# Food Delivery Time Prediction

**Team members :**
Anuja Surana 002644394 T Navya Sushma 002708292

Pratik Poojari 002645173 Vaidehi Tawde 002643847

**(a) Definition of the Problem:**

The food delivery industry hinges on the precision and timeliness of its deliveries. An integral part of enhancing customer satisfaction and the overall experience is the ability to predict delivery times accurately. The goal of this project was to construct a model capable of predicting the time taken for a delivery, given a set of features associated with the order and the delivery process.

**(b) Rational of Target Variable Selection:**

The chosen target variable for this project is "Time_taken (minutes)," representing the duration taken to complete a delivery. Predicting this variable is vital as it directly influences customer satisfaction. If businesses can estimate delivery times with higher accuracy, they can manage expectations better, allocate resources more efficiently, and optimize the delivery process.

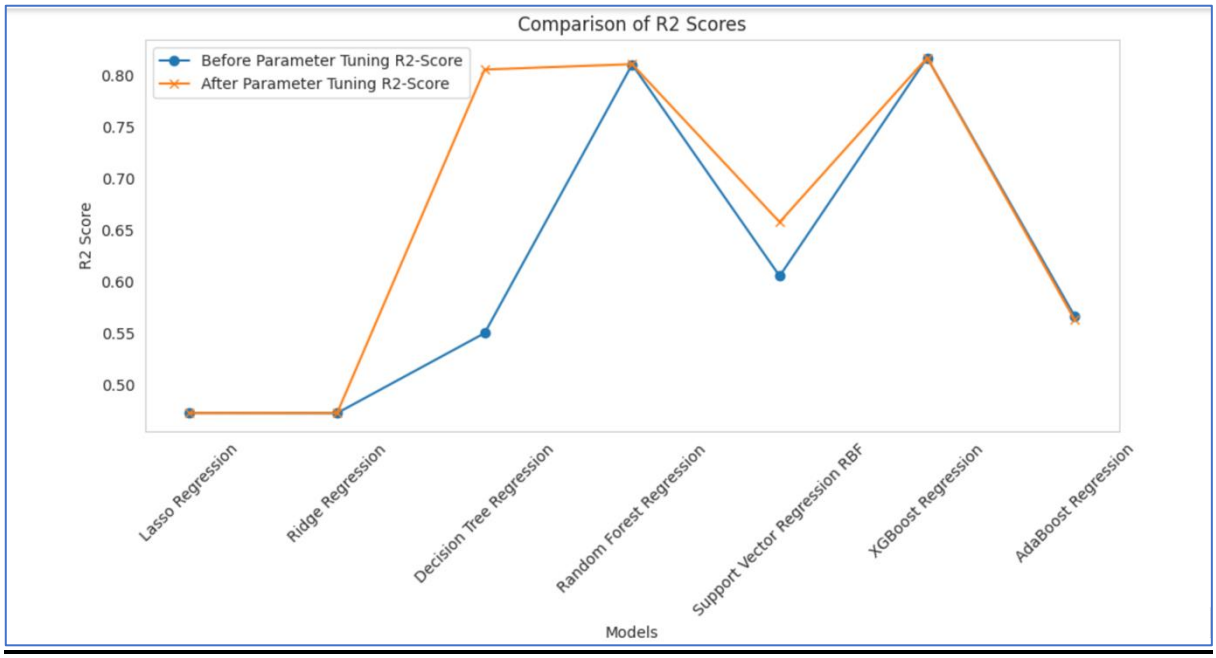
**(c) A Short Note on Which Kind of Machine Learning is Suitable for This Problem:**

Given that the aim is to predict a continuous numeric value (delivery time in minutes) based on a set of input features, this problem is best suited for supervised regression. Multiple regression algorithms were employed, including linear regression, Lasso, Ridge, Decision Tree, Random Forest, Support Vector Regression, XGBoost, and AdaBoost Regression. Regression techniques are apt for this problem as they allow for the prediction of a continuous outcome variable from one or more predictor variables.

**(d) Conclusion:**

Through rigorous data cleaning, feature engineering, and regression modeling, a predictive model was built to forecast the food delivery time. This model has the potential to significantly enhance the delivery experience by providing accurate time estimates. Based on the evaluation of various regression models using the engineered features, we have observed the following results: Xgboost stands out as the top-performing model with an impressive R2 score of 0.81. This indicates a strong ability to capture the variance in the target variable.

For Artificial Neural Network (ANN), the R2 score was best when using 2 hidden layers, each with 128 neurons. This configuration achieved the highest predictive performance among the ANN models tested. By implementing Grid Search CV for hyperparameter tuning, we achieved notable performance improvements for the Decision tree regressor and Support vector regression models. However, other models showed limited enhancement even with hyperparameter tuning




**(e) Scope for Future Work:**

While the current model is a strong starting point, there are areas for enhancement: Incorporate real-time traffic data for more dynamic predictions.
Use deep learning techniques to capture complex relationships in the data.
Integrate feedback loops from delivery personnel and customers to continually refine the model. Real-Time Updates by integrating real-time data feeds into the prediction, model to provide dynamic and up-to-the-minute delivery time estimates. Personalized Predictions: Create personalized delivery time predictions for each customer based on their order history, location, and preferred delivery times. Customer Feedback Integration: Incorporate customer feedback and reviews into the prediction model. Demand Forecasting: Combine delivery time predictions with demand forecasting to optimize staffing and resource allocation for different time periods, ensuring timely deliveries even during peak hours.

These are just a few examples of the potential future scope for the food delivery time prediction dataset. The key is to leverage data-driven insights to improve the accuracy, efficiency, and overall experience of food delivery services.

**(f) Did You Go Beyond the Expectation and Deserve the Extra 5 Points?**

Based on the insights from the Jupyter notebook:
Multiple regression models were deployed, showcasing a comprehensive approach to the problem. An exhaustive EDA process including maps was undertaken, ensuring that the model was built on a solid foundation of understanding. Also, Model Training implementation was done from scratch and Corelation Heatmaps were developed and visualized. The project not only focused on modelling but also on data quality through cleaning and feature engineering. Given these points, it can be argued that there was an extra effort put into ensuring the project's success, potentially warranting the extra 5 points. This summary provides a concise overview of the project. You can further refine or expand upon any section as you see fit, based on additional details or specific results from your Jupyter notebook.