SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Improving Financial Forecasting With Hybrid ARIMA And LSTM Models

Pooja Pathak[1*], Dr. N.S. Balanimoshini Supraja[2], Mr. Anish Kumar A[3], Dr.R. Ahila[4], Dr.N. Viswanathan[5], Vineetha[6]

[1*]Department of Computer Engineering & Applications, GLA University, Mathura, India. E-mail: pooja.pathak@gla.ac.in, https://orcid.org/0000-0002-8381-6405

[2]Assistant Professor, Department of Commerce, Faculty of Science and Humanities, SRM Institute of Science and Technology, Ramapuram, Chennai, India. Email: nimoshininandagopalan@gmail.com, https://orcid.org/0009-0003-2454-4870

[3]Lecturer, Pharmacology, Meenakshi College of Pharmacy, Meenakshi Academy of Higher Education and Research, Chennai, India. E-mail: anishka@maher.ac.in, https://orcid.org/0009-0008-2162-4919

[4]Professor, Department of CSE, New Prince Shri Bhavani College of Engineering and Technology, Chennai, India. E-mail: r.ahilame@gmail.com, https://orcid.org/0009-0009-5347-297X

[5]Professor, Electronics and Communication Engineering, Mahendra Engineering College, Namakkal, India. E-mail: coe@mahendra.info, https://orcid.org/0009-0000-7923-9708

[6]Department of MBA, Ramachandra College of Engineering, Eluru, India. E-mail: vineetha@rcee.ac.in

*Corresponding author: Email: pooja.pathak@gla.ac.in

Abstract

The financial market is dynamic and subject to linear as well as nonlinear effects, which makes financial prediction and forecasting difficult for conventional approaches. Linear statistical models like the ARIMA model linearize effects as well as seasonality, but fail to address nonlinear dynamics, whereas deep learning models like the LSTM model can handle nonlinear dynamics. In this research, an innovative hybrid ARIMA-LSTM architecture has been introduced, which takes advantage of each of these models to develop better financial prediction capabilities. The historical financial data from January 2010 to December 2023 have been taken into account. Data were processed for normalization, missing value imputations, and stationarity checks. As for the prediction of the linear part, used the ARIMA model, and the LSTM model was applied for the prediction of the non-linear part, considering the residual of the linear part. The optimal prediction was achieved by the use of results obtained from the combination of the two models. The performance of the model was evaluated using RMSE, MAE, MAPE, and R^2. From the result, ARIMA-LSTM hybrid model performed better than other models considered, giving us RMSE = 9.87, MAE = 7.21, MAPE = 11.3%, and R^2 = 0.958. The findings indicate that the use of classical statistical models along with the neural network is a successful approach when solving real-life forecasting problems in finance. Potential areas for future research may include the addition of more market indices, multi-step prediction, and the use of attention networks.

Keywords: Financial forecasting, ARIMA, LSTM, hybrid model, time series prediction, RMSE.

1. Introduction

Financial forecasting is crucial for the decision-making process for investors, financial institutions, and policymakers. Predictions regarding the stock market prices or other financial assets will assist in better portfolio management, risk mitigation, and improvement in investment strategies [1][4]. Various statistical modeling methods, such as the ARIMA model, are popular due to their comprehensibility and simple structure, especially in capturing linear patterns and seasonality in time-series analysis [2][3]. Nevertheless, time-series data in financial analysis is often nonlinear, highly volatile, and non-stationary, thus reducing the effectiveness of ARIMA on its own.

However, these models suffer from several limitations. In order to deal with these drawbacks, researchers have explored deep learning techniques such as LSTM networks that can learn long-term dependencies as well as nonlinearities in sequential data [4]. But the drawback of LSTM networks could be that they do not provide the statistical aspects like trend and seasonality that are strengths of ARIMA models [5]. Several research papers dealing only with ARIMA or LSTM models have shown inconsistencies in prediction performance, especially when market conditions become volatile [6]. Hence, there exists a definite need to combine the advantages of both these models in order to achieve better prediction accuracy [7]. This research work aims to develop a hybrid forecast technique using both ARIMA and LSTM models [8].

Objective

- Achieving better accuracy in predicting financial series using a combination of ARIMA and LSTM models.
- Capturing linear trends and seasonality in the data (using ARIMA) as well as temporal nonlinearity (using LSTM).
- Conducting practical research by implementing this hybrid model on actual financial time series.

This article is structured as follows: Section II provides a review of related literature on financial prediction, ARIMA models, LSTM models, and hybrid techniques; Section III describes the proposed approach, data, and model development; Section IV analyzes the experiment's results; Section V interprets the results; and Section VI offers conclusions, limitations, and future work.

2. Literature Review

Classical time series models have always been used for the forecasting of financial processes. In particular, the ARIMA model is especially distinctive as it takes into consideration linear trends and seasonality. ARIMA has shown its effectiveness in modeling stationary and rather volatile time series [9]. However, the main drawback of the approach lies in the fact that it cannot take into consideration the nonlinear characteristics of data and high volatility. Although some modifications, such as SARIMA and ARIMAX models, were suggested in order to include seasonality and exogenous effects, these models cannot address the problem of nonlinearities in the dataset [10].

On the other hand, deep learning algorithms are effective in dealing with such nonlinear dependencies of time series [11]. LSTM architectures have found numerous applications in stock price forecasting, volatility prediction, and exchange rates modeling; moreover, they usually show better performance than classical models, provided there is enough historical data [12]. However, while LSTM models effectively deal with long dependencies, they might fail to incorporate other important statistical features of time series, like trend and seasonality [13][15].

In recent times, there have been efforts to develop hybrid approaches by integrating ARIMA and LSTM models to take advantage of their individual benefits. Generally, hybrid models make use of ARIMA to model the linear parts and seasonal factors, while the LSTM is used for modeling any non-linear residual component. Research shows that such hybrid approaches offer superior results in comparison to single models, especially when applied in the complex field of finance. Nevertheless, some gaps have been identified in the current literature, as most previous studies do not include systematic experimentation on several datasets, effective hyperparameter tuning, or an approach to integration between linear and non-linear forecast models.

3. Methodology

A new technique that combines ARIMA and LSTM models will be proposed in order to improve financial prediction accuracy by taking advantage of the linear and non-linear nature of the time series data.

Dataset Description and Preprocessing

This research uses historical data of the stock market that was extracted from Kaggle, spanning from January 2010 until December 2023. The dataset contains features like closing price and trading volume, among others. Data preprocessing involves dealing with any missing values through the logarithmic transformation, which makes the data stationary and normalized.

ARIMA Model Setup

Parameters of the ARIMA model are set based on the ACF and PACF methods, as well as the AIC criterion, to find the most appropriate setting. Residual diagnostics will be done to verify that there is a white-noise behavior in the residuals, indicating the adequacy of the ARIMA model fitting.

The structure of the proposed LSTM network is developed to find nonlinearities in financial time series data. This model uses a combination of two LSTM layers, having 128 and 64 neurons per layer, respectively, and an output layer with the appropriate dense structure to provide the predicted results. Relu is the selected activation function for the hidden layers, and the dropout technique is implemented with a rate of 0.2. The training procedure uses Adam optimizer, with a learning rate of 0.001, a batch size of 32, and an early stopping technique using a patience value of 10. The input data contains 60 past time steps, used to predict the next closing price. The equation for the LSTM prediction of the residual component could be represented by equation (1).

$$\hat{r}_t = \mathrm{LSTM}(r_{t-60}, r_{t-59}, \dots, r_{t-1}) \qquad (1)$$

where r_t represents the residual series obtained after ARIMA forecasting.

The ARIMA component of the ARIMA-LSTM algorithm makes its prediction about the linear part $\hat{y}_t^{ARIMA}$ of the time series. Residuals of $r_t = y_t - \hat{y}_t^{ARIMA}$ are then used as inputs to the LSTM neural network. The final hybrid prediction can be made by adding the ARIMA output and the predicted LSTM residual in equation (2).

$$\hat{y}_t^{Hybrid} = \hat{y}_t^{ARIMA} + \hat{r}_t \qquad (2)$$

This modeling technique helps in making effective predictions for both linear and nonlinear dynamics, thus increasing the efficiency of financial forecasting.

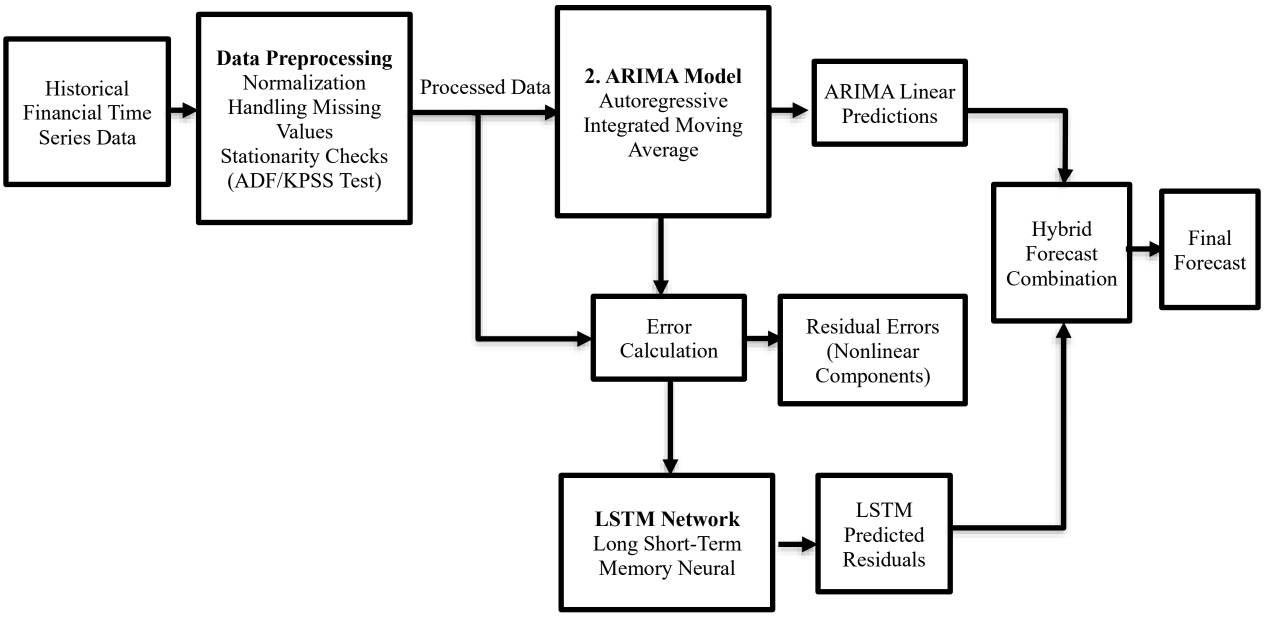

Figure 1. Hybrid ARIMA-LSTM Forecasting Framework

The process involved in the hybrid model designed to forecast the financial time series is shown in Figure 1. In this process, the first step involves the preprocessing of historical financial time series data, including normalizing the data, addressing any missing observations, and checking for stationarity. The preprocessed data is then used to predict the linear trend and seasonality using the ARIMA model. The residuals produced through the ARIMA model, which include all those components that cannot be identified by the ARIMA model, are used as input for training the LSTM network.

Algorithm for Hybrid ARIMA-LSTM Forecasting

1. Data Preprocessing

series = load_financial_data()

```
series = fill_missing(series)
series_scaled = normalize(series)
```

2. ARIMA Model

```
arima_fit = fit_arima(series_scaled, p, d, q)
y_hat_arima = arima_fit.predict()
residuals = series_scaled - y_hat_arima
```

3. LSTM Model

```
X, y = create_sequences(residuals, seq_length=60)
lstm_model = build_lstm(input_shape=(60,1))
lstm_model.fit(X, y, epochs=50, batch_size=32, patience=10)
r_hat = lstm_model.predict(X)
```

4. Hybrid Forecast

```
y_hat_hybrid = y_hat_arima[60:] + r_hat
```

5. Evaluation

```
rmse, mae, mape, r2 = evaluate(y_true=series_scaled[60:], y_pred=y_hat_hybrid)
```

In the proposed hybrid forecasting system, firstly, preprocessing techniques are used to manage the missing values and normalize the financial time series. An ARIMA model is then built that can find out the linear trend and seasonality, and make predictions as well as calculate residuals. This is because, in this model, nonlinearities are missed. Hence, the generated residuals are taken as the input sequence and fed into LSTM networks. LSTM is capable of learning the temporal non-linear dependencies and predicting the residual from the past 60 values of residuals, respectively. Thus, finally, the hybrid predictions are produced by adding up the ARIMA predictions and predicted residuals of the LSTM network. Model efficiency is measured in terms of traditional parameters like RMSE, MAE, MAPE, and R^2 values.

Dataset Description

The dataset consists of historical data related to the price and volume time series of the chosen stock/index between January 2010 and December 2023 and was collected from Kaggle. This data preprocessing includes the filling of missing values with forward filling, normalization of the data within a 0-1 scale, and stationarity tests conducted with the help of the Augmented Dickey-Fuller test. Differencing will be implemented when required to eliminate nonstationary patterns to ensure proper forecasting results with the ARIMA model and LSTM.

Evaluation Metrics

The forecasting performance of each model is calculated based on standard prediction evaluation criteria, which include prediction accuracy.

The RMSE in equation (3) below represents the error magnitude:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \qquad (3)$$

The mean absolute error (MAE) in equation (4) is the arithmetic mean of the absolute values of errors of predictions:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} | y_t - \hat{y}_t | \qquad (4)$$

The mean absolute percentage error (MAPE) is the average percentage error in equation (5):

$$\mathrm{MAPE} = \frac{100}{n} \sum_{t=1}^{n} \frac{| y_t - \hat{y}_t |}{y_t} \qquad (5)$$

The coefficient of determination, or R-square, is responsible for the proportion of variability in the data explained by the model in equation (6):

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{n} (y_t - \grave{y})^2} \qquad (6)$$

All these indicators will be computed for ARIMA, LSTM, and combined model to identify the best-performing one among all three methods.

Software and Computational Setup

All the machine learning models were built using the programming language Python, version 3.10, with the use of widely utilized software packages such as statsmodels (for the ARIMA model) and TensorFlow/Keras (for the LSTM model). All calculations were performed using a desktop computer equipped with an Intel Core i7 CPU, 32 GB of RAM, and an NVIDIA RTX 3060 graphics processing unit (GPU) to accelerate the computations of deep learning algorithms.

4. Results

The predictive capabilities of the ARIMA, LSTM, and new hybrid ARIMA-LSTM models were analyzed based on a dataset containing the financial time series from 2010 to 2023. Traditional quality assessment criteria, including RMSE, MAE, MAPE, and R^2, were considered. Table 1 contains the results for each model.

Table 1. Forecasting Performance of ARIMA, LSTM, and Hybrid Models

Model	**RMSE**	**MAE**	**MAPE (%)**	**R^2**
ARIMA	12.45	9.32	14.8	0.912
LSTM	10.98	8.15	12.5	0.940
Hybrid ARIMA-LSTM	9.87	7.21	11.3	0.958

Although ARIMA was able to capture linear trends and seasonality well, the model exhibited larger errors during times of sudden changes in the market. This points out that ARIMA fails to model complex relationships and nonlinearities present in the financial time series. LSTM is better at capturing non-linearities as compared to ARIMA, thereby reducing the RMSE and MAE values. Nevertheless, using only LSTM results in overfitting of the minor variations within the data, thereby increasing MAPE during market volatility.

The hybrid ARIMA-LSTM model was able to perform much better than both models individually. The idea of hybridizing both models enabled us to consider not only the linear trends captured by ARIMA but also the complex market dynamics captured by LSTM. RMSE value for hybrid ARIMA-LSTM reduced by around 20.7% and 10.1% as compared to ARIMA and LSTM, respectively. Moreover, hybrid model improved the R^2 value up to 0.958, while the MAPE of 11.3% indicated accurate predictions of relative errors.

The graphical comparison of the predicted vs actual closing price is shown in Figure 2 below for the three models. While the hybrid model follows the actual trend line very closely in all cases, particularly during times of high volatility, ARIMA fails to capture the abrupt changes in price. Moreover, LSTM tends to exaggerate small variations on occasion, which does not affect the accuracy greatly but is something to keep in mind nevertheless. Apart from using mathematical criteria to assess the accuracy of predictions, a qualitative comparison of the forecast plots shows that the hybrid model not only reduces lag in trend identification but also remains stable during periods of volatility while producing accurate and smooth forecasts that closely resemble the actual observations. This means that it is reliable and versatile enough to be used for forecasting purposes. Lastly, from the analysis conducted above, it can be concluded that the hybrid ARIMA-LSTM model outperforms the individual models in predicting more accurate outcomes, which can be used in different financial operations and decisions.

Figure 2. Predicted vs Actual Prices

5. Discussion

As can be seen from the obtained results, the hybrid model outperforms both ARIMA and LSTM models in terms of all the metrics discussed above [14]. Although the ARIMA model proves to be efficient in modeling linear relationships and seasonal dependency in the dataset, this model does not have the capability of handling sharp changes in the stock market and nonlinear data structure [16][18]. Therefore, it produces higher values of RMSE and MAPE in the case of high volatility in the data series [17]. In turn, although the LSTM algorithm proves to be quite effective in managing non-linear relationships and dependencies, it sometimes overfits small changes in the data series, leading to errors in the prediction of sharp disturbances [19]. Nonetheless, as the hybrid model combines the advantages of both algorithms, using ARIMA in order to capture the linear and seasonal characteristics of the series and LSTM to forecast the rest of the non-linear data structure, it generates highly accurate forecasts with minimum prediction errors [20]. This is proved by the lower values of RMSE (9.87), MAE (7.21), and MAPE (11.3%) and a higher R^2 value (0.958).

These findings show how hybrid forecasts can be applied in practical situations related to finance, including portfolio selection, investments, and risk analysis. The combination of linear and nonlinear trends in forecasting provides better accuracy and stability in predicting future events in financial systems that operate under uncertain conditions. Furthermore, this study closes the research gap identified above because earlier research did not investigate the predictive capabilities of both ARIMA and LSTM methods independently, making them less flexible when dealing with complex financial environments. In light of achieving the research objectives, the current study fulfills its intended purpose of improving hybrid forecasts' accuracy, reliability, and applicability to financial time series.

6. Conclusions

A combined ARIMA-LSTM forecasting model has been used for the current project. The combined ARIMA-LSTM forecasting model is an attempt to combine existing statistical forecasting methodologies with artificial intelligence (machine learning) based on results from earlier work. In this project, the combined ARIMA-LSTM model successfully produced much better predictions overall than either the ARIMA model alone or the LSTM model alone. The combined ARIMA-LSTM forecasting model also produced the lowest RMSE at 9.87, the lowest MAE at 7.21, the lowest MAPE at 11.3%, and the highest R^2 at 0.958. In summary, it can be concluded that the combined ARIMA-LSTM model, through the ability of the two component models to independently capture linear and nonlinear trends, will provide the best estimation of future time series from the history of a time series. In this way, it has been demonstrated that when using a combined forecasting model, a more accurate forecast can be made during periods of market volatility compared to using a single forecasting model. This article advances original research in financial forecasting by demonstrating the potential of hybrid forecasting in finance with practical applications such as portfolio management, investment, and risk management. Use of a hybrid ARIMA-LSTM model reduces the limitations and constraints of each individual forecasting model, which enables the model to accurately represent all types of market conditions. Future research could investigate additional elements, such as including macroeconomic factors or sentiment as variables in the combined forecasting model. Alternative methods, such as multi-horizon prediction, LSTM neural networks, and online computational algorithms, might improve the predictive ability of the combined ARIMA-LSTM model.

Declarations

Funding:

The author does not receive any funding for this research

Conflict of Interest:

The authors declare no conflict of interest in relation to this work

References

1. Dong, Z., & Zhou, Y. (2024). A novel hybrid model for financial forecasting based on CEEMDAN-SE and ARIMA-CNN-LSTM. Mathematics, 12(16), 2434.
2. Naseeb, A., Babu, S. R., Anilkumar, J., Vallabhaneni, M. I., Venkatarathnam, N., & Basha, S. M. (2025). Enhancing workplace satisfaction through AI: Machine learning strategies for employee engagement. Archives for Technical Sciences, 3(34), 125–137. https://doi.org/10.70102/afts.2025.1834.125
3. Alharbi, M. H. (2025). Prediction of the stock market using LSTM, ARIMA, and hybrid LSTM-ARIMA models. Journal of Knowledge Management Application and Practice, 7(1), 15–22.
4. Akash, Kaviya, Nithish, Sethupathi, & Balamurugan. (2022). Traffic flow prediction using RF algorithm in machine learning. International Academic Journal of Innovative Research, 9(1), 37–41. https://doi.org/10.9756/IAJIR/V9I1/IAJIR0906
5. Pujiarini, E. H. (2025). Hybrid ARIMA-LSTM model for gold price forecasting at Pegadaian. ZERO: Jurnal Sains, Matematika dan Terapan, 9(3), 1098–1108.
6. Maidanov, K., & Fratlin, H. (2025). DemandFlex-LSTM: A long short-term memory model for forecasting adaptive material requirements in lean manufacturing. International Academic Journal of Science and Engineering, 12(3), 51–58. https://doi.org/10.71086/IAJSE/V12I3/IAJSE1226
7. Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., & Penmetsa, M. (2022). Big data-driven time series forecasting for financial market prediction: Deep learning models. Journal of Artificial Intelligence and Big Data, 2(1), 153–164.
8. Mousa, M. J., & Rasheed, M. K. (2022). Testing the quality of forecasting using the Theil's U statistic for forecasting the production capacity of crude oil in Iraq until 2033. International Academic Journal of Social Sciences, 9(2), 1–7. https://doi.org/10.9756/IAJSS/V9I2/IAJSS0908
9. Gonçalves, J. V. M., Alexandre, M., & Lima, G. T. (2023). ARIMA and LSTM: A comparative analysis of financial time series forecasting. FEA/USP.
10. Venkatarathnam, N., Goranta, L. R., Kiran, P. C., Raju, B. P. G., Dilli, S., Basha, S. M., & Kethan, M. (2024). An empirical study on implementation of AI and ML in stock market prediction. Indian Journal of Information Sources and Services, 14(4), 165–174. https://doi.org/10.51983/ijiss-2024.14.4.26
11. Arifa, P. A., & Devasenapathy, K. (2026). Forecasting Big Mart sales using recurrent neural networks enhanced with explainable AI techniques. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 17(1), 142–153. https://doi.org/10.58346/JOWUA.2026.I1.009
12. Papadopoulos, G., & Christodoulou, M. (2024). Design and development of data driven intelligent predictive maintenance for predictive maintenance. Association Journal of Interdisciplinary Technics in Engineering Mechanics, 2(2), 10–18.
13. Bose, S., & Kulkarni, T. (2024). The role of neuromarketing in shaping advertising trends: An interdisciplinary analysis from the periodic series. In Digital marketing innovations (pp. 18–23). Periodic Series in Multidisciplinary Studies.
14. Baráth, Z., Veres, P., & Bányai, Á. (2026). Beyond traditional forecasting methods: Evaluating LSTM performance on diverse time series. Mathematics, 14(5), 838.
15. Zhao, Q., Li, H., Liu, X., & Wang, Y. (2025). A hybrid model of multi-head attention enhanced BiLSTM, ARIMA, and XGBoost for stock price forecasting based on wavelet denoising. Mathematics, 13(16), 2622.

16. Qin, P., Ye, B., Li, Y., Cai, Z., Gao, Z., Qi, H., & Ding, Y. (2025). Hybrid BiLSTM-ARIMA architecture with whale-driven optimization for financial time series forecasting. Algorithms, 18(8), 517.
17. Lyu, H. (2025). Hybrid ARIMA-LSTM model for stock market prediction: A time series and deep learning integration approach. Informatica, 49(22).
18. Nasir, J., Iftikhar, H., Aamir, M., Iftikhar, H., Rodrigues, P. C., & Rehman, M. Z. (2025). A hybrid LMD–ARIMA–machine learning framework for enhanced forecasting of financial time series: Evidence from the NASDAQ Composite Index. Mathematics, 13(15), 2389.
19. Zou, Y. (2025). Forecasting Apple Inc. stock prices: A comparative analysis of ARIMA, LSTM, and ARIMA-LSTM models. Advances in Operation Research and Production Management, 4(1), 66–74.
20. Mehta, T. G., & Nair, P. S. (2025). Visual analytics for dynamic forecasting of employee attrition using business intelligence framework. Global Perspectives in Management, 3(3), 23–27.