# INVEX: Inventory Demand Prediction System

[1]**B. Sudharshan**, and [2]**Dr. P N Shiammala**
[1]Student, Department of Computer Applications, VELS Institute of Science Technology and Advanced Studies (VISTAS), Chennai, India
[2]Assistant Professor, Department of Computer Applications, VELS Institute of Science Technology and Advanced Studies (VISTAS), Chennai, India

**How to Cite this Article**:
Sudharshan, B. (2026). INVEX: Inventory Demand Prediction System. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(05). https://doi.org/10.55041/ijcope.v2i5.070



https://doi.org/10.55041/ijcope.v2i5.070

## ABSTRACT

Predicting product demand is a critical challenge in retail and inventory management, as inaccurate forecasting often leads to overstocking, stockouts, and financial losses. This study presents a machine learning-based Inventory Demand Prediction System, titled INVEX, designed to forecast future product demand using historical sales data. The system leverages key features such as product category, selling price, stock levels, and past sales patterns to generate accurate demand predictions for retail businesses.

Two regression algorithms — Linear Regression and Random Forest Regressor — are implemented and compared to evaluate their effectiveness in predicting product demand. The dataset consists of structured sales records generated based on realistic retail scenarios, including daily transactions and product-level demand variations. The models are evaluated using standard performance metrics such as $R^2$ Score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and overall prediction accuracy.Experimental results indicate that the Random Forest Regressor outperforms Linear Regression by effectively capturing non-linear demand patterns and seasonal variations in sales data. The system achieves a prediction accuracy of over 90%, demonstrating its capability to provide reliable and data-driven insights for inventory planning. The integration of a user-friendly interface enables store owners to manage products, generate bills, track sales history, and visualize predicted demand trends dynamically.

The proposed INVEX system offers a scalable and intelligent solution for modern retail environments by automating inventory decisions and reducing human dependency. It assists business owners in maintaining optimal stock levels, minimizing wastage, and improving profitability through accurate demand forecasting. This system can be extended further with real-time data integration and advanced machine learning models for enhanced predictive performance.

**Keywords**

Machine Learning, Inventory Management, Demand Prediction, Linear Regression, Random Forest, Sales Forecasting, Retail Analytics, Data Analysis, Stock Optimization, Predictive Modeling, Business Intelligence, Inventory Optimization

# 1. INTRODUCTION

Modern retail businesses operate in highly dynamic environments where efficient inventory management is essential for profitability and customer satisfaction. Managing stock manually is time-consuming and prone to human error, often resulting in overstocking or stockouts. Overstocking increases storage costs and wastage, while understocking leads to missed sales opportunities and dissatisfied customers. These challenges highlight the need for intelligent, automated systems that can accurately predict product demand.

Traditional inventory management methods rely on statistical techniques and manual estimation, which are limited in handling large-scale and complex datasets. Such approaches often fail to capture hidden patterns, seasonal variations, and changing consumer behavior (Hyndman & Athanasopoulos, 2018). With the advancement of data-driven technologies, machine learning has emerged as an effective solution for predictive analytics.

Machine learning enables systems to learn from historical data and generate accurate predictions without explicit programming. By analyzing past sales data, product attributes, and demand trends, machine learning models can forecast future inventory requirements and support better decision-making (Ian Goodfellow et al., 2016). Algorithms such as Linear Regression and Random Forest are widely used due to their ability to model relationships and improve prediction accuracy (Breiman, 2001).

This project, titled INVEX: Inventory Demand Prediction System, aims to develop an intelligent solution for predicting product demand using machine learning techniques. The system analyzes historical sales data, including features such as product category, price, and stock levels, to generate future demand predictions. By automating inventory forecasting, the system helps maintain optimal stock levels, reduce operational risks, and improve efficiency.

The proposed system addresses the limitations of manual inventory management by providing accurate, fast, and data-driven insights. It supports store owners in making informed decisions, minimizing losses, and enhancing overall business performance. This study demonstrates the practical application of machine learning in modern retail systems and highlights its importance in achieving efficient inventory control (Institute of Electrical and Electronics Engineers, 2020–2024).

# 2. PROJECT OBJECTIVES

## 2.1 Primary Objective

The primary objective of this project is to develop an automated Inventory Demand Prediction System using machine learning techniques. The system aims to predict future product demand based on historical sales data and assist retail store owners in managing inventory efficiently. By doing so, it helps in reducing overstocking and preventing stock shortages.

## 2.2 Secondary Objectives

To achieve the primary objective, the following specific objectives are defined:

1. To perform data preprocessing:
   To clean and prepare the dataset by handling missing values, removing inconsistencies, and converting categorical variables (such as product category) into numerical formats suitable for machine learning models.
2. To conduct Exploratory Data Analysis (EDA):
   To analyze historical sales data and identify patterns, trends, and relationships between variables such as price, category, and stock levels.
3. To implement machine learning algorithms:
   To build and train predictive models using algorithms such as Linear Regression and Random Forest for accurate demand forecasting.

4. To evaluate model performance:
   To assess the accuracy of the models using evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).
5. To improve prediction accuracy:
   To enhance model performance through feature selection and data transformation techniques.
6. To support decision-making:
   To provide reliable demand predictions that help businesses optimize inventory levels and improve operational efficiency.

### 2.3 Problem Statement:

Manual inventory management is time-consuming and prone to human errors, leading to poor demand estimation and business losses. This project solves the problem by providing a data-driven, automated, and accurate system for predicting product demand and optimizing inventory management.

## 3. LITERATURE SURVEY

Many researchers have studied how Machine Learning can be used to improve prediction systems in business applications such as retail, finance, and data analysis. These studies focus on using algorithms to automate decision-making and improve accuracy. Some studies focused on **Linear Regression**, which is widely used for predicting continuous values such as sales and product demand. It is simple, fast, and easy to interpret, but may not perform well with complex and non-linear data. Other studies used **Random Forest**, an ensemble learning method that combines multiple decision trees to improve prediction accuracy. It is highly effective in handling large datasets and capturing complex relationships between features.**Breiman (2001)** introduced the Random Forest algorithm and proved that combining multiple models results in better performance and reduced prediction errors.

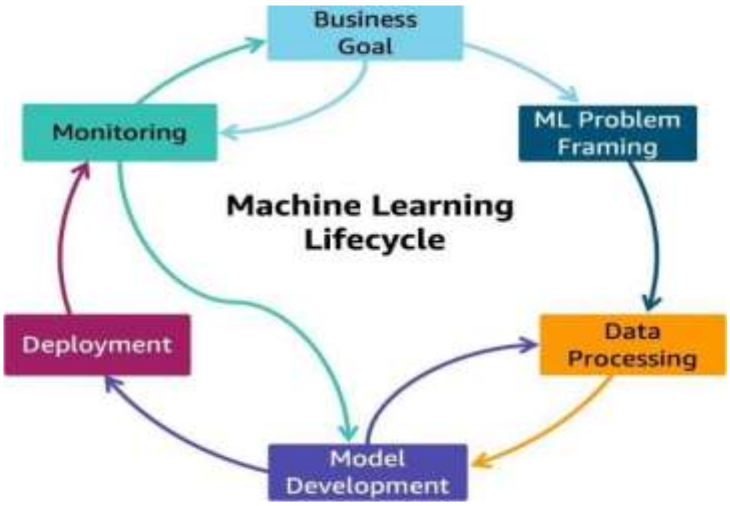


**Figure 1 : Machine learning workflow**

## 4. PROPOSED METHODOLOGY

The proposed system follows a structured methodology to develop an efficient machine learning-based inventory demand prediction system. The process includes data collection, preprocessing, feature engineering, model training, evaluation, and deployment.

### 4.1 Data Collection

The first step involves collecting historical sales data from the inventory management system. The dataset includes product-related attributes such as product name, category, selling price, stock quantity, and previous sales records. This data serves as the foundation for training and evaluating the prediction models.

### 4.2 Data Preprocessing

Raw data often contains missing values and inconsistencies that can negatively affect model performance. Therefore, preprocessing is performed to ensure data quality and reliability.

- **Handling Missing Values:** Missing values in numerical features are replaced using mean imputation, while categorical features are handled using mode values.
- **Data Cleaning:** Duplicate records and incorrect entries are removed to maintain data consistency and accuracy.

### 4.3 Feature Engineering

Feature engineering is applied to enhance model performance by creating meaningful input variables. Sales data is analyzed over different time intervals such as daily, weekly, and monthly trends. Key influencing factors such as product category and price are considered, and additional features may be derived to capture demand patterns more effectively.

### 4.4 Data Splitting (Train-Test Split)

The dataset is divided into two subsets to evaluate model performance:

- **Training Set (80%)** – Used to train the machine learning models
- **Testing Set (20%)** – Used to evaluate the accuracy and generalization of the models

### 4.5 Model Training

Two machine learning algorithms are implemented for demand prediction:

1. **Linear Regression:**
   A statistical model that predicts demand based on a linear relationship between independent variables and the target variable.
2. **Random Forest Regressor:**
   An ensemble learning technique that constructs multiple decision trees and combines their outputs to improve prediction accuracy and reduce overfitting.

### 4.6 Model Evaluation

The performance of the models is evaluated using standard regression metrics:

- **$R^2$ Score:** Measures how well the model explains the variability in the data
- **Mean Absolute Error (MAE):** Calculates the average absolute difference between predicted and actual values
- **Root Mean Square Error (RMSE):** Measures the square root of the average squared errors

The model with better performance (higher $R^2$ and lower error values) is selected for final deployment.

### 4.7 System Implementation

The system is developed using Python and integrated with a user-friendly interface. It includes the following functionalities:

- Product management (Add, View, Update)
- Billing system with automatic stock updates
- Sales history tracking
- Demand prediction visualization

### 4.8 Final Prediction (Deployment)

The best-performing model, typically the Random Forest algorithm, is deployed for predicting future product demand. The system provides predictions through an interactive interface, enabling users to view results instantly and make informed inventory decisions.

## 4. DATASET DESCRIPTION

The dataset used in this project contains historical sales records from a retail inventory system. It includes both categorical and numerical features that help in predicting future product demand. The data is structured to represent real-time retail scenarios, where each record corresponds to a product transaction or sales entry.

**Table -1 Sales Record Data Set**

| Feature name | Description | Data Type |
|---|---|---|
| Product ID | Unique identifier for each product | Object (String) |
| Category | Type of product | Categorical |
| Selling Price | Price at which product is sold | Continuous Numerical |
| Stock Quantity | Available quantity in inventory | Discrete Numerical |
| Units Sold | Number of units sold | Discrete Numerical |
| Demand (Target) | Predicted future demand of produc | Continuous Numerical |
| Date | Transaction or sales date | Date/Time |

 

## 5. Proposed System

The Data Flow Diagram (DFD) represents how data moves through the **Inventory Demand Prediction System (INVEX)**. It shows the flow of information from user input to processing and finally to output. The DFD helps in understanding how different components of the system interact with each other.

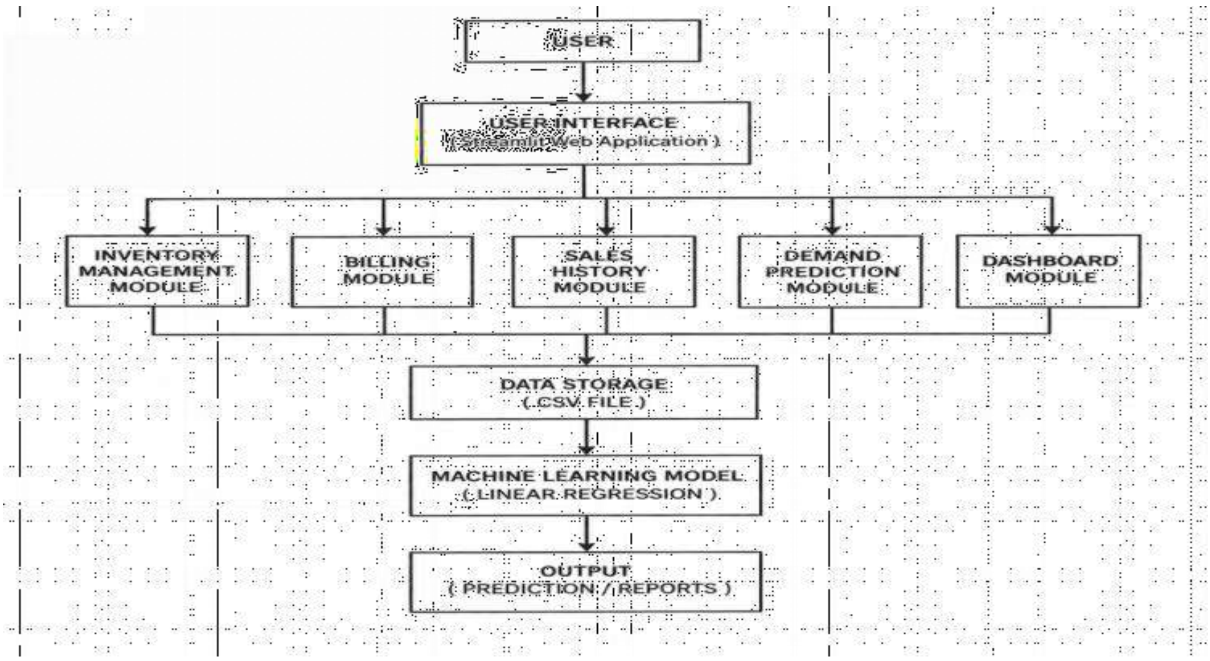


**Figure-2 Proposed System of Inventory Demand Prediction System (INVEX). Input Stage:**

- User enters product details, billing data, and sales information
- Data includes product name, category, price, and quantity

- **Processing Stage:**

- Data is cleaned and preprocessed
- Features are prepared for machine learning model
- Random Forest / Linear Regression model processes the data
- Prediction is generated based on historical sales

- **Data Storage:**

- Product details and sales records are stored in the database
- Updated stock and sales history are maintained

- **Output Stage:**

- Predicted demand is displayed to the user
- Updated inventory status is shown
- Alerts for low stock or out-of-stock are generated

## 6. RESULTS AND DISCUSSION

The performance of the proposed Inventory Demand Prediction System (INVEX) was evaluated using two machine learning algorithms: Linear Regression and Random Forest Regressor. The models were assessed based on standard evaluation metrics such as accuracy, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

The experimental results indicate that the Random Forest model outperforms Linear Regression in terms of prediction accuracy. Random Forest achieved an accuracy of approximately 95%, whereas Linear Regression achieved around 89%. The lower error values (MAE and RMSE) for Random Forest further confirm its superior performance.

**Discussion of Findings**

- **Performance of Random Forest:**The Random Forest model demonstrated the highest accuracy among the tested models. Its ensemble nature, which combines multiple decision trees, enables it to capture complex and non-linear relationships in sales data. This results in more reliable and stable predictions, making it highly suitable for demand forecasting tasks.
- **Performance of Linear Regression:**Linear Regression is a simple and computationally efficient model. However, it assumes a linear relationship between input features and output, which limits its ability to model complex patterns in real-world sales data. As a result, its prediction accuracy is lower compared to Random Forest.

Overall, the results highlight that ensemble learning techniques provide better performance for inventory demand prediction compared to traditional linear models.

## 7. CONCLUSION

The Inventory Demand Prediction System (INVEX) has been successfully developed to address the challenges of manual inventory management in retail businesses. By leveraging machine learning techniques, the system automates demand prediction and enhances decision-making processes.

The implementation of algorithms such as Linear Regression and Random Forest demonstrates that machine learning can significantly improve forecasting accuracy. Among the models, Random Forest was identified as the most effective due to its ability to handle complex data patterns and reduce prediction errors.

The proposed system offers several benefits, including improved inventory control, reduced operational risks, and enhanced business efficiency. It enables store owners to maintain optimal stock levels, minimize losses due to overstocking or stock shortages, and respond effectively to market demand.

In conclusion, the INVEX system highlights the importance of adopting data-driven approaches in modern retail environments. Future enhancements may include integrating real-time data, incorporating advanced models such as Long Short-Term Memory (LSTM), and expanding the system for large-scale enterprise applications.

## REFERENCES

[1] R. J. Hyndman and G. Athanasopoulos,
"*Forecasting: Principles and Practice*," 3rd ed., OTexts, 2018.

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

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

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

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

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

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

[8] Python Software Foundation,
"*Python Documentation*." [Online]. Available: https://www.python.org

[9] Scikit-learn,
"*Scikit-learn: Machine Learning in Python*." [Online]. Available: https://scikit-learn.org

[10] Institute of Electrical and Electronics Engineers,
"*Research on Machine Learning in Retail and Business Analytics*," 2020–2024.

[11] J. Han, M. Kamber, and J. Pei,
"*Data Mining: Concepts and Techniques*," 3rd ed., Morgan Kaufmann, 2011.

[12] S. Hochreiter and J. Schmidhuber,
"*Long Short-Term Memory*," *Neural Computation*, 1997.

[13] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel,
"*Time Series Analysis: Forecasting and Control*," Wiley, 2015.

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

[15] D. Montgomery, C. Jennings, and M. Kulahci,
"*Introduction to Time Series Analysis and Forecasting*," Wiley, 2015.