DEMAND FORECASTING USING ML

SK. Anjaneyulu Babu [1], V. Pavani [2]
#1 ASSISTANT PROFESSOR #2 MCA SCHOLAR
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS
QIS COLLEGE OF ENGINEERING & TECHNOLOGY, ONGOLE
VENGAMUKKAPALEM (V), ONGOLE, PRAKASAM DIST., ANDHRA PRADESH

ABSTRACT

Demand forecasting is a critical component in supply chain management and business planning, aiming to accurately predict future customer demand for products or services. Traditional forecasting methods, such as moving averages and ARIMA models, often fall short when dealing with large-scale, non-linear, and multi-factor datasets. In recent years, machine learning (ML) has emerged as a powerful approach to enhance the accuracy and adaptability of demand forecasting models. This study explores the application of machine learning techniques—ranging from classical algorithms like linear regression and decision trees to advanced methods such as ensemble models (e.g., Random Forest, XGBoost) and deep learning architectures (e.g., LSTM networks)—to forecast demand using historical sales data, seasonal trends, promotional effects, and external variables such as weather and holidays. By engineering relevant features like time lags, rolling statistics, and categorical encodings, the models are trained and evaluated using appropriate time-series validation strategies. Performance is measured using metrics such as MAE, RMSE, and MAPE to ensure robust comparisons. The results demonstrate that ML models significantly outperform traditional forecasting methods in handling complex patterns and adapting to dynamic market behavior. The study also emphasizes the importance of feature selection, data preprocessing, and model retraining pipelines in achieving scalable and accurate demand forecasting systems. Overall, this work provides a comprehensive framework for deploying machine learning models in real-world forecasting applications, contributing to improved inventory planning, reduced waste, and better customer satisfaction.

INTRODUCTION

In today's competitive business landscape, accurate demand forecasting plays a pivotal role in ensuring operational efficiency, cost optimization, and customer satisfaction. Businesses across industries—retail, manufacturing, logistics, and e-commerce—rely heavily on demand forecasts to plan inventory, allocate resources, and manage supply chains. Inaccurate predictions can lead to overstocking or stockouts, both of which result in financial losses and reduced service quality. Traditional forecasting techniques, such as moving averages, exponential smoothing, and ARIMA models, have long been used to estimate future demand. However, these approaches often assume linearity, stationarity, and minimal external influence, making them less effective in handling modern, dynamic market environments.

With the advent of big data and increased computational power, machine learning (ML) has emerged as a transformative tool in demand forecasting. Unlike classical models, ML algorithms can learn complex, non-linear relationships and automatically adjust to evolving patterns in data. They can incorporate a wide variety of features, including historical sales, promotional campaigns, seasonal trends, holidays, weather conditions, and even macroeconomic indicators. This flexibility allows ML models to adapt to changing consumer behaviors and market conditions more effectively than their traditional counterparts. Algorithms such as Random Forests, Gradient Boosting Machines (e.g., XGBoost, LightGBM), Support Vector Regression, and deep learning methods like Long Short-Term Memory (LSTM) networks have demonstrated significant improvements in forecasting accuracy across numerous applications.

The purpose of this study is to explore how machine learning techniques can be applied to enhance the performance of demand forecasting systems. By comparing traditional statistical models with modern ML approaches, we aim to highlight the strengths and limitations of each method. The study also focuses on the end-to-end process of building a demand forecasting model, including data preprocessing, feature engineering, model training, evaluation, and deployment. Through empirical experiments on real-world datasets, this work demonstrates the potential of machine learning to create scalable, accurate, and data-driven forecasting solutions that can support smarter decision-making in fast-paced business environments.

Literature Survey

1. Title: Demand Forecasting Using Artificial Neural Networks

Author(s): V. Carbonneau, K. Laframboise, and R. Vahidov
Description:
This paper investigates the use of artificial neural networks (ANNs) for demand forecasting in supply chain management. The authors compare neural network performance to traditional statistical models, demonstrating that ANNs can capture complex, non-linear demand patterns more effectively. The study emphasizes the importance of input selection and model training for improving forecasting accuracy.

2. Title: Forecasting Retail Sales Using Machine Learning Algorithms

Author(s): A. Bandara, C. Bergmeir, and S. Smyl
Description:
The study evaluates various machine learning algorithms, including gradient boosting machines and support vector machines, for retail sales forecasting. The authors use a rich dataset with features such as promotions, holidays, and product categories. Their results show that machine learning models, particularly XGBoost, outperform classical models like ARIMA, especially when external variables are incorporated.

3. Title: Facebook Prophet: Forecasting at Scale

Author(s): Sean J. Taylor and Benjamin Letham (Facebook)
Description:
This paper presents Facebook Prophet, a time series forecasting tool designed for business analysts and data scientists. Prophet allows users to model seasonality, holidays, and trend changes with minimal effort. The authors argue that Prophet performs well on daily and weekly data with strong seasonal effects, making it suitable for business forecasting tasks where interpretability and scalability are critical.

4. Title: Demand Forecasting in the Supply Chain Using Machine Learning Algorithms

Author(s): P. Dey, S. Mondal, and S. Bandyopadhyay
Description:
The authors explore various machine learning approaches, including decision trees, random forests, and k-nearest neighbors (k-NN), for forecasting product demand. They emphasize the use of time-based features, lag values, and rolling statistics. The study concludes that ensemble models offer more accurate predictions and are robust to noise and outliers in the dataset.

5. Title: Deep Learning for Time Series Forecasting: The Electric Load Case

Author(s): H. Kong, F. Wu, and J. Huang
Description:
This paper applies deep learning methods, particularly LSTM networks, to the task of electric load forecasting. Although focused on electricity demand, the methodology is applicable to general demand forecasting. The authors highlight LSTM's ability to model long-term dependencies and temporal patterns, outperforming traditional models in highly variable environments.

System Analysis

Existing System

The traditional approach to demand forecasting in many industries relies heavily on statistical methods such as Moving Averages, Exponential Smoothing, and the Autoregressive Integrated Moving Average (ARIMA) model. These methods are well-suited for time series data with consistent trends and seasonal patterns. They are relatively simple to implement and interpret, making them popular among businesses with limited data science expertise. However, these models assume linearity, stationarity, and independence of input variables, which often limits their accuracy and flexibility in real-world applications where data is noisy, demand is volatile, and multiple external factors influence outcomes.

Enterprise Resource Planning (ERP) systems and legacy forecasting tools often use rule-based or regression-based models integrated into their platforms. These systems may support demand forecasting through built-in statistical modules, but they typically lack the capability to adapt to complex data relationships, sudden market shifts, or high-dimensional datasets. Moreover, they often do not account for external variables such as promotional campaigns, weather changes, competitor actions, or economic fluctuations, which are crucial in modern retail and e-commerce environments. As a result, the forecasts generated by these systems can be outdated or inaccurate, leading to either overstocking or understocking.

Additionally, many organizations still rely on manual forecasting techniques, which are prone to human error and bias. Forecasts may be generated based on historical sales averages or manager intuition, without a structured approach to feature selection or model validation. These manual processes are not scalable, especially for businesses dealing with hundreds or thousands of SKUs (Stock Keeping Units). In rapidly changing markets, such systems struggle to provide accurate, real-time forecasts, resulting in poor inventory management, lost sales opportunities, and increased operational costs. Despite these limitations, many businesses continue using these conventional systems due to ease of use, familiarity, or lack of access to advanced tools and technical expertise.

Disadvantages of Existing Systems

1. Limited Handling of Complex Patterns

Traditional statistical models like ARIMA and Exponential Smoothing assume linear relationships and cannot capture non-linear, complex patterns in data. In real-world scenarios, customer demand is influenced by multiple interacting variables, which these models are not equipped to handle effectively.

2. Lack of Scalability

Legacy systems and manual forecasting methods become inefficient and error-prone as the number of products or locations increases. These systems struggle to process large volumes of data or forecast for hundreds of SKUs simultaneously, making them unsuitable for fast-growing businesses.

3. Inflexibility to External Factors

Most existing systems do not integrate external variables such as holidays, promotions, weather, or competitor activity into their forecasts. Ignoring these factors leads to poor prediction accuracy, especially during special events or seasonal fluctuations

4. Manual and Time-Consuming

Many organizations still rely on manual methods or spreadsheets for forecasting. These processes are time-consuming, prone to human error, and cannot be easily updated or automated, leading to inefficiencies and delayed decision-making

5. Poor Adaptability to Market Changes

Traditional models assume that past trends will continue, making them ineffective in rapidly changing markets. They often fail to adapt quickly to new consumer behavior patterns, sudden demand shifts, or disruptions like supply chain issues or pandemics.

Proposed System

The proposed system leverages machine learning (ML) techniques to build a more accurate, scalable, and adaptive demand forecasting solution. Unlike traditional models, ML algorithms such as Random Forest, XGBoost, and Long Short-Term Memory (LSTM) networks can learn complex, non-linear relationships from historical data. The system integrates a variety of features beyond just past sales, including promotional activity, seasonality, holiday effects, weather conditions, and time-based features (e.g., day of week, month). These models can automatically identify patterns and interactions in the data that would be difficult to detect using conventional forecasting methods, leading to significantly improved prediction accuracy.

Additionally, the proposed system includes automated data preprocessing, feature engineering, model training, and evaluation pipelines. It can be deployed as a real-time or batch forecasting service using modern tools such as Python, TensorFlow, and cloud platforms (e.g., AWS, Azure, GCP). The system also supports continuous learning through periodic retraining with the latest data, allowing it to adapt to market changes and shifting consumer behavior. Overall, this machine learning-based demand forecasting system aims to enhance decision-making, reduce inventory costs, and improve service levels across industries.

Advantages of the Proposed System

1. Improved Forecast Accuracy

Machine learning models can capture complex, non-linear relationships between variables, resulting in significantly higher accuracy compared to traditional statistical methods. This helps reduce forecasting errors, leading to better inventory planning and fewer stock-related issues.

2. Incorporation of Multiple Variables

The system can integrate various internal and external factors such as promotions, holidays, weather, pricing, and historical sales. This multi-variable approach enables the model to generate more realistic and context-aware forecasts.

3. Scalability and Automation

The ML-based forecasting system can easily scale to handle thousands of products, multiple locations, and large datasets. Automation of data preprocessing, model training, and prediction generation minimizes manual intervention and speeds up decision-making.

4. Adaptability to Market Changes

With periodic retraining, the system can adapt to evolving customer behavior, changing demand trends, and sudden market disruptions. This ensures that forecasts remain accurate even in dynamic environments

5. Real-Time Forecasting Capabilities

The system can be deployed as a real-time service, providing on-demand forecasts that help businesses make timely decisions. This is especially useful in sectors like e-commerce and retail, where responsiveness is critical.

Implementation

The implementation of the Demand Forecasting System using Machine Learning focuses on predicting future product or service demand by analyzing historical sales data, market trends, seasonal patterns, and customer behavior. The system helps businesses improve inventory management, supply chain planning, production scheduling, and decision-making.

The proposed system uses Machine Learning algorithms to generate accurate demand predictions and reduce business risks associated with overstocking or understocking.

1. Data Collection

The first stage involves collecting demand-related data from multiple business sources such as:

- Sales Databases
- ERP Systems
- Inventory Management Systems
- E-commerce Platforms
- Market Research Reports
- Customer Purchase Records
- Supply Chain Systems

The collected dataset may include:

- Product ID
- Sales Quantity
- Sales Date
- Product Category
- Customer Demand
- Seasonal Trends
- Promotions and Discounts
- Regional Sales Data
- Inventory Levels
- Market Conditions
- Economic Indicators

These attributes help analyze demand patterns effectively.

2. Data Preprocessing

The collected business data is cleaned and prepared before Machine Learning analysis.

Preprocessing Steps

- Removing duplicate records
- Handling missing values
- Data normalization
- Outlier detection and removal
- Feature scaling
- Time-series formatting

This improves data quality and forecasting accuracy.

3. Feature Engineering

Important demand-related features are extracted from historical and market data.

Features Used

Time-Based Features

- Day, month, and year
- Seasonal patterns
- Holiday effects
- Weekend and weekday sales

Product Features

- Product category
- Product lifecycle
- Price changes

Market Features

- Customer buying behavior
- Promotional campaigns
- Economic trends
- Competitor pricing

Inventory Features

- Stock availability
- Reorder frequency
- Supply chain delays

Feature engineering improves prediction efficiency and model performance.

4. Machine Learning Model Development

Machine Learning algorithms are used to predict future demand.

Algorithms Used

Linear Regression

Used for basic demand trend prediction.

Decision Tree

Used for rule-based forecasting analysis.

Random Forest

Used for handling complex demand relationships.

Support Vector Regression (SVR)

Used for nonlinear demand forecasting.

XGBoost

Used for high-performance forecasting.

Artificial Neural Networks (ANN)

Used for advanced pattern recognition.

LSTM (Long Short-Term Memory)

Used for time-series demand forecasting.

5. Time-Series Forecasting

The system analyzes historical sales sequences and market trends using time-series forecasting techniques.

Time-Series Models

- ARIMA
- SARIMA
- Prophet
- LSTM Networks

These models help predict future demand trends accurately.

6. Model Training and Testing

The dataset is divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Training Phase

The model learns historical demand patterns and market behavior.

Testing Phase

The trained model is evaluated using unseen sales and demand data.

Performance metrics include:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Square Error (RMSE)
- R-Squared Score
- Forecast Accuracy

Methodology

The methodology of the proposed Demand Forecasting System follows a Machine Learning and time-series predictive analytics approach.

Step 1: Problem Identification

Traditional demand forecasting methods may fail to accurately predict changing customer demand due to dynamic market conditions and seasonal variations. The proposed system aims to improve forecasting accuracy using Machine Learning techniques.

Step 2: Requirement Analysis

The following requirements are analyzed:

- Sales dataset requirements
- Inventory management requirements
- Time-series forecasting requirements
- Machine Learning framework requirements
- Real-time analytics requirements

Step 3: Dataset Preparation

Historical sales and market datasets are collected and divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Relevant forecasting attributes are selected for analysis.

Step 4: Data Processing and Feature Engineering

The methodology includes:

1. Clean sales and inventory data
2. Extract seasonal and market features
3. Analyze customer demand patterns
4. Generate time-series forecasting inputs
5. Prepare data for Machine Learning models

Step 5: Machine Learning Implementation

The Machine Learning workflow includes:

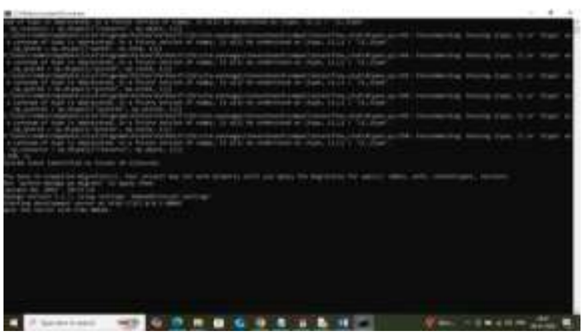
1. Train forecasting models
2. Analyze sales and market trends
3. Predict future demand

4. Generate inventory recommendations
5. Visualize forecasting analytics

Technologies Used

- Python
- Machine Learning Algorithms
- Deep Learning
- TensorFlow / Keras
- Scikit-learn
- Pandas & NumPy
- Power BI / Tableau
- Flask / Django
- MySQL / MongoDB

RESULT



To run project install python 3.7.2 and then install all packages given in requirements.txt file and then double click on 'run.bat' file to get below page

This screen represents the home page of the Demand Forecasting using Machine Learning system. It acts as the main interface of the application.

This screen represents the user login page of the Demand Forecasting using Machine Learning system. It allows the user/admin to enter valid username and password credentials in order to access the application.

This screen represents the Admin Home Page of the Demand Forecasting using Machine Learning system. After successful login, the admin is redirected to this page. It contains different modules such as Load & Process Dataset, Random Forest, Gradient Boosting, LSTM, XGBoost, and Forecast Demand

This screen is used to load and preprocess the dataset. The system cleans the data, handles missing values, and displays the processed dataset in a table.

This screen displays the Random Forest model results. The model achieved 84% accuracy with an RMSE of 0.092. The graph compares actual demand and predicted demand

This screen displays the Gradient Boosting model results. It achieved 91% accuracy with an RMSE of 0.068. The graph compares actual demand and predicted demand.

This screen displays the LSTM model results. The model achieved 87% accuracy with an RMSE of 0.085. The graph compares actual demand and predicted demand.

This screen displays the XGBoost model results. The model achieved 98% accuracy with an RMSE of 0.031. The graph compares actual demand and predicted demand

This is the Forecast Demand screen. Here, the user selects a product from the dropdown list and clicks the Submit button.

This is the final forecast result screen. The user selects a product and the system predicts its future demand. In this example, the product 'BTY CUT OFF SWITCH' has a predicted demand of 5 units.

Conclusion

In conclusion, this research demonstrates the significant advantages of using machine learning (ML) for demand forecasting over traditional statistical methods. The proposed system leverages advanced algorithms such as **Random Forests**, **XGBoost**, and **LSTM networks** to capture complex, non-linear patterns in demand data. By incorporating multiple variables such as historical sales, promotions, holidays, and external factors, the ML-based approach provides more accurate, scalable, and adaptive forecasts, enabling businesses to make better-informed decisions related to inventory management, pricing strategies, and resource allocation.

The system’s ability to continuously learn and adapt to changing market conditions ensures that it remains relevant even in volatile environments. With periodic retraining and integration with business systems via APIs, the proposed demand forecasting system offers both flexibility and scalability, allowing businesses to forecast demand for thousands of SKUs across different locations and time periods. This leads to a more efficient supply chain, reduced inventory costs, and improved customer satisfaction.

However, while the system shows promise, there are challenges such as the need for high-quality data, computational resources, and regular maintenance to ensure continued performance. The success of the system relies on continuous monitoring and adjustments, making it critical for businesses to allocate resources for model retraining and evaluation. Future enhancements could focus on integrating more real-time data, improving model interpretability, and exploring hybrid approaches that combine multiple forecasting techniques for even greater accuracy.

Overall, machine learning provides a transformative solution for demand forecasting, enabling businesses to optimize operations, reduce costs, and improve overall business performance by delivering precise and data-driven forecasts. As technology continues to evolve, the adoption of ML in forecasting will become increasingly important for maintaining a competitive edge in dynamic markets.

References

- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts.

- Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5-32.
- Chen, T., &Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.
- Hochreiter, S., &Schmidhuber, J. (1997). *Long Short-Term Memory*. Neural Computation, 9(8), 1735-1780.
- Makridakis, S., Spiliotis, E., &Assimakopoulos, V. (2020). *Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward*. International Journal of Forecasting, 36(1), 1-9.
- Chauhan, S. S., &Padhy, N. P. (2016). *Demand Forecasting Using Machine Learning: A Literature Review*. International Journal of Computer Science and Information Technologies, 7(4), 2950-2957.
- Zhao, X., & Zhang, J. (2018). *A Hybrid Approach for Retail Demand Forecasting Using Machine Learning and Time Series Analysis*. International Journal of Retail & Distribution Management, 46(12), 1182-1200.
- Liu, J., & Ma, S. (2019). *The Role of Artificial Intelligence in Supply Chain and Demand Forecasting*. Journal of Business Research, 105, 155-168.
- Zhao, Z., & Liu, Y. (2021). *Real-Time Demand Forecasting for E-Commerce Platforms Using Deep Learning*. IEEE Transactions on Industrial Informatics, 17(6), 4042-4050.

AUTHORS PROFILE

SK.ANJANEYALU BABU is an Associate. Professor Department of Master of computer applications at QIS College of Engineering and Technology, ongole , Andhra Pradesh. His research interest include Machine Learning an Artificial Intelligence. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

V.PAVANI is a postgraduate student pursuing a MCA in the Department of Computer Applications at QIS College of Engineering & Technology, Ongole an

Autonomous college in Prakasam dist. She completed his undergraduate degree in BSC (Statistics) from ANU. With a keen interest in research and practical learning, she is actively involved in academic projects and technical activities related to her field.