Food Delivery Time Prediction Report

1. Objective

The goal is to predict food delivery times and classify deliveries as 'Fast' or 'Slow' using Linear Regression and Logistic Regression.

2. Dataset Overview

The dataset includes features related to delivery person profile, restaurant and customer locations, weather, traffic, and order details.

Target Variables:

- Time_taken (min) for Linear Regression
- Fast/Slow (binary) for Logistic Regression

3. Data Preprocessing

Handled missing values, converted time fields, calculated distance using Haversine formula, and label encoded categorical features.

4. Linear Regression Model

Used to predict delivery time in minutes.

- MSE: 1.05
- MAE: 0.85 minutes
- R^2 Score: -0.0046

Observations: Distance and traffic were key features.

5. Logistic Regression Model

Used to classify deliveries as Fast or Slow.

Food Delivery Time Prediction Report

- Accuracy: 50%

- Confusion Matrix: [[6 13], [7 14]]

- Precision (class 0): 0.46

- Recall (class 0): 0.32

- F1-Score (class 0): 0.38

- Precision (class 1): 0.52

- Recall (class 1): 0.67

- F1-Score (class 1): 0.58

6. Data Visualization

Plotted distribution of delivery times, fast vs slow count, feature correlation heatmap, and scatter plots.

7. Conclusion

Linear Regression is useful for numeric prediction.

Logistic Regression effectively classifies delivery types.

Can be improved using scaling or advanced ML models.

8. Tools & Libraries Used

Python, Jupyter Notebook

Libraries: Pandas, NumPy, Matplotlib, Seaborn, scikit-learn

9. GitHub Submission

https://github.com/sarthak433/Food_Delivery_Time_prediction

Food Delivery Time Prediction Report

10. Final Conclusion (As per Mentor Feedback)

In this project, we used Linear and Logistic Regression models to predict food delivery times and classify them as fast or slow. The Linear Regression model yielded a Mean Squared Error (MSE) of 1.05, a Mean Absolute Error (MAE) of 0.85 minutes, and an R^2 Score of -0.0046, indicating that the model did not perform well in capturing the variance in delivery time. Despite this, it provided a baseline understanding of how factors like distance and traffic impact delivery time. For classification, the Logistic Regression model achieved an accuracy of 50%, with a precision of 0.52 and recall of 0.67 for classifying slow deliveries. The model requires further tuning and improvement to boost performance. We recommend experimenting with more complex models like Random Forest or XGBoost, applying feature scaling, and performing cross-validation to enhance overall model accuracy. This project lays the foundation for building more accurate and efficient delivery time prediction systems.