Delivery Time Prediction Project - Detailed Report & SOP

1. INTRODUCTION

The Delivery Time Prediction Project aims to predict delivery times based on various features like distance, weather, traffic, and courier experience.

This will help logistics companies reduce delays, optimize resources, and enhance customer satisfaction.

2. PROBLEM STATEMENT

Given a dataset with order details and delivery conditions, we aim to:

- Perform Exploratory Data Analysis (EDA) to identify patterns.
- Build a predictive machine learning model.
- Generate actionable insights.

3. DATASET OVERVIEW

Source: Provided CSV file.

Shape: 1000 rows x 9 columns.

Features:

Order_ID: Unique identifier for each order.

Distance_km: Delivery distance in kilometers.

Weather: Weather condition (Sunny, Rainy, Cloudy).

Traffic_Level: Traffic status (Low, Medium, High).

Time_of_Day: Morning, Afternoon, Evening, Night.

Vehicle_Type: Bike, Car, Van, etc.

Preparation_Time_min: Time taken to prepare the order.

Courier_Experience_yrs: Courier's delivery experience.

Delivery_Time_min: Target variable (delivery time).

4. TOOLS & TECHNOLOGIES

Python (Pandas, NumPy, Matplotlib, Seaborn, scikit-learn, fpdf).

Jupyter Notebook for development.

Matplotlib & Seaborn for visualization.

Random Forest Regressor for modeling.

5. PROJECT EXECUTION STEPS

Step 1: Data Loading

We use pandas to read the CSV dataset into a DataFrame.

Delivery Time Prediction Project - Detailed Report & SOP

Step 2: Data Inspection
Check shape, data types, and sample records.
Identify missing values and inconsistencies.

Step 3: Data Cleaning
Fill missing categorical values with mode.
Fill missing numerical values with mean.
Rename inconsistent columns.

Step 4: Exploratory Data Analysis (EDA)
Univariate analysis: Distribution of delivery times.
Bivariate analysis: Impact of weather, traffic, distance on delivery time.
Multivariate analysis: Correlations between features.

Step 5: Feature Encoding
Use One-Hot Encoding for categorical variables.

Step 6: Train-Test Split
80% training, 20% testing split.

Step 7: Model Training
Random Forest Regressor is chosen due to its robustness with mixed data types.

Step 8: Model Evaluation
Metrics:
MAE: 5.90 min
RMSE: 8.82 min
R^2 Score: 0.83

Step 9: Insights
Traffic and weather majorly affect delivery time.
Longer distances and preparation times increase delays.

6. SOP (STANDARD OPERATING PROCEDURE)
1. Install dependencies using pip.

Delivery Time Prediction Project - Detailed Report & SOP

2. Place dataset.csv in working directory.
3. Open Jupyter Notebook and run cells sequentially.
4. Perform EDA using plots to understand patterns.
5. Encode categorical variables.
6. Split data and train model.
7. Evaluate model performance using metrics.
8. Generate reports for documentation.

7. CONCLUSION

This project effectively predicts delivery time with high accuracy and provides valuable insights for improving delivery operations.