Comparative Analysis of Traditional and LSTM Models for Netflix Stock Movement Classification

He Bai

Faculty of Engineering, The University of Sydney, Sydney, New South Wales (NSW), 2050, Australia
hbai0922@uni.sydney.edu.au

Abstract. With the continuous development of machine learning methods, various types of models are increasingly used in stock prediction. However, the systematic comparison between traditional machine learning methods and deep learning methods is still relatively insufficient for classification prediction tasks such as upward and downward direction. In this study, a series of technical indicators were constructed as input features with the upward and downward direction of Netflix stock as the classification prediction target, and Logistic Regression, Decision Tree, Random Forest, eXtreme Gradient Boosting, Voting Classifier, and Long Short-Term Memory Network models were trained and evaluated, respectively. The results show that the logistic regression model and Extreme Gradient Boosting (XGBoost) perform well overall in all evaluation metrics, indicating their strong generalization ability in this classification task. And although Long Short Term Memory (LSTM) is good at dealing with time series, its performance on this dataset did not exceed that of traditional methods. After the study, it was found that traditional classification models have strong stability and practicality in certain situations. This paper provides empirical support to explore the applicability of traditional models in financial time series classification.

Keywords: Stock Movement Classification, Machine Learning, Logistic Regression, LSTM.

1 Introduction

Stocks are significant in the economic market, characterized by complex and unpredictable changes. Stock data contains a wealth of linear and nonlinear complex information, and the nonlinear and nonparametric capabilities of machine learning are well-suited to uncover the intricate relationships between stock data and stock prices. Netflix was founded in 1997 and publicly listed in 2002, and the company's exponential growth has not only revolutionized the way people consume entertainment, but it has also made it a staple in many portfolios [1]. Therefore, studying the stock movement of Netflix helps to grasp the running trend and investment value of high-growth technology companies in the market.

Depending on the prediction objective, stock forecasting can be broadly categorized into price prediction and trend prediction, with most existing research focusing on the

A. J. Moshayedi (ed.), *Proceedings of the 2025 International Conference on Hybrid Commerce, Human Capital, and Economic Dynamics (ICHCH 2025)*, Advances in Economics, Business and Management Research 374,
https://doi.org/10.2991/978-2-38476-585-0_20

former. A wide range of models has been employed for price prediction, including linear regression (LR), artificial neural networks, single-layer long short-term memory networks, and other artificial intelligence techniques [1]. In the early stages of applying machine learning to quantitative finance, computational constraints led researchers to favor relatively simple models such as logistic regression, SVM, and decision trees. However, given the highly complex and non-linear nature of financial markets, it is widely acknowledged that simple models are insufficient to capture market dynamics or achieve consistent excess returns. Consequently, more sophisticated models have been increasingly adopted in the field, among which random forests and neural networks are considered representative [2].

In the study conducted by Geng, logistic regression, decision trees, and multilayer neural networks were employed to predict stock indices. Among the three individual forecasting models, the multilayer neural network demonstrated the highest accuracy and stability [2]. Singh and Khanna further explored hybrid techniques and reported that the combination of neural networks and random forests (NN+RF) outperformed all single models and other hybrid approaches, achieving an accuracy of 78.57% [3]. In Wu's study, six machine learning models were assessed through ten-fold cross-validation, with the random forest model achieving the lowest mean squared error of 0.000563, indicating superior predictive performance [4]. Similarly, Chen applied random forest, XGBoost, and support vector regression models for rolling prediction tasks with parameter optimization and found that XGBoost consistently outperformed the other two models across all evaluation metrics [5]. Jia's research revealed that the LSTM model achieved a mean square error of 0.003, which was significantly lower than that of the BP neural network (0.0036), suggesting LSTM's stronger capability in data fitting [6].

In addition to single-model approaches, recent studies have proposed more sophisticated and integrated architectures for stock prediction. For instance, combining convolutional neural networks with long short-term memory networks has been proven to substantially improve prediction accuracy [7]. Furthermore, hybrid models such as ANN-MLP and GARCH-MLP, which integrate artificial neural networks with statistical and deep learning techniques, have been developed. These models, often incorporating the backpropagation algorithm and multi-layer feedforward structures, have demonstrated promising results in empirical applications [8].

This study focuses on Netflix stock (2013-2023) as the subject of analysis and the intraday downward and upward direction of its stock price through machine learning methods. Seven technical indicators are constructed, and the returns of the past five trading days are used as input features. Traditional models, including logistic regression, decision tree, random forest, and XGBoost, are compared with the LSTM model.Subsequent paragraphs, however, are indented.

2 Method

2.1 Dataset

The dataset Netflix Stock Price is sourced from Kaggle and contains historical stock price data for Netflix from 2013 to 2023 [9]. The dataset includes six technical indicators, including opening price, high price, low price, closing price, adjusted closing price, and trading volume.Subsequent paragraphs, however, are indented.

2.2 Method Introduction

Feature Engineering. Based on the original six indicators, seven additional technical indicators were selected and calculated, including Moving Average, Relative Strength Index, On-Balance Volume, Volatility, Momentum, Moving Average Convergence Divergence, and Lagging features. These technical indicators can transform raw price data into structured characteristic variables that can enhance the understanding of trends in multiple dimensions, thus improving the accuracy and stability of classification forecasts.

Moving Average averages prices over a period and moves them over time to create a continuous “average curve”. It is calculated by the specified time window. In this study, moving averages over 5 days, 10 days, 30 days, and 60 days were computed.

Relative Strength Index (RSI) evaluates the extent of recent price fluctuations and their relative strength, with a range from 0 to 100. The indicator determines the level of strength and weakness by calculating the average upside and average downside over the past fourteen days.

Volatility is commonly measured by the standard deviation of the stock price over a certain period. It reflects the variability of stock price movement. In this study, the period was set to 30 days.

Momentum captures short-term price changes by comparing the current price to that of several days earlier.10 days were chosen in this study as the time interval for calculation.

Moving Average Convergence Divergence measures the gap between short-term and long-term exponential moving averages, which shows the strength of market trends. The study was set for 12 days and 26 days.

To capture short-term fluctuations in stock prices, this paper introduces a new feature that calculates the price change between the closing price of the current day and the closing price of the past 5 days, i.e., the change in the closing price over the last 5 trading days.

Model Introduction. Logistic Regression, Decision Tree, XGBoost, and LSTM are models that used in this study.

Logistic regression is a statistical method employed for binary classification, specifically to predict stock market trends (such as upward or downward movement) [10].

Decision tree algorithm follows a top-down recursive process, aiming to build a tree that reduces entropy at each step, prioritizing splits that result in the fastest decrease in overall entropy. The entropy at the leaf nodes is zero, meaning that all instances in each leaf node belong to the same class [11].

XGBoost is a gradient boosting framework that approximates the residuals of the model's loss function using a Taylor expansion. It incorporates a regularization term based on the complexity of the tree to fit the data residuals. This method allows for controlling the model's complexity, thereby reducing model variance. During the optimization process, XGBoost implements pre-pruning to improve the accuracy of the predictions [12].

LSTM is a specialized architecture of recurrent neural networks (RNNs) designed to overcome the vanishing gradient problem in traditional RNNs, allowing for efficient learning and retention of information across long data sequences [1].

In this study, six machine learning models are compared. All the models are evaluated on the same dataset using Accuracy, Precision, Recall, and F1score as evaluation metrics. And the experimental workflow is illustrated in Fig. 1.

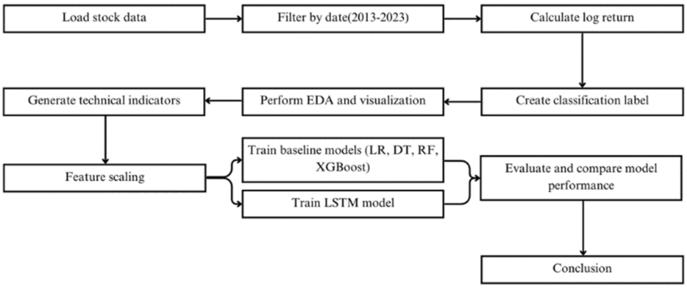

Fig. 1. Workflow (Photo/Picture credit: Original).

3 Result

To improve the scientific validity of the experiments, 10-fold cross-validation was performed for traditional models. Among them, XGBoost was tuned and cross-validated using the best parameters. LSTM, on the other hand, was subjected to 20 repetitions of the experiments to count the average performance and the optimal results to enhance the robustness and comparability of the results. The model training results are presented in Table 1.

Table 1. Training Results.

Name	Accuracy	Precision	Recall	F1 score
Logistic Regression	**0.5849**	**0.5872**	**0.6326**	**0.6054**
Decision Tree	0.4951	0.5054	0.4663	0.4821
Random Forest	0.5395	0.5536	0.4974	0.4985
XGBoost (Tuned)	0.5830	0.5780	0.6736	0.6153
VotingClassifier	0.5060	0.5155	0.4788	0.4922
LSTM(average)	0.5002	0.4818	0.4986	0.4891
LSTM(best)	0.5233	0.5051	0.5792	0.5396

Logistic regression has the best performance in Accuracy and Precision, and higher recall and F1 scores, indicating that it is the most stable model and the most suitable model for the present data. The XGBoost model achieves the best results in Recall and F1 Score, and although Accuracy is slightly lower than that of the logistic regression, its performance is especially outstanding in Recall, which is more suitable for dealing with data with uneven category distribution. Although the Accuracy is slightly lower than that of logistic regression, the XGBoost model performs better in Recall and F1 Score, which is more suitable for the data with uneven distribution of categories. In contrast, the Decision Tree model performs poorly in all indicators and has the weakest overall performance, while the other traditional models have average overall performance. For the LSTM model, its average Precision and F1 Score are low, and its Recall performance is still good, its overall performance is inferior to that of traditional models. Although the Recall of the optimal LSTM is higher than the average, its overall performance is still inferior to that of XGBoost and logistic regression.

4 Discussion

Even though LSTM is a complex model suitable for processing time series data, it did not show significant advantages in the evaluation metrics. LSTM is good at capturing long-term time-series dependencies and usually requires a large amount of training data to learn effective time-series patterns. The samples in this dataset are only 2,000 items, so it is possible that the sample size of the data is not large enough, and the advantages of LSTM may not be able to be realized. At the same time, LSTM contains various hyperparameters, including the number of layers and the learning rate, etc., and the tuning parameter is complicated and time-consuming, and costly, which may lead to unstable training effects or fall into local optimization, thus affecting the final model performance.

As logistic regression relies on the assumption of a linear correlation between input features and the predicted output, if the feature variables themselves have a certain linear relationship, the logistic regression model tends to capture this relationship well. For example, there is usually a certain linear relationship between relative strength index and stock prices trend: when relative strength index is greater than 70, the stock tends to be in an overbought state, and the price may fall back; when relative strength index is less than 30, the stock is in an oversold state, and the price may rebound. On the contrary, the model of logistic regression is simpler and can still show good results with small samples.

Although Voting Classifier often demonstrates superior performance in many studies, in this experiment, its accuracy is only 0.5060, which is weak among all models and not as effective as Logistic Regression, XGBoost, or LSTM (best). The base models used by Voting Classifier include Logistic Regression, Decision Tree, Random Forest, and tuned XGBoost, of which Decision Tree, with an accuracy of 0.4951, performs poorly and may affect the overall integration performance. Although the soft voting strategy is used in this experiment, the weights of the models are equal by default. Under this setting, the prediction probability of the Decision Tree with poorer performance may dilute the judgment ability of the strong models, resulting in unsatisfactory integration.

To improve the predictive performance of the Voting Classifier, customized weights can be assigned to different base models by increasing the weights of better-performing models, such as Logistic Regression and XGBoost. Another possible improvement is to remove the Decision Tree model, which has relatively poor predictive performance and may negatively affect the ensemble results.

5 Conclusion

In this study, based on the original features of the dataset, new technical indicators and a total of six classification models are constructed to complete the binary prediction of the rise and fall of Netflix stock. Each model was compared in the accuracy, precision, recall, and F1 score. By comparing the traditional model with the LSTM model, the effectiveness of several models in the financial time series classification task is verified, which provides practical experience for stock market prediction.

Logistic regression performs well when the feature variables themselves have a certain linear relationship, indicating its stability and reasonableness in the time series classification task. Due to the weak performance of the decision tree model, resulting in the Voting Classifier results of the four traditional models for integrated training are not as good as the single model, so the combination optimization training of the Voting Classifier can be considered to explore the optimal way of combining the models with the collocation strategy. The LSTM model, which is good at dealing with time series data, has poorer results, indicating that this model needs the support of a larger sample size. And more features can be added in future research, which can be considered to include incorporating macroeconomic indicators (e.g., GDP, CPI). The model can be tuned with more complex hyperparameters, or more complex network architectures can be designed to improve the accuracy of the model.

In addition, the data used in this study only comes from the historical trading data of a single stock, i.e. Netflix, and the results obtained can only reflect the predictive performance of the models on that stock and therefore do not yet have the basis for direct generalization to other individual stocks or the broader stock market. If the research objective is expanded to include the entertainment stock market, the stock data of other representative companies in the industry can be further introduced to enhance the generalizability and reference value of the research findings.

References

1. Sharma, U., Kumar, A., Kumar, R., Manchanda, M., Hooda, S.: Netflix stock market price prediction. In: AIP Conference Proceedings, vol. 3224, no. 1, p. 020061 (2025)
2. Geng, E., Man, Y., Hu, K.: Research on integrated forecasting methods for the stock market. Statistics and Applications 10(6), 1009–1013 (2021)
3. Singh, S., Khanna, D.: Stock market trends analysis using various machine learning techniques. In: AIP Conference Proceedings, vol. 3209, no. 1, p. 050005 (2024)

4. Wu, R.: Forecast analysis of Shanghai composite index based on machine learning method. Hans Journal of Data Mining 6(1), 1–8 (2016)
5. Chen, X.: Machine learning based stock return prediction and portfolio research. Operations Research and Fuzziology 14(2), 599–609 (2024)
6. Jia, C.: Comparison of advantages of multiple machine learning methods in stock prediction. Finance 15(1), 238–245 (2025)
7. He, K., Jiang, Q.: Research on stock prediction algorithm based on CNN and LSTM. Academic Journal of Computing and Information Science 5(12), 98–106 (2022)
8. Kurani, A., Doshi, P., Vakharia, A., Shah, M.: A comprehensive comparative study of artificial neural networks (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science 10, 1–26 (2021)
9. Shan, H.: Netflix Stock Price. https://www.kaggle.com/datasets/henryshan/netflix-stock-price. Accessed: July 16, 2025
10. Singh, S., Khanna, D.: Stock market trends analysis using various machine learning techniques. In: AIP Conference Proceedings, vol. 3209, no. 1, p. 050005 (2024)
11. Wang, Y., Chen, D., Tang, Y.: Stock prediction based on CART decision tree and boosting methods. Journal of Harbin University of Science & Technology 24(6) (2019)
12. Chen, X.: Stock return prediction and portfolio optimization based on machine learning. Operations Research and Fuzziology 14(2), 599–609 (2024)