerce Products Based on Machine Learning. arXiv preprint arXiv:2403.19345.
- [15] Zhou, H., Xu, K., Bao, Q., Lou, Y., & Qian, W. (2024). Application of Conversational Intelligent Reporting System Based on Artificial Intelligence and Large Language Models. *Journal of Theory and Practice of Engineering Science*, 4(03), 176-182.
- [16] Li, L., Xu, K., Zhou, H., & Wang, Y. (2024). Independent Grouped Information Expert Model: A Personalized Recommendation Algorithm Based on Deep Learning.
- [17] Che, C., Lin, Q., Zhao, X., Huang, J., & Yu, L. (2023, September). Enhancing Multimodal Understanding with CLIP-Based Image-to-Text Transformation. In *Proceedings of the 2023 6th International Conference on Big Data Technologies* (pp. 414-418).
- [18] Huang, Zengyi, et al. "Research on Generative Artificial Intelligence for Virtual Financial Robo-Advisor." *Academic Journal of Science and Technology* 10.1 (2024): 74-80.
- [19] Huang, Zengyi, et al. "Application of Machine Learning-Based K-Means Clustering for Financial Fraud Detection." *Academic Journal of Science and Technology* 10.1 (2024): 33-39.
- [20] Wu, Y., Jin, Z., Shi, C., Liang, P., & Zhan, T. (2024). Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis. arXiv preprint arXiv:2403.08217.
- [21] Shi, C., Liang, P., Wu, Y., Zhan, T., & Jin, Z. (2024). Maximizing User Experience with LLMops-Driven Personalized Recommendation Systems. arXiv preprint

arXiv:2404.00903.

- [22] Huang, Zengyi, et al. "Application of Machine Learning-Based K-Means Clustering for Financial Fraud Detection." *Academic Journal of Science and Technology* 10.1 (2024): 33-39.
- [23] Xu, Z., Gong, Y., Zhou, Y., Bao, Q., & Qian, W. (2024). Enhancing Kubernetes Automated Scheduling with Deep Learning and Reinforcement Techniques for Large-Scale Cloud Computing Optimization. arXiv preprint arXiv:2403.07905.
- [24] Xu, X., Xu, Z., Ling, Z., Jin, Z., & Du, S. (2024). Comprehensive Implementation of TextCNN for Enhanced Collaboration between Natural Language Processing and System Recommendation. arXiv preprint arXiv:2403.09718.
- [25] Song, B., Xu, Y., & Wu, Y. (2024). ViTCN: Vision Transformer Contrastive Network For Reasoning. arXiv preprint arXiv:2403.09962.