iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73814

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Comparative Study of Sales Forecasting using LSTM and ARIMA Methods

Rudy Adipranata[1], Leo Willyanto Santoso[2], Kenny Manuel Evan[3]
Informatics Department, Petra Christian University, Surabaya, Indonesia

Abstract: The current competitive manufacturing environment drives many companies to respond quickly to demand. One effective approach is to predict future market conditions. In this research, sales predictions were carried out using a case study of a lunch box manufacturing company. The company requires a method to predict lunch box sales to estimate the number of products to be produced. This aims to prevent excessive overproduction or underproduction. The prediction methods used in this research were Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA). The LSTM method tends to be better suited to non-linear data such as market conditions, while ARIMA was used for comparison. Based on the prediction results for the company's two products, the LSTM method performed better than ARIMA in all assessment types: Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE).
Keywords: sales prediction, long short term memory, autoregressive integrated moving average

I. INTRODUCTION

Today's competitive manufacturing environment forces companies to respond quickly to changes in demand. Companies are beginning to implement a supply chain approach that is more influenced by market demand. Therefore, forecasting calculations based on market demand, such as sales, are really needed by the company. The accurate prediction of future sales is a cornerstone for strategic business operations, directly influencing inventory optimization, production scheduling, and financial planning within an organization [1]. This necessitates the employment of robust forecasting methodologies capable of discerning intricate patterns and dependencies within historical sales data. Among the various approaches, time series forecasting models, ranging from classical statistical methods to advanced machine learning and deep learning algorithms, have emerged as prominent tools for this purpose [2].

In recent years, the emergence of deep learning techniques has revolutionized the field of time-series forecasting. Among various deep learning architectures, Long Short-Term Memory (LSTM), a type of Recurrent Neural Network (RNN), stands out for its unique capabilities. This is primarily due to its specialized architecture, which allows it to effectively capture and learn long-term dependencies in sequential data, a critical aspect often present in sales time series [3]. Conversely, Autoregressive Integrated Moving Average (ARIMA) models, a class of classical statistical models, have long been a staple in time series analysis due to their interpretability and proven effectiveness in modeling linear relationships within sequential data [4]. However, while traditional models like ARIMA have demonstrated considerable utility, especially in scenarios with well-defined linear trends and seasonality, they often struggle with the complex, non-linear, and often chaotic patterns inherent in real-world sales data [2].

Numerous research have investigated various methodologies for time series forecasting, ranging from traditional statistical approaches to more contemporary machine learning and deep learning models. Specifically, research has explored the application of time-series analysis and regression models, alongside advanced learning machines, for tasks such as stock price forecasting [5]. The established utility of statistical models like the Autoregressive Integrated Moving Average for time series forecasting is well-documented, often serving as a benchmark due to its interpretability and robust performance in capturing linear dependencies [6, 7]. However, the increasing complexity and non-linearity observed in real-world time series data, such as sales figures, often necessitate more sophisticated models capable of capturing intricate non-linear relationships [8]. Consequently, deep learning models, particularly Long Short-Term Memory networks, have gained prominence for their ability to model long-term dependencies and complex non-linear patterns within sequential data, offering significant advantages over traditional methods in diverse forecasting applications [9, 10]. For instance, research have shown that deep learning algorithms, including LSTM, can significantly outperform traditional models like ARIMA, with reported error rate reductions as high as 84-87% [11].

In this research, we conducted a case study on a company specializing in manufacturing lunch boxes. The accuracy of sales forecasting is crucial for this company to optimize raw material inventory management, daily production planning, and marketing strategies, thereby efficiently meeting customer demand and minimizing waste. We used the LSTM method to forecast sales and compare it with the ARIMA. This comparative study aims to thoroughly evaluate the performance of LSTM networks against ARIMA models in the context of sales forecasting, thereby determining which methodology offers superior predictive accuracy for sales datasets [12]. Measurement used for comparison will include Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE), providing a comprehensive assessment of each model's predictive capabilities.

II. LITERATURE REVIEW

Early research in sales forecasting largely relied on classical statistical models, which include exponential smoothing, moving averages, and Autoregressive Integrated Moving Average (ARIMA) models, demonstrating their utility in capturing linear trends and seasonality within sales data [13]. The ARIMA model is a classic and powerful tool for time series forecasting. It's a method that works by analyzing and modeling the patterns present in past time series data to predict future values. The process is famously known as the Box-Jenkins methodology, which consists of four main steps: model identification, parameter estimation, model validation and forecasting. In model identification, it determines the best parameters (p, d, q) for the ARIMA model. p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model. This is done by analyzing the time series plot, and more importantly, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, which help reveal underlying correlation patterns in the data. In second step, the parameter estimation phase involves fitting the chosen model to the historical data, often employing techniques such as maximum likelihood estimation to determine the optimal coefficients for the autoregressive, integrated, and moving average components. The third step is model validation phase. It involves evaluating the fitted model's adequacy by examining the residuals to ensure they resemble white noise (completely random), indicating that no further information can be extracted. If they are not random, it means the model hasn't captured all the patterns, and it needs to go back to the identification step. The last step is the forecasting phase. It applies the validated ARIMA model to extrapolate future prediction, accompanied by prediction intervals that quantify the uncertainty of these projections, thereby providing a complete probabilistic forecast [14]. The general formula for an ARIMA model can be written as equation 1:

$$\left(1-\sum_{i=1}^{p} \emptyset_i L^i\right)(1-L)^d y_t = \left(1+\sum_{j=1}^{q} 0_j L^j\right)\varepsilon_t \qquad (1)$$

where L is the lag operator, the α_i are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average part and the ε_t are error terms. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

Despite their interpretability and proven effectiveness in modeling linear relationships, these classical methods often fall short in accurately capturing the complex, non-linear patterns and long-term dependencies frequently observed in modern sales data [15]. This limitation has propelled the development and adoption of advanced machine learning techniques, such as recurrent neural networks and their variants like Long Short-Term Memory networks, which are specifically designed to address sequential data challenges by capturing intricate temporal dynamics [16, 17].

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) specifically designed to handle and predict time-series data [18]. Unlike traditional feedforward neural networks, RNNs have a memory of past information through a feedback loop, allowing them to process sequences of data. However, standard RNNs suffer from the vanishing gradient problem, where they struggle to learn and remember dependencies that span over a long time a crucial limitation for complex sequences like financial data or speech. LSTM was created to overcome this exact problem. It introduces a sophisticated internal mechanism, called a cell state, that acts as a conveyor belt for information. This cell state runs through the entire network, allowing it to selectively add or remove information, ensuring that important long-term patterns aren't forgotten. This architectural innovation enables LSTMs to effectively process and predict sequential data, making them particularly suitable for tasks requiring the understanding of context over extended periods, such as sales forecasting (Abdulkadir et al., 2018).

LSTM has memory and multiple gate types: input gates, forget gates, and output gates. LSTM can learn over 1,000 steps in advance, depending on the network's complexity. The input gate will receive information in the form of hidden states originating from the previous cell and new information originating from the current input, then the information will be combined and processed using the sigmoid function and the tanh function. The result of the sigmoid function will change the value to 0 to 1 to determine which information will be updated. The closer to 0 means the information is not important, the closer to 1 means the information is important. The result of the tanh function in the form of a value of -1 to 1 is used to support the cell to learn the information better. The forget gate decides which information to keep or discard. The forget gate receives information in the form of hidden states from the previous cell and new information from the current input. These are then combined and processed using a sigmoid function, which produces a result ranging from 0 to 1. The closer the result is to 0, the more information will be discarded. Conversely, the closer it is to 1, the more information will be retained. The output gate determines which hidden state will be sent to the next cell. The output gate receives information in the form of the hidden state from the previous cell and new information from the current input. This information is then combined and processed with the sigmoid function. The new cell state is then processed through the tanh function. The result of the tanh function is multiplied by the result of the sigmoid function to obtain the information that will be stored in the new hidden state. The hidden state and the new cell state are then passed on to the next cell. Architecture of LSTM is shown in Fig. 1.

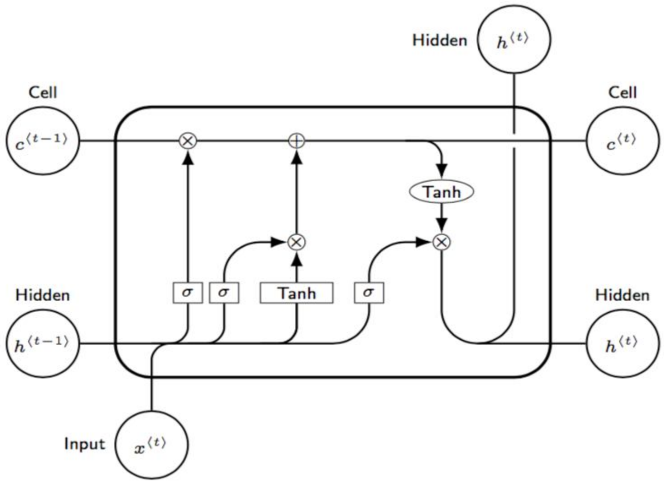


Fig. 1 Architecture of Long Short-Term Memory

To calculate the performance of the prediction results, Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE) are used. These metrics are crucial for a thorough evaluation of predictive models, offering insights into different facets of prediction error. MAD is the average of the absolute values of the differences between the actual and predicted values. A larger value indicates a greater distance between the predicted and actual data, and vice versa. The formula of MAD is shown in Equation 2.

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad (2)$$

Y_i is actual value, $\hat{Y}_i$ is prediction value and n is the number of data.

MAPE is the average absolute percentage of the reduction of the actual value to the predicted value, divided by the predicted value. MAPE interprets the error value as a percentage that describes the average difference in the distance between the actual value and the predicted value. The MAPE formula is shown in Equation 3.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{\hat{Y}_i}\right| \times 100\% \quad (3)$$

Y_i is actual value, $\hat{Y}_i$ is prediction value and n is the number of data.

MSE is the average of the squared differences between the predicted and actual values. The squared is used to magnify the difference and clarify the distance between the actual and predicted values. The MSE formula is shown in Equation 4.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (4)$$

Y_i is actual value, $\hat{Y}_i$ is prediction value and n is the number of data.

III.METHODOLOGY

Achieving an accurate predictive model hinges on correctly identifying the right parameters and steps. The process for forecasting with ARIMA requires a series of stages, as follows: data processing, finding the best parameter and prediction evaluations. Data preprocessing is used to prepare and process raw data into usable and actionable information. In the ARIMA method, data must first be checked for stationary behavior.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Stationary data fluctuates around the mean, where the value is constant and does not tend to fluctuate sharply. A common test for stationarity is the Augmented Dickey-Fuller (ADF) test. Data is considered stationary if its p-value is less than 0.05. If the data is non-stationary, differencing or adjusting the data, is necessary to ensure it can be processed. This process is usually performed a maximum of two times and is not necessary if the data is already stationary.

Once the data is stationary, it is divided into three parts: train, validation, and test. Train data is used to fit the model. Validation data is used to validate the parameter values p, d, and q in the loop. The best parameters based on the validation results will be used to predict the testing data. The parameters used in the ARIMA method are p, d, and q. The p value is the number of lagged observations in the model, also known as the model order. The d value is the number of times the observations are differencing, also known as the differencing level. The q value is a measure of the moving average. There are several ways to find the right parameters for the prediction process, such as using the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) methods. However, one of the most effective methods is to use a looping method and try a series of combinations of the p, d, and q parameters. Once the data and parameters are ready, the final step is prediction evaluation. In this step, the model is fitted based on the p, d, and q parameters that produce the smallest error value. Then, predictions are made in the testing process using the model that has been created. After that, the predictions are evaluated. To perform sales prediction using LSTM, the following steps are taken: data preprocessing, hyperparameter determination and multistep prediction. Data preprocessing is a method used to prepare and process raw data into usable and processable information. In the LSTM method, data is normalized using MinMaxScaler with a range of (0,1). Data normalization is used to improve program speed and accuracy. After normalization, the data is divided into input sequences (x) and target sequences (y) based on the window size or time range desired. After data preprocessing, the next step is to determine the hyperparameters. The LSTM uses one dense layer with 64 neurons, one dense layer with 32 neurons, and one dropout layer. The parameters used in this research can be seen in Table 1.

TABLE I
LSTM HYPERPARAMETER

Hyperparameter	**Value**
Activation	Relu
Optimizer	Adam
Loss	MSE
Epoch	100
Batch size	32
Validation_split	0.2
Window size	5

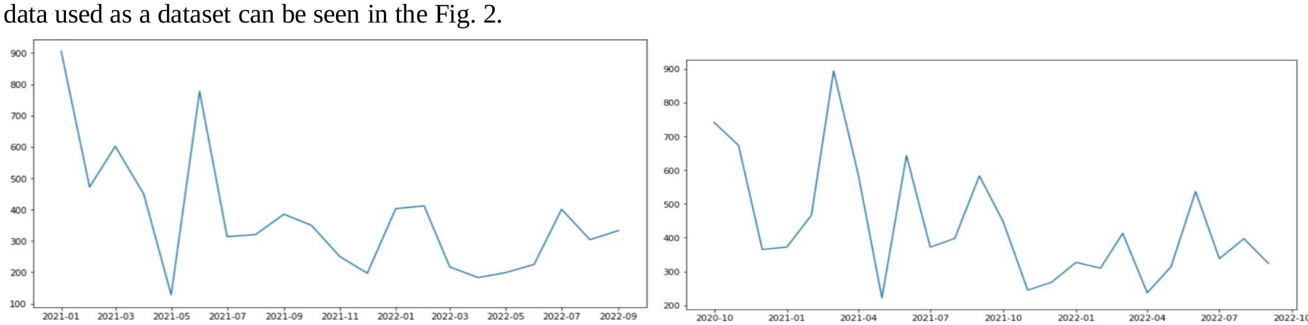
In the final step, once the data and parameters are ready and the model has been compiled, the next step is to fit the model based on the predetermined parameters. The prediction process in this research uses a multi-step method, where the forecast data obtained is reused to predict further sales.

IV. EXPERIMENTAL RESULT AND DISCUSSION

In the sales prediction process, the dataset used is sales data for about two years, around year 2020 until 2022. Two products will be used for prediction, 7200ml sealware and 3500ml sealware. Each product will be tested using ARIMA and LSTM methods. The evaluation results will then be compared to determine which method is better at predicting sales.

A. ARIMA Testing

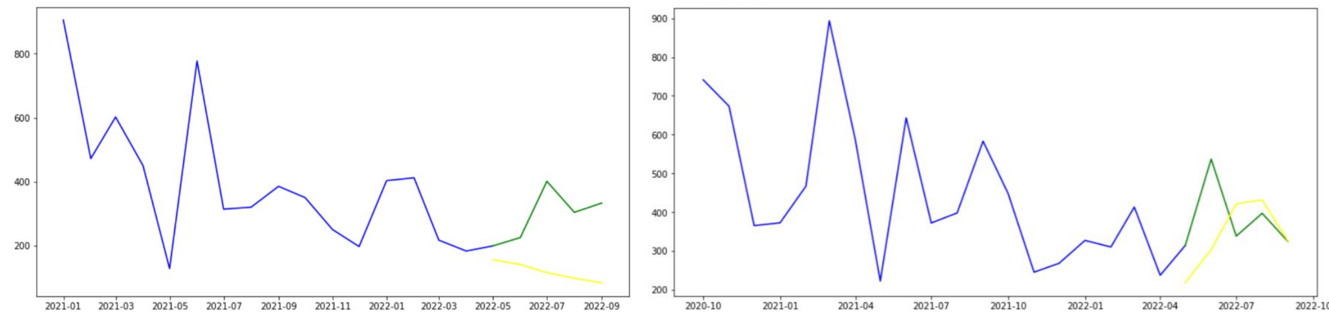
Sales data used as a dataset can be seen in the Fig. 2.



(a) (b)

Fig. 2 Dataset of sales data (a) 7200ml sealware (b) 3500ml sealware

Based on the ADF test results, it was found that the sales data is stationary, because the p-value is <0.05. Therefore, the data does not need to be simplified by using differencing. The dataset is divided into data training 14 months, data validation 5 months, and data testing 5 months. There are 128 parameter combinations of p, d, q. The p value ranges from 0 to 7, the d value ranges from 0 to 2, and the q value ranges from 0 to 7. Based on the results of this analysis, model fitting will be performed with parameters p,d,q is (6,1,5) for 7200ml sealware and (5,1,7) for 3500ml sealware. The graph of actual and predicted data is shown in Fig. 3.



(a) (b)

Fig. 3 Prediction result of (a) 7200ml sealware (b) 3500ml sealware using ARIMA

Several months of predicted values can be seen in Table 2, and the evaluation results consisting of MAD, MAPE, and MSE can be seen in Table 3.

TABLE II
PREDICTION RESULTS OF ARIMA

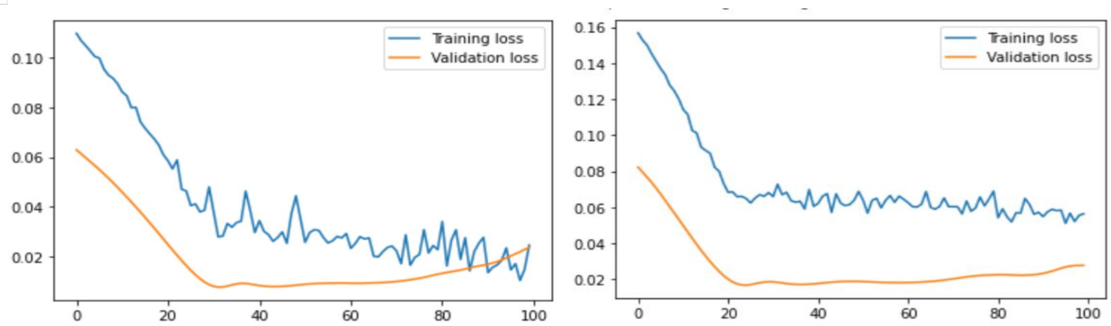
Sales Months	Sales of 7200ml sealware		Sales of 3500 ml sealware	
	Actual	Prediction	Actual	Prediction
2022-05	199	156	314	218
2022-06	225	141	537	303
2022-07	401	115	338	421
2022-08	304	98	397	432
2022-09	333	83	324	325

TABLE III
MAD, MAPE AND MSE RESULTS OF ARIMA

	MAD	MAPE	MSE
7200ml sealware	173.69	54.61%	39030.22
3500ml sealware	89.76	21.56%	14400.52
Average	131.72	38.08%	26715.37

B. LSTM Testing

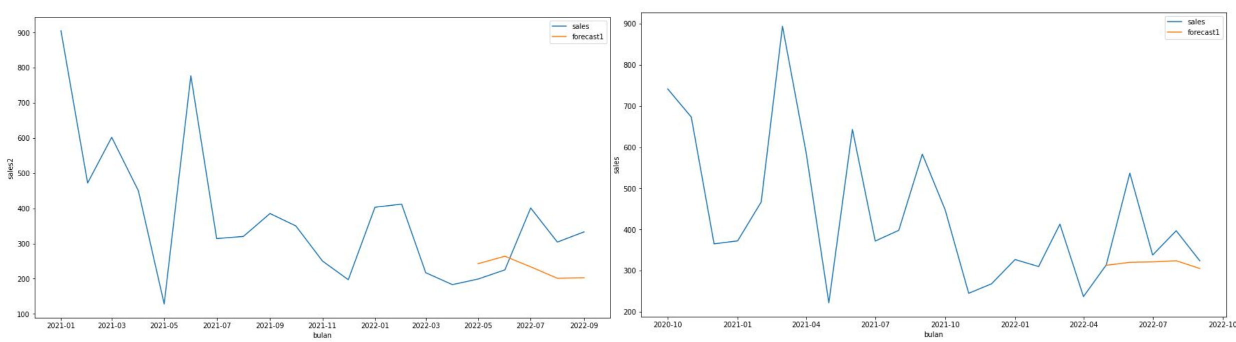
The dataset used in the LSTM test is the same as that used in the ARIMA test, sales data for 7200ml sealware and 3500ml sealware, as seen in Fig. 2. Data loss per epoch during the LSTM training process can be seen in Fig. 4.



(a) (b)

Fig. 4 Training and validation loss of (a) 7200ml sealware (b) 3500ml sealware using LSTM

Losses will continue to decrease until they reach a point where they cease to decrease. However, Fig. 4a shows that for 7200ml sealware product, the loss value decreases significantly until the final epoch, but for 3500ml sealware product, after epoch 20, the loss value does not decrease significantly anymore, as seen in Fig. 4b. After training phase, prediction testing was conducted using the last five months of data. The prediction results can be seen in Fig. 5.



(a) (b)

Fig. 5 Prediction results of (a) 7200ml sealware (b) 3500ml sealware using LSTM

Some months of predicted values using LSTM can be seen in Table 4, and the evaluation results consisting of MAD, MAPE, and MSE can be seen in Table 5.

TABLE IV
PREDICTION RESULTS OF LSTM

Sales Months	Sales of 7200ml sealware		Sales of 3500 ml sealware	
	Actual	Prediction	Actual	Prediction
2022-05	199	243	314	312
2022-06	225	263	537	320
2022-07	401	233	338	321
2022-08	304	201	397	324
2022-09	333	202	324	305

TABLE V
MAD, MAPE AND MSE RESULTS OF LSTM

	MAD	MAPE	MSE
7200ml sealware	96.57	30.77%	11795.66
3500ml sealware	56.54	11.76%	9144.60
Average	76.55	21.26%	10470.13

C. Comparison Between ARIMA and LSTM

The comparison between the ARIMA and LSTM methods was conducted by comparing the average of MAD, MAPE, and MSE values of each method. The method with the lower average is the better method. The results of this comparison can be seen in Table 6.

TABLE VI
COMPARISON OF ARIMA AND LSTM

	Avg of MAD	Avg of MAPE	Avg of MSE
ARIMA	131.72	38.08%	26715.37
LSTM	76.55	21.26%	10470.13

Table 6 shows that for all measurements (MAD, MAPE, and MSE), the LSTM results are lower than ARIMA. These results indicate that for the case study of lunch box sales forecasting at the company used in this research, using the LSTM method will produce predictions with higher accuracy.

V. CONCLUSIONS

This research predicted the sales of lunch boxes produced by a company using ARIMA and LSTM methods. The experiment was conducted using a dataset of sales data for two years. From the experimental results and measurements using MAD, MAPE, and MSE, it was shown that the prediction results using LSTM had better performance than the results using ARIMA. This shows that the LSTM method is more suitable for use on datasets for this case study because the lunch box sales data is non-linear.

REFERENCES

[1] Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C., & Marconi, M., “Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices,” Forecasting, vol. 3, no. 3, pp. 644-662, 2021. https://doi.org/10.3390/forecast3030040.

[2] Ghienne, M., & Limare, A., “Learning structural stress virtual sensors from on-board instrumentation of a commercial aircraft,” Computers & Structures, vol. 289, 107155, 2023. https://doi.org/10.1016/j.compstruc.2023.107155.

[3] Zhou, D., Zhuang, X., Zuo, H., Wang, H., & Yan, H., “Deep Learning-Based Approach for Civil Aircraft Hazard Identification and Prediction,” IEEE Access, vol. 8, pp. 103665-103683, 2020. https://doi.org/10.1109/ACCESS.2020.2997371.

[4] Siami-Namini, S., & Namin, A. S., “Forecasting Economics and Financial Time Series: ARIMA vs. LSTM,” arXiv (Cornell University), 2018. https://doi.org/10.48550/arxiv.1803.06386.

[5] Alamu, O. S., & Siam, M. K., “Stock Price Prediction and Traditional Models: An Approach to Achieve Short, Medium and Long-Term Goals,” Journal of Intelligent Learning Systems and Applications, vol. 16, no. 4, pp. 363-383, 2024. https://doi.org/10.4236/jilsa.2024.164018.

[6] Verano, D. A., Husnawati, H., & Ermatita, E., “Implementation of Autoregressive Integrated Moving Average Model to Forecast Raw Material Stock in The Digital Printing Industry,” Journal of Information Technology and Computer Science, vol. 5, no. 1, 2020. https://doi.org/10.25126/jitecs.202051117.

[7] Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K., “A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks,” Future Internet, vol. 15, no. 8, 255, 2023. https://doi.org/10.3390/fi15080255.

[8] Hadwan, M., Al-Maqaleh, B. M., Al-Badani, F. N., Khan, R. U., & Al-Hagery, M. A., “A Hybrid Neural Network and Box-Jenkins Models for Time Series Forecasting,” Computers, Materials & Continua/Computers, Materials & Continua, vol. 70, no. 3, pp. 4829-4845, 2021. https://doi.org/10.32604/cmc.2022.017824.

[9] Gasparin, A., Luković, S., & Alippi, C., “Deep Learning for Time Series Forecasting: The Electric Load Case,” arXiv (Cornell University), 2019. https://doi.org/10.48550/arxiv.1907.09207.

[10] Islam, M. R., Zamil, M. Z. H., Rayed, M. E., Kabir, M. M., Mridha, M. F., Nishimura, S., & Shin, J., “Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes,” IEEE Access vol. 12, pp. 121449-121479, 2024. https://ieeexplore.ieee.org/abstract/document/10663422.

[11] Lee, S. I., & Yoo, S. J., “A Deep Efficient Frontier Method for Optimal Investments,” arXiv (Cornell University), 2017. https://arxiv.org/pdf/1709.09822.

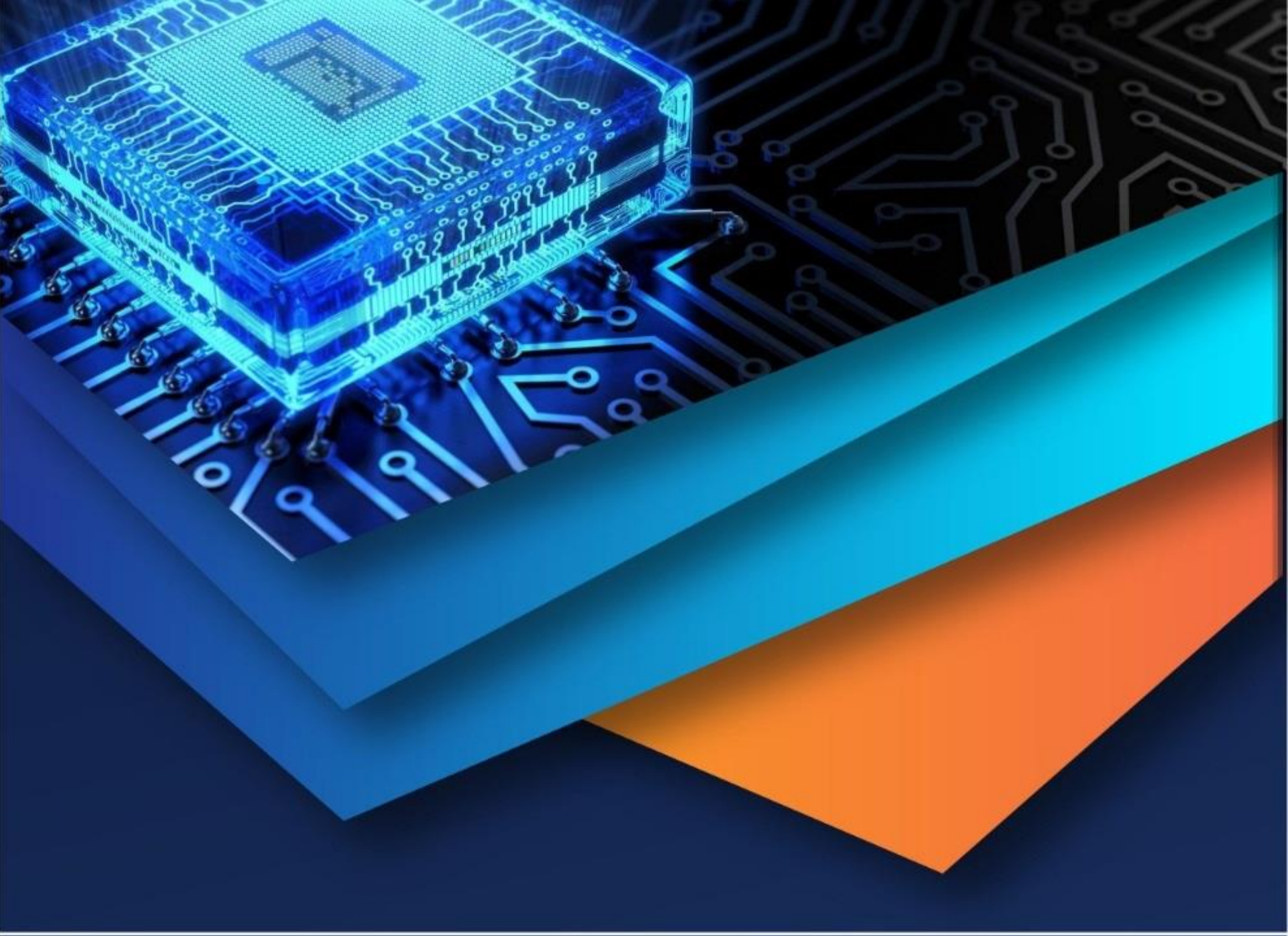
[12] Albeladi, K., Zafar, B., & Mueen, “A. Time Series Forecasting using LSTM and ARIMA,” International Journal of Advanced Computer Science and Applications, vol. 14, no. 1, 2023. https://doi.org/10.14569/ijacsa.2023.0140133.

[13] Singh, D. A. A. G., Leavline, E. J., Muthukrishnan, S., & Yuvaraj, R., “Machine Learning based Business Forecasting,” International Journal of Information Engineering and Electronic Business, vol. 10, no. 6, pp. 40-51, 2018. https://doi.org/10.5815/ijieeb.2018.06.05.

[14] Skenderi, G., Joppi, C., Denitto, M., & Cristani, M., “On the Use of Learning-Based Forecasting Methods for Ameliorating Fashion Business Processes: A Position Paper,” In Lecture Notes in Computer Science vol. 13644, pp. 647-659, 2023. https://doi.org/10.1007/978-3-031-37742-6_50.

[15] Ganguly, P., & Mukherjee, I., “Enhancing Retail Sales Forecasting with Optimized Machine Learning Models,” 4th International Conference on Sustainable Expert Systems (ICSES), 2024. https://doi.org/10.1109/icses63445.2024.10762950.

[16] Ghienne, M., & Limare, A. "Learning structural stress virtual sensors from on-board instrumentation of a commercial aircraft," Computers & Structures, vol. 289, 2023. https://doi.org/10.1016/j.compstruc.2023.107155
[17] Wang, J., Guo, Y., & Wang, Y., "A sequential random forest for short-term vessel speed prediction," Ocean Engineering, vol. 248, 2022. https://doi.org/10.1016/j.oceaneng.2022.110691
[18] Abdulkadir, S. J., Alhussian, H., Nazmi, M. A. M., & Elsheikh, A. A. "Long Short Term Memory Recurrent Network for Standard and Poor's 500 Index Modelling," International Journal of Engineering & Technology, vol. 7, no, 25, 2018. https://doi.org/10.14419/ijet.v7i4.15.21365



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY