# THE INVENTORY MANAGEMENT AND DEMAND FORECASTING SYSTEM

Rajkumar.S 1, Abinesh.S 2,Sanjith.E.S 3

 rajkumar@velammal.edu.in1,abineshsg2@gmail.com2, Sanjithesacc1@gmail.com 3

Department of Information Technology, Velammal Engineering College.

# ABSTRACT

Accurate demand forecasting and efficient inventory management are the secret to maximizing supply chain operations. Demand forecasting is an estimate of future product demand based on historical data and other inputs. Accurate forecasts help companies to prevent stockouts and overruns, which enables improved planning for production, inventory, and logistics. Companies hold additional stock as a buffer to respond to fluctuations in demand and supply. Finding the correct amount of additional stock is the secret to making customers satisfied with minimal excess inventory and carrying costs. Traditional forecasting models are concerned primarily with minimizing forecast errors but do not consider the direct impact they have on inventory costs. This research demonstrates a new demand forecasting and safety stock optimization method using the Kaggle-Walmart Sales Forecasting Dataset. Our method involves a unique Key . Performance Indicator (KPI) that gives more precise forecasting and takes into account the cost of inventory. Our research also develops a more efficient safety stock optimization approach by taking into account supply chain reliability and seasonality in the demand trends derived from historical sales data. Based on real data application, we demonstrate that our method makes forecasting more accurate and inventory more effective and reduces stockouts and overstocking hazards.

**KEYWORDS** : Supply Chain ,Planning Demand, Forecasting Statistical ,Forecasting Methods ,Inventory Optimization

# 1. INTRODUCTION

Accurate demand forecasting is extremely vital in inventory management in the modern supply chain systems. With the capacity to predict future demand, firms can improve production planning, manage inventory levels more effectively, and make distribution easier, saving costs and making customers happier. Supply chains are typically faced with unexpected changes in demand, especially during peak seasons, which may result in shortages of stock, overstocking, or loss of money if not handled effectively. We apply the Kaggle-Walmart Sales Forecasting Dataset in this research to develop a more effective demand forecasting and stock control approach. We apply Linear Regression and Random Forest Regressor, robust machine learning models, to achieve higher accuracy levels in the prediction and identify subtle patterns of sales. We also apply data analysis techniques such as finding trends, seasonality testing, and statistical modeling to gain insightful conclusions from historical sales. Our system integrates these techniques to provide companies accurate weekly sales forecasts for Walmart. This enables them to make informed decisions, prevents excess waste due to overstocking, and maintains products on shelves when demand is highest. Overall, this process enhances supply chain efficiency, minimizes financial risk, and allows for improved inventory planning.

**2. LITERATURE SURVEY**

Big data and machine learning have revolutionized supply chain management tremendously, presenting new opportunities as well as challenges. Machine learning is very critical in demand forecasting, which helps companies to operate more effectively by handling knowledge tasks. An examination of 79 studies showed that neural networks (27%) and support vector machines (10%) are among the most widely used algorithms. Industrial application (65%) prevails in most uses, whereas agriculture is not yet explored to a great extent (5%). This study points to the need for utilizing machine learning in making data-driven decisions in supply chain operations [1]. The integration of machine learning (ML) with operations research and management science (OR/MS) is an effective paradigm for data-driven decision-making in optimization problems. In contrast to conventional approaches, the method considers auxiliary data outside direct cost-related variables, solving conditional stochastic optimization with incomplete observations. The paper builds the coefficient of prescriptiveness (P) to quantify decision-making efficiency and illustrates its effect on inventory control. In a media distribution firm, the method enhanced operational efficiency by 88%, reflecting its wide applicability and computational effectiveness in reality [2]. Predictive analytics (PA) is playing an increasingly important role in supply chain management (SCM) these days, specifically in demand forecasting. This study compares SARIMA and LSTM models based on 37 months of an Austrian retailer's retail sales data. The findings are that LSTM works better with stable-demand products, while SARIMA works on seasonal products. Moreover, with external variables in SARIMAX, forecasts of promotional products significantly improve. A hybrid model by clustering stores with similar attributes and training SARIMA(X) and LSTM models is proposed to improve forecasting accuracy at the store level [3]. Mass customization is among the most significant characteristics of smart manufacturing, and demand forecasting is one of the most essential drivers of successful Sales and Operations Planning (S&OP). Proper forecasting ensures that the inventory is maximized and the cost of running multiple warehouses is minimized. This research compares ARIMA (an old time series model) with nonlinear network models to identify the best forecasting method. A mathematical modeling-based algorithm is also presented to investigate sales processes and maximize production planning based on anticipated inventory levels[4]. Pricing and revenue optimization of brick-and-mortar retail stores, particularly high-end luxury goods, is an urgent challenge based on seasonal changes in demand and the high level of investment risk. This paper presents a decision-support system for price-based prediction of weekly demand in terms of price, holiday seasons, offers/discounts, and stock keeping unit levels based on regression tree/random forest machine learning algorithms. Integer linear programming is further employed to minimize prices with respect to demand-price interdependencies. The branch & bound and branch & cut are additionally used together with heuristic algorithms for optimizing revenue maximization processes in luxury retailing products [5]. Demand forecasting is important to industries in order to avoid oversupply or stockouts and thereby minimize revenue loss. The study compares ARIMA and LSTM models in forecasting demand for eight dairy products of five production facilities with five years of data. The study compares statistical (ARIMA) and deep learning (LSTM) methods on parameters such as time interval (monthly versus weekly), univariate versus multivariate data, and error measures of models. The study reveals that ARIMA works effectively in stable patterns of demand, and LSTM performs well in seasonality, so both models can be relied on in demand forecasting [6]. Artificial Intelligence (AI) has greatly enhanced decision-making and efficiency through learning business processes, pattern recognition, and data analysis. Though successful in most sectors, AI implementation in Supply Chain Management (SCM) is low. This research examines AI subdomains appropriate for SCM applications, examining previous success and areas where AI can contribute the most. The research indicates possible AI-based solutions for supply chain process optimization and decision-making [7]. Electricity is a major driver of economic growth and social cohesion, needing efficient supply-demand equilibrium. Accurate power consumption prediction is based on identification of influential drivers. In this research, a Convolutional Neural Network (CNN) for Short-Term Load Forecasting (STLF) with a two-dimensional input architecture is proposed to improve prediction accuracy. The suggested CNN model is superior to conventional AI approaches in one-quarter-ahead and 24-hour-ahead prediction, proving to be efficient in energy demand prediction [8]. Weather forecasting is still a difficult research issue, but improved deep learning and big data have made forecasting more accurate. This paper proposes a light deep learning model with stacked LSTM layers to forecast several weather parameters in a multi-input multi-output (MIMO) system. The model is experimented with wind speed, humidity, dew point, and temperature, and cascaded models are more accurate for short-term forecasting than regular LSTM and 1D convolution networks [9]. Demand forecasting is an essential component of contemporary supply chain management as suppliers shift from order-based to forecast-based systems. ARIMA and LSTM methods are used in this research to create rolling forecast models that enhance demand and inventory accuracy. The models are tested using actual IC tray orders in semiconductor production and demonstrate the superiority of LSTM over short-term forecasting and a dramatic decrease in inventory following implementation [10] . This paper proposes a closed-loop supply chain (CLSC) integrated inventory model of a retailer and manufacturer with random demand and returns, where a carbon tax policy is imposed to minimize transport, manufacturing, and storage emissions and the manufacturer makes green technology and take-back investments to capitalize on maximum sustainability. The mathematical model is created to minimize supply chain costs and sensitivity analysis is used to illustrate how variable production rates and optimum collection rates can properly balance cost and emission reductions [11]. This paper proposes a prescriptive business process control model through event-driven process predictions within the process manufacturing sector. Drawing upon predictive analytics and big data, the study also outlines the specificity of nonlinear and cyclical manufacturing processes. It uses an exemplar study from a German steel factory company to show the feasibility of utilizing sensor technology in obtaining and processing data in real time. The paper also stipulates seven core requirements for using this idea and suggests a general prescriptive enterprise system architecture for improving production planning and control[14] . This paper introduces a new approach to customer demand forecasting based on Support Vector Regression (SVR). The approach merges statistical learning theory with nonlinear and linear programming models to develop a general regression function. A process in steps enhances the accuracy of predictions by unearthing underlying demand patterns in historical sales data. The new approach is exemplified and shown to perform effectively in customer demand forecasting[15]. This study illustrates a technique of predicting the quantity of engine oil to be needed in the automotive and industrial lubricant market using an artificial neural network (ANN). The model takes into account important parameters like quality, price, and delivery time to ascertain the impact on demand. It trains and validates the ANN using past sales data and checks the predictions using root mean square error. The results show that the ANN model works well with actual demand, proving it to be efficient in predicting future needs [16]. This research demonstrates a smart approach to enhance blood supply chain management to counteract uncertainty of blood demand, waste, and shortages. The system incorporates machine learning and time series forecasting and has three crucial components: blood demand forecasting, donor classification, and appointment scheduling. The proposed AI solution allows for better management of blood inventory, increasing the volume of blood collected by 11% and reducing waste by 20%. The findings outline how predictive analytics can play a key role in enhancing blood supply chains [18]. This research evaluates the use of the Random Forest model to forecast the stocks of India's leading three fintech companies—Policy Bazar, Paytm, and Niyogin Ltd.—based on precise data. With 293,280 data points from sources such as Moneycontrol and Kotak Securities, this research demonstrates that the model performs optimally, achieving a score of above 95%. This research is unique in its nature since there are no previous studies that employed Random Forest for making fintech stock predictions in India [19]. This research examines ARIMA and LSTM models in forecasting time series, particularly handling changes and incomplete data. LSTM performs better when dealing with small datasets and in the presence of missing data, but ARIMA performs better when dealing with large datasets. The evaluation employs RMSE and runtime to demonstrate that ARIMA is faster to execute on computers, while LSTM performs better in handling incomplete data [20].

# 3. PROPOSED WORK

The proposed system is designed to enhance demand forecasting and inventory optimization for Walmart stores using machine learning techniques. By leveraging historical sales data, economic indicators, and store-specific features, the system aims to accurately predict future weekly sales and support data-driven inventory decisions.

**Fig(1): System Architecture Diagram**



**3.1 Data Collection**

The dataset used in this study is the Walmart Sales Forecasting dataset available on Kaggle, which contains transactional and auxiliary data from multiple stores. The following files are used:

train.csv: Weekly sales data by store and department (used for training)

test.csv: Future sales data to be predicted

features.csv: Contains external features such as temperature, CPI, fuel price, holidays, and markdowns

stores.csv: Store metadata such as store type and size

Data was accessed through the Kaggle API and loaded using Pandas. All files were merged on common columns like Store, Dept, and Date to build a single, structured dataset for modeling.

**3.2 Data Preprocessing and Feature Engineering**

Preprocessing was conducted to clean and enrich the dataset:

Missing Values Handling: Missing values in features such as MarkDown1–5 were imputed with 0. Datetime Features: Year, Month, and Week were extracted from the Date column to capture seasonal trends. Low Stock Flagging: A binary flag was introduced based on store Size (<1000 = low stock). EOQ Calculation: Economic Order Quantity (EOQ) was calculated using standard inventory formula to recommend replenishment quantities.

No categorical encoding was required as the model selected (Random Forest) handles categorical numerical values directly.

**3.3 Model Development**

Two regression models were developed for predicting Weekly\_Sales:

Random Forest Regressor: An ensemble tree-based method suitable for non-linear data. It builds multiple decision trees and averages their output to reduce overfitting and improve generalization. Ridge Regression: A linear model with L2 regularization to handle multicollinearity and prevent overfitting.

Both models were trained using the same features and compared based on evaluation metrics.

**3.4 Hyperparameter Optimization**

To enhance model performance, GridSearchCV was used to perform exhaustive hyperparameter tuning. The grid search explored combinations of the following parameters:

**Random Forest Regressor:**

1. n\_estimators: Number of trees in the forest
2. max\_depth: Maximum depth of trees
3. min\_samples\_split: Minimum samples required to split a node

**Ridge Regression:**

1. alpha: L2 regularization strength

GridSearchCV was executed with 5-Fold Cross-Validation, ensuring that each model was trained and validated on different subsets of the data to avoid overfitting and improve reliability.



**3.5 Model Training and Evaluation**

The dataset was split into 80% training and 20% testing. After hyperparameter tuning, the best models were trained and evaluated using the test set.

Evaluation Metrics:

1. Mean Squared Error (MSE): Measures average squared error
2. R-squared (R²): Indicates how well variance is explained
3. MAPE: Measures prediction error as a percentage (after removing zero actuals)
4. Confusion Matrix (Simulated): Evaluated the model’s ability to classify Low Stock vs. Sufficient Stock using a threshold on predicted sales

**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **MSE** | **R² Score** | **Fixed MAPE** |
| Random Forest | 13,690,506 | 0.9737 | 1.75% |
| Ridge Regression | 475,933,409 | 0.0873 | Invalid (due to extreme values) |

**Inference**: Random Forest outperformed Ridge Regression in all metrics and was selected for the final system implementation.

**3.6 Tools and Implementation**

Google Colab: Cloud platform for development and testing ;Pandas, NumPy: Data processing ;Matplotlib, Seaborn: Visualization ;scikit-learn: Model building, hyperparameter tuning ;Gradio: Lightweight UI for real-time user interaction

**3.7 Interactive UI and Decision Support System**

A Gradio-based user interface was developed to allow non-technical users to interact with the model. Users input a Store and Department number, and the system returns:

● Predicted Weekly Sales

● Recommended EOQ

● Low Stock Alert

****

**This enables store managers and supply chain analysts to:**

**● Predict future demand**

**● Plan purchases efficiently**

**● Avoid stockouts and reduce inventory holding costs**

**3.8 Summary of Final Model Selection**

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Random Forest** | **Ridge Regression** |
| Prediction Power | High (R² = 0.97) | Very Low (R² = 0.08) |
| Error (MSE) | Low |  High |
| Realistic MAPE | 1.75% | Invalid |
| Usability | Non-linear & robust | Too simple for this task |

**Final model used in UI** : Random Forest Regressor

# 4. RESULTS AND DISCUSSION

 Implementation of this demand forecasting and inventory management system has provided valuable insights regarding how to stock, predict sales, and how much to stock. Based on machine learning algorithms such as Linear Regression and Random Forest Regressor, and inventory concepts such as Economic Order Quantity (EOQ), the system enables improved decision-making for a smooth supply chain. The findings indicate that the system effectively prevents stock shortages, reduces overstocking, and restocks stock in the optimal manner.

One major implication of this project is the Low\_Stock flag, which shows which stores are below a certain level of stock. The flag alerts store managers and supply chain planners which stores need to be replenished right away to prevent shortages and meet customers' needs. The system calculates this flag based on the store size, making it True if the store size is below 1000 units and False if above this amount. This useful yet simple feature facilitates better inventory control by keeping stores from having periods where customers cannot find products. Another significant outcome of the project is EOQ calculation. EOQ calculation determines the optimal order quantity for each store. Based on historical sales, the EOQ formula assists stores in minimizing holding costs and ordering costs. This results in improved inventory management.



Fig(4)output screen shots

 Stores with higher sales receive higher EOQ values, so they never run out of stock for customers. Stores with lower sales receive lower EOQ values to avoid overstock. For example, a store selling 50,000 units per week has an EOQ of 316.23, but a store selling 10,000 units per week has an EOQ of 141.42. This balances inventory among stores such that each store has the appropriate amount of stock depending on how much they sell. The system also employs machine learning to predict sales. It considers historical sales data, seasonal patterns, promotions, and discounts to predict future sales. Accurate sales predictions enable stores to predict inventory ahead of time, prevent supply chain issues, and control stock more effectively. Stores can keep top-selling products available at all times by predicting future sales patterns, which minimizes lost sales and keeps customers satisfied. This system operates efficiently, as is evident in a sample output table. The table illustrates that various stores have varying sales trends and require varying inventory management schemes. For example, a store with sales of 42,000 units a week has an EOQ of 289.22 and appears as Low\_Stock = False, indicating that it has sufficient inventory. Contrarily, a store with sales of 9,500 units a week has an EOQ of 137.41 and appears as Low\_Stock = True, which indicates that it needs to be restocked immediately. This is a systematic approach, which means that each store has an appropriate balance between having sufficient stock and controlling costs. In addition, by providing data-driven information, the system enables supply chain managers to make data-driven decisions rather than assumptions. The combination of inventory need forecasting and automated stock management enables Walmart stores to operate effectively, with little stockouts and excess inventory costs. The combination of Economic Order Quantity (EOQ) and demand forecasting models enables managers to adjust in real-time based on actual sales, thereby making inventory management more responsive and dynamic. In conclusion, this project shows how machine learning and inventory optimization techniques can improve supply chain management to a large extent. The system determines low-inventory stores, the ideal order quantities, and the projected future sales trends, allowing businesses to manage their inventory more effectively. Using this forecasting model, companies can eliminate wasted inventory, improve order fulfillment rates, and make sure customers can always access the products they require. Future improvements can be made by integrating seasonality patterns, real-time sales information, and outside economic factors to improve demand forecasting. This model is a great resource for retail firms that want to manage their inventory more effectively and run more efficiently in a more competitive environment.

PROJECT LINK : [ML-Powered Sales Forecasting and EOQ Optimization in Retail - Colab](https://colab.research.google.com/drive/1-y6SSgiC6Sh2skaooerFJKCuOkZ3EHjR?usp=sharing)

# 5. CONCLUSION

This research demonstrates how crucial it is to accurately predict demand and control inventory in order to make supply chain activities more efficient. Utilizing machine learning algorithms such as Linear Regression and Random Forest Regressor, and sophisticated data analysis techniques, the proposed system makes predictions more accurate, minimizes stockouts and overstocking, and facilitates simpler decision-making. Adding historical sales data, seasonal patterns, and key performance indicators (KPIs) makes inventory management more data-centric. Further work can explore more deep learning techniques, hybrid models, and external inputs such as economic indicators to make predictions more accurate and dynamic in shifting retail environments.

# REFERANCE

[1]Data analytics in the supply chain management: Review of machine learning applications in demand forecasting ;A Aamer, LP Eka Yani, IM Alan Priyatna ;Operations and Supply Chain Management: An International Journal, 2020•journal.oscm-forum.org

[2]From predictive to prescriptive analytics ;D Bertsimas, N Kallus ;Management Science, 2020•pubsonline.informs.org

[3] Predictive analytics for demand forecasting–a comparison of SARIMA and LSTM in retail

SCM ;T Falatouri, F Darbanian, P Brandtner, C Udokwu ;Procedia Computer Science, 2022•Elsevier

[4]Demand forecasting based on machine learning for mass customization in smart manufacturing

M Kim, J Jeong, S Bae ; Proceedings of the 2019 International Conference on Data Mining and Machine …, 2019•dl.acm.org

1. Demand prediction and price optimization for semi-luxury supermarket segment ; T Qu, JH Zhang, FTS Chan, RS Srivastava, MK Tiwari, WY Park ;Computers & industrial engineering, 2017•Elsevier
2. Demand forecasting in production planning for dairy products using machine learning and statistical method ; C Vithitsoontorn, P Chongstitvatana ; 2022 international electrical engineering congress (iEECON), 2022•ieeexplore.ieee.org
3. Min, H. (2010), “Artificial intelligence in supply chain management: theory and applications”, International Journal of Logistics Research and Applications, Vol. 13 No. 1, pp. 13-39.
4. Using deep learning for short-term load forecasting ; NMM Bendaoud, N Farah ; Neural computing and applications, 2020•Springer
5. A deep learning approach to predict weather data using cascaded LSTM network; Z Al

Sadeque, FM Bui ; 2020 IEEE Canadian conference on electrical and computer …, 2020•ieeexplore.ieee.org

1. On the application of ARIMA and LSTM to predict order demand based on short lead time and on-time delivery requirements ; CC Wang, CH Chien, AJC Trappey ;

Processes, 2021•mdpi.com

1. A manufacturer-retailer inventory model with remanufacturing, stochastic demand, and green investments ; WA Jauhari, ID Wangsa ; Process Integration and Optimization for

Sustainability, 2022•Springer

1. On the application of ARIMA and LSTM to predict order demand based on short lead time and on-time delivery requirements ; CC Wang, CH Chien, AJC Trappey ;

Processes, 2021•mdpi.com

1. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model ; AA Ghobbar, CH Friend ; Computers & operations research, 2003•Elsevier [14] Prescriptive control of business processes: new potentials through predictive analytics of big data in the process manufacturing industry ; J Krumeich, D Werth, P Loos ; Business & Information Systems Engineering, 2016•Springer
2. Customer demand forecasting via support vector regression analysis ; AA Levis, LG Papageorgiou - Chemical Engineering Research and Design, 2005 – Elsevier
3. Demand forecasting of engine oil for automotive and industrial lubricant manufacturing company using neural network ; R Sharma, P Singhal ; Materials Today:

Proceedings, 2019•Elsevier

1. Inventory management for new products with triangularly distributed demand and lead-time

P Wanke, H Ewbank, V Leiva, F Rojas -Computers & Operations Research, 2016•Elsevier

1. Ben Elmir, W.; Hemmak, A.; Senouci, B. Smart Platform for Data Blood Bank Management: Forecasting Demand in Blood Supply -Chain Using Machine Learning. Information 2023
2. [Forecasting stock prices of fintech companies of India using random forest with high-frequency data](https://www.sciencedirect.com/science/article/pii/S2199853123002822) ;[BK Meher](https://scholar.google.com/citations?user=fCEUTegAAAAJ&hl=en&oi=sra), M Singh, [R Birau](https://scholar.google.com/citations?user=wh0Z3IYAAAAJ&hl=en&oi=sra), [A Anand](https://scholar.google.com/citations?user=9HQ4QngAAAAJ&hl=en&oi=sra) - Journal of Open Innovation …, 2024 – Elsevier
3. [Comparative analysis of ARIMA and LSTM for predicting fluctuating time series data](https://beei.org/index.php/EEI/article/view/6034)

[DG Taslim](https://scholar.google.com/citations?user=87iSE_8AAAAJ&hl=en&oi=sra), [IM Murwantara](https://scholar.google.com/citations?user=Q9ld6p8AAAAJ&hl=en&oi=sra) ;Bulletin of Electrical Engineering and Informatics, 2024•beei.org