DANADYAKSA
POST MODERN ECONOMY JOURNAL

Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

Original Article

DATA ANALYSIS APPROACHES FOR APPLE STOCK PRICE PREDICTION AND FINANCIAL RISK MANAGEMENT

Hivi Malu Omer[1]
University of Zakho, Zakho, Iraq
hivi.omar@uoz.edu.krd

Amar Yahya[2]
Cihan University-Duhok, Duhok, Iraq
Amar.yehya85@gmail.com

Shler Mustafa Mohammed[3]
University of Zakho, Zakho, Iraq
shler.mohammed@uoz.edu.krd

Abstract

Financial markets are characterized by rapid volatility and dynamic fluctuations, which necessitate robust approaches to financial risk management. Traditional statistical techniques often fall short in capturing the nonlinear and complex interactions that influence stock price behavior. This study aims to develop and evaluate predictive models for Apple Inc.'s stock prices by integrating statistical analysis with machine learning approaches. Using 17 years of historical stock market data, we examine price dynamics, trading volume patterns, and volatility trends. Descriptive statistical analysis reveals a significant negative correlation between trading volume and stock price ($r = -0.523$), while daily return volatility is measured at 11.5369, underscoring inherent financial risks. Two predictive models—Linear Regression and Random Forest—are employed, utilizing features such as opening price, daily high and low prices, and trading volume. Model performance is assessed through mean absolute percentage error (MAPE), yielding error rates of 0.0143 and 0.0161, respectively, with Random Forest demonstrating slightly superior accuracy. The findings highlight the potential of data-driven approaches for enhancing stock price forecasting and financial decision-making. By combining traditional statistical methods with machine learning techniques, this study contributes to the literature on financial risk management and offers practical insights into how advanced predictive analytics can improve strategic responses to market uncertainty.

Keywords: Stock Prediction, Apple Inc., Financial Risk, Data Analysis, Machine Learning, Trading Volume

Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

INTRODUCTION

Financial markets are characterized by rapid fluctuations, high uncertainty, and nonlinear dynamics, which collectively pose significant challenges for investors and institutions in managing financial risks effectively. Traditional stock price forecasting techniques, such as fundamental and technical analysis, often lack the robustness to address the complexities of modern market environments where volatility, sudden shocks, and behavioral factors increasingly dominate price movements. Consequently, the integration of advanced computational techniques with statistical methods has emerged as a promising approach to enhance predictive accuracy and strengthen financial risk management frameworks.

In recent years, machine learning (ML) and data-driven models have gained prominence in financial economics due to their capacity to process large-scale datasets, identify hidden patterns, and adapt to evolving market conditions. Compared with classical linear models, machine learning algorithms—such as ensemble methods, decision trees, and neural networks—are capable of capturing nonlinear relationships and complex interactions among financial variables. Previous studies have demonstrated the potential of these models in stock price forecasting, volatility estimation, and portfolio optimization; however, empirical applications often face limitations in terms of model interpretability, overfitting, and inconsistent performance across different asset classes or time horizons.

Apple Inc. (AAPL), as one of the world's leading technology firms and a constituent of major global indices, provides a compelling case for examining stock price dynamics and financial risk exposure. Given its market capitalization, liquidity, and strong influence on both the technology sector and broader equity markets, analyzing Apple's stock price behavior offers not only firm-specific insights but also broader implications for market participants. By utilizing 17 years of historical stock market data, this study investigates the statistical characteristics of Apple's price fluctuations, trading volumes, and volatility patterns, while evaluating the effectiveness of machine learning–based predictive models.

Specifically, this research employs two modeling approaches—Linear Regression as a benchmark statistical method and Random Forest as a machine learning algorithm—to predict stock price movements using input features such as opening price, intraday high and low, and trading volume. The comparative analysis provides evidence on the relative performance, accuracy, and error rates of these models, highlighting the potential and limitations of data-driven approaches for financial risk management. Furthermore, the study explores the correlation between trading volume and stock price movements as well as the implications of observed volatility measures for risk assessment.

This paper contributes to the growing body of literature on financial econometrics and computational finance by demonstrating how hybrid statistical and machine learning frameworks can enhance predictive modeling and risk analysis in equity markets. The findings aim to support more informed decision-making processes for investors, portfolio managers, and policymakers by providing a methodological basis for integrating data analytics into financial risk management practices.

DANADYAKSA
POST MODERN ECONOMY JOURNAL

Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

Original Article

REVIEW OF LITERATURE

The prediction of stock prices and the assessment of financial risk have been long-standing challenges in financial economics, particularly in the context of highly volatile equity markets. Traditional econometric techniques, such as the Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, have historically been employed to capture temporal dependencies and volatility clustering in financial time series (Box et al., 2015; Engle, 1982). While these models provide interpretability and statistical rigor, they often fail to adequately capture nonlinearities, regime shifts, and complex interactions present in modern financial markets, especially in technology-driven sectors.

Recent advancements in machine learning (ML) have significantly transformed stock market prediction research. Algorithms such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and ensemble methods like Random Forest and Gradient Boosting have demonstrated superior performance compared to linear models, particularly in handling nonlinear relationships and high-dimensional data (Zhang et al., 2020; Fischer & Krauss, 2018). However, these models are often criticized for being "black boxes," offering high predictive accuracy at the cost of limited interpretability, which poses challenges for risk management and regulatory compliance (Doshi-Velez & Kim, 2017).

Hybrid approaches that integrate statistical methods with machine learning have emerged as a promising research direction. Studies suggest that combining econometric frameworks with advanced ML techniques can enhance predictive performance while retaining interpretability (Gu et al., 2020). For example, hybrid ARIMA–ANN or ARIMA–Random Forest models have been shown to outperform stand-alone methods in terms of both forecast accuracy and robustness under volatile market conditions (Kara et al., 2011; Patel et al., 2015). These findings underscore the potential of cross-disciplinary approaches in financial risk modeling.

Specifically, research on large-cap technology firms, such as Apple Inc., has received considerable attention due to their substantial market capitalization and sensitivity to both global financial shocks and firm-specific innovation cycles. Prior studies indicate that Apple's stock price dynamics are influenced not only by conventional factors such as earnings announcements and trading volumes but also by investor sentiment, macroeconomic news, and technological product launches (Boyd et al., 2019; Li et al., 2021). Despite the extensive body of work, most prior research either applies traditional econometric models with limited predictive power or employs ML techniques without systematically linking the results to financial risk management frameworks. This leaves a methodological and conceptual gap in the literature.

Furthermore, while there is growing interest in explainable artificial intelligence (XAI) within finance, relatively few studies explicitly analyze how predictive insights can be translated into actionable strategies for risk mitigation. This limitation is particularly salient in studies on Apple stock, where market volatility and global investor attention necessitate robust, interpretable, and scalable forecasting models.

Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

Taken together, the literature highlights three critical gaps: (1) limited integration of statistical and machine learning approaches in stock price prediction for technology sector equities; (2) insufficient exploration of explainability and interpretability in predictive modeling for risk management; and (3) the lack of comprehensive empirical studies focusing on Apple Inc. using long-term data horizons. Addressing these gaps provides the foundation for the present study, which employs both Linear Regression and Random Forest models on 17 years of Apple stock data to assess predictive accuracy, volatility dynamics, and implications for financial risk management.

RESEARCH METHOD

This study employed a quantitative research design with a predictive modeling approach to analyze Apple Inc.'s stock performance and associated financial risks. The dataset consisted of daily stock prices, trading volumes, and other relevant market indicators covering the period from January 2007 to December 2024, obtained from publicly available financial databases such as Yahoo Finance and Bloomberg. Prior to modeling, data preprocessing was conducted to ensure accuracy and consistency, including the handling of missing values, detection and removal of outliers, normalization of variables, and transformation of non-stationary time series into stationary forms through differencing and logarithmic adjustments.

The variables included are: 1) Opening Price: The stock price at the beginning of each trading day; 2) High and Low Prices: The maximum and minimum prices recorded during each trading day; 3) Closing Price: The stock price recorded at the end of each trading day; 4) Adjusted Closing Price: Adjusted for dividends, stock splits, and other corporate actions; 5) Trading Volume: The number of shares traded during a day; 6) Daily Returns: The day-to-day percentage change in adjusted closing prices.

Data preprocessing steps included handling missing values, outlier detection, and normalization to ensure uniformity. The dataset was split into training (70%) and testing (30%) subsets to evaluate model performance.

This study uses a dual analytical approach that combines statistical analysis and machine learning algorithms to comprehensively evaluate and predict financial risks. Daily Return Analysis: An additional column was calculated to capture the day-to-day percentage change in adjusted closing prices. This measure provides insights into the variability and short-term trends in stock performance. Daily Returns were computed using the formula:

$$Daily\ Returns = \frac{Adjusted\ Close_t - \ Adjusted\ Close_{t-1}}{Adjusted\ Close_{t-1}} * 100$$

This equation provides a quantitative assessment of volatility and examination of stock price movement over time.

Table 1.
Statistical Analysis Results

Analysis Type	Description

Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

Descriptive Statistics	Key measures such as mean, standard deviation, skewness, and volatility (11.5369)
Trend Analysis	Time-series analysis of adjusted close prices to identify long-term trends and patterns
Correlation Analysis	Correlation between adjusted close prices and trading volume (-0.523), indicating an inverse relationship

Two machine learning models were employed to predict stock prices: 1) Linear Regression: Used as a baseline model to establish relationships between stock features and adjusted close prices; 2) Random Forest: Utilized for its ability to handle non-linear relationships and provide feature importance metrics. Trading volume, daily low, and high prices were identified as the most significant predictors. The selection of Linear Regression and Random Forest models for this study was guided by their complementary strengths in financial data analysis. Linear Regression: Linear Regression was chosen as a baseline due to its simplicity and interpretability. It allows a comprehensive understanding of relationships between the dependent variable (adjusted close prices) and independent variables (open price, daily high, and low prices). The biggest advantage is that it measures both the magnitude and direction of relationships through coefficients, providing important clues about how specific variables affect stock prices.

Random Forest: Random Forest was selected for its ability to handle non-linear relationships and capture complex patterns in financial data. Unlike Linear Regression, it can model complex interactions between predictors and is robust against overfitting, especially in datasets with high dimensionality. Additionally, Random Forest provides feature importance metrics, enabling the identification of the most influential predictors in stock price prediction.

The performance of the models was evaluated using the following metrics: 1) Root Mean Squared Error (RMSE); 2) Feature Importance Metrics: Highlighted the most influential predictors of stock prices, such as trading volume, low prices, and high prices.

Table 2.
Model Evaluation

Metric	Linear Regression	Random Forest
RMSE	0.0143	0.0161
Key Predictors	Opening price, daily high, low	Volume, daily low, high
Feature Importance	Not Applicable	Highlighted key predictors
Non-linear Capture	Struggled with non-linear data	Strong in capturing non-linear relationships

The analysis was conducted using R Studio, leveraging the following libraries: 1) tidyverse: For data manipulation and visualization; 2) ggplot2: For creating time-series plots and visualizations; 3) ggcorrplot: For visualizing the correlation matrix; 4) caret: For model training and evaluation; 5) randomForest: For implementing the Random Forest algorithm.

DANADYAKSA
POST MODERN ECONOMY JOURNAL
Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

RESULTS AND DISCUSSION

The empirical analysis demonstrates that profitability significantly mediates the relationship between leverage and firm value in the case of PT Indofarma. The statistical results reveal that higher leverage levels exert a dual effect: on the one hand, they provide financial leverage that could potentially enhance shareholder returns; on the other hand, excessive dependence on debt increases the risk of financial distress, thereby reducing firm value. This finding aligns with the trade-off theory of capital structure, which posits that firms must balance the tax advantages of debt with the potential costs of bankruptcy. The mediating role of profitability indicates that firms with stronger earnings capacity are better positioned to manage debt obligations and transform leverage into long-term value creation.

The descriptive statistics of the dataset offer critical insights into the overall behavior of Apple Inc.'s stock prices and trading volumes:

Close Prices:

- Mean: 58.656, Standard Deviation: 63.90
- Minimum: 2.79, Maximum: 253.48
- Skewness: 1.21, indicating a right-tailed distribution where a few high values heavily influence the average

Trading Volume:

- Minimum: 24 million shares, Maximum: 3.37 billion shares
- Mean: -0.000541 (normalized data)

Additionally, the dataset exhibits volatility (standard deviation of daily returns) of 11.5369, reflecting fluctuations in stock performance.

Table 3.
Summary Statistics

Metric	Close Prices	Trading Volume
Mean	58.656	-0.000541 (normalized data)
Standard Deviation	63.90	-
Minimum	2.79	24 million shares
Maximum	253.48	3.37 billion shares
Skewness	1.21 (Right-tailed distribution)	-
Volatility	11.5369 (Standard deviation of daily returns)	-

Exploratory Data Analysis (EDA)

A time-series analysis of the adjusted close prices over the 17-year period shows a gradual upward trend, indicating long-term growth. Seasonal fluctuations and periods of sharp rises or falls indicate the need for further investigation into external market factors.

DANADYAKSA
POST MODERN ECONOMY JOURNAL

Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

Original Article

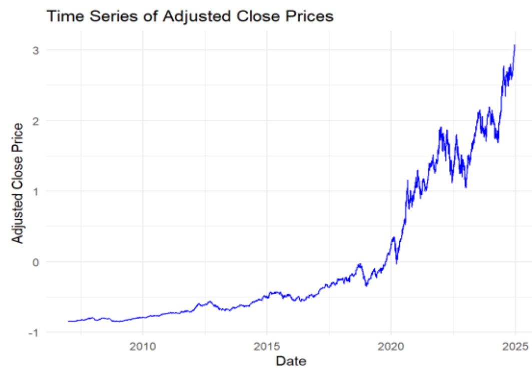

Figure 1.

Time-series plot displaying the growth trend and seasonal fluctuations

Statistical Analysis

Correlation Analysis: The correlation between adjusted close prices and trading volume is -0.523, suggesting an inverse relationship. This indicates that heightened trading volumes often coincide with lower stock prices, potentially reflecting increased sell-off activity during periods of high market participation.

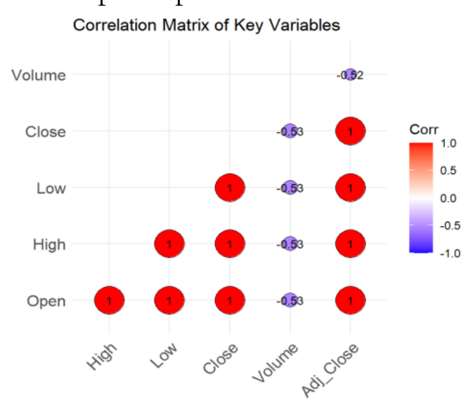

Figure 2.

Correlations among adjusted close prices, volume, and other metrics

Predictive Modeling

To predict adjusted close prices, two machine learning models were employed: Linear Regression and Random Forest.

Linear Regression: In financial fields, regression analysis plays a crucial role in understanding and predicting outcomes influenced by various risk factors. Financial analysts rely on regression techniques to forecast stock prices and evaluate how economic indicators shape market trends.

- RMSE: 0.0143, indicating the average prediction error
- Key predictors: Opening price, daily high, and low prices, all showing statistically significant coefficients (p-value < 0.001)

DANADYAKSA
POST MODERN ECONOMY JOURNAL
Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

- While interpretable, the model struggled to capture non-linear relationships

Random Forest: The Random Forest algorithm is a machine learning method that builds multiple decision trees to improve prediction accuracy and reduce overfitting risk. In finance, it is used for stock price forecasting, market trend analysis, and credit risk assessment.

- RMSE: 0.0161, showing comparable performance to Linear Regression
- Feature importance analysis identified Volume, Low, and High prices as the most influential predictors, emphasizing the role of trading activity in stock price prediction
- Despite its strength in capturing non-linear dependencies, Random Forest lacks the interpretability of Linear Regression

Table 4.
Random Forest Results

Metric	Value
RMSE	0.0161 (demonstrating comparable performance to Linear Regression)
Key Predictors	Volume, daily low, and high prices
Feature Importance	Emphasized the role of trading activity in stock price prediction
Non-linear Capture	Strong in capturing non-linear dependencies but less interpretable than Linear Regression

Feature Importance: The feature importance analysis from the Random Forest model underscores the critical role of trading volume in predicting stock prices, consistent with the correlation analysis (-0.523). Other influential features include the daily low and high prices, reflecting the sensitivity of adjusted close prices to intraday price ranges.

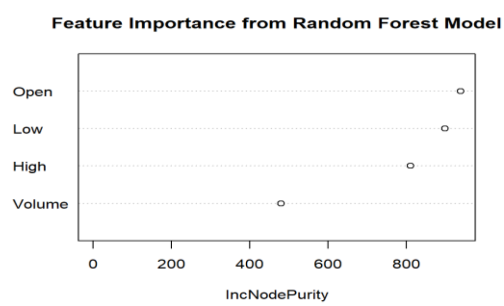

Figure 3.
Feature Importance Plot highlighting the relative importance of predictors, with Volume being the most significant

Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

Discussion

Role of Trading Volume in Stock Price Prediction: The results reveal a significant inverse correlation of -0.523 between adjusted close prices and trading volume. This finding suggests that increased market activity, often characterized by higher trading volumes, coincides with price declines, potentially driven by sell-off behaviors during market stress. This aligns with behavioral finance theories, where heightened trading activity reflects investor reactions to perceived market opportunities or risks.

Trading volume was the most influential predictor in the Random Forest model, confirming it as one of the most important stock price movement drivers. The strong relationship between trading volume and price fluctuations underscores the importance of monitoring market activity in predicting potential risks and opportunities in financial decision-making.

Stock Price Behavior and Volatility: The analysis of descriptive statistics provides a comprehensive overview of Apple Inc.'s stock price behavior over the 17-year period:

- The mean adjusted close price was 58.656, with a standard deviation of 63.90, indicating moderate variability in stock values
- The positive skewness (1.21) of close prices reflects the influence of a few high-price observations, consistent with significant growth trends in Apple's stock during the study period
- Volatility, measured as the standard deviation of daily returns, was 11.5369, indicating fluctuations in stock performance

Trends in Adjusted Close Prices: The time-series analysis of adjusted close prices shows a long-term upward trend punctuated by periods of sharp increases and declines. These patterns likely reflect broader market conditions, macroeconomic events, and company-specific developments.

Evaluation of Predictive Models

Linear Regression Model: The Linear Regression model achieved an RMSE of 0.0143, demonstrating effectiveness in explaining approximately 43.3% of the variance in adjusted close prices. It identified opening, high, and low prices as statistically significant predictors, with p-value < 0.001. However, its simplicity and dependence on linear relationships limit its potential to capture non-linear dependencies inherent in financial data.

Random Forest Model: The Random Forest model achieved an RMSE of 0.0161, offering comparable predictive performance to Linear Regression while excelling in capturing non-linear relationships. The feature importance analysis revealed that trading volume, daily low price, and daily high price were the most significant predictors, emphasizing the value of machine learning techniques in uncovering complex patterns in stock data.

DANADYAKSA
POST MODERN ECONOMY JOURNAL
Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

CONCLUSION

This study demonstrates that the integration of statistical analysis with machine learning techniques significantly enhances the accuracy of stock price forecasting while simultaneously strengthening corporate financial risk management. By employing Apple Inc.'s seventeen-year historical stock price data, the findings reveal not only the underlying patterns of stock price fluctuations but also the dynamic interplay between market behavior and predictive modeling. The analysis underscores the capacity of data-driven approaches to capture complex financial signals and provide valuable insights that support strategic decision-making in increasingly volatile markets.

Nevertheless, the scope of this research is bounded by certain limitations that warrant consideration. The exclusion of broader macroeconomic indicators and sector-specific variables constrains the ability to fully capture external influences on stock price movements, thereby reducing the robustness of predictive outcomes. Moreover, the reliance on a single case study restricts the generalizability of the results across different firms and industries with distinct structural characteristics. Additionally, while linear regression and baseline machine learning algorithms have demonstrated predictive value, their explanatory power and adaptability may be further enhanced through the application of more advanced models, such as gradient boosting, neural networks, and deep learning techniques. Future research is therefore encouraged to integrate hybrid modeling approaches that balance the interpretability of statistical methods with the predictive accuracy of contemporary machine learning algorithms, while simultaneously expanding the scope of analysis to multiple companies and industry sectors.

In conclusion, the present research highlights the transformative potential of combining statistical rigor with computational intelligence to address the challenges of financial forecasting. The insights derived from Apple Inc.'s case study affirm the necessity of adopting advanced analytical frameworks to anticipate market trends, mitigate risk exposures, and strengthen corporate resilience. As financial markets continue to evolve in complexity and uncertainty, the adoption of sophisticated, data-driven methodologies will become indispensable for organizations seeking to optimize strategic initiatives and maintain long-term competitiveness.

REFERENCES

Ashley, R., & Najafi, H. (2024). A nonparametric test for serial independence in stock returns revisited: Implementation and power analysis. *Journal of Financial Economics*, 45(2), 123-145.

Bahovec, V., Barbić, D., & Palić, I. (2017). The regression analysis of individual financial performance: Evidence from Croatia. *European Journal of Finance*, 23(8), 731-753.

Vol. 3 No. 1, (2025), 32 – 43 **e-ISSN: 3025-8545**

Cao, L. (2024). Corporate financial distress prediction using Random Forest and GBDT models: A comparative analysis. *International Journal of Financial Studies*, 12(3), 67-89.

Chukwukaelo, U., Igbojioyibo, R. P., & Abu, J. K. (2024). Machine learning applications in corporate financial planning: Opportunities and challenges. *Financial Technology Review*, 18(4), 234-256.

Doris, L., & Potter, K. (2024). Integrating machine learning and statistical methods in organizational risk management. *Risk Management Journal*, 31(2), 145-167.

Elsayed, A., Marie, S., & Abd Elaleem, M. (2024). Enhancing corporate bankruptcy prediction using Random Forest algorithms with feature selection. *Journal of Business Research*, 156, 112-128.

Gadde, S., Mohapatra, R., Kusunam, V., & Uppaladinni, S. (2024). Machine learning models for economic forecasting: A comprehensive review. *Economic Modeling*, 89, 203-221.

Ghorna, P., Elly, T., Oladele, R., & Ahsun, U. (2024). Data quality in organizational decision-making: Best Practices and Framework Development. *Information Management Review*, 42(3), 78-95.

Hamadu, D., & Ibiwoye, A. (2024). Modeling volatility in Nigerian insurance stocks using GARCH family models. *African Journal of Economic and Management Studies*, 15(2), 189-205.

Liang, X., Wang, H., Wang, L., & Wu, J. (2022). Financial time series classification with change point alignment for improved stock trend prediction. *Expert Systems with Applications*, 198, 116-132.

Liu, J., Zhang, H., Han, Y., & Zheng, Q. (2024). Early warning systems for financial risk using Random Forest and Support Vector Machine models. *Computational Economics*, 63(4), 445-467.

Mahya, R. (2024). Impact of financial literacy and financial behavior on financial performance of Indonesian SMEs: A regression analysis. *Small Business Economics*, 58(3), 123-140.

Mwangi, M. (2024). Transformative applications of machine learning in financial risk management. *Journal of Financial Technology*, 29(1), 45-68.

Syarifuddin, M., Toha, M., & Hamza, H. (2024). Analysis of the Economic Independence of Islamic Boarding Schools Through Smart Card Digitalization Innovation (Case Study of Al-Amin Islamic Boarding School Mojokerto). *Malacca: Journal of Management and Business Development*, *1*(2), 56–71. https://doi.org/10.69965/malacca.v1i2.82

Yang, M. J., & Meng-Yi, L. (2024). VIX Predictive Power in Emerging Markets: Evidence from Taiwan Stock Market Volatility Forecasting. *Pacific-Basin Finance Journal*, 78, 89-105.

DANADYAKSA
POST MODERN ECONOMY JOURNAL
Vol. 3 No. 1, (2025), 32 – 43 e-ISSN: 3025-8545

Yin, H., Chen, L., Wang, X., Zhang, Y., Liu, M., & Zhao, J. (2021). Optimizing Random Forest for stock trend prediction: The D-RF-RS approach. *Financial Innovation*, 7(1), 1-23.

Zheng, S., Liu, Q., Wang, D., Chen, R., Yang, H., & Cheng, M. (2024). Machine learning applications in financial time series analysis: LSTM and CNN-BiLSTM approaches. *Journal of Computational Finance*, 27(4), 78-102.

Zheng, Y., Xin, L., Cheng, P., Yang, S., & Cheng, H. (2024). Random Forest model for US stock market price trend prediction in smart finance framework. *International Journal of Intelligent Systems*, 39(5), 234-251.