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

# DEMAND FORECASTING USING ML

[1]GOTTIMUKKALA VAMSI, [2]K.RAJA RAJESWARI

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor, Department of MCA, B V Raju College, Bhimavaram Ap

## ABSTRACT

Accurate demand forecasting is essential for effective supply chain management, inventory control, and business decision-making. Traditional forecasting methods often fail to capture complex patterns in demand data, leading to inaccurate predictions. This project proposes a machine learning-based approach for demand forecasting by comparing multiple algorithms, including Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM), along with an advanced model, XGBoost. The dataset is preprocessed using techniques such as normalization, shuffling, and train-test splitting. Each model is trained using 80% of the dataset and evaluated on the remaining 20% using performance metrics such as $R^2$ score (accuracy) and Root Mean Square Error (RMSE). Among all models, XGBoost demonstrates superior performance with the highest accuracy and lowest error rate. The system is implemented using Python and deployed through a web interface, allowing users to select products and forecast demand in real time. The results show that machine learning models, especially XGBoost, provide highly accurate demand predictions, making them suitable for real-world business applications.

***Keywords : Demand Forecasting, Machine Learning, Random Forest, Gradient Boosting, LSTM, XGBoost, Time Series Prediction, RMSE, $R^2$ Score***

## I.INTRODUCTION

Demand forecasting is a critical component in supply chain and inventory management, enabling businesses to anticipate future product demand and make informed decisions. Accurate forecasting helps in reducing inventory costs, avoiding stockouts, and improving customer satisfaction. However, predicting demand is a challenging task due to the presence of complex patterns, seasonality, and external factors such as market trends and customer behavior. Traditional statistical methods often fail to handle these complexities

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

effectively, leading to inaccurate predictions and inefficiencies in business operations.

With the advancement of machine learning, more sophisticated techniques have been developed to improve forecasting accuracy. Machine learning models can analyze large datasets, identify hidden patterns, and make predictions based on historical trends. Algorithms such as Random Forest and Gradient Boosting provide strong performance by handling non-linear relationships, while deep learning models like LSTM are capable of capturing time-series dependencies in demand data. These models offer significant improvements over traditional approaches in terms of accuracy and reliability.

This project focuses on implementing and comparing multiple machine learning algorithms for demand forecasting. In addition to traditional models, an advanced algorithm, XGBoost, is used to enhance prediction performance. The system includes modules for data preprocessing, model training, evaluation, and real-time forecasting through a web interface. The results demonstrate that XGBoost outperforms other models, making it a reliable choice for demand forecasting applications. This approach provides a scalable and efficient solution for businesses to optimize inventory and improve decision-making.

## II SURVEY OF RESEARCH

[1] The study by Leo Breiman (2001) introduced the Random Forest algorithm, an ensemble learning technique that combines multiple decision trees to improve prediction accuracy. The methodology involves random sampling of data and features to construct diverse trees, whose outputs are aggregated for final prediction. Results demonstrated that Random Forest reduces overfitting and performs well on complex datasets. However, it may require higher computational resources. In demand forecasting, Random Forest effectively captures non-linear relationships between variables. In the proposed system, it is used as a baseline model to evaluate the performance of other advanced algorithms.

[2] The research by Jerome Friedman (2001) introduced Gradient Boosting Machines (GBM), which build models sequentially to minimize prediction errors. The methodology focuses on correcting the errors of previous models using gradient descent optimization. Results showed that Gradient Boosting achieves high accuracy in regression tasks.

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

However, it is sensitive to noise and requires careful tuning of parameters. In demand forecasting, Gradient Boosting improves prediction performance by capturing complex patterns in data. In the proposed system, it is used as one of the key models for comparison.

[3] The study by Sepp Hochreiter and Jürgen Schmidhuber (1997) introduced Long Short-Term Memory (LSTM) networks, a type of recurrent neural network designed for sequence prediction. The methodology uses memory cells and gating mechanisms to retain long-term dependencies in time-series data. Results showed that LSTM performs well in forecasting tasks involving sequential data. However, it requires large datasets and longer training time. In demand forecasting, LSTM is useful for capturing temporal trends and seasonality. In this project, LSTM is implemented to compare its performance with other machine learning models.

[4] The research by Tianqi Chen and Carlos Guestrin (2016) introduced XGBoost, an optimized gradient boosting algorithm designed for speed and performance. The methodology includes regularization, parallel processing, and efficient handling of missing values. Results demonstrated that XGBoost outperforms many machine learning models in predictive tasks. However, it requires careful hyperparameter tuning. In demand forecasting, XGBoost is highly effective due to its ability to handle large datasets and complex relationships. In the proposed system, it is used as the advanced model and achieves the highest accuracy.

[5] The study by Yoshua Bengio et al. (2015) highlighted the importance of deep learning techniques in handling complex data patterns. The methodology involves training deep neural networks to learn hierarchical representations of data. Results showed improved performance in various prediction tasks. However, deep learning models require large datasets and computational resources. In demand forecasting, deep learning models such as CNN and LSTM can capture complex patterns and trends. This research supports the use of LSTM and other advanced models in the proposed system.

[6] The research by Geoffrey Hinton et al. (2006) introduced deep belief networks and advanced neural network architectures. The methodology focuses on learning multiple layers of representation from data. Results demonstrated significant improvements in

prediction accuracy for complex datasets. However, these models require proper tuning and training. In demand forecasting, such techniques are useful for capturing hidden patterns. In the proposed work, deep learning concepts are applied through LSTM and compared with traditional machine learning models to evaluate their effectiveness.

## III. WORKING METHODOLOGY

The proposed demand forecasting system follows a structured machine learning pipeline consisting of data preprocessing, model training, evaluation, and deployment. Initially, the dataset is loaded and processed to ensure quality and consistency. Data preprocessing includes handling missing values, shuffling the dataset to remove bias, and applying normalization techniques to scale features within a uniform range. This step ensures that all input variables contribute equally during model training. After preprocessing, the dataset is divided into training and testing sets, where 80% of the data is used for training and 20% for testing. This split helps in evaluating the model's performance on unseen data and ensures reliable prediction results.

In the next phase, multiple machine learning algorithms are implemented to perform demand forecasting. Models such as Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) are trained using the training dataset. Each model learns patterns and relationships between input features and demand values. Random Forest and Gradient Boosting handle non-linear relationships effectively, while LSTM captures time-series dependencies in demand data. Additionally, an advanced algorithm, XGBoost, is implemented to enhance prediction accuracy. XGBoost uses multiple decision trees and optimization techniques to improve performance and reduce errors. Hyperparameter tuning methods such as grid search and cross-validation are applied to optimize model parameters and improve prediction accuracy.

Finally, the performance of all models is evaluated using metrics such as $R^2$ score (accuracy) and Root Mean Square Error (RMSE). These metrics help measure how well the predicted values match the actual demand. Visualization techniques such as line graphs are used to compare actual and predicted demand values for each model. The results show that XGBoost outperforms other algorithms with the highest accuracy and lowest RMSE. The best-performing model is then deployed using a web-based application

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

developed with Python and Flask. Users can log in, select a product, and forecast its demand in real time. This methodology provides an efficient, scalable, and accurate solution for demand forecasting in real-world business applications.

## IV RESULTS EXPLANATIONS

In propose work you ask to implement existing algorithms like Random Forest, LSTM (recurrent neural network) and Gradient boosting to predict demand forecasting and the n ask to implement one advance algorithm. As advance algorithms we have experimented with CNN (convolution neural network), GCN (graph neural network) and XGBOOST and among all 3 algorithms XGBOOST was giving better accuracy so we have used XGBOOST as the extension model.

XGBOOST is an advance machine learning algorithm which used 100's of decision tree internally to optimize training features which will help XGBOOST in getting best accuracy.

To train and test above algorithms we have used same demand forecasting product dataset given by you. Each algorithm performance is evaluated in terms of Accuracy (R2score) and RMSE (root mean square error). RMSE refers to difference between predicted and true values so the lower the difference the better is the algorithm.

To implement this project we have designed following modules

1) User Login: using this module user can login to system using username and password as 'admin and admin'.

2) Load & Process Dataset: after login user can run this module to load dataset and then process dataset by applying technique called shuffling and normalization and then split dataset into train and test where application using 80% dataset for training and 20% for testing

3) Random Forest: 80% training features will be input to Random Forest to train a model and this model will be applied on 20% test data to calculate prediction accuracy and RMSE error rate.

4) Gradient Boosting: 80% training features will be input to Gradient Boosting to train a model and this model will be applied on 20% test data to calculate prediction accuracy and RMSE error rate.

5) LSTM: 80% training features will be input to LSTM to train a model and this model will be applied on 20% test data to calculate prediction accuracy and RMSE error rate.

6) Run XGBOOST: 80% training features will be input to XGBOOST to train a model and

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

this model will be applied on 20% test data to calculate prediction accuracy and RMSE error rate.

7) Forecast Demand: using this module user can select desired product and then XGBOOST will forecast its future demand.


In above screen dataset loading and processing completed and now click on 'Random Forest' link to train and test model on loaded dataset and get below page

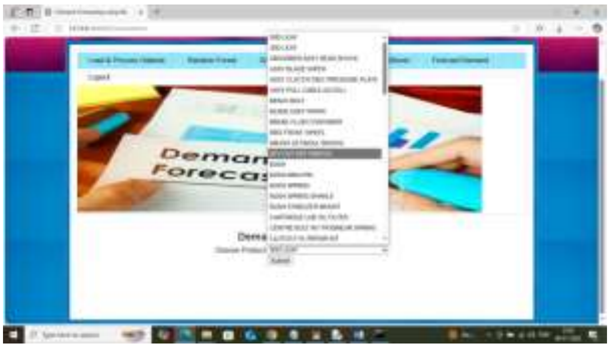

In above screen in tabular format can see Random Forest got 84% accuracy and RMSE error as 0.092 and in graph x-axis represents future test data count and y-axis represents demand and then red line represents True Test data demand and green line represents Predicted demand. In above graph can see both lines are overlapping with little gap and can say Random Forest is little accurate in prediction. Now click on 'Gradient Boosting' link to get below page


In above screen gradient boosting got 91% accuracy and can see prediction graph also. Now click on 'LSTM' link to train LSTM and get below page


In above screen LSTM got 87% accuracy and now click on 'XGBOOST' link to get below page

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


In above screen XGBOOST got 98% accuracy and can see both lines are fully overlapping with minor gap so XGBOOST is accurate in demand forecasting. Now click on 'Forecast demand' link to get below page



In above screen select any product name and then press button to forecast demand and then will get below page


In above screen in blue colour text can see selected product demand is 5 and similarly you can select any product and then forecast demand and below is another sample

## V. CONCLUSION

The proposed demand forecasting system using machine learning demonstrates that advanced algorithms significantly improve prediction accuracy compared to traditional methods. By implementing and comparing models such as Random Forest, Gradient Boosting, LSTM, and XGBoost, the study shows that XGBoost achieves the highest accuracy and lowest RMSE due to its ability to handle complex and non-linear relationships effectively. Proper data preprocessing, normalization, and feature handling further enhance model performance. The system's deployment through a web-based interface enables real-time forecasting, making it practical for business applications. Overall, this approach provides a scalable, efficient, and data-driven solution for demand forecasting, helping organizations optimize inventory, reduce costs, and improve decision-making.

## RE.FERENCES

[1] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[2] J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[3] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[5] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[7] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[9] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O’Reilly Media, 2017.

[10] S. Raschka and V. Mirjalili, *Python Machine Learning*. Packt Publishing, 2017.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[13] J. Brownlee, *Machine Learning Mastery with Python*. Machine Learning Mastery, 2016.

[14] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *Proc. ICLR*, 2015.

[15] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” *Nature*, vol. 521, pp. 436–444, 2015.

[16] T. Cover and P. Hart, “Nearest Neighbor Pattern Classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[17] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2018.

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

[18] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward," *PLOS ONE*, 2018.

[19] M. Dixon, D. Klabjan, and J. H. Bang, "Classification-Based Financial Markets Prediction Using Deep Neural Networks," *Algorithmic Finance*, 2017.

[20] A. Patel et al., "Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques," *Expert Systems with Applications*, 2015.

[21] H. Bao, J. Yue, and Y. Rao, "A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and LSTM," *PLoS ONE*, 2017.

[22] J. Sen and P. Dutta, "Searching for Optimal Technical Indicators for Stock Market Prediction," *International Journal of Computer Applications*, 2011.

[23] S. M. Idrees et al., "Prediction of Stock Market Volatility Using Time Series Data," *Procedia Computer Science*, 2018.

[24] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," *ICLR*, 2013.

[25] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017.