P-ISSN : 2715-2448 | E-ISSN : 2715-7199
Vol.7 No.1 January 2026
Buana Information Technology and Computer Sciences (BIT and CS)

Arabica Coffee Sales Forecasting Using ARIMA Neural Network (Case Study: KOTEM Bondowoso)

Muhammad 'Ariful Furqon[1]*, Saudi Efendi[2], Yanuar Nurdiansyah[3]
[1]Informatics Department, Universitas Jember
[2,3]Information Technology Department, Universitas Jember
E-mail: ariful.furqon@unej.ac.id[1]*, 182410102022@mail.unej.ac.id[2], yanuar_pssi@unej.ac.id[3]

***Received:** 2025-05-01| **Revised:** 2026-01-21 | **Accepted:** 2026-01-30*

Abstract

Kopi Tembakau (KOTEM) Bondowoso is a local coffee producer specializing in Arabica ground coffee, sourced directly from nearby farmer cooperatives. A major hurdle they face is accurately forecasting sales, a critical factor for optimizing production and inventory. To tackle this, a hybrid forecasting model blending ARIMA (for linear/seasonal trends) and Neural Networks (for non-linear patterns) was developed. The study analyzed KOTEM's sales data from September 2019 to August 2022, preprocessed to address non-stationarity via differencing and normalization. Results revealed the hybrid model outperformed standalone ARIMA, achieving a 1.0% MAPE (vs. ARIMA's 1.3%). It also better captured sales volatility and seasonal shifts, offering more dependable forecasts. ARIMA-NN could significantly enhance KOTEM production scheduling and stock management.

Keywords: ARIMA, Coffee sales, Forecasting, Neural Network, MAPE.

I. Introduction

The business world is growing rapidly, requiring companies to analyze their business environment and carefully anticipate potential future developments [1]. One such industry is coffee, a valuable plantation product that plays an important role in the economy and serves as a significant source of foreign exchange due to Indonesia's position as the fourth largest coffee producer worldwide [2]. Coffee is also known for its health benefits, such as lowering the risk of diabetes, boosting stamina, relieving headaches, and improving respiratory function [3], [4]. Bondowoso Regency in East Java has great potential for coffee production since nearly 30% of its plantations are dedicated to coffee cultivation [5]. Among these varieties, Arabica coffee stands out with its distinctive flavor favored by many coffee lovers.

The growth of this sector has led to an increase in companies supplying coffee products. This has intensified competition within the business world, especially in the café sector. The coffee business attracts many entrepreneurs because it is considered both essential and profitable. However, fierce competition makes it challenging for companies to adjust their sales levels effectively. A successful company is one that can align sales with production and consumer demand so that products sell smoothly without obstacles [6]. Problems often arise when sales forecasts are inaccurate, leading to production errors such as overproduction or underproduction. This results in unsold inventory or excess stock, which can cause financial losses, particularly with coffee products, where excess stock cannot be reused due to sensitivity in aroma and taste changes over time. Such issues frequently occur at companies like Kopi Tembakau (KOTEM) Bondowoso.

KOTEM, a business owned by Mr. Bambang Suwito and established in 2012 in Kupang Village, Pakem District, Bondowoso Regency, focuses on producing high-quality Arabica coffee powder. The company benefits from its strategic location near local coffee plantations, allowing direct access to fresh green coffee beans. KOTEM independently manages the entire production chain—from selecting and drying green beans to roasting with ovens and grinding into powder using specialized equipment. Despite having a well-controlled production process, KOTEM faces significant challenges in accurately forecasting sales demand. This forecasting is critical because errors in predicting sales volumes can lead to either overproduction, resulting in excess inventory and increased holding costs, or underproduction, causing stockouts and lost sales opportunities [7], [8]. Both scenarios negatively impact operational efficiency and profitability.

Technically, forecasting sales involves analyzing historical sales data to identify patterns and trends that can predict future demand [8], [9], [10]. Traditional methods like Auto Regressive Integrated Moving Average (ARIMA) are widely used for time series forecasting due to their strong statistical foundation in modeling linear patterns [11]. However, ARIMA has limitations in capturing nonlinear relationships and complex patterns in data [12]. On the other hand, Neural Networks excel at learning nonlinear dependencies but may struggle with seasonality and trend components without proper preprocessing [13]. To overcome these individual limitations, this study applies a hybrid forecasting approach combining ARIMA and Neural Networks (Hybrid ARIMA-NN). This method leverages ARIMA's strength in modeling linear components and Neural Networks' ability to capture nonlinear patterns, aiming to improve forecast accuracy [12].

Despite the growing use of hybrid models in various industries, there remains a research gap in applying such advanced forecasting techniques specifically to the Arabica coffee powder market in regions like Bondowoso. Most existing studies focus on broader agricultural products or different coffee varieties, leaving a need for localized, product-specific forecasting models that consider unique market dynamics and production constraints. This research addresses this gap by developing and testing a Hybrid ARIMA-NN model tailored to KOTEM's sales data, aiming to provide more reliable sales forecasts that support better production planning and inventory management.

II. Method

The main objective of this research is to develop an accurate sales forecasting model for Arabica coffee powder using a hybrid ARIMA-Neural Network. This study was carried out in several steps, as shown in Figure 1.

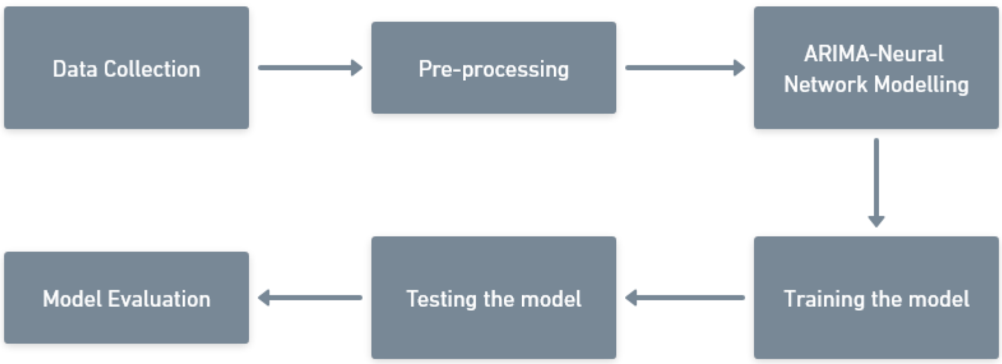


Figure 1. Arabica Coffee Sales Forecasting Methods

The research process begins with a comprehensive collection of data from KOTEM Bondowoso, ensuring that the dataset accurately captures historical sales trends as well as seasonal fluctuations. Once gathered, the data undergoes careful cleaning to address missing entries, inconsistencies, and any anomalies, thereby maintaining its quality and reliability [14]. To prepare the data for analysis, techniques such as differencing are applied to remove trends and stabilize the series, making it stationary and suitable for modelling [15]. Additionally, normalization is performed to scale the data uniformly, which is particularly important for the Neural Network to learn effectively without bias caused by varying data ranges. The hybrid model combines the strengths of ARIMA, which excels at modeling linear trends and seasonal patterns in time series data, with the Neural Network's ability to detect complex, nonlinear relationships that may exist within the sales figures [13]. This integrated approach aims to produce forecasts that are more accurate than those generated by either method alone. After training the model on historical data, its predictive capability is tested using new, unseen data. The model's accuracy is evaluated using the Mean Absolute Percentage Error (MAPE), a metric that expresses the average prediction error as a percentage, providing an intuitive measure of how close the forecasts are to actual sales [16].

III. Results and Discussion

This The data collection process involved conducting interviews to gather relevant information for this study. During these interviews, sales data of Arabica coffee powder were obtained. The collected sales data from KOTEM Bondowoso, covering the period from September 2019 to August 2022, are presented in Figure 2.

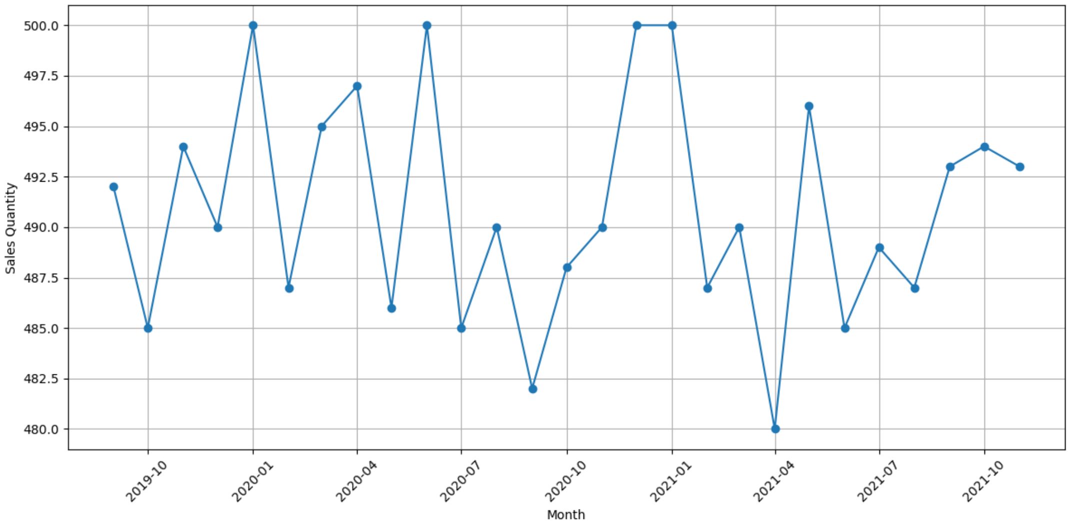


Figure 2. Sales of Arabica coffee (Sep 2019- Nov 2021)

Figure 2 shows that the data appears to be stationary. However, to confirm stationarity, tests must be conducted on both the variance and the mean of the data. The stationarity in variance can be assessed using the Box-Cox transformation plot [17]. The stationarity in mean can be evaluated through methods such as the Box-Cox test, as well as Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots [18].

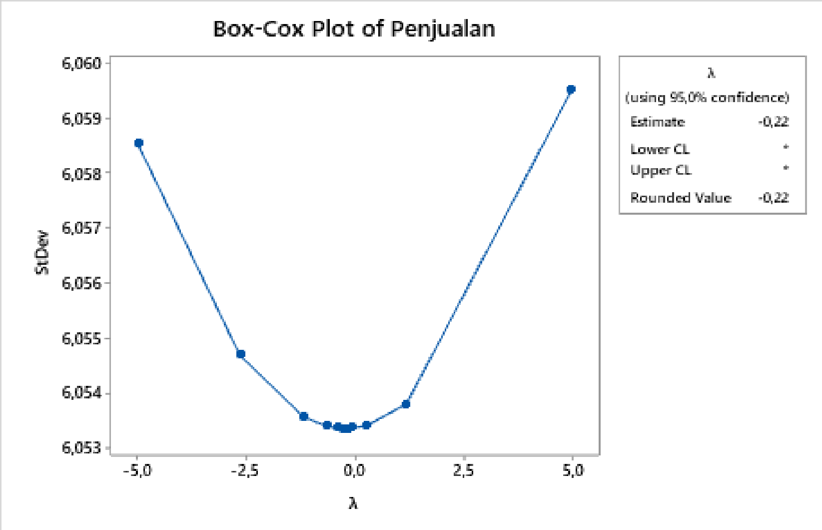
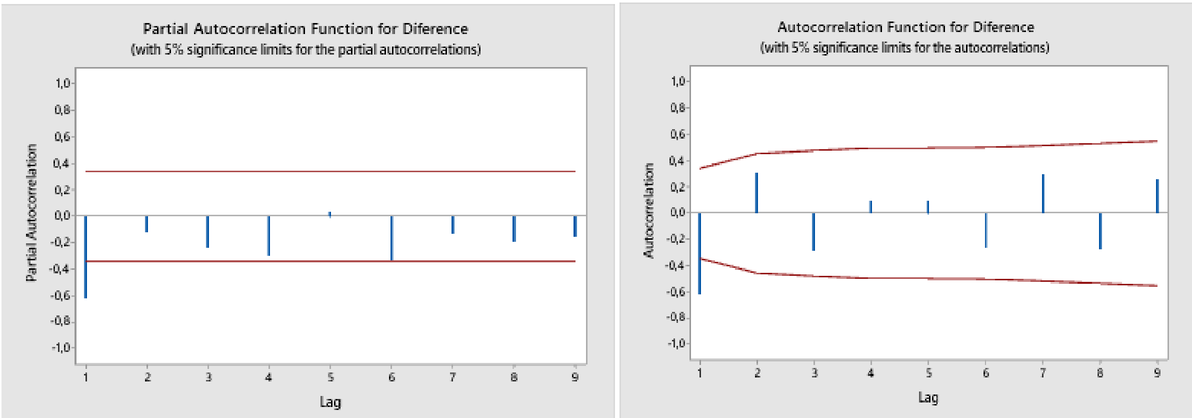


Figure 3. Stationary test of data through the Box-Cox plot



(a) (b)

Figure 4. Stationary data test through (a) ACF and (b) PACF Plots

The Box-Cox plot test is performed to determine the rounded value of the data. In Figure 3, the rounded value is -0.22, indicating that the data is not yet stationary. Data is considered stationary if the rounded value equals 1, which means differencing will be required to achieve stationarity [17]. In the subsequent stationarity test using ACF and PACF plots as shown in Figure 4, it is observed that only lag one falls outside the confidence bounds. Since only one lag exceeds the limit, the data can be considered stationary with respect to the mean. According to this test, data is deemed stationary if no more than three lags fall outside the bounds. However, differencing is still necessary because, based on the Box-Cox test results, the data does not yet meet the criteria for stationarity [18]. Differencing is a process used to transform a non-stationary time series into a stationary one [19]. After applying differencing, the Box-Cox test results show that the rounded value has reached 1, indicating that the data is now stationary in terms of variance.

After the data becomes stationary, the next step is model estimation or selecting potential ARIMA models. From this process, several candidate ARIMA models were identified, including (0,1,1), (1,1,0), (1,1,1), (1,1,2), and (2,1,1). Among these candidates, the best-fitting ARIMA model was determined to be the (2,1,1) model with a p-value of 0 based on Table 1. The optimal ARIMA model, (2,1,1), was used for forecasting, producing the predictions shown in Table 2.

Table 1. ARIMA final model's parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-1,620	0,207	-7,81	0,000
AR 2	-0,701	0,156	-4.48	0,000
MA 1	-0,945	0,205	-4,61	0,000

Table 2. Arabica coffee powder prediction using ARIMA

No	Time	Actual Data	Prediction	Residual ARIMA
1	September 2019	492	487,422	-4,086404240
2	October 2019	485	494,984	-3,000871600
3	November 2019	494	488,249	-1,202103680
4	December 2019	490	494,052	1,002318439
5	January 2020	500	489,575	6,215606579
…	…	…	…	…
…	…	…	…	…
35	July 2022	488	493,325	-1,774414620
36	August 2022	495	493,376	5,279546218

Table 2 shows the predictions of Arabica coffee powder prices using the ARIMA model from September 2019 to August 2022. The predicted values are quite close to the actual prices, with small differences (residuals) between them. This suggests that the ARIMA model does a good job of capturing the main trends and linear patterns in the data. However, the residuals, both positive and negative, indicate that the ARIMA model doesn't fully explain all the fluctuations in the data. These leftover differences might be due to more complex, non-linear patterns that ARIMA alone can't capture. To address this, the residuals from the ARIMA model can be further analyzed using a Neural Network. This hybrid approach aims to improve the accuracy by modeling the non-linear parts that ARIMA misses, combining the strengths of both methods. Hybrid ARIMA-NN is designed to leverage the strengths of both methods. Generally, the hybrid model can be expressed as Equation 1 [20].

$$Y_t = L_t + N_t \tag{1}$$

Where:

L_t : Residual values obtained from the ARIMA model

N_t : Predicted values from the Neural Network that model the non-linear components

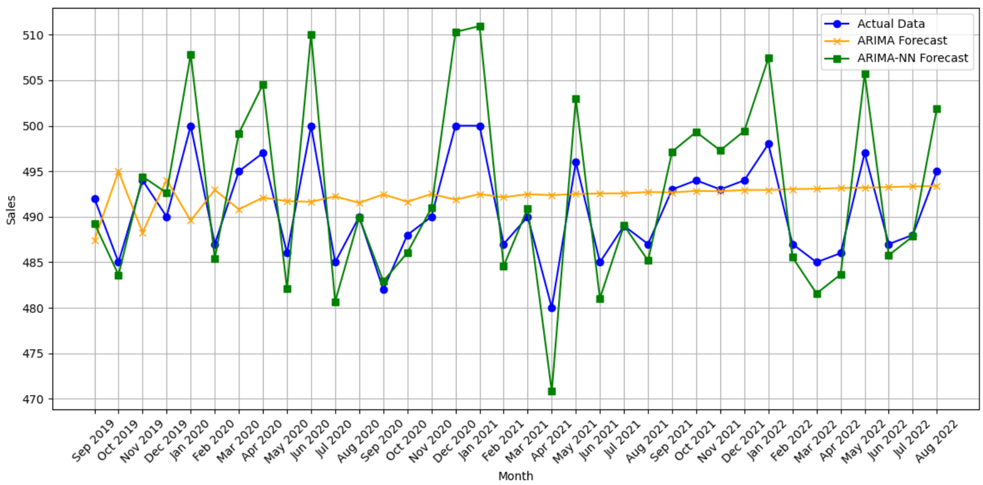


Figure 5. Comparison of Arabica coffee prediction using ARIMA and ARIMA-NN

Figure 5 presents the forecasted prices of Arabica coffee powder using the hybrid ARIMA-NN model, compared alongside the ARIMA model predictions and the actual data. The figure illustrates a clear comparison between actual sales data, ARIMA forecasts, and hybrid ARIMA-NN forecasts from September 2019 to August 2022. The actual sales data shows noticeable fluctuations, reflecting real market variability. The ARIMA model produces a relatively smooth and stable forecast, capturing the general trend but missing many of the short-term ups and downs. In contrast, the hybrid ARIMA-NN forecast closely follows the actual data's peaks and valleys, indicating its ability to model both linear and non-linear patterns effectively.

In this study, to assess the accuracy of several prediction methods used, we calculated the MAPE. MAPE is a commonly used metric in model evaluation because it provides an average percentage error between actual values and predicted values, making it easy to interpret. The formula for MAPE can be expressed in Equation 2.

$$MAPE = \frac{1}{n}\sum\nolimits_{t=1}^{n} \left|\frac{A_t - F_t}{A_t}\right| \times 100\% \qquad (2)$$

Where:

A_t : The actual value at time t,
F_t : The predicted value at time t,
n : The number of observations.

The actual data and predictions were gathered using two forecasting methods: ARIMA and Hybrid ARIMA-NN. We calculated the MAPE for each method. The results showed that the ARIMA and Hybrid methods had MAPE values of about 1.3% and 1.0%, respectively. Based on the MAPE calculation, the hybrid ARIMA-NN is a better approach in this study, demonstrating better accuracy than the ARIMA.

IV. Conclusion

This study successfully developed a hybrid ARIMA-Neural Network model for forecasting Arabica coffee powder sales, which demonstrated improved accuracy compared to the standalone ARIMA model. The hybrid approach effectively captures linear, seasonal, and complex non-linear patterns in the sales data, as reflected by the lower MAPE. These results indicate that combining statistical and machine learning methods can significantly enhance forecasting performance, providing valuable support for production planning and inventory management at KOTEM Bondowoso. For future research, it would be beneficial to explore other hybrid models by integrating different machine learning techniques, such as Support Vector Machines or Long Short-Term Memory (LSTM) networks, to improve forecasting accuracy further. Additionally, incorporating external factors like weather conditions, market trends, or promotional activities could provide a more comprehensive model. Expanding the dataset to include more extended time periods or multiple locations may also help generalize the model's applicability.

V. Acknowledgment

We would like to express our sincere gratitude to Universitas Jember for their unwavering support and provision of resources throughout the course of this research.

References

[1] E. O. Eboigbe, O. A. Farayola, F. O. Olatoye, O. C. Nnabugwu, and C. Daraojimba, "Business intelligence transformation through AI and data analytics," *Engineering Science & Technology Journal*, vol. 4, no. 5, pp. 285–307, 2023.

[2] D. Apriani,; Feny, M.; Alghifari, and M. Igamo, "Indonesian Coffee at The International Market," *Jurnal Paradigma Ekonomika*, vol. 17, no. 2, pp. 261–272, Sep. 2022, doi: 10.22437/jpe.v17i2.13983.

[3] R. Naqvi, A. Mehdi, and A. Zehra, "Coffee as a Functional Drink: Coffee-drinking and health benefits that support the concept of coffee as a functional food," *Prog Nucl Energy 6 Biol Sci*, vol. 3, no. 4, pp. 516–524, Dec. 2023, doi: 10.55006/BIOLSCIENCES.2023.3405.

[4] L. Barrea *et al.*, "Coffee consumption, health benefits and side effects: a narrative review and update for dietitians and nutritionists," *Crit Rev Food Sci Nutr*, vol. 63, no. 9, pp. 1238–1261, 2023, doi: 10.1080/10408398.2021.1963207.

[5] Perhutani, "Khofifah: Bondowoso penghasil kopi terbesar di Jatim." Accessed: Apr. 29, 2025. [Online]. Available: https://www.perhutani.co.id/khofifah-bondowoso-penghasil-kopi-terbesar-di-jatim/

[6] G. S. Day, "Aligning the Organization with the Market," *MIT Sloan Manag Rev*, Oct. 2006, Accessed: Apr. 29, 2025. [Online].

[7] R. Sesario, T. Duha, A. Alfiah, S. A. Pramono, P. A. Cakranegara, and P. N. Pontianak, "SINGLE EXPONENTIAL SMOOTHING IN FORECASTING TOOLS AND MEDICINE STOCKS," *INFOKUM*, vol. 10, no. 4, pp. 27–32, Oct. 2022, Accessed: Jan. 05, 2024. [Online].

[8] A. S. Pranata, N. O. Adiwijaya, and M. Furqon, "Screen Printing T-shirt Stock Forecasting System with Weight Moving Average," *Jurnal Komputer Terapan*, vol. 9, no. 1, pp. 50–57, Jun. 2023, doi: 10.35143/jkt.v9i1.5834.

[9] V. Komaria, N. El Maidah, and M. A. Furqon, "Prediksi Harga Cabai Rawit di Provinsi Jawa Timur Menggunakan Metode Fuzzy Time Series Model Lee," *Komputika : Jurnal Sistem Komputer*, vol. 12, no. 2, pp. 37–47, Sep. 2023, doi: 10.34010/KOMPUTIKA.V12I2.10644.

[10] M. Furqon, E. R. Fahlefi, and N. O. Adiwijaya, "Drug Sales Forecasting Using Single Exponential Smoothing (Case Study: NDM Pharmacy)," in *2nd International Conference on Neural Networks and Machine Learning 2023 (ICNNML 2023)*, 2024, pp. 25–31.

[11] R. Hidayat and B. H. Mustawinar, "peramalan jumlah wisatawan asing dengan model arima," *Infinity: Jurnal Matematika dan Aplikasinya*, vol. 2, no. 2, pp. 104–115, Mar. 2022, doi: 10.30605/27458326-100.

[12] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Appl Soft Comput*, vol. 11, no. 2, pp. 2664–2675, Mar. 2011, doi: 10.1016/J.ASOC.2010.10.015.

[13] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *Eur J Oper Res*, vol. 160, no. 2, pp. 501–514, Jan. 2005, doi: 10.1016/J.EJOR.2003.08.037.

[14] V. Gudivada, A. Apon, and J. Ding, "Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations," *International Journal on Advances in Software*, vol. 10, no. 1, pp. 1–20, 2017.

[15] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[16] M. Estrada *et al.*, "Evaluation of several error measures applied to the sales forecast system of chemicals supply enterprises," *International Journal of Business Administration*, vol. 11, no. 4, pp. 39–51, 2020.
[17] A. C. Atkinson, M. Riani, and A. Corbellini, "The box–cox transformation: Review and extensions," 2021.
[18] S. Yadav and K. P. Sharma, "Statistical Analysis and Forecasting Models for Stock Market," *ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications*, pp. 117–121, Jul. 2018, doi: 10.1109/ICSCCC.2018.8703324.
[19] Z. Hossain, A. Rahman, M. Hossain, and J. H. Karami, "Over-Differencing and Forecasting with Non-Stationary Time Series Data," *Dhaka University Journal of Science*, vol. 67, no. 1, pp. 21–26, Jan. 2019, doi: 10.3329/DUJS.V67I1.54568.
[20] A. A. Alsuwaylimi, "Comparison of ARIMA, ANN and Hybrid ARIMA-ANN models for time series forecasting," *Information Sciences Letters*, vol. 12, no. 2, pp. 1003–1016, 2023.