Proceedings of CONF-BPS 2026 Symposium: GenAI, Labour Markets, and the Economics of Human and Financial Capital
DOI: 10.54254/2754-1169/2026.LD30724

Comparative Analysis of Machine Learning Models for Predicting Apple Stock Price Direction: From Baseline to Optimization

Chenfeng Wang

School of Statistics and Data Science, Shanghai University of International Business and Economics, Shanghai, China
24076064@suibe.edu.cn

Abstract. Stock price prediction is a significant challenge in financial industry due to market complexity and volatility. Current research on Apple stock forecasting often uses limited model comparisons and lacks systematic optimization, like Long Short-Term Memory (LSTM), resulting in constrained robustness across varying market conditions. This study employs and optimizes machine learning models for predicting the direction of Apple Inc.'s stock price. The research used daily trading data from 2013 to 2023 and implemented four baseline models—Logistic Regression, Random Forest, Extreme Gradient Boosting (XGBoost), and LSTM after data preprocessing and feature selection. The baseline evaluation revealed significant fitting issues, including underfitting in linear models and overfitting in tree-based models and LSTM. To address these limitations, this study applied optimization framework, including incorporating feature engineering techniques such as creating lagged variables and rolling statistics, and hyperparameter tuning. This application substantially improved model performance and generalization capability. The optimized LSTM and XGBoost models emerged as top performers, achieving accuracies of 71% and 70%, with AUC scores of 0.78. The findings demonstrate that predictive performance depends not only on algorithm selection but also on integrated working steps including data preprocessing, model-specific adjustments, and experimental design. This research provides a methodological framework for addressing complex forecasting challenges in financial time series.

Keywords: Apple Stock Price Prediction, Model Optimization, Feature Engineering, LSTM, XGBoost

1. Introduction

Stock prices are influenced by macroeconomic conditions, corporate performance, investor sentiment, and sudden events. As a global technology leader, Apple Inc.'s stock performance reflects both company-specific factors, like product innovation, and broader sector-wide and macroeconomic trends [1]. Annual product launches can trigger sharp price movements, while its high liquidity and strong fundamentals consistently attract investor interest. Its pattern of long-term

DOI: 10.54254/2754-1169/2026.LD30724

growth and short-term uncertainty makes accurate prediction particularly valuable [2]. Recent studies combine statistical methods with machine learning to improve the accuracy of Apple stock predictions. For example, Horák & Kaisle analyzed long-term Apple stock price time series, demonstrating its resilience during economic crises and the pandemic and highlighting how external events drive price movement [3]. Similarly, Omer et al. used Linear Regression and Random Forest models, confirming machine learning's ability to capture nonlinear relationships and identifying trading volume as a important indicator [4]. These works provide a methodological foundation but also lack comprehensive model comparison and systematic optimization.

A closer look at the literature reveals more limitations in current forecasting methodologies. Traditional time-series models like Autoregressive Integrated Moving Average (ARIMA) effectively capture short-term trends but struggle with long-term nonlinear shocks due to their linear assumptions [3,5]. Though hybrid approaches like Discrete Wavelet Transform-Logistic Regression (DWT-LR) and Long Short-Term Memory Neural Network (LSTM) improved accuracy through denoising, they often lack generalizability across different market conditions. Fundamental analysis provide structural insights but are rarely effectively used in practical forecasting frameworks [6,7]. The emergence of machine learning introduces new potential but also reveals distinct challenges. For instance, Wang demonstrated the value of multi-model frameworks by combining Random Forest, Linear Regression, and Extreme Gradient Boosting (XGBoost) [8]. However, such studies typically focus on predicting exact price levels rather than directional movements, limiting their practical utility for investment decisions where classifying price direction is paramount.

Consequently, existing research presents two critical gaps. Most studies compare only narrow sets of modeling paradigms, and there is insufficient optimization of complex architectures like LSTM. These limitations collectively constrain model robustness and generalizability across varying market environments.

By systematically comparing four representative models and conducting in-depth optimization, this research not only addresses the shortcomings of existing research and provides more reliable methodological references for Apple stock prediction but also offers a research framework that can be referenced for similar financial time series forecasting problems.

This research processes data and builds baseline models (Logistic Regression, Random Forest, XGBoost, LSTM), then improves performance through feature engineering and hyperparameter tuning to optimize Apple stock direction forecasting.

2. Methodology

2.1. Data

This research utilizes daily trading data of Apple Inc. (AAPL) sourced from Kaggle, covering the period from January 2013 to December 2023 [9]. The dataset contains 2,466 observational samples with 14 feature variables. Key variables include date, open price, high price, low price, close price, and trading volume. Technical indicators such as Moving Average 20 (MA20), Moving Average 50(MA50), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and volatility are also included. The prediction task is formulated as a binary classification problem. The target variable is set to 1 when the daily stock price shows an increase, and 0 when it decreases. This binary label serves as the prediction target for all models.

DOI: 10.54254/2754-1169/2026.LD30724

2.2. Exploratory data analysis and preprocessing

An exploratory data analysis was conducted to examine data characteristics. The analysis revealed that price variables exhibited similar distribution patterns, while trading volume displayed a right-skewed distribution. Technical indicators were found to conform to their expected theoretical characteristics. Correlation analysis indicated strong positive correlations among price variables and a negative correlation between trading volume and prices. Figure 1 reflects the distribution of variables. Figure 2 reflects the correlations among various variables.

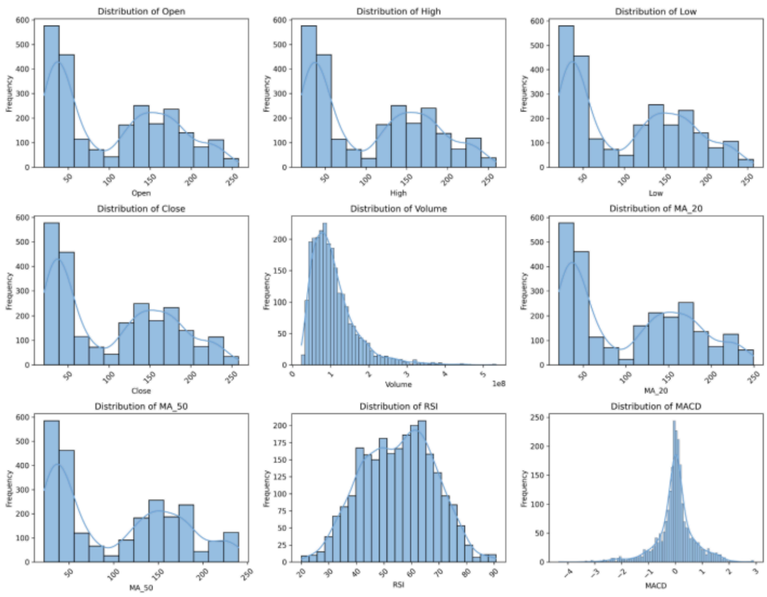

Figure 1. Distribution plots of key stock and technical variables (original)

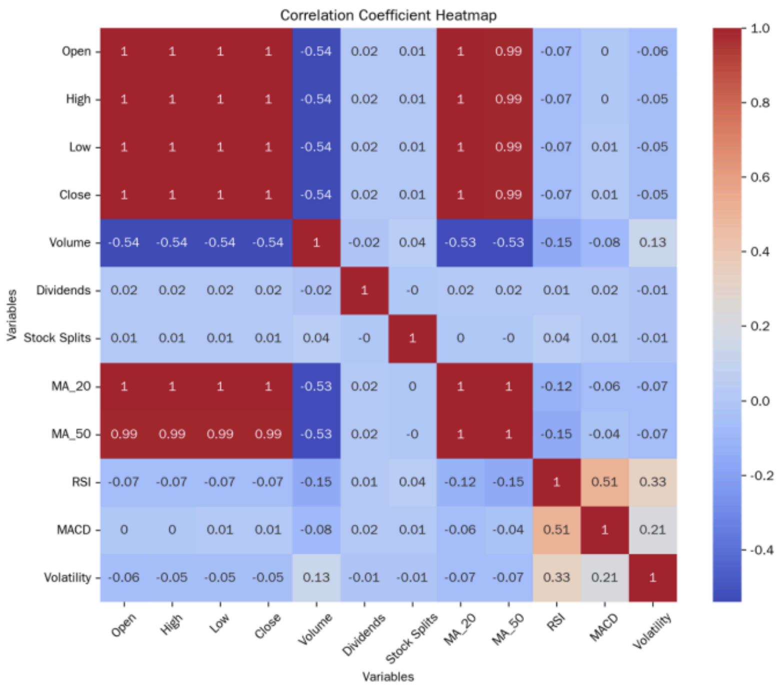

Figure 2. Correlation coefficient heatmap of all features (original)

DOI: 10.54254/2754-1169/2026.LD30724

Based on these findings, the following preprocessing steps were implemented: Forward-filling was applied to handle missing data, maintaining time series continuity. The IQR method was used to detect and process outliers, particularly extreme values in volume and volatility. Technically meaningful indicators were selected as model inputs, avoiding raw price data. Variables with limited predictive value, such as dividends and stock splits, were excluded.

2.3. Model principles and advantages

Four machine learning models were selected to represent different algorithmic approaches for stock direction prediction. Each model possesses distinct characteristics suitable for financial forecasting tasks.

Logistic Regression (LR) establishes a linear decision boundary to classify stock movements. As a linear model, its main strengths include computational efficiency and straightforward interpretability. However, this linear nature also limits its capacity to identify complex nonlinear patterns in market data.

Random Forest (RF) operates as an ensemble of decision trees. By combining predictions from multiple trees, the model could reduce prediction deviation and enhance generalization. This results in great avoidance of overfitting.

XGBoost employs a sequential boosting technique so that each new tree corrects errors from previous ones. It uses regularization terms to prevent overfitting and maintain high prediction accuracy. This contributes to capturing complex relationships between features in financial data.

LSTM networks specialize in processing sequential information through specialized gating mechanisms. These gates selectively preserve or discard information across time steps, enabling the model to recognize long-term dependencies in price sequences.

These four models cover linear models, ensemble methods and deep learning, forming a comparison of different models for stock price direction forecasting.

2.4. Experimental design

The experimental design followed two stages: Firstly, after data preprocessing, performance of baseline models was evaluated. Secondly, performance of optimized models through feature engineering and hyperparameter tuning was evaluated again.

Considering the characteristics of financial time series, this study employs a chronological split method, allocating the first 80% of data to the training set and the remaining 20% to the test set. This strategy effectively prevents data leakage and simulates real-world investment prediction situation.

Five standard classification metrics were used to evaluate baseline and optimized models' performance. Accuracy represents the percentage of correct predictions. Precision measures of how often rising price predictions were correct. Recall indicates the model's success in identifying actual price increases. F1-score provides a balance between precision and recall. Finally, AUC-ROC evaluates the model's overall ability to distinguish between rising and falling price days across all classification thresholds.

The baseline models exhibited great fitting issues. Logistic Regression showed significant underfitting and high errors across both training and test sets. Random Forest and XGBoost showed severe overfitting, performing perfectly on training set but poorly on test set. The LSTM model also demonstrated moderate overfitting tendencies.

DOI: 10.54254/2754-1169/2026.LD30724

These diagnostic results guided the optimization strategy: enhancing feature representation for Logistic Regression, applying stronger regularization to tree-based models, and adjusting architectural parameters for LSTM to improve generalization capability.

To address these challenges, a structured optimization framework was implemented. This involved feature engineering techniques including lagged variable creation, technical indicator combinations, and temporal feature construction. Simultaneously, hyperparameter tuning was conducted using grid search and cross-validation methods. The specific optimization strategies for each model are detailed in Table 1.

Table 1. Details of feature engineering and hyperparameter tuning

Model	Feature Engineering	Hyperparameter Tuning
Logistic Regression	Price return features Technical indicator interactions Log transformation of Volume	Regularization strength & type Solver algorithm optimization
Random Forest	Removal of low-correlation features Retention of key technical indicators	Maximum tree depth constraint Minimum samples for split Number of estimators
XGBoost	SHAP-based feature analysis Dynamic feature selection Time-lagged features	Learning rate reduction Tree depth constraint Subsampling rates
LSTM	Time-series window construction Feature normalization Rolling statistical features	Network units adjustment Dropout regularization Batch size & Early stopping

3. Result

3.1. Baseline model performance and fitting status

The baseline models showed clear limitations in both prediction accuracy and fitting quality. Initial accuracy scores were low, ranging from 42% to 53% (Table 2). Logistic Regression performed worst with only 42% accuracy and 0.50 AUC, indicating nearly random predictions.

Table 2. Baseline model performance metrics

Metric	Logistic Regression	LSTM	Random Forest	XGBoost
Accuracy	42%	53%	48%	52%
Precision	45%	55%	50%	55%
Recall	40%	52%	45%	50%
F1-Score	42%	53%	47%	52%
AUC	0.50	0.61	0.58	0.60

Fitting analysis revealed more problems (Table 3). Logistic Regression had severe underfitting, with just a 2% gap between training and validation errors. This suggests the model was too simple. In contrast, Random Forest and XGBoost showed serious overfitting. Their training and validation error gaps reached 50% and 38% respectively. LSTM also overfit, though less severely, with a 17% error gap that indicated need for better architecture tuning.

DOI: 10.54254/2754-1169/2026.LD30724

Table 3. Baseline model fitting status analysis

Model	Training Error Rate	Cross-Validation Error Rate	Error Rate Gap	Fitting Diagnosis
Logistic Regression	58%	60%	2%	Severe Underfitting
Random Forest	2%	52%	50%	Severe Overfitting
XGBoost	10%	48%	38%	Severe Overfitting
LSTM	30%	47%	17%	Mild Overfitting

3.2. Optimized model performance and improvement analysis

LSTM and XGBoost became the best performers (Table 4). LSTM reached 71% accuracy with 0.78 AUC, while XGBoost achieved 70% accuracy with the same AUC score. Random Forest improved to 68% accuracy, and Logistic Regression remained weakest at 56% accuracy despite some gains.

Table 4. Optimized model performance metrics

Metric	Logistic Regression	LSTM	Random Forest	XGBoost
Accuracy	56%	71%	68%	70%
Precision	58%	72%	67%	74%
Recall	53%	70%	63%	72%
F1-Score	55%	71%	65%	73%
AUC	0.62	0.78	0.75	0.78

The optimization also greatly improved model fitting (Table 5). LSTM achieved excellent generalization with only a 3% error gap. Random Forest and XGBoost substantially reduced overfitting, cutting their error gaps from 50% to 26% and from 38% to 19% respectively. Logistic Regression continued to show underfitting, consistent with its linear limitations.

Table 5. Optimized model fitting status analysis

Model	Training Error Rate	Cross-Validation Error Rate	Error Rate Gap	Fitting Diagnosis
LR	43%	45%	2%	Severe Underfitting
Random Forest	8%	34%	26%	Moderate Overfitting
XGBoost	12%	31%	19%	Mild Overfitting
LSTM	27%	30%	3%	Good Fit

After systematic optimization, all models showed clear improvement. (Figure 3)

DOI: 10.54254/2754-1169/2026.LD30724

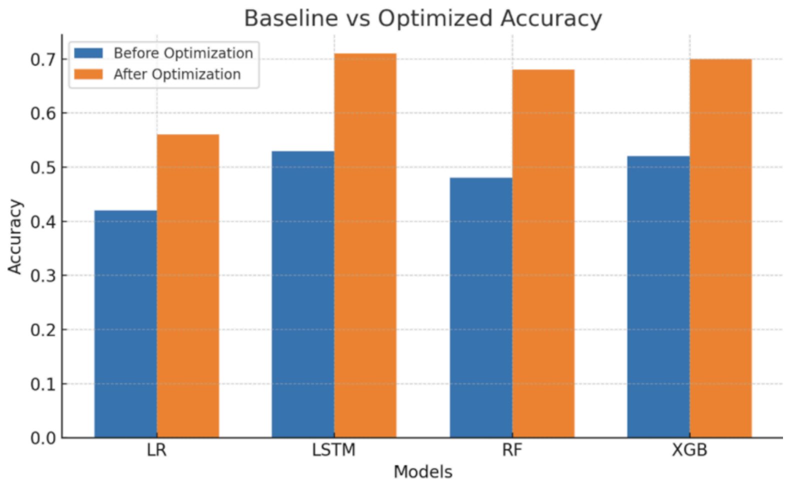

Figure 3. Baseline vs optimized accuracy (original)

4. Discussion

The core finding of this study is that the predictive capability of the models was significantly improved through optimization. The strong performance of LSTM and XGBoost, along with the general enhancement of all models, not only supports the effectiveness of the optimization strategies but also reveals the logic of improvement from the initial baseline to the optimized versions [10]. The improvement in model performance fundamentally results in the optimization of data representation, model architecture, and experimental design.

The limitations of the baseline models originated from the mismatch between data and models. On the data level, the initial feature set was static and flat, lacking temporal context information essential for effectively capturing market changes. This shortcoming constrained the models' ability to learn continuous patterns in price movements. On the model level, the linear model suffered from underfitting due to its simple assumptions, while the three models and LSTM network overfitted in the training data due to a lack of necessary regularization constraints [11]. On the experimental design level, the use of same and simple hyperparameters in different models locks the full potential of each model, thereby influencing the models' true predictive capability.

The optimization strategies implemented in this study provided effective solutions. On the data level, by introducing lagged variables and rolling statistics, the static data was transformed into dynamic sequences with temporal information, which provides crucial market dynamics context for the models. On the model level, model-specific hyperparameter tuning played a key role. Through techniques such as introducing regularization and applying structural constraints, model complexity was greatly controlled, allowing them to learn generalizable patterns.

This study confirms that the improvement of prediction performance derives from a systematic optimization process that integrates data, models, and experimental design [12]. This methodology provides a useful framework for addressing complex time-series forecasting problems.

5. Conclusion

This study established a framework for predicting the direction of Apple's stock price, systematically constructing the process from baseline model to optimized models.

Methodologically, the study implemented data preprocessing, exploration data analysis to handle missing values and outliers and chronological split methods. Four models—Logistic Regression,

DOI: 10.54254/2754-1169/2026.LD30724

Random Forest, XGBoost, and LSTM—were selected, followed by an optimization phase incorporating targeted feature engineering and hyperparameter tuning.

The baseline models exhibited significant fitting issues: Logistic Regression suffered from severe underfitting, while the tree-based models and LSTM displayed overfitting, resulting in low accuracy and poor generalization capability. After optimization, all models showed substantial improvement. LSTM and XGBoost emerge as the top performers. The optimization process successfully corrected these fitting problems, thereby enhancing the models' generalization capacity.

These findings reveal the reasons behind the performance differences. The superior performance of LSTM and XGBoost can be attributed to their capabilities in capturing temporal dependencies and complex feature interactions. The optimization process proved to be key to performance enhancement, effectively addressing limitations during the baseline phase. By enriching data representation to provide essential temporal context and controlling each model's complexity through hyperparameter tuning, the process allows the models to learn robust and generalizable patterns.

Stock price direction forecasting depends not only on selecting strong algorithms, but also on an integrated optimization that incorporates data preparation, targeted model adjustments and rigorous experimental design. This framework provides a methodological reference for addressing similar financial time-series forecasting challenges. Future research directions include exploring the integration of additional dynamic data and validating models' generalizability across different market conditions.

References

[1] Bai, M. R. (2023). Macro analysis on Apple stock in order to determine investment value of Apple. Journal of Intelligence and Knowledge Engineering, 1(1).
[2] Cao, E. S. (2024). Predicting Apple stock price based on ARIMA model. In Proceedings of ICFTBA 2024 Workshop: Finance's Role in the Just Transition. Department of Statistics, University of Connecticut.
[3] Horák, J., & Kaisler, D. (2022). Evaluation of development of Apple Inc. stock price time series. SHS Web of Conferences, 132, 01012.
[4] Jiang, H. Q. (2023). The application of the ARIMA-GARCH hybrid model for forecasting the Apple stock price. Journal of Intelligence and Knowledge Engineering, 1(1).
[5] Jin, C. (2024). Forecasting models for Apple Inc. stock price using regression smoothing and Box Jenkins time series analysis. In Proceedings of Science, Technology and Social Development Proceedings Series (Vol. 2).
[6] Kaggle. (2025). Apple Inc. stocks dataset (10 years) [Data set]. Retrieved from https: //www.kaggle.com/datasets/abdulwadood11220/apple-stocks-dataset-10-years
[7] Lee, S. (2021). Feature investigation for stock returns prediction using XGBoost and deep learning sentiment classification (CMC Senior Theses, 2715). Claremont McKenna College.
[8] Li, Z. (2023). Stock market analysis and prediction using LSTM: A case study on technology stocks. Innovations and Applications of Engineering Technology, 2(1), 1-4.
[9] Omer, H. M., Yahya, A., & Mohammed, S. M. (2025). Data analysis approaches for Apple stock price prediction and financial risk management. Danadyaksa Post Modern Economy Journal, 3(1), 32–43.
[10] Wang, Y. (2025). Predicting Apple Inc. stock prices using machine learning techniques. Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025), 455–461.
[11] Yang, Y. X. (n.d.). Forecasting Apple stock closed prices by LR and LSTM with discrete wavelet transformation. Woodsworth College, University of Toronto.
[12] Zhao, Y., & Bai, T. (n.d.). Analysis of Apple stock based on centroid clustering analysis algorithm. Academic Journal of Computing & Information Science, 3(4), 95-101.