SAN DIEGO STATE UNIVERSITY

Optimizing Food Delivery Operations through Data-Driven Delivery Time Forecasting

Done By: Saiprashanth Paladugula
Poojitha Pakkeeru ,Siddharth Suresh Babu
Siva Gadiparthi
Department: Computer Science

Introduction

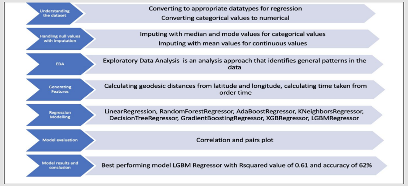
This project aims to develop a machine learning model for predicting food delivery times using historical data on past deliveries. Fast and reliable food delivery is critical for customer satisfaction, and unpredictable delays negatively impact experience. An accurate delivery time prediction model would empower food delivery businesses to set realistic expectations, optimize operations, and improve customer service.

Problem Motivation

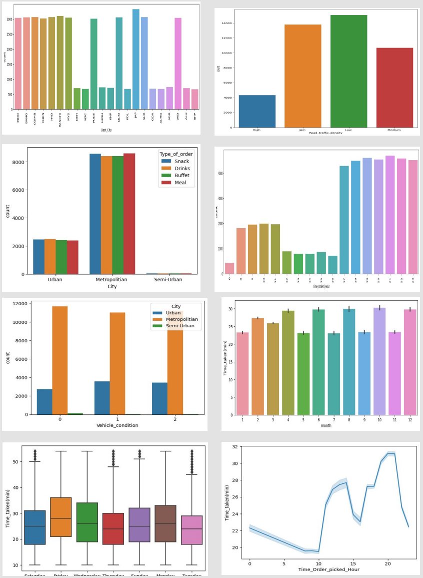
We have leveraged a dataset of over 45,000 past food deliveries with relevant features like driver details, order attributes, and geographic coordinates.

The framework of leveraging historical delivery data,engineering relevant predictors, and applying machine learning can extend to tackle last mile delivery challenges is widely applicable and it can be easily extended across several diverse industries. With custom optimization, such as accommodating for leet capacity, frequency, and stop density, can enable efficient, reliable fulfillment everywhere.

Approach



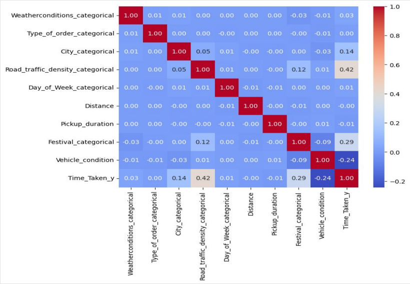
Approach Details

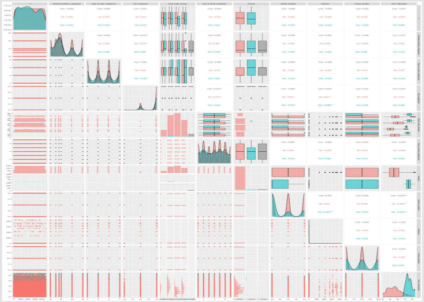


When you take a look at the count of orders in cities, it is highest in Jaipur and lowest in Bhopal. Metropolitan cities in most cases have the highest orders. All food items have the same number of orders. Weather conditions have a reverse effect as fog and stormy conditions have more food orders.

Low and Jam conditions in traffic density have more number of orders. It can be seen that when there is no festival in the country,the ordering frequency of food is high. Also the number of orders during festival is very less but the time taken to deliver the food is high, this could be because of the high traffic during festivals. We can also say that most orders are placed during 5PM to 11 PM during the day. The time taken to deliver food is higher at night times as per our line plot.

Evaluation





From the correlation plot we can say that few variables like traffic density, festival and city have correlation with our response variable Time Taken.

We can also note that Vehicle condition has a negative correlation with Time Taken and Festival variables.

We have done the ggpairs plot in R as we thought it depicts the relationship between all variables extremely well along with giving us correlation values.

Results

index	Modelling Name	MSE	R_2	RSE	Accuracy	RMSE
0	LinearRegression	60.5453	0.29744	7.78227	29.7444	7.78108
1	RandomForestRegressor	38.2967	0.55561	6.18938	55.5612	6.18843
2	AdaBoostRegressor	44.8627	0.47942	6.69898	47.9421	6.69797
3	KNeighborsRegressor	45.7624	0.46898	6.76582	46.8981	6.76479
4	DecisionTreeRegressor	69.384	0.19488	8.33097	19.4881	8.32971
5	GradientBoostingRegressor	35.2306	0.59119	5.93644	59.1191	5.93554
6	XGBRegressor	34.3252	0.6017	5.85966	60.1697	5.85877
7	LGBMRegressor	32.7635	0.61982	5.72481	61.9818	5.72394

LGBM Regressor has the highest R^2 value and the lowest RMSE value, indicating the best fit for our model with the chosen features.

Conclusions

Based on the Correlations plot and the Pairs plot, the predictors: road traffic density, vehicle condition, and festivals have the highest impact the time taken for delivery. This means that there is a strong correlation between the above mentioned predictors and the outcome. To improve customer satisfaction we can improve vehicle condition and maintain a bigger inventory than expected during festivals.

References

https://www.kaggle.com/code/jyotigupta86/delivery-time-eda
https://medium.com/@uma.bollikonda/6-types-of-regression-models-in-machine-learning-you-should-know-about-ed1e83dd0c4d
https://dl.acm.org/doi/abs/10.1145/3447548.3467068
https://ieeexplore.ieee.org/abstract/document/9346325