© 2025 IJRTI | Volume 10, Issue 1 January 2025 | ISSN: 2456-3315

ML Driven Inventory Management System & Supply Chain Optimization: A Survey Paper

[1] Sanket Pimprikar, [2] Neha Patkulkar, [3] Om Varma, [4] Abhinav Ughade, [5] Dr. Vandana Rohokale

[1,2,3,4] Student, Bachelor of Engineering
[5] Head of Department,
[1,2,3,4,5] Department of Electronics and Telecommunication,
[1,2,3,4,5] Sinhgad Institute of Technology and Science, Pune, Maharashtra, India
[1]sanketpimprikar2811@gmail.com, [2]patkulkarneha04@gmail.com, [3]omvarma13@gmail.com,
[4]abhiughade1422@gmail.com, [5]hodetc_sits@sinhgad.edu

***Abstract*—Precise Demand Forecasting of Components in the Manufacturing industry is critical to supply chain management, as various factors affect the demand for the product. To regulate and maintain the buffer stock of components in the Inventory is necessary. This project focuses on reducing the downtime in the manufacturing process by predicting the demand for components and providing analysis on the buffer stock to be maintained to avoid downtime and overspending of company resources to acquire the components, which have a volatile demand in the industry. The project focuses on inventory optimization, cost reduction & reducing downtime. This paper aims to present an integrated forecasting strategy for intermittent or volatile demand of components in the manufacturing industry by comparing the accuracy of various machine learning models like Random Forest, XGBoost, and LSTM. Enhancing supply chain strategies by providing invaluable insights into demand forecasting of the components is the aim of this machine learning model for informed decision-making.**

***Index Terms*—Inventory Optimization, Inventory Management System, Machine Learning, XGBoost, Random Forest, LSTM, Demand Forecasting, Supply chain management, Time Series Forecasting, Cost & Downtime Reduction, ensemble learning.**

I. INTRODUCTION

In today's competitive business environment, time series forecasting methods enable companies to make informed decisions, boost productivity, minimize downtime, and drive overall profitability [1]. Through predictive analysis, businesses gain critical insights into inventory needs, ensuring they meet production demands while avoiding excess holding costs. Holding costs refer to the expenses associated with storing inventory that is not currently active in the production or sales cycle. By leveraging demand forecasting and predictive analysis, companies can make data-driven decisions to enhance operational efficiency, maximize profits, and ensure a seamless supply chain. Effective inventory management plays a key role in improving operational performance, maintaining production continuity, and enhancing customer satisfaction.

The manufacturing industries often face challenges like stockouts, due to the volatile demand in the industry. Many factors come into play when predictive analysis is to be done, like trends & seasonality, volatile demand, excess or shortage of stock. And these consequences can become even more serious when they are amplified by the "Bullwhip effect" This effect, illustrated in the Fig. 1, describes the phenomenon whereby a small variation in demand at customer level will tend to increase throughout the Supply Chain, consequently, their operation inefficiency [2]. An AI-driven demand forecasting system is crucial for managing demand fluctuations and mitigating the Bullwhip effect. This effect arises from the lead time between placing and receiving orders, as well as potential forecast adjustments. To be effective, each order must accommodate not only the current period's demand variations but also any shifts in predicted demand occurring during the lead time [2].

Traditional forecasting methods struggle while dealing with non-linear or complex data, which in turn affects the prediction negatively, leading to cause a Bullwhip effect. The authors compare these traditional forecasting methods with the new machine learning forecasting methods. Review of recent studies revealed a common theme of exploring diverse ensemble learning approaches to improve forecast performance compared to traditional and machine learning methods [1]. To address the challenge of data non-linearity, complexity, high-dimensionality, and at the same time taking trends and seasonality into consideration, using of ensemble learning techniques are further discussed in this survey paper [3].

The machine learning models discussed and proposed in this model addresses these challenges and also produces better results compared to the traditional approaches. The hybrid of Random Forest, XGBoost and LSTM give off better results while addressing the challenges. Ensemble of these methods have shown effective, hence leading to inventory optimization, cost reduction, and downtime mitigation [1]-[3].

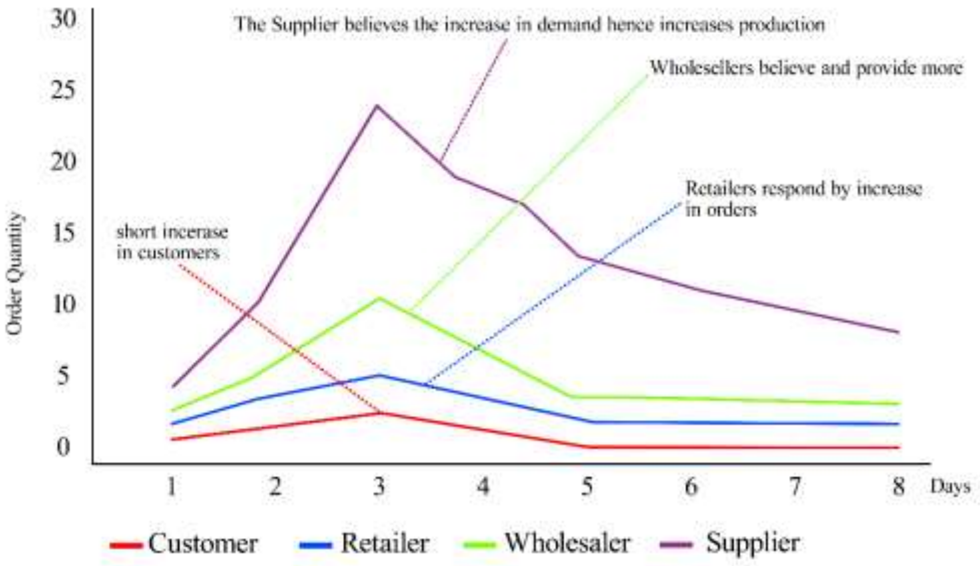

Figure. 1 Illustration of the Bullwhip Effect Across Supply Chain Tiers

II. Background And Related Work

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts. True Type 1 or Open Type fonts are required. Please embed all fonts, in particular symbol fonts, as well, for math, etc.
The recent studies have shown that statistical models for time series, like ARIMA (autoregressive integrated moving average), Holts-Winters moving average, exponential smoothing that are appropriate and satisfactory in many applications, since they can model data and make predictions on relatively large time spaces (year, month...), provided that the time series is totally linear [4].

Machine learning models used in previous studies included Random Forest, XGBoost, LSTM which performed better in most cases, specifically XGBoost. Random Forests provide several advantages, including improved accuracy, resistance to overfitting, and the ability to interpret results through feature importance analysis. They are versatile, effectively handling both classification and regression tasks. This approach is especially well-suited for complex problems, outperforming individual decision trees and delivering reliable performance across a wide range of machine learning applications [1]. Forecasting accuracy when compared with multiple models like ELM, and Decision Tree Random Forest and XGBoost showed significant accuracy and hence were chosen as the predictive analysis models [5]. The studies have also showed that grid search method to select the best combinations of model hyperparameters to better capture the characteristics of the time series. And proved that the fitted LSTM model is capable of producing more accurate and efficient predictions, and discussed the future scope of implementing LSTM model with another Machine Learning model to improve accuracy [4].

LSTM model can be used as a substitute for Prophet model, where the factors of trends and seasonality are to be considered, where the hybrid of Prophet and XGBoost were proved to be most effective models in terms of evaluation metrics [6]. In the supply chain demand prediction, it is observed that, feature engineering is not given much priority, which if done efficiently can affect the accuracy of model in a positive way. Accurately pre-processed data along with the hybrid of the Machine Learning models, such as XGBoost and LightGBM, outperform the models when implemented individually [7].

The extensive description of research gaps identified during this study are discussed in the next section.

III. Research Gap And Analysis

In this section, the authors present key findings derived from an extensive study of demand forecasting models and methodologies. Through a detailed review of existing literature and comparative analyses, the authors identified several research gaps that highlight opportunities for improvement in current forecasting techniques. The tables below provide a comprehensive summary of the findings, showcasing specific gaps in the integration of hybrid models, real-time forecasting capabilities, and advanced feature engineering. Additionally, the authors outline the scope for addressing these gaps, offering actionable insights and strategies to enhance demand forecasting accuracy and supply chain optimization.

TABLE 1. A SUMMARY OF RESEARCH REVIEWED IN LITERATURE SURVEY

Paper No.	Research Focus	Key Findings	Insights
[1]	B2B demand prediction	Ensemble models beat SARIMA; robust data prep needed	Test in diverse sectors, add real-time features
[2]	Logistics demand	LSTM better handles complex patterns than ARIMA	Consider adding external data sources
[3]	Warehouse stock forecasting	LSTM reduces stock issues; outperforms ARIMA	Explore hybrid neural models
[4]	SCM optimization	LSTM outperforms simpler models; optimized via grid search	Combine LSTM with other ML techniques
[5]	ML models for SCM	XGBoost and Random Forest are highly accurate	Emphasizes need for real-time models
[6]	Indian supply chains	ML methods paired with FMEA for risk management	Integrate real-time risk handling
[7]	Aircraft supply chain	Ensemble methods improve forecasting by 18-40%	Broaden application beyond aerospace

Table 1 describes the research focus and work done in the papers. The key insights and notes are drawn out from this table which shows the correct approach to address the problem of demand forecasting. Usage of Hybrid models and use of LSTM along with other Machine Learning techniques is suggested in almost all the papers. The need for real time models is also reflected in most of the papers.

TABLE 2. SUMMARY OF METHODS & KEY INSIGHTS

Paper No.	XGBoost	LSTM	Random Forest	Traditional Methods	Metrics	Notes
[1]	✔	✖	✔	SARIMA	RMSE, WMAPE	Ensemble models outperform SARIMA; highlights preprocessing.
[2]	✖	✔	✖	ETS	RMSE, SMAPE	LSTM is superior to ARIMA; hybrid models suggested.
[3]	✖	✔	✖	ARIMA	RMSE, MAE	LSTM outperforms ARIMA; no ensemble methods explored.
[4]	✖	✔	✖	Exp. Smoothing	RMSE, SMAPE	Optimized LSTM outperforms RNN; hybrid techniques proposed.
[5]	✔	✖	✔	Decision Tree, ELM	R^2, RMAE	XGBoost shows high accuracy; emphasizes feature engineering.
[6]	✔	✖	✔	FMEA	WMAPE, MAE	Combines ML and risk management; ensemble models excel.
[7]	✔	✖	✖	None	MSE, MAE	Non-overlapping windows enhance accuracy; useful for aerospace.

Table 2 focuses on the methods used in the surveyed papers and the key findings include that Hybrid of XGBoost/Random Forest and LSTM is not used in any case and suggestion of attempting this Hybrid model is a recurrent theme in all of the surveyed papers.

List of key research gaps and objectives that addresses these gaps:

1. **Hybrid Models Integration:** Existing studies focus on either ensemble methods like XGBoost or deep learning models like LSTM, but not both together.

   Scope for improvement: By integrating ensemble methods and LSTM, a hybrid approach could offer a better balance between handling high-dimensional data and capturing non-linear patterns in volatile demand.

2. **Real-Time Demand Forecasting:** Current research primarily uses historical data for demand forecasting without addressing real-time applications.

   Scope for improvement: Implement real-time forecasting models that can adapt dynamically to fluctuating demand signals, helping to reduce downtime and optimize inventory in real-time.
3. **Feature Engineering and Custom Variables**: Limited attention is given to advanced feature engineering tailored to manufacturing component demand, particularly in capturing external factors like seasonal trends, weather, or economic conditions.

Scope for improvement: Enhancing the feature engineering process by creating custom variables that reflect the volatility of manufacturing demand could significantly improve prediction accuracy.

4. **Hyperparameter Tuning for Model Optimization:** Papers rely on grid search for tuning hyperparameters, which is computationally expensive and time-consuming.

   Scope for improvement: Implement advanced hyperparameter tuning techniques like Bayesian optimization to find the most effective configurations for both LSTM and ensemble methods.

5. **Inventory Optimization and Buffer Stock:** Correct analysis of predicted data to interpret the demand and buffer stock with use of Economic Order Quantity (EOQ) Model is not provided in the papers.

   Scope for improvement: Implement appropriate formulas for interpretation of the predicted data to obtain more accurate and efficient results for inventory optimization.

IV. Machine Learning Models For Demand Forecasting

To develop a **Machine Learning-driven Inventory Management System and optimize the supply chain**, various Machine Learning models need to be implemented and evaluation metrics are to be set to compare the operability and accuracy of the Hybrid Models.

Various Machine Learning Models studied and that are to be implemented are:

1. **Random Forest:**
   Random Forest is an ensemble learning algorithm that constructs multiple decision trees using randomly selected subsets of data and features. The final prediction is obtained by averaging the outputs of these trees in regression tasks, making the model stable and robust. It effectively handles high-dimensional data, captures non-linear patterns, and mitigates overfitting through its bagging technique. This algorithm is particularly suited for demand forecasting in datasets with intricate feature interactions, such as seasonality and economic factors, and works well for moderately volatile demand without significant time dependencies.

2. **XGBoost:**
   XGBoost, short for Extreme Gradient Boosting, is a gradient boosting framework designed for efficiency and performance. It builds trees sequentially to minimize prediction errors, incorporating L1 and L2 regularization to reduce overfitting and improve generalization. Its highly optimized implementation allows for parallel processing and handles missing values effectively. XGBoost is especially suitable for structured datasets with non-linear relationships and is well-suited for capturing complex interactions between variables in demand forecasting scenarios.

3. **LSTM (Long Short-Term Memory):**
   LSTM is a specialized recurrent neural network (RNN) architecture designed to process sequential data by retaining long-term dependencies through its memory cell and gated mechanism. It is particularly effective for time-series data, capturing patterns and trends influenced by sequential factors. LSTM addresses the vanishing gradient problem common in traditional RNNs, making it highly suitable for forecasting demand fluctuations that exhibit seasonality or temporal dependencies, as well as incorporating external time-dependent factors like holidays or promotions.

4. **Hybrid Models:**
   4.1 Hybrid of LSTM & Random Forest:
   This hybrid approach combines LSTM's ability to model sequential dependencies in time-series data with Random Forest's strength in analysing complex non-temporal features. LSTM captures long-term temporal trends, while Random Forest handles categorical and numerical variables effectively. This combination enhances prediction accuracy and is well-suited for datasets that feature both temporal and non-temporal factors influencing demand.

   4.2 Hybrid of LSTM & XGBoost:
   This hybrid integrates LSTM's time-series modelling capabilities with XGBoost's proficiency in managing non-linear feature interactions. LSTM processes sequential patterns and dependencies, while XGBoost captures intricate variable relationships with precision. The synergy of these two methods makes this hybrid model particularly effective for demand forecasting in scenarios with highly volatile demand patterns influenced by both temporal and feature-driven factors.

V. Methodology

In this section, the authors outline the proposed methodology for implementing hybrid models to address the challenges of demand forecasting in the manufacturing industry. The methodology encompasses data preprocessing, feature engineering, model training, and evaluation strategies tailored for volatile and intermittent demand patterns. By combining the strengths of different machine learning algorithms, such as LSTM and ensemble models like Random Forest and XGBoost, the hybrid approach aims to leverage both temporal dependencies and non-linear feature interactions. This approach ensures robust predictions and optimizes inventory management.

The methodology is structured to systematically handle the complexities of demand forecasting, including managing large datasets, addressing missing values, and tuning hyperparameters to enhance model performance. Drawing insights from existing

studies (e.g., Javed et al., 2024; El Khaili et al., 2022), the hybrid model leverages time-series patterns through LSTM and augments them with ensemble learning techniques for feature-driven accuracy.
This integrated strategy ensures that the hybrid model adapts to both historical trends and external influences, enabling precise and actionable forecasts for supply chain optimization.

1. **Data Preparation**
Effective demand forecasting begins with comprehensive data preparation, which involves cleaning, handling missing values, and performing feature engineering. This process ensures the dataset is ready for training machine learning models. Key steps include removing irrelevant or redundant features and creating derived variables to capture seasonality or external influences like weather or promotions (as seen in Javed et al. [5]).

2. **Model Training**
Model training involves exposing machine learning algorithms to a training dataset to enable them to learn relationships between input features and target variables. Algorithms such as Random Forest, XGBoost, and LSTM are particularly effective for demand forecasting due to their ability to model complex, non-linear relationships. Parameter tuning is performed to optimize each model, ensuring accurate and reliable predictions (inspired by Hodžić et al. [3] and Shao & Song [7]).

3. **Evaluation Metrics**
To assess the performance of forecasting models, metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) are commonly used. These metrics provide insights into prediction accuracy and model fit. The importance of these metrics is underscored in the work of Agrawal et al. [6], where they were instrumental in evaluating the suitability of models for Indian SCM.

4. **Implementation of Hybrid Models**
Hybrid models combine the strengths of multiple algorithms to enhance prediction accuracy. For instance:
- Hybrid LSTM & Random Forest: Leverages LSTM for capturing temporal dependencies and Random Forest for handling categorical and numerical data interactions.
- Hybrid LSTM & XGBoost: Combines LSTM's time-series capabilities with XGBoost's gradient-boosting techniques to capture complex relationships and ensure robust predictions.

These approaches enable models to handle both temporal patterns and complex interactions, as evidenced in studies like El Khaili et al. [2] and Lim et al. [1].

VI. Results And Comparative Analysis

In this section, the authors present the results of the study, including a comparative analysis of various models and their performance metrics. The authors also discuss the performance of both individual and hybrid models, highlighting key observations and insights derived from the analysis.

TABLE 3. COMPARATIVE TABLE OF MODEL PERFORMANCE & METRICS

Model	Metrics Used	Performance	Observations
ARIMA	MSE, RMSE, MAE, MAPE	- MSE: 1.67 - MAE: 317,530.90 - MAPE: 0.75	Performs well for linear time-series patterns but struggles with capturing non-linear and volatile demand, resulting in higher error metrics.
Random Forest	R^2, RMAE	- R^2: 0.965 - RMAE: 8.5%	Strong at handling complex feature interactions but lacks specialization for sequential dependencies in time-series data.
XGBoost	R^2, RMAE	- R^2: 0.979 - RMAE: 8%	High accuracy due to advanced gradient-boosting techniques; excels in capturing complex, non-linear patterns in structured datasets.
LSTM (Monolayer)	MSE, RMSE, MAE, MAPE	- RMSE: 4,707.78 - SMAPE: 0.027	Effective for time-series data, capturing sequential dependencies better than Random Forest and ARIMA but slightly less accurate than multilayer.
LSTM (Multilayer)	MSE, RMSE, MAE, MAPE	- RMSE: 4,487.32 - SMAPE: 0.026	Demonstrates superior performance in modelling complex sequential patterns and dependencies, achieving the lowest error metrics overall.

Table 3. provides a comparative analysis of commonly used models for demand forecasting, highlighting their performance and key observations. ARIMA is effective for capturing linear trends but struggles significantly with highly volatile or non-linear demand patterns, resulting in higher error metrics. On the other hand, Random Forest and XGBoost excel at modelling non-linear relationships and managing complex, high-dimensional data. However, both models are less effective for time-series forecasting because they lack the capability to model sequential dependencies. In contrast, LSTM, which is specifically designed for sequential data, is highly effective in capturing long-term dependencies and handling volatile demand patterns. Despite its advantages, LSTM requires substantial computational resources and meticulous hyperparameter tuning to achieve optimal performance.

Hybrid models, such as combining LSTM with Random Forest or XGBoost, can outperform individual models in volatile demand situations. LSTM captures temporal dynamics, while Random Forest or XGBoost handles complex non-linear interactions, making these hybrid approaches more suitable for datasets that involve both time-dependent and non-linear elements.

VII. Discussion

1. **Model Suitability**
   - Random Forest excels at handling high-dimensional data and capturing non-linear relationships, but struggles with sequential dependencies.
   - XGBoost provides high accuracy for structured datasets with complex non-linear relationships but lacks temporal modelling capabilities.
   - LSTM is ideal for volatile and complex time-series data due to its ability to capture long-term dependencies and sequential patterns, making it highly effective in forecasting demand spikes.
2. **Comparative Advantages of Hybrid Models:**

   Hybrid models, such as combining LSTM with Random Forest or XGBoost, leverage the strengths of both methods. LSTM models sequential dependencies, while ensemble methods like Random Forest and XGBoost capture non-linear feature relationships. This synergy enhances forecast accuracy, particularly in supply chain scenarios where both temporal and external factors (e.g., promotions or economic fluctuations) influence demand. By incorporating both sequential and non-sequential data, hybrid models can more effectively address volatile demand, improving supply chain forecasting and optimization.

3. **Limitations**
   - Random Forest and XGBoost require substantial feature engineering and struggle with modelling sequential data effectively.
   - LSTM demands high computational resources and careful tuning, particularly with smaller or noisy datasets, making it resource-intensive.
   - Hybrid models, while powerful, introduce added complexity and require significant computational power and expertise, which may reduce interpretability compared to simpler models, presenting challenges for actionable decision-making in supply chain management.

VIII. Conclusion

Machine Learning (ML)-driven demand forecasting is crucial for optimizing inventory management and supply chains, especially where traditional models fail to handle non-linear and volatile demand efficiently. By leveraging ML algorithms like XGBoost, Random Forest, and LSTM, businesses can achieve greater forecasting accuracy, reduce operational costs, and minimize the risk of stockouts or excess inventory.

XGBoost and Random Forest excel at managing complex, high-dimensional data with robustness and interpretability, while LSTM effectively captures sequential dependencies, making it ideal for time-series data. Hybrid models that integrate these strengths offer even higher accuracy, cost savings, and adaptability to real-world complexities, thus providing a comprehensive solution for supply chain optimization.

Future work should focus on real-time forecasting capabilities, adaptive algorithms, and refined feature engineering to further enhance model performance. Incorporating external variables, such as economic indicators, will also be key to developing more resilient and efficient supply chain systems.

References

[1] P. Y. Lim, W. Y. Chin and Y. H. Kong, "Exploring the Effectiveness of Ensemble Learning for Business-to-Business (B2B) Demand Forecasting: An Empirical Analysis," 2024 3rd International Conference on Digital Transformation and Applications (ICDXA), Kuala Lumpur, Malaysia, 2024, pp. 226-230, doi: 10.1109/ICDXA61007.2024.10470870

[2] Terrada, L., El Khaili, M., & Ouajji, H. (2022). Demand forecasting model using deep learning methods for supply chain management 4.0. *International Journal of Advanced Computer Science and Applications*, *13*(5). K. Hodžić, H. Hasić, E. Cogo and Ž. Jurić, "Warehouse Demand Forecasting based on Long Short-Term Memory neural networks," 2019 XXVII International Conference on Information, Communication and Automation Technologies (ICAT), Sarajevo, Bosnia and Herzegovina, 2019, pp. 1-6

[3] K. Hodžić, H. Hasić, E. Cogo and Ž. Jurić, "Warehouse Demand Forecasting based on Long Short-Term Memory neural networks," 2019 XXVII International Conference on Information, Communication and Automation Technologies (ICAT), Sarajevo, Bosnia and Herzegovina, 2019, pp. 1-6, doi: 10.1109/ICAT47117.2019.8939050.

[4] El Filali, A., Lahmer, E. H. B., El Filali, S., Kasbouya, M., Ajouary, M. A., & Akantous, S. (2022). Machine Learning Applications in Supply Chain Management: A Deep Learning Model Using an Optimized LSTM Network for Demand Forecasting. *International Journal of Intelligent Engineering & Systems*, *15*(2).

[5] S. Javed and R. Akhtar, "Data driven Approaches for Demand Forecasting in Supply Chain for Business Decisions," 2024 5th International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 2024, pp. 1-7, doi: 10.1109/ICACS60934.2024.10473234

[6] P. N. Agrawal, S. V. Bhutare, D. Pramod and V. B. S, "Empowering Supply Chains in India: Precision Demand Forecasting and Machine Learning Strategies for Resilience and Sustainability in Dynamic Environments," 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2024, pp. 1-6, doi: 10.1109/ICRITO61523.2024.10522458

[7] C. Shao and J. Song, "An Adaptive Time Series Forecasting Model for Aircraft Component Supply Chain Demand Prediction," 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China, 2024, pp. 1020-1023, doi: 10.1109/AINIT61980.2024.10581522