SCM optimization using AI

Banana Problem

Problem Statement:

Predicting Delivery Delays in Multi-Modal Logistics

Background:

A logistics company handles shipments using **multiple transportation modes** — trucks, rail, and air. However, they are experiencing frequent **delivery delays**, leading to dissatisfied customers and increased operational costs.

The company wants to **forecast whether a shipment will be delivered on time or delayed** based on historical shipment data and key factors like distance, vehicle type, traffic, weather, and driver experience.

Your task as a **data analyst** is to build a **predictive model** that determines the probability of delay for upcoming shipments and identifies the **major factors contributing to delays**.

Objective:

1. Predict whether a shipment will be **delayed (1) or on time (0)**.
2. Identify **key factors** (distance, vehicle type, weather, traffic, driver experience) that contribute to delays.
3. Provide insights to improve **logistics efficiency and route planning**.

Expected Outputs

1. **Prediction File (CSV)**:

Shipment_ID	Predicted_Delay
S006	1
S007	0
S008	1

2. **Performance Metrics Table:**
| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
| Logistic Regression | 0.87 | 0.85 | 0.88 | 0.86 | 0.91 |

3. **Feature Importance Plot:**
Visual representation of top variables influencing shipment delays.

Columns Explained:

·**Shipment_ID:** Unique identifier for each shipment

·**Origin / Destination:** Start and end cities

·**Distance_km:** Distance of shipment in kilometers

·**Vehicle_Type:** Mode of transport (Truck, Rail, Air)

·**Route_Traffic_Index:** Traffic congestion score (1–10, higher = worse traffic)

·**Weather:** Weather condition during shipment (Clear, Rainy, Foggy, Stormy)

·**Driver_Experience_Years:** Years of experience of the driver (for road transport)

·**Scheduled_Delivery_Hours:** Planned delivery time in hours

·**Actual_Delivery_Hours:** Actual delivery time in hours

·**Delayed:** Target variable (0 = On-time, 1 = Delayed)

Instructions

1. Open the dataset Shipment_Delivery_Dataset

2. Handle missing values and clean the data

3. Choose a classification model

4. Train the model

5. Evaluate performance (Accuracy, Precision, Recall, F1-Score)

6. Predict delays for the test shipments

7. Save predictions in CSV: Shipment_ID, Predicted_Delay

8. Submit: report, predictions CSV, and code/script