**Case Study: Predicting Delivery Delays in Last-Mile Logistics**

**Background**

A leading logistics company is facing increasing customer complaints regarding late deliveries. Timely delivery is a critical factor in customer satisfaction, brand reputation, and operational efficiency. However, deliveries are often influenced by multiple operational factors such as traffic, weather, route conditions, and driver performance.

To improve service levels, the company wants to develop a predictive model that can forecast whether a delivery will be delayed before it leaves the warehouse. This will help in proactive decision-making, such as reallocating drivers, adjusting routes, or updating customers in advance.

**Business Problem**

How can we predict whether a delivery will be delayed based on operational and environmental factors?

The company wants to:

* Identify the key drivers of delivery delays.
* Build a model that predicts the likelihood of a delay.
* Use these insights to optimize operations and improve customer satisfaction.

**Data Description**

The dataset consists of multiple variables related to delivery operations:

1. Delivery Distance – Distance from the warehouse to the customer (in km).
2. Traffic Congestion – Level of traffic congestion during delivery (categorical: Low, Medium, High).
3. Weather Condition – Weather during delivery (categorical: Clear, Rainy, Stormy, Foggy, etc.).
4. Delivery Slot – Scheduled time slot of delivery (Morning, Afternoon, Evening, Night).
5. Driver Experience – Number of years of delivery driving experience.
6. Number of Stops – Total number of stops before reaching the final delivery location.
7. Vehicle Age (new) – Age of the delivery vehicle in years.
8. Road Condition Score (new) – Rating of road conditions on the route (1–5 scale, 1 = Poor, 5 = Excellent).
9. Package Weight (new) – Total weight of the package in kilograms.
10. Fuel Efficiency (new) – Fuel efficiency of the delivery vehicle (km per liter).
11. Warehouse Processing Time (new) – Time taken to process and dispatch the package at the warehouse (in minutes).
12. Delivery Delay (Target) – Whether the delivery was delayed (Yes/No).

**Objective**

The goal is to build a classification model that predicts whether a delivery will be delayed based on the above factors.

**Expected Outcome**

* A predictive model that assigns a probability of delay to each delivery.
* Identification of the most influential factors (e.g., warehouse processing time, traffic congestion).
* Business recommendations to reduce delays (e.g., improve warehouse processes, use newer vehicles, adjust routes in poor road conditions).