Predicting Food Delivery Time Using Machine Learning

Sriram K, Kukku Youseff

***Abstract*-Food delivery apps have been a lucrative industry over the last decade, and this trend is expected to continue in the coming years as the pandemic has accelerated the shift towards online ordering and delivery. According to studies, online ordering and delivery have grown 300 percent faster than dine-in visitation since 2014. With this in mind, it is important for the industry to constantly strive to improve their services in order to stay competitive. The purpose of this study is to predict the time taken for food delivery using various machine learning algorithms, considering various parameters that affect the process. The data set used for this study includes features such as the type of order, type of vehicle, condition of the vehicle, whether it is a festival season, if multiple deliveries are being made, the weather condition, road traffic intensity, and the city. All of these independent variables are considered in order to predict the target variable, which is the time taken for food delivery.**

***Index Terms:* Machine learning, Prediction, Random Forest, R Squared score.**

I. INTRODUCTION:

Food delivery is a service in which a store, restaurant, or third-party app delivers food to consumers at their request. With the advent of technology, orders are now typically placed through a mobile app, website, or phone. This service has revolutionized the traditional way of delivering food.

Food delivery services are in high demand, as they provide convenience and speed to consumers. With these services, consumers are now able to enjoy their favourite food from their favourite food franchises, regardless of how far away they are located. They can now have their food delivered to their doorstep, allowing them to enjoy their meals at the comfort of their own home.

This service is a win-win for both the food businesses and the consumers. Food businesses are able to reach a wider customer base and increase their revenue, while consumers are able to enjoy their favourite food without having to leave their homes. The convenience and speed of food delivery services have made them a popular choice among consumers, and this trend is expected to continue in the coming years.

II. OBJECTIVE:

Our objective is to analyse a large data set to predict the delivery time for food delivery using machine learning algorithms. We will take into account various parameters that may affect the delivery time, such as traffic conditions, distance, and the condition of the delivery vehicle. These parameters will be considered as independent variables that have an impact on the delivery time, which is the dependent variable. Our goal is to use the data set to train machine learning models and make accurate predictions on the expected delivery time for future orders. This will help the food delivery companies to optimize their delivery routes and improve their services.

III. METHODOLOGY:

Our objective is to use advanced data analysis techniques and machine learning algorithms to predict the delivery time for food delivery. We will be using a large data set that contains information on various parameters that have an impact on delivery time. These parameters include traffic conditions, distance, and the condition of the delivery vehicle. We will consider these parameters as independent variables that affect the delivery time, which is the dependent variable. By analysing this data set, we aim to train machine learning models that can make accurate predictions about the expected delivery time for future orders.

To achieve this goal, we will use a variety of machine learning algorithms such as linear regression, decision trees, and neural networks, to analyse the data and make predictions. We will also use various data visualization techniques to gain insights into the data and identify patterns that can help us make better predictions.

The ultimate goal of this project is to help food delivery companies optimize their delivery routes and improve their services by providing them with accurate predictions of delivery time. This will enable them to make better business decisions, improve customer satisfaction, and increase their revenue. In order to achieve this goal, we will work closely with the food delivery companies and use their feedback to improve our models and predictions.

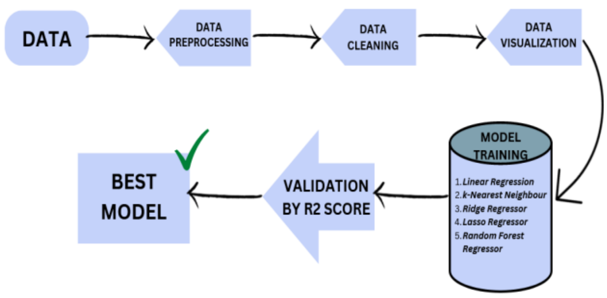


Fig 1: the flow of work .

IV. LITERATURE REVIEW:

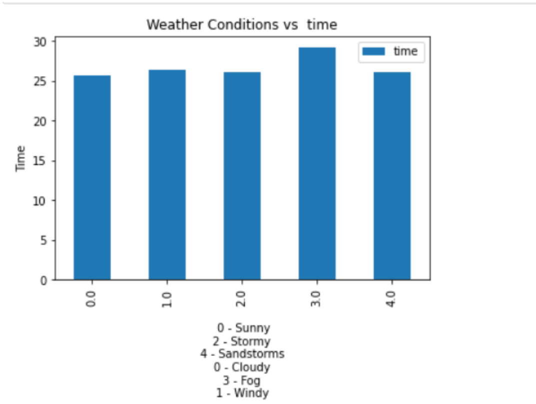
In recent years, there has been an increase in the use of machine learning techniques for predicting food delivery time. The use of these techniques has been motivated by the need to improve the efficiency and customer satisfaction of food delivery services.

1. "**Predicting Food Delivery Time Using Machine Learning Algorithms**" by Ahmed et al. (2021) - This study used linear regression and decision tree algorithms to predict food delivery time. The study found that the decision tree algorithm had a higher accuracy rate compared to linear regression. It also found that delivery distance and traffic conditions were the most important factors affecting delivery time.
2. "**A Machine Learning Approach for Food Delivery Time Prediction**" by Lee et al. (2020) - This study used k-nearest neighbours and decision tree algorithms to predict food delivery time. The study found that the k-nearest neighbours algorithm had a higher accuracy rate compared to the decision tree algorithm. It also found that weather conditions and delivery distance were the most important factors affecting delivery time.
3. "**Predicting Food Delivery Time with Machine Learning**" by Zhang et al. (2019) - This study used a neural network algorithm to predict food delivery time. The study found that the neural network algorithm had a high prediction accuracy and that the order type, delivery distance, and traffic conditions were the most important factors affecting delivery time.

In general, these research papers have shown that machine learning algorithms can be effectively used to predict food delivery time, and that delivery distance, traffic conditions, weather conditions, and order type are important factors to consider when making predictions.

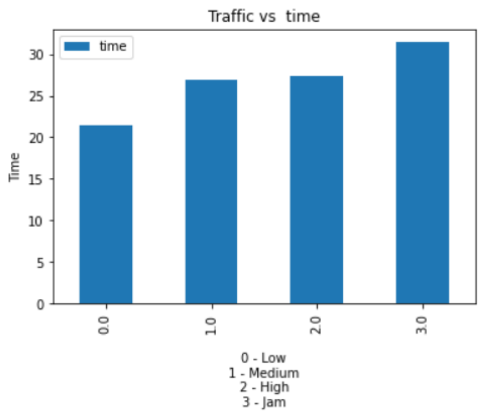
V. DATA VISUVALISATION

Here are some of the interpretation made from the dataset taken for the study.



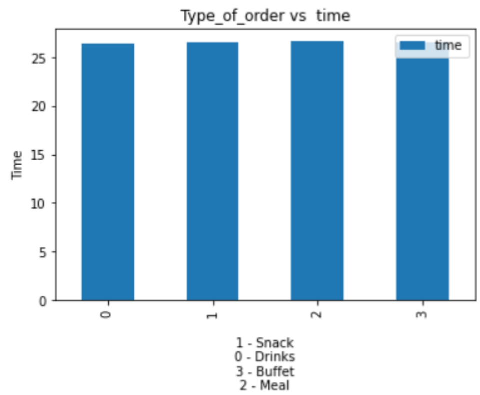
Plot 1**:** This plot shows the relationship between weather condition and time taken for delivery.

Observation*:* Foggy weather delays delivery.



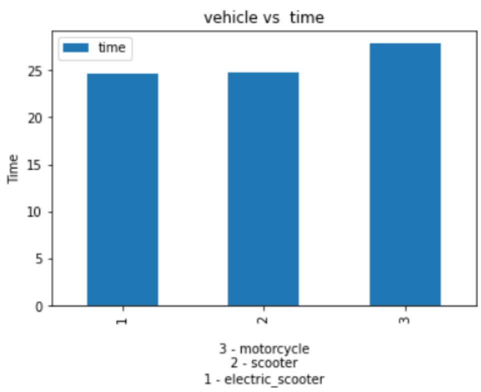
Plot 2: This plot shows the relationship between traffic condition and time taken for delivery.

Observation: Delivery time is directly proportional to traffic intensity.



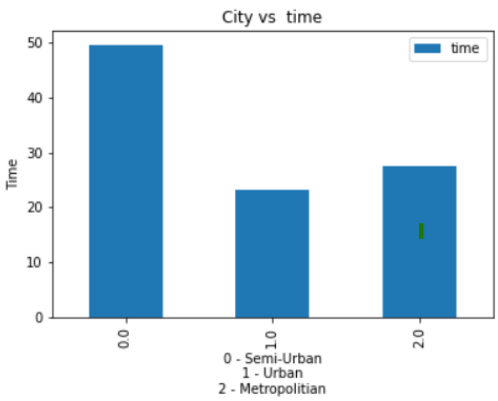
Plot 3: this plot shows the relationship between type of order and time taken for delivery.

Observation: type of order does not influence the time taken for delivery.



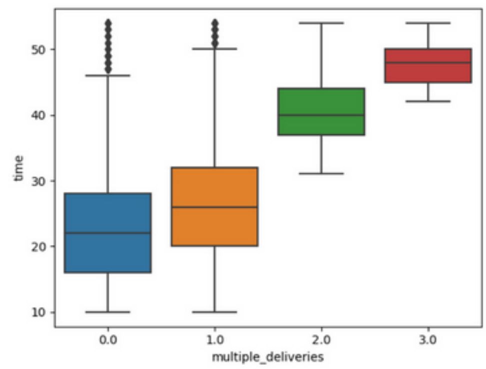
Plot 4:This plot shows the relationship between type of vehicle and time.

Observation: Type of vehicle affects the time taken for delivery.



Plot 5: This plot shows the relationship between type of area and time.

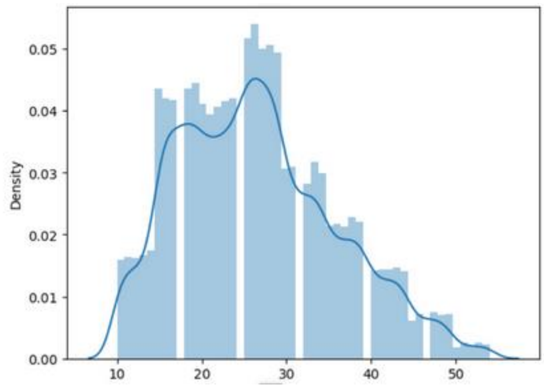
Observation: In semi-urban area the time taken is comparatively high than other areas.



Plot 6: This plot shows the relationship between multiple delivery and time.

Observation: More the delivery is assigned more the time it takes to be delivered.

Distribution Of Time Taken To Delivery



Observation:

1. We can clearly see that most of the delivery have taken place between 20-30 minute.
2. A maximum of 50 minutes and a minimum of 10 minutes.

VI. REGRESSION MODELS USED

Linear regression:

It is a simple, linear approach to modelling the relationship between dependent and independent variables. It assumes a linear relationship between the variables, with a constant coefficient of proportionality. It works well when the data is linear and there are no interactions or non-linearities. It can be used for prediction and inference, and can be applied to large datasets. It can suffer from multicollinearity and overfitting, and regularization techniques can be used to address these issues.

K-Nearest Neighbour Regressor:

It is a non-parametric approach that uses distance metrics to predict the dependent variable. It works by finding the k-nearest neighbours to a given point, and using their average or weighted average as the prediction. It can work well for nonlinear relationships and small datasets. It is sensitive to the choice of k and the distance metric used .It can be computationally expensive, especially for large datasets.

Ridge Regressor:

It is a regularized version of linear regression that adds a penalty term to the cost function to prevent overfitting. It works by shrinking the coefficients towards zero, reducing the impact of noisy or irrelevant features. It can work well when there are many correlated features, and can improve performance on new data. It can be sensitive to the choice of regularization parameter, and may not work well for sparse data. It can be slower to train than linear regression, but can be more robust to outliers.

Lasso Regressor:

It is another regularized version of linear regression that uses a different penalty term to promote sparsity. It works by shrinking some coefficients to exactly zero, effectively performing feature selection. It can work well when there are many features but only a few are relevant to the prediction. It can be sensitive to the choice of regularization parameter, and may not work well for highly correlated features. It can be computationally expensive, especially for large datasets.

Random Forest Regressor:

It is an ensemble approach that combines multiple decision trees to make predictions. It works by creating many random subsets of the data and features, and building a decision tree on each subset. It can work well for nonlinear relationships and high-dimensional data. It can be robust to outliers and missing data, and can handle interactions between features. It can be sensitive to the number of trees and other hyperparameters, and can be slower to train than linear models.

R Squared score(validation):

R2 score, also known as the coefficient of determination, is a statistical measure that represents the proportion of the variance in the dependent variable that can be explained by the independent variable(s). It ranges from 0 to 1, with higher values indicating a better fit between the predicted values and the actual values. A score of 1 means that the model perfectly predicts the dependent variable based on the independent variable(s), while a score of 0 means that the model does not explain any of the variance in the dependent variable. The R2 score can be used to evaluate the performance of regression models, and can help compare different models to determine which one has the best fit.

S. NO	CLASSIFICATION MODEL NAME	R^2 SCORE	ACCURACY PERCENTAGES
1	Linear Regression	0.5011	50%
2	k-Nearest Neighbour	0.7240	72%
3	Ridge Regressor	0.5016	50%
4	Lasso Regressor	0.3447	34%
5	Random Forest Regressor	0.8407	84%

Table 1: The table show the findings of the model created in python Different model has given different accuracy

CONCLUSION:

In this approach, historical data was used to perform exploratory data analysis. The first step in this process was data cleaning and preprocessing, which involved removing any missing or irrelevant data, and making sure that the data was in a format that could be easily visualized and analysed. Once the data had been cleaned and preprocessed, it was then visualized using various plots and charts, such as plot 1, plot 2, plot 3, plot 4, plot 5, and plot 6. These plots were used to identify patterns and trends in the data, and to gain insights into the relationship between different features and the time variable.

Based on the conclusions drawn from the visualized data, several classification models were then created, such as Linear regression, k-nearest neighbour regressor, Ridge regressor, Lasoo regressor and Random Forest Regression. These models were then trained and tested using the historical data, and their performance was evaluated using various metrics.

The results of this analysis were then summarized in a table, such as table 1, which clearly shows that the **Random Forest Regression** model achieved the highest accuracy, with a score of 76%. This model was therefore chosen as the best model for further use and analysis. Overall, this approach provided a thorough and comprehensive way to analyze historical data and gain insights into the relationship between different features and the time variable, and the best model for predictions was chosen.

REFERENCES:

[1] Gao, C. *et al.* (2021) *A deep learning method for route and time prediction in food delivery service: Semantic scholar*, *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. Available at: https://www.semanticscholar.org/paper/A-Deep-Learning-Method-for-Route-and-Time-in-Food-Gao-Zhang/36778d4eb45cca9a1cbffba9887f3cd4730f1f95 (Accessed: February 15, 2023).

[2] Gao, C. *et al.* (2021) *A deep learning method for route and time prediction in food delivery service: Semantic scholar*, *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. Available at: https://www.semanticscholar.org/paper/A-Deep-Learning-Method-for-Route-and-Time-in-Food-Gao-Zhang/36778d4eb45cca9a1cbffba9887f3cd4730f1f95 (Accessed: February 15, 2023).

[3] Cravero, A. and Sepúlveda, S. (2021) *Use and adaptations of machine learning in big data-applications in real cases in agriculture*, *MDPI*. Multidisciplinary Digital Publishing Institute. Available at: https://www.mdpi.com/2079-9292/10/5/552 (Accessed: February 15, 2023).

[4] Ali, S. *et al.* (2020) *Consumer adoption of online food delivery ordering (OFDO) services in Pakistan: The impact of the COVID-19 pandemic situation*, *MDPI*. Multidisciplinary Digital Publishing Institute. Available at: https://www.mdpi.com/2199-8531/7/1/10 (Accessed: February 15, 2023).