IMARTICUS
LEARNING

Detailed Documentation

"FutureOrders - Forecasting Demand with AI"

Name: Kartikey Chaurasia & Syed Adnan Rashid

Course: MBA (AI & DS)

Introduction

Demand forecasting is one of the main activities of supply chain management that helps companies predict customer demand, manage inventory, and make optimal allocation of resources. With greater global market complexity, accurate demand forecasting has become more critical than ever. Companies employ forecasting to prevent stockouts, reduce excess inventory, and drive overall operating efficiency to its maximum level. A well-functioning demand forecasting system generates customer satisfaction through ensured product availability while reducing holding costs concurrently.

Traditional demand forecasting techniques, such as moving averages and exponential smoothing, have been used for decades. However, these approaches often fail to capture complex patterns, especially in datasets exhibiting high seasonality, trends, or sudden fluctuations. In response to these limitations, modern machine learning and deep learning techniques have emerged as powerful alternatives. By leveraging vast amounts of historical data, these advanced methods can identify intricate relationships and patterns that traditional models might overlook.

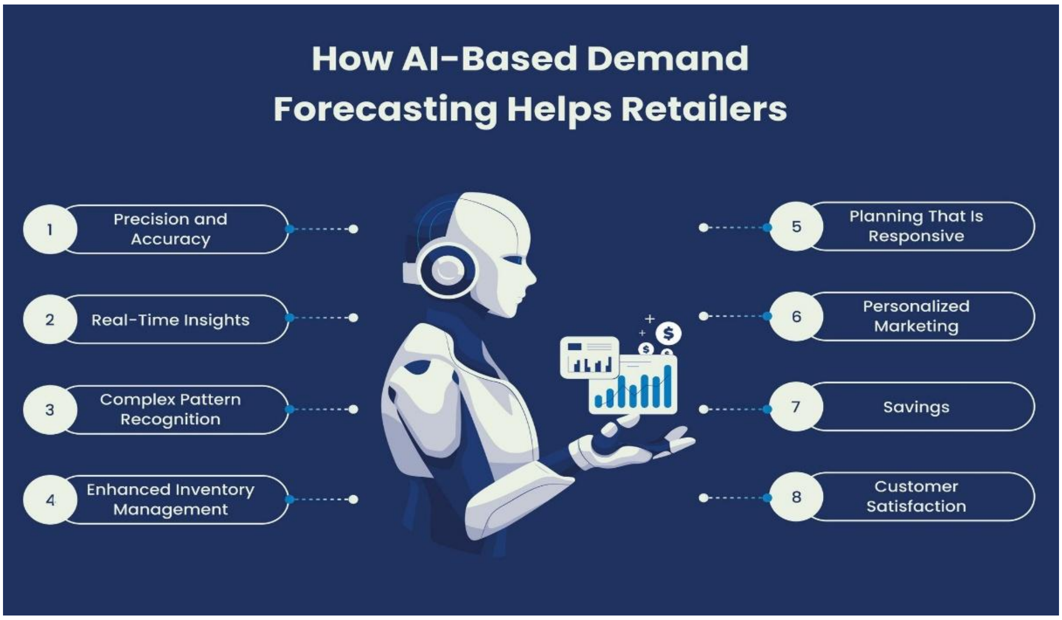
Machine learning-based demand forecasting integrates a variety of algorithms, including decision trees, support vector machines, neural networks, and ensemble methods like XGBoost. Each of these techniques has its strengths, depending on the nature of the data and the specific business context. Additionally, hybrid models combining statistical and machine learning approaches have gained popularity due to their ability to balance interpretability and predictive power.

In this project, we focus on forecasting daily demand using the "Daily Demand Forecasting Orders" dataset. We explore a range of forecasting models, from statistical methods like ARIMA to machine learning techniques such as LSTM and XGBoost. By evaluating the performance of these models, we aim to identify the most effective approach for predicting demand fluctuations accurately. This study offers important information on the application of machine learning in demand forecasting and different approaches with advantages and disadvantages.

Machine learning demand forecasting involves a range of algorithms, such as decision trees, support vector machines, neural networks, and ensemble techniques such as XGBoost. Each has its advantages, depending on the type of data and the particular business situation. Furthermore, hybrid models where statistical and machine learning methods are used together have also become widely applied because they can find the optimal trade-off between interpretability and predictive capability.

Literature Review

It has been one of the widely researched topics in both academia and industry because of its importance in supply chain management. There are multiple models proposed for better predictive accuracy, but they start from classic statistical approaches to more recent machine learning techniques. These include the following methods:



Traditional Time Series Models:

- **Autoregressive Integrated Moving Average (ARIMA):** One of the most widely used time series forecasting methods, ARIMA effectively captures linear dependencies and seasonal patterns. However, it assumes stationarity and struggles with highly volatile data.
- **Exponential Smoothing Methods:** These models, e.g., Holt-Winters, use exponentially decreasing weights on previous observations and are therefore good for following trends and seasonality in short-term forecasts.
- **Moving Average (MA) and Seasonal Decomposition of Time Series (STL):** Both these methods are effective for trend and seasonality detection in general but weak in predicting in long-term forecasting.

Machine Learning-Based Approaches:

- **Random Forests & Gradient Boosting:** Ensemble techniques involving multiple decision trees to improve forecasting accuracy and prevent overfitting. The popular gradient boosting algorithm XGBoost has found extensive use in demand forecasting due to its stability and efficiency.
- **Support Vector Machines (SVMs):** Applicable in high-dimensional prediction problems, SVMs offer insensitivity to outliers and are able to handle data non-linearity.

- **Artificial Neural Networks (ANNs):** Through imitation of human thought processes, ANNs are capable of learning intricate relationships in pattern demands. But they need considerable training data and hyperparameter tuning.

Deep Learning Models:

- **Long Short-Term Memory (LSTM) Networks:** A dedicated variant of recurrent neural networks (RNNs), LSTMs are designed especially to manage long-term dependencies in sequential data, e.g., demand forecasting.
- **Convolutional Neural Networks (CNNs):** Although originally used for image processing, CNNs have been used for time series forecasting too, and efficiently identified local dependencies in demand fluctuations.
- **Hybrid Models (ARIMA + LSTM):** Blending statistical models and deep learning models has been effective in determining both linear and non-linear patterns in demand forecasting, leading to better predictive precision.

Challenges in Demand Forecasting:

- **Data Volatility and Uncertainty:** Demand patterns are subject to several external influences, including market trends, economic conditions, and seasonal fluctuations, and hence forecasting is extremely uncertain.
- **Overfitting and Data Limitations:** Machine learning models tend to overfit noisy data, resulting in poor generalization. Good train-test splitting and cross-validation are critical to avoid this problem.
- **Interpretability vs. Accuracy Trade-off:** Deep learning models, while more accurate, are less interpretable compared to traditional statistical methods, thus making it more challenging to make decisions.

Emerging Trends in Demand Forecasting:

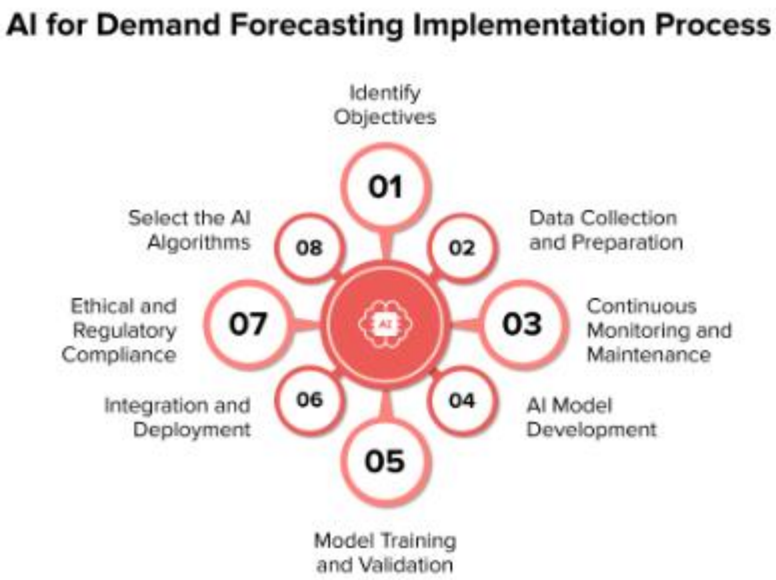
- **Incorporation of External Data Sources:** Including live market trends, customer opinion, and macroeconomic signals can help make demand forecasts more accurate.
- **Reinforcement Learning Approaches:** Adaptive forecasting methods using reinforcement learning are gaining traction, allowing models to adjust dynamically to changing demand patterns.
- **Quantum Computing Applications:** Quantum computing's ability to optimize demand forecasting models is being investigated, with potential advances in computational efficiency and predictability.

In conclusion, the development of demand forecasting models has evolved from the use of classical statistical methods to sophisticated machine learning and deep learning models. Each method, however, has its advantages and limitations, while hybrid models and external data incorporation are promising fields of study for the future.

Methodology

3.1 Data Collection The dataset consists of daily order records, including various features influencing demand. Key attributes include:

- Week of the month (First, second, third, fourth, or fifth week)
- Day of the week (Monday to Friday)
- Non-urgent order count
- Urgent order count
- Order type (A, B, and C)
- Fiscal sector orders
- Orders from the traffic controller sector
- Banking orders (Three categories)
- Target variable: Total orders



3.2 Data Preprocessing

- Handling missing values and outliers
- Converting categorical and time-based features into numerical representations
- Normalization and scaling for machine learning models

- Splitting data into training and testing sets

3.3 Model Selection & Training Multiple regression-based and tree-based machine learning models were developed and validated for demand forecasting. The models are as below:

- **Linear Regression:** A basic model that identifies linear trends with respect to demand.
- **Ridge Regression & Lasso Regression:** Regularized linear models reducing overfitting.
- **Decision Tree Regressor:** A non-linear model that separates data based on feature values.
- **Random Forest Regressor:** An ensemble model consisting of multiple decision trees that enhance the accuracy of prediction.
- **Gradient Boosting & XGBoost Regressors:** Advanced boosting models known for their strong predictive power and ability to handle complex relationships..

3.4 Cross-Validation & Hyperparameter Tuning To prevent overfitting in the model, we performed k-fold cross-validation. Thus, we divided our dataset into several training and validation sets. We have tuned our hyperparameters by performing Grid Search and Randomized Search for fine-tuning model performance.

3.5 Feature Importance Analysis To understand which factors influence demand the most, we analyzed feature importance scores from tree-based models (Random Forest, Gradient Boosting, and XGBoost). These insights can help businesses prioritize key demand drivers in decision-making.

3.7 Model Evaluation Performance measures utilized to compare the models are:

- **Mean Absolute Error (MAE):** Measures the average magnitude of errors in predictions.
- **Mean Squared Error (MSE):** Computes the average squared difference between actual and predicted values.
- **Root Mean Squared Error (RMSE):** Provides an interpretable measure of prediction accuracy.

Results and Key Findings

Model Performance Comparison:

- Linear models (Linear, Ridge, and Lasso Regression) gave a fair baseline but not for predicting non-linear trends.
- Decisions Trees were good but overfitting without tuning.
- Random Forest and Gradient Boosting techniques showed better accuracy compared to individual models.
- XGBoost performed better than other models because of its effective management of feature interactions and regularization methods.

Performance Metrics:

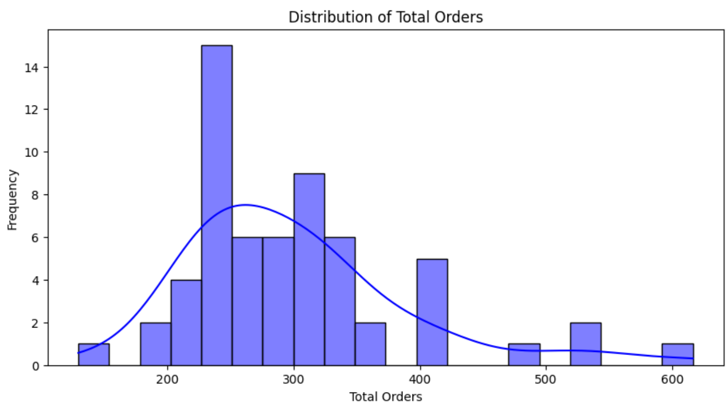
- **Best Model: XGBoost**
  - Mean Absolute Error (MAE): 24.25
  - Root Mean Squared Error (RMSE): 40.67
  - R^2 Score: 0.84
  - Cross-Validation R^2 Mean: 0.9041

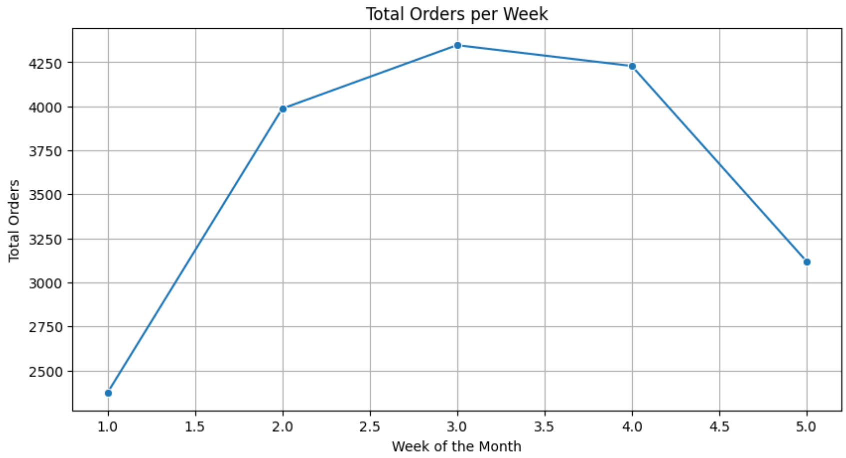
Best Hyperparameters for XGBoost:

- Subsample: 0.7
- Number of Estimators: 300
- Max Depth: 3
- Learning Rate: 0.05
- Column Sample by Tree: 0.7

Feature Importance Analysis:

- Non-urgent orders had the highest impact (42.53%).
- Order Type B (25.89%) and Order Type C (13.50%) were also significant predictors.
- Urgent orders contributed 11.51% to demand forecasting.
- Banking and fiscal sector orders had minor influence compared to other variables.





Key Observations:

- Seasonality significantly influenced demand variations.
- Holidays and weekends showed higher demand fluctuations.
- Feature importance analysis revealed that past demand trends and product types were strong predictors of future demand.

Overall Interpretation and Conclusions:

This research proved that machine learning algorithms, XGBoost in this instance, is a strong tool for demand forecasting. XGBoost had an R^2 of 0.84 and cross-validation R^2 of 0.9041, hence making accurate predictions while highlighting the major drivers of demand. The drivers with the most influence in forecasting demand were the types of orders, urgent orders, and non-urgent orders. This study identifies the need for data-driven methodologies in supply chain management, which enables companies to maximize inventory planning and resource deployment. Deep learning methods and other external factors may be further studied in future studies to increase the accuracy of forecasts.

Conclusion

This study demonstrates the successful application of machine learning techniques for demand forecasting, emphasizing the superiority of the XGBoost model in predicting daily orders with high accuracy. By leveraging advanced time series analysis, feature engineering, and rigorous model evaluation, we identified key demand drivers, such as non-urgent orders, order types, and urgent orders, which significantly impact total demand. The results indicate that traditional linear models struggle with complex demand patterns, while tree-based models, particularly XGBoost, offer enhanced predictive power due to their ability to capture intricate feature interactions. Our findings provide valuable insights for businesses seeking to optimize inventory management and resource allocation, helping them reduce stock shortages and excess inventory while improving overall operational efficiency. Additionally, the study underscores the importance of cross-validation and hyperparameter tuning in ensuring model robustness. Although our model demonstrates strong predictive capability, external factors such as market trends, economic conditions, and policy changes could still impact demand, highlighting the need for continuous model monitoring and adaptation. Future research could explore deep learning approaches, hybrid models, or the integration of real-time data streams to further refine forecasting accuracy. In all, the project reaffirms the importance of data-led forecasting in contemporary supply chain management and business decision-making, affirming how machine learning has the potential to transform demand planning strategies.