Final Report

Food Delivery Time Prediction

1□ Project Overview

The objective of this project is to predict food delivery time using machine learning techniques based on various operational and environmental factors. The project also aims to classify deliveries as fast or delayed using logistic regression.

The analysis includes data preprocessing, exploratory data analysis (EDA), regression modeling, classification modeling, and generation of business insights.

2□ Dataset Description

The dataset contains 200 records of food delivery orders with the following features:

- Distance (km)
- Weather Conditions
- Traffic Conditions
- Delivery Person Experience
- Order Priority
- Order Time
- Vehicle Type
- Restaurant Rating
- Customer Rating
- Order Cost
- Tip Amount
- Delivery Time (Target variable)

Each row represents one food delivery instance.

3□ Data Preprocessing Steps

The following preprocessing steps were performed:

1. Data Cleaning

- Checked for missing values.
- No missing values were found.
- Removed unnecessary columns:
    - Order_ID
    - Customer_Location
    - Restaurant_Location

2. Encoding Categorical Variables

Categorical variables such as:

- Weather Conditions
- Traffic Conditions
- Order Priority
- Order Time
- Vehicle Type

were converted into numerical format using one-hot encoding.

3. Feature Scaling

Continuous numerical features such as:

- Distance
- Delivery Person Experience
- Restaurant Rating
- Customer Rating
- Order Cost
- Tip Amount

were standardized using StandardScaler to ensure uniform scaling.

4□ Exploratory Data Analysis (EDA)

Descriptive Statistics

Statistical measures such as mean, standard deviation, minimum, maximum, and quartiles were calculated to understand data distribution.

Correlation Analysis

A heatmap was generated to analyze correlations between features and delivery time.

Key observations:

- Distance has a noticeable impact on delivery time.
- Traffic conditions influence delivery delay.
- Delivery person experience affects delivery speed.

Outlier Detection

A boxplot was generated for Delivery Time to visualize potential outliers.

5□ Predictive Modeling

Two models were implemented:

A. Linear Regression (Delivery Time Prediction)

Objective:
Predict exact delivery time in minutes.

Evaluation Metrics:

- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- R^2 Score

Results:

- MSE: 1021.93
- MAE: 27.18
- R^2: -0.10

Interpretation:
The negative R^2 indicates that the linear model does not explain the variance effectively. This suggests that delivery time may not follow a strictly linear pattern and may require more advanced models.

B. Logistic Regression (Fast vs Delayed Classification)

Objective:
Classify deliveries as:

- Fast (≤ 60 minutes)
- Delayed (> 60 minutes)

Evaluation Metrics:

- Accuracy
- Precision
- Recall
- F1 Score
- Confusion Matrix
- ROC Curve

Results:

- Accuracy: 52.5%
- Precision: 60.6%
- Recall: 76.9%
- F1 Score: 67.8%

Interpretation:
The classification model performs moderately well. High recall indicates that the model successfully identifies most delayed deliveries.

6□ Model Comparison

Model	Purpose	Performance
Linear Regression	Predict exact delivery time	Weak performance
Logistic Regression	Classify delayed deliveries	Moderate performance

Logistic regression performed better in practical terms because identifying delays is often more important than predicting exact time.

7□ Actionable Insights & Recommendations

Based on model analysis, the following operational improvements are suggested:

1. Route Optimization
   Optimize delivery routes for long-distance orders.
2. Traffic-Based Staffing
   Increase delivery staff during high-traffic periods.
3. Driver Training
   Improve training programs for less experienced delivery personnel.
4. Peak Hour Management
   Allocate more resources during evening and night peak hours.
5. Data Expansion
   Collect more detailed route and real-time traffic data to improve prediction accuracy.

8□ Conclusion

This project successfully implemented a complete machine learning pipeline including:

- Data preprocessing
- Exploratory data analysis
- Regression modeling
- Classification modeling
- Performance evaluation
- Business insight generation

While linear regression showed limited predictive power, logistic regression provided meaningful insights into delivery delays. Further improvements can be achieved using advanced machine learning models such as Random Forest or Gradient Boosting.