Food Demand Forecasting

Executive Summary

The project "Food Demand Forecasting" was undertaken to assist a meal delivery company in forecasting demand for upcoming weeks across various fulfillment centers in multiple cities. Accurate demand forecasting is vital for effective planning of raw material stocking and staffing of centers, considering the perishable nature of the materials and the weekly replenishment cycle.

Three regression models were employed to fit the data and forecast demand: Linear Regression, Random Forest Regression, and XGBoost Regressor. Each model was evaluated based on performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

The results indicated that Random Forest Regression and XGBoost Regression outperformed Linear Regression significantly, exhibiting lower errors across all metrics. These models demonstrated superior predictive capabilities, making them suitable for demand forecasting in fulfillment centers.

Based on the analysis, recommendations were made to implement Random Forest Regression or XGBoost Regression models for demand forecasting, regularly update the models with new data, and integrate the forecasting results into procurement planning and staffing strategies.

In conclusion, the project successfully demonstrated the effectiveness of machine learning regression models in forecasting food demand, offering valuable insights for optimizing operations and enhancing customer satisfaction for the meal delivery company.

Dataset Review

The dataset used for this project had historical data on food orders and related variables across multiple fulfillment centers and cities. It includes features such as date/time of order, type of meal, city, fulfillment center, and other relevant factors influencing demand.

- **Data Wrangling:** The dataset underwent thorough cleaning and preprocessing to handle missing values, outliers, and inconsistencies, ensuring the quality and reliability of the data used for modeling.
- **Exploratory Data Analysis (EDA):** EDA was performed to gain insights into the distribution of variables, identify correlations, and understand the underlying patterns in the data.
- **Dataset Split:** The dataset was split into training and testing sets to facilitate model training and evaluation, ensuring unbiased performance assessment.

Introduction

The project titled "Food Demand Forecasting" aims to assist a meal delivery company in forecasting demand for upcoming weeks across various fulfillment centers in multiple cities. Accurate demand forecasting is crucial for effective planning of raw material stocking and staffing of centers. The replenishment of raw materials occurs weekly, and precise forecasts are essential due to the perishable nature of the materials.

Objectives

The primary objectives of the project are as follows:

- Develop accurate demand forecasting models for fulfillment centers.
- Assist in raw material procurement planning based on forecasted demands.
- Facilitate efficient staffing of fulfillment centers through accurate demand forecasts.

Methodology

Three regression models were employed to fit the data and forecast demand:

1. **Linear Regression**
2. **Random Forest Regression**
3. **XGBoost Regressor**

Model Details and Performance

A. Linear Regression

Conceptual Details:

- Linear regression assumes a linear relationship between the input variables (features) and the target variable (demand).
- It fits a straight line to the data, minimizing the sum of squared differences between the observed and predicted values.
- Despite its simplicity, linear regression can provide baseline performance for comparison with more complex models.

Performance Metrics:

- Mean Absolute Error (MAE): 161.33
- Mean Squared Error (MSE): 88997.90
- Root Mean Squared Error (RMSE): 298.33

B. Random Forest Regression

Conceptual Details:

- Random Forest Regression is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction of the individual trees.
- It addresses overfitting by introducing randomness in the tree-building process, such as bootstrap sampling of data and random feature selection.
- Random Forests can capture complex nonlinear relationships and interactions between variables, making them suitable for forecasting tasks.

Performance Metrics:

- Mean Absolute Error (MAE): 68.72
- Mean Squared Error (MSE): 21534.75
- Root Mean Squared Error (RMSE): 146.75

C. XGBoost Regression

Conceptual Details:

- XGBoost (Extreme Gradient Boosting) is a scalable and efficient implementation of gradient boosting machines.
- It sequentially builds an ensemble of weak learners (decision trees) to minimize a predefined loss function.
- XGBoost incorporates regularization techniques to prevent overfitting and enhance generalization performance.
- It is known for its high accuracy and speed, making it popular for industry applications.

Performance Metrics:

- Mean Absolute Error (MAE): 70.11
- Mean Squared Error (MSE): 18634.90
- Root Mean Squared Error (RMSE): 136.51

Results and Discussion

- **Random Forest Regression** and **XGBoost Regression** outperformed **Linear Regression** significantly in terms of all performance metrics.
- Both Random Forest and XGBoost models exhibited lower errors (MAE, MSE, RMSE), indicating their superior predictive capabilities.
- Random Forest Regression and XGBoost Regression can provide more accurate demand forecasts, enabling better procurement planning and staffing decisions.

Recommendations

Based on the analysis and model performances, the following recommendations are proposed:

- Implement Random Forest Regression or XGBoost Regression models for demand forecasting at fulfillment centers.
- Regularly update the models with new data to enhance forecasting accuracy.
- Integrate the forecasting results into procurement planning and staffing strategies for optimal resource utilization.

Conclusion

The project successfully demonstrated the effectiveness of machine learning regression models in forecasting food demand for fulfillment centers. By leveraging advanced techniques such as Random Forest Regression and XGBoost Regression, the meal delivery company can make informed decisions regarding raw material procurement and staffing, ultimately improving operational efficiency and customer satisfaction.

Future Scope

- Explore advanced machine learning algorithms to further enhance demand forecasting accuracy.
- Incorporate additional variables such as weather patterns, holidays, and special events to refine the forecasting models.
- Extend the application of demand forecasting to other areas of the meal delivery business, such as delivery route optimization and inventory management.