

Delivery Time Prediction - Project Report

Introduction

This project aimed to build a regression model to predict delivery time for orders placed through Porter. The model uses various features including restaurant type, order details, and delivery partner availability. Our goal was to improve delivery time estimates to help optimize operational efficiency.

Data Overview

The dataset includes features such as order timestamps, item count, price range, number of delivery partners on shift, and distance between restaurant and customer. Categorical variables like store type and order protocol were also included.

Preprocessing & Feature Engineering

Timestamps were converted to datetime, and delivery duration was calculated in minutes. Categorical features were encoded, and additional features like hour of day, day of week, and a weekend indicator were derived. Outliers were handled using the IQR method.

Exploratory Data Analysis

EDA revealed that 'distance' had the strongest correlation with delivery duration. Other influential features included 'total_busy_dashers' and 'order_protocol'. Residual analysis showed a good model fit with minimal outliers after cleaning.

Model Building

We used a Random Forest Regressor, splitting the data into 80% training and 20% testing. The model was trained using scaled numerical features, and feature importance was analyzed. The top 10 features were used in the final model.

Delivery Time Prediction - Project Report

Evaluation Metrics

The model was evaluated using Mean Absolute Error (MAE) and R^2 Score. Residual plots, histograms, and Q-Q plots were used for error analysis. MAE was within acceptable range, indicating a reasonably accurate model.

Insights and Conclusion

Distance was the most critical factor affecting delivery time. Busy dashers and order protocol also played significant roles. The model can be used to support operational decisions and improve delivery efficiency. Future improvements could include incorporating traffic or weather data.