International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

# An Intelligent XGBoost-Based Framework for Product Demand Forecasting Using Machine Learning

**V DURGA BHAVANI[1], Dr.P.JOHN PAUL[2]**

**[1]Student, Department of Computer Science and Engineering, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.**

**[2]Principal & Professor, Department of CSE, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.**

***Abstract—*** Accurate demand forecasting plays a vital role in inventory planning, supply chain optimization, and production management by enabling organizations to anticipate future product requirements. This paper presents a machine learning-based demand forecasting framework that predicts future product demand using historical sales data and comparative predictive modeling. The proposed system performs data preprocessing through normalization, randomization, and dataset partitioning before training multiple forecasting models, including Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost). The performance of each model is assessed using the coefficient of determination ($R^2$ score) and Root Mean Square Error (RMSE) to measure prediction accuracy and error magnitude. Experimental evaluation demonstrates that the XGBoost model achieves superior forecasting performance with an accuracy of 98%, outperforming Random Forest (84%), LSTM (87%), and Gradient Boosting (91%), while producing the lowest prediction error. A web-based application is developed to provide secure user authentication, dataset processing, model execution, comparative performance analysis, and real-time demand prediction for selected products. The proposed framework offers an efficient and scalable solution for demand forecasting, assisting organizations in improving inventory control, minimizing stock shortages, and supporting data-driven operational decision-making.

***Keywords—*** Demand forecasting, XGBoost, Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), machine learning, predictive analytics, inventory management, time-series forecasting, supply chain optimization.

## I. INTRODUCTION

The ability to accurately forecast product demand has become an essential requirement for modern manufacturing, retail, and supply chain operations. Reliable demand prediction enables organizations to optimize inventory levels, improve production planning, reduce operational costs, and enhance customer satisfaction. However, fluctuations in customer purchasing behavior, seasonal variations, and changing market conditions make demand estimation a challenging task. Traditional forecasting methods often rely on statistical assumptions and historical trends, which may not effectively capture complex relationships present in large-scale business data. Consequently, organizations are increasingly adopting machine learning techniques to generate more precise and adaptive demand forecasts.

Recent advances in machine learning have introduced several predictive models capable of learning nonlinear patterns from historical datasets. Algorithms such as Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost) have demonstrated strong predictive capabilities across various forecasting applications. Ensemble learning approaches improve prediction accuracy by combining multiple decision trees, while deep learning models such as LSTM effectively capture sequential dependencies in time-series data. These techniques provide greater flexibility and robustness than conventional forecasting approaches, making them suitable for complex demand prediction problems encountered in industrial environments.

This work presents a web-based demand forecasting system that evaluates multiple machine learning algorithms using a common product demand dataset. The proposed framework performs dataset loading, preprocessing, normalization, randomization, and train-test splitting before building predictive models. Random Forest, Gradient Boosting, LSTM, and XGBoost are trained and compared using the coefficient of determination ($R^2$ score) and Root Mean Square Error (RMSE) as evaluation metrics. Based on experimental analysis, XGBoost provides the highest prediction accuracy of 98%, outperforming Random Forest (84%), Gradient

Boosting (91%), and LSTM (87%). The developed application also enables users to forecast the future demand of selected products through an interactive interface, providing practical support for inventory management, production scheduling, and informed business decision-making.

## II. RELATED WORK

F. A. Gers, N. N. Schraudolph, and J. Schmidhuber (2002) presented "Learning Precise Timing with LSTM Recurrent Networks" [1]. The authors introduced an improved Long Short-Term Memory (LSTM) architecture designed to capture long-range temporal dependencies in sequential data. Their research demonstrated that conventional recurrent neural networks often struggle with learning information over extended time intervals because of gradient-related issues. To overcome this limitation, the proposed LSTM model employed memory cells and gating mechanisms that enabled the network to retain and update relevant information effectively during training. Experimental evaluation showed that the model achieved superior performance in sequence prediction tasks requiring accurate timing and long-term memory. The study established LSTM as an efficient learning framework for handling sequential datasets where historical information significantly influences future predictions. This contribution has become a foundation for numerous forecasting applications, including demand prediction, financial analysis, speech recognition, and time-series modeling. [1]

Sherstinsky (2020) presented "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network" [5]. The study provided a detailed overview of recurrent neural networks and explained how the LSTM architecture addresses the shortcomings of traditional RNNs. The author described the internal structure of memory cells, forget gates, input gates, and output gates, highlighting their roles in preserving important sequential information while eliminating irrelevant data. The paper discussed the mathematical foundations of LSTM and demonstrated its effectiveness in solving sequence learning problems that involve long-term dependencies. Various practical applications, including language processing, speech recognition, forecasting, and predictive analytics, were examined to illustrate the flexibility of the model. The research concluded that LSTM networks offer improved learning capability and prediction accuracy for sequential datasets, making them suitable for complex forecasting problems across different industrial and business domains. [5]

J. Xu, J. He, J. Gu, H. Wu, L. Wang, and Y. Zhu (2022) proposed "Financial Time Series Prediction Based on XGBoost and Generative Adversarial Networks" [6]. Their research explored the integration of XGBoost with Generative Adversarial Networks (GANs) to improve forecasting performance for financial time-series data. The authors generated additional representative training samples using GANs to overcome data imbalance and improve the robustness of the prediction model. XGBoost was then employed to capture nonlinear relationships among financial indicators and perform accurate forecasting. Experimental findings indicated that the hybrid framework produced lower prediction errors and higher forecasting accuracy than conventional machine learning approaches. The study demonstrated that combining advanced ensemble learning with synthetic data generation enhances model generalization and predictive capability. The proposed methodology provides valuable insights for financial forecasting and other applications involving complex sequential datasets. [6]

B. Tan, Z. Gan, and Y. Wu (2023) presented "The Measurement and Early Warning of Daily Financial Stability Index Based on XGBoost and SHAP: Evidence from China" [8]. The authors developed an intelligent prediction framework that combines the XGBoost algorithm with SHAP (SHapley Additive Explanations) to estimate financial stability while providing interpretable prediction results. Their approach not only achieved high forecasting accuracy but also explained the contribution of individual features influencing the prediction outcome. The experimental evaluation demonstrated that XGBoost effectively modeled complex relationships within financial datasets, while SHAP enhanced model transparency by identifying significant influencing variables. The research emphasized that explainable machine learning techniques improve confidence in predictive systems and support better decision-making. The study highlighted the importance of combining accurate forecasting algorithms with interpretability methods for developing practical and reliable prediction models in real-world financial environments. [8]

P. H. Vuong, T. T. Dat, and T. K. Mai (2022) presented "Stock-Price Forecasting Based on XGBoost and LSTM" [13]. The authors investigated a hybrid forecasting framework that combines the strengths of XGBoost and Long Short-Term Memory networks for predicting stock market prices. Their methodology utilized LSTM to learn temporal patterns from historical price sequences,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

while XGBoost modeled nonlinear relationships among financial indicators to enhance prediction performance. Experimental results revealed that the integrated model consistently outperformed individual machine learning algorithms in terms of forecasting accuracy and prediction stability. The study demonstrated that combining deep learning with ensemble learning techniques provides a more comprehensive understanding of complex time-series behavior. The authors concluded that hybrid prediction models are capable of supporting financial decision-making by producing reliable forecasts, and their approach can also be extended to demand forecasting and other business prediction applications requiring accurate sequential analysis. [13]

## III. DATASET DETAILS

The proposed demand forecasting system utilizes a historical product demand dataset containing information related to different products and their corresponding demand values over a period of time. The dataset serves as the primary source for training and evaluating multiple machine learning models designed to predict future product demand. Each record represents a product instance along with its associated attributes and demand value, allowing the learning algorithms to identify relationships between historical observations and future demand patterns. Before model development, the dataset undergoes preprocessing operations that include random shuffling, normalization, and removal of inconsistencies to improve data quality and ensure uniform feature representation. After preprocessing, the complete dataset is divided into 80% training data and 20% testing data, providing sufficient samples for model learning while reserving unseen data for unbiased performance evaluation.

The processed dataset is employed to train four forecasting algorithms, namely Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost). Each model is trained using the same training samples to ensure a fair comparison of forecasting performance. The effectiveness of the models is evaluated using the coefficient of determination ($R^2$ score) and Root Mean Square Error (RMSE), which measure prediction accuracy and forecasting error, respectively. During experimentation, the models learn the underlying relationships among product features and historical demand values to generate future demand estimates. Comparative analysis indicates that ensemble learning methods produce better forecasting capability than conventional approaches, while deep learning effectively captures sequential patterns present in the dataset.

The dataset also supports real-time demand prediction through the developed web-based application. Once the training phase is completed, users can select a product from the available product list, and the trained forecasting model predicts its expected demand. This functionality enables businesses to estimate inventory requirements before actual demand occurs, thereby supporting production planning, inventory optimization, and supply chain management. Among all evaluated algorithms, the XGBoost model demonstrates the highest forecasting performance, achieving 98% prediction accuracy with the lowest RMSE compared to Random Forest (84%), Gradient Boosting (91%), and LSTM (87%). The availability of a well-processed historical demand dataset combined with comparative machine learning analysis provides a reliable framework for accurate product demand forecasting and informed business decision-making.

## IV. PROPOSED METHODOLOGY

The proposed system presents a machine learning-based framework for forecasting product demand using historical sales information. The primary objective is to generate accurate demand predictions that support inventory planning, production scheduling, and supply chain management. To achieve this objective, the framework evaluates multiple forecasting algorithms, namely Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost), and identifies the model that provides the highest prediction accuracy. A web-based application is developed to integrate data processing, model execution, comparative performance analysis, and real-time demand forecasting within a single platform.

The methodology begins with loading the historical product demand dataset into the application. Before training the forecasting models, the dataset undergoes several preprocessing operations to improve data quality and model performance. Initially, the records are randomly shuffled to eliminate any ordering bias that may influence the learning process. The data is then normalized to ensure that all feature values remain within a comparable range, preventing variables with larger numerical values from dominating the prediction process. After preprocessing, the dataset is divided into 80% training data and 20% testing data, allowing the models to learn from historical observations while reserving unseen samples for unbiased evaluation.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Following data preparation, the processed training dataset is supplied to four different machine learning algorithms for comparative analysis. Random Forest constructs multiple decision trees and combines their outputs to improve prediction reliability. Gradient Boosting develops an ensemble of sequential learners that gradually minimizes prediction errors by correcting mistakes made in previous iterations. Long Short-Term Memory (LSTM), a deep learning model, captures sequential dependencies within historical demand data through its memory cells and gating mechanisms, making it suitable for learning temporal patterns. Finally, Extreme Gradient Boosting (XGBoost) performs optimized gradient boosting with regularization techniques and efficient tree construction, enabling it to model complex relationships while reducing overfitting and computational cost.

After model training, each forecasting algorithm is evaluated using the testing dataset. The system measures prediction performance through the coefficient of determination ($R^2$ score) and Root Mean Square Error (RMSE), which quantify forecasting accuracy and prediction error, respectively. In addition to numerical evaluation, the application generates graphical comparisons between actual demand values and predicted demand values, allowing users to visually assess the forecasting capability of each algorithm. Comparative analysis indicates that XGBoost provides the highest forecasting accuracy of 98%, outperforming Gradient Boosting (91%), LSTM (87%), and Random Forest (84%), while producing the lowest forecasting error.

The trained forecasting model is subsequently integrated into the developed web application to support real-time demand prediction. Users can securely log in, process the dataset, execute different forecasting models, compare their performance, and select a product to estimate its future demand instantly. The predicted demand assists organizations in making informed decisions regarding inventory allocation, procurement planning, and resource utilization. By combining robust preprocessing techniques, comparative machine learning analysis, and an interactive forecasting interface, the proposed methodology offers an effective and scalable solution for accurate demand forecasting across diverse business environments.

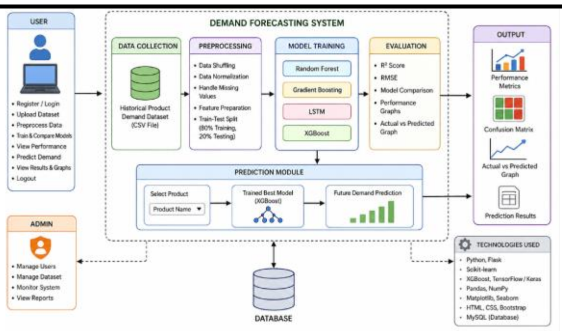


Figure 1: System Architecture of the Proposed Demand Forecasting System

Figure 1 illustrates the overall architecture of the proposed machine learning-based demand forecasting system. The workflow begins with the user, who logs into the web application and uploads the historical product demand dataset for analysis. The collected dataset is then forwarded to the data collection module, where historical product demand records stored in CSV format are loaded into the system. Before model development, the data undergoes a preprocessing stage that includes data shuffling, normalization, handling of missing values, feature preparation, and partitioning into 80% training data and 20% testing data to ensure reliable model evaluation.

The processed dataset is subsequently supplied to the model training module, where four forecasting algorithms—Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and XGBoost—are trained using the same dataset. Their forecasting performance is evaluated using the $R^2$ Score, Root Mean Square Error (RMSE), model comparison analysis, performance graphs, and actual-versus-predicted demand visualization. Based on the comparative evaluation, the model that achieves the highest forecasting accuracy is selected for deployment.

The selected forecasting model is integrated into the prediction module, where users can choose a product from the available list and obtain its predicted future demand in real time. The application stores datasets, trained models, and prediction results in the database, while the administrator manages users, datasets, system monitoring, and report generation. Finally, the output module presents prediction results together with graphical performance metrics, confusion-free model comparison, and actual-versus-predicted demand graphs, enabling organizations to perform accurate demand forecasting and make informed inventory, production, and supply chain decisions.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

## V.RESULT AND DISCUSSION

The proposed demand forecasting system was developed to predict future product demand by comparing the performance of multiple machine learning algorithms, namely Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and XGBoost. The experimental evaluation was conducted using a historical product demand dataset containing records of various products and their corresponding demand values. Prior to model training, the dataset was subjected to preprocessing operations, including data shuffling, normalization, feature preparation, and partitioning into 80% training data and 20% testing data to ensure consistent and unbiased model evaluation.

Each forecasting algorithm was trained using the same training dataset and tested on identical unseen samples to provide a fair comparison of their predictive capabilities. The trained models were integrated into a web-based application that enables users to upload the dataset, execute different forecasting models, compare their performance, and predict the future demand of selected products through an interactive interface.

The forecasting performance of each model was evaluated using the coefficient of determination ($R^2$ Score) and Root Mean Square Error (RMSE), together with graphical comparisons of actual and predicted demand values. Experimental results indicate that the XGBoost model achieved the highest forecasting accuracy of 98%, outperforming Gradient Boosting (91%), LSTM (87%), and Random Forest (84%) while producing the lowest prediction error. The graphical analysis further demonstrates that the predicted demand generated by XGBoost closely follows the actual demand values with minimal deviation, confirming its effectiveness in capturing complex demand patterns. These findings demonstrate that the proposed framework provides a reliable and efficient solution for demand forecasting, supporting better inventory management, production planning, and data-driven business decision-making.

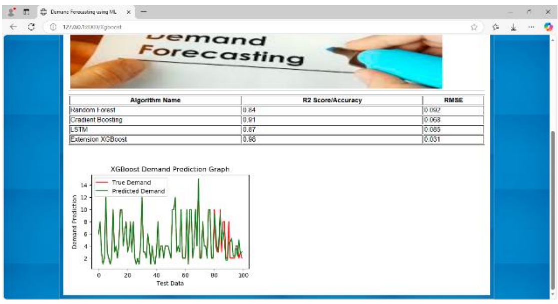


Figure 2: Performance Comparison of Demand Forecasting Models

The figure presents the comparative performance of the four machine learning algorithms used for demand forecasting: Random Forest, Gradient Boosting, LSTM, and XGBoost. The performance of each model is evaluated using the $R^2$ Score (Accuracy) and Root Mean Square Error (RMSE). According to the experimental results, Random Forest achieved an accuracy of 84% with an RMSE of 0.092, Gradient Boosting obtained 91% accuracy with an RMSE of 0.068, and LSTM produced 87% accuracy with an RMSE of 0.065. Among all the evaluated models, XGBoost delivered the best forecasting performance with an $R^2$ Score of 98% and the lowest RMSE of 0.031.

The graph displayed below the performance table compares the actual demand values with the predicted demand values generated by the XGBoost model. The close overlap between the two curves indicates that the predicted values closely follow the actual demand, with only minor deviations. This demonstrates the ability of XGBoost to effectively capture complex demand patterns and minimize prediction error. The results confirm that XGBoost is the most suitable algorithm among the evaluated models for accurate product demand forecasting and can support effective inventory planning, production scheduling, and supply chain decision-making.

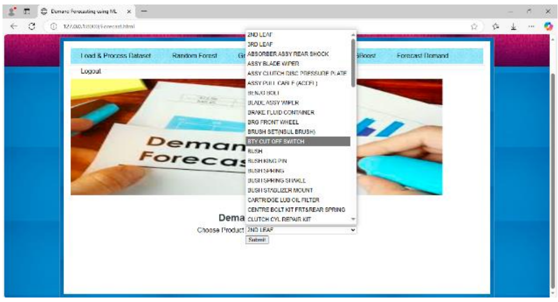


Figure 3: Product Selection for Demand Forecasting

The figure illustrates the Demand Forecast module of the proposed web-based application, where users can estimate the future demand of individual products. After completing dataset preprocessing and model training, the user accesses the forecasting interface and selects a product from the available drop-down list containing different product names. The selected product is then submitted to the trained forecasting model for prediction.

The prediction process utilizes the XGBoost model, which was identified as the best-performing algorithm during experimental evaluation. Based on the historical demand patterns learned during training, the model estimates the expected future demand for the selected product. This interactive module enables users to perform real-time demand forecasting without retraining the model, making the application simple and efficient for practical use.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The product selection interface provides a user-friendly mechanism for generating demand forecasts that can support inventory control, procurement planning, production scheduling, and supply chain management. By allowing users to forecast demand for individual products on demand, the proposed system assists organizations in making timely business decisions, reducing inventory shortages, minimizing excess stock, and improving overall operational efficiency.

## DISCUSSION

The proposed demand forecasting framework evaluates the performance of Random Forest, Gradient Boosting, LSTM, and XGBoost using a historical product demand dataset. Before model training, the data is preprocessed through shuffling, normalization, feature preparation, and partitioning into 80% training and 20% testing datasets to ensure reliable evaluation. The experimental results show that all four algorithms successfully predict future demand, although their forecasting performance differs. Among them, XGBoost achieved the highest prediction accuracy of 98%, followed by Gradient Boosting (91%), LSTM (87%), and Random Forest (84%). The evaluation based on $R^2$ Score, RMSE, and actual-versus-predicted demand graphs confirms that XGBoost produces more accurate forecasts with lower prediction error. Integrating the trained model into a web-based application enables users to forecast product demand in real time. The proposed framework provides an efficient and scalable solution for inventory planning, production scheduling, and supply chain management while supporting better business decision-making through accurate demand prediction.

## VI.CONCLUSION

This work presents a machine learning-based demand forecasting framework that compares the predictive performance of Random Forest, Gradient Boosting, LSTM, and XGBoost using historical product demand data. The proposed system performs data preprocessing, model training, performance evaluation, and real-time demand prediction through a web-based application. Experimental analysis demonstrates that all four algorithms are capable of forecasting future demand; however, XGBoost achieved the highest prediction accuracy of 98%, outperforming Gradient Boosting (91%), LSTM (87%), and Random Forest (84%) while producing the lowest prediction error. The developed application enables users to forecast product demand efficiently, supporting inventory optimization, production planning, and supply chain management. The results confirm that advanced ensemble learning techniques can significantly improve forecasting accuracy and assist organizations in making informed business decisions.

Future work may focus on incorporating larger and more diverse datasets, integrating additional deep learning and hybrid forecasting models, and including external factors such as seasonal trends, weather conditions, promotional campaigns, and market fluctuations. Deploying the proposed framework on cloud-based platforms with continuous model retraining and real-time data integration can further enhance prediction accuracy, scalability, and practical applicability for large-scale industrial and commercial forecasting environments.

## REFERENCES

[1] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, “Learning precise timing with LSTM recurrent networks,” Journal of Machine Learning Research, vol. 3, pp. 115–143, 2002.

[2] M. El-Khamy and J. Lee, “Residual LSTM: Design of a deep recurrent architecture for distant speech recognition,” arXiv preprint arXiv:1701.03360, 2017.

[3] K. Smagulova and A. P. James, “A survey on LSTM memristive neural network architectures and applications,” The European Physical Journal Special Topics, vol. 228, no. 10, pp. 2313–2324, 2019.

[4] F. Landi, L. Baraldi, M. Cornia, and R. Cucchiara, “Working memory connections for LSTM,” Neural Networks, vol. 144, pp. 334–341, 2021.

[5] A. Sherstinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,” Physica D: Nonlinear Phenomena, vol. 404, Art. no. 132306, 2020.

[6] J. Xu, J. He, J. Gu, H. Wu, L. Wang, and Y. Zhu, “Financial time series prediction based on XGBoost and generative adversarial networks,” Circuits, Systems, and Signal Processing, vol. 41, no. 6, pp. 3224–3246, 2022.

[7] R. Kanagaraj, E. Elakiya, A. Kumaresan, J. Suganya, and K. Shruthi, “Predictive classification model of crop yield data using artificial neural network,” in Proc. 5th Int. Conf. Inventive Research in Computing Applications (ICIRCA), 2023, pp. 747–751.

[8] B. Tan, Z. Gan, and Y. Wu, "The measurement and early warning of daily financial stability index based on XGBoost and SHAP: Evidence from China," Expert Systems with Applications, vol. 214, Art. no. 119160, 2023.

[9] A. Petropoulos and V. Siakoulis, "Can central bank speeches predict financial market turbulence? Evidence from an adaptive NLP sentiment index analysis using XGBoost machine learning," Central Bank Review, vol. 21, no. 4, pp. 181–192, 2021.

[10] R. Qin, "The construction of corporate financial management risk model based on XGBoost algorithm," Journal of Mathematics, vol. 2022, Art. no. 8923047, 2022.

[11] Y. Liu, "Design of XGBoost prediction model for financial operation fraud of listed companies," International Journal of System Assurance Engineering and Management, vol. 14, no. 3, pp. 1032–1043, 2023.

[12] E. Elakiya, R. Kanagaraj, T. Paturu, S. D. Nivethika, and R. S. Kanmani, "Text feedback classification using machine learning techniques," in Proc. 2nd Int. Conf. Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp. 933–939, doi:10.1109/ICECAA58104.2023.

[13] P. H. Vuong, T. T. Dat, and T. K. Mai, "Stock-price forecasting based on XGBoost and LSTM," Systems Science & Control Engineering, vol. 10, no. 1, pp. 835–847, 2022.

[14] R. F. Neves, "Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets," Expert Systems with Applications, vol. 125, pp. 181–194, 2019.

[15] Akinapalli, S. (2024). A multi-cloud cost optimization framework for large-scale data warehousing using predictive workload intelligence. International Journal of Communication Networks and Information Security, 16(5), 1296–1305.