

Operational Insights

Based on descriptive analysis of the provided dataset the following are the operational conditions to enhance the business profitability and flow:

1. Prepare more on-call drivers during cloudy weather.
2. Have more call center agents during cloudy weather to receive complaints.
3. Since app usage is high during cloudy weather, provide higher storing and computing capacity during this season.

Data & Model Insights

The provided data consists of information with regards to delivery, it contains date time information, spatial and order size data, weather and traffic conditions. Categorical data were transformed to numerical data by label encoding and the time to delivery was calculated as the difference between the time the products were delivered and the time the products were ordered.

The goal of this experiment is to propose an accurate model that can predict the total time to deliver individual orders. Two models were developed for this problem, the linear regression model was used as a comparative baseline model and the XGBoost regression model was the proposed model. The models were fine-tuned using pre-processed datasets, for the detailed process on data cleaning and feature engineering please see the DS_INTERVIEW.ipynb. Parzan tree was utilized to determine the best parameters for XGBoost, this is to ensure that only optimal values of the parameters will be used in the modelling phase.

Using root mean squared error, RMSE, as the performance metric, k-fold validation was conducted, this is a process of repeatedly testing the model on the various combination of subsamples from the original data. Results showed that the XGBoost regression perform better with a k-fold cross validation score of 5.85 minutes, while the linear regression has a validation score of 9.19 minutes, in this case the lower the validation score the more accurate is the model. Hence, XGBoost regression was used for this experiment.

Model Recommendations

- The provided data were limited to 10 samples, to gain a better perspective of the data behavior, as well as to enhance the model performance, it is recommended to provide more datasets, both for business insighting and model training.
- If time is not a constraint, it would also be recommended to explore other complex models, like deep learning at determine its performance against the two existing models.