

# Design System Phase for a Food Delivery Company

Imagine working for a food delivery company such as Glovo, Just Eat, or Deliveroo. Your task is to design a system capable of estimating the total delivery time (in minutes) of an order. This estimate should be displayed to the customer immediately after confirming the purchase. Given the availability of three possible models—Linear Regression, Logistic Regression, and Decision Trees—the following describes the complete design process, from problem definition to production deployment.

## 1. Problem Definition and Success Metric

The goal is to predict the total delivery time of an order, expressed in minutes. This is a **supervised regression problem**, where the target variable is the actual delivery time recorded from historical data. The success of the model can be measured using the following metrics:

- **Mean Squared Error (MSE)** — to quantify the average squared difference between predicted and actual delivery times.
- **Accuracy (tolerance-based)** — defined as the percentage of predictions that fall within an acceptable range of the true delivery time (for instance, within  $\pm 5$  minutes).

## 2. Data and Feature Engineering

To build an accurate predictive model, several factors that influence delivery time must be considered. The proposed features are:

- a) Distance between the client and the restaurant (km);
- b) Distance between the rider and the restaurant at the moment of assignment (km);
- c) Estimated food preparation time (minutes);
- d) Traffic intensity (numerical scale from 1 to 10);
- e) Average vehicle speed (km/h);

- f) Type of day (binary variable: festive vs. ordinary);
- g) Order load ratio (current number of orders / restaurant's maximum order capacity);
- h) Time of day (categorized into 12 two-hour intervals).

Feature engineering may include normalization of numerical variables (e.g., distances, speeds), encoding of categorical features (e.g., festive day, time interval), and removal of outliers from extreme delivery cases.

### 3. Model Selection

Among the available options, **Linear Regression** is selected as the preferred model. This choice is motivated by its simplicity, interpretability, and effectiveness in estimating continuous values such as time. The linear model provides a clear understanding of how each feature affects the delivery time, making it easier to monitor and improve the system over time. Although Decision Trees could capture nonlinear relationships, the linear regression model offers greater stability and generalization capabilities, especially in the initial design phase.

### 4. Model Evaluation

The dataset would be divided into training, validation, and testing subsets. Model performance would be assessed using:

- **Mean Squared Error (MSE)** on the test set to evaluate accuracy and detect potential overfitting.
- A tolerance-based accuracy metric to evaluate the proportion of predictions close to the real delivery time.

Residual analysis would also be performed to identify systematic errors and potential biases (e.g., underestimation during peak hours).

### 5. System Operation in Production

Once deployed, the system would operate as follows:

- **Input collection:**
  - Restaurants provide static or semi-static information such as preparation time, day type, operating hours, and order load ratio.
  - External APIs (e.g., Google Maps) supply dynamic data such as distances and traffic estimates.
  - The company system (e.g., Glovo backend) provides the average vehicle speed based on fleet data.

- **Prediction:** When an order is confirmed, the system automatically gathers all relevant features, processes them through the trained linear regression model, and generates an estimated delivery time.
- **Output:** The estimated delivery time is displayed to the user in real time, and the actual outcome is stored for continuous model retraining and performance monitoring.

This system would thus allow the company to provide accurate and adaptive delivery time estimates, improving customer satisfaction and operational efficiency.