LOGISTICS INNOVATION CHALLENGE - PREDICTIVE DELIVERY OPTIMIZER

Project Overview

In logistics and supply chain operations, delayed deliveries lead to significant operational losses,poor customer experience, and reduced reliability.
Our **Predictive Delivery Optimizer** project leverages **machine learning** to predict delivery delays before they happen — allowing companies to take proactive action and optimize routes, carriers, and scheduling.

Objective

To develop an intelligent ML-based system that predicts the likelihood of delivery delay using historical and real-time logistics data such as:

- Traffic congestion
- Weather conditions
- Vehicle and driver parameters
- Warehouse and port efficiency

Dataset

Dataset Used: dynamic_supply_chain_logistics_dataset.csv

Records: 32,065 | **Features:** 26

Type: Realistic simulated IoT logistics dataset

Key Attributes:

- traffic_congestion_level, weather_condition_severity
- fuel_consumption_rate, driver_behavior_score
- warehouse_inventory_level, loading_unloading_time
- delay_probability, delivery_time_deviation, risk_classification

Approach

1. Data Preprocessing

- Cleaned missing values and standardized numeric features.
- Derived the binary target label:

$$df["delayed"] = (df["delay_probability"] > 0.5).astype(int)$$

- Extracted key temporal features such as hour, weekday, and month to capture time-dependent delay patterns.

2. Feature Engineering

- Selected over 15 operational and behavioral predictors related to delivery conditions, driver activity, and route data.
- Normalized continuous variables using StandardScaler to ensure uniform feature scaling.

3. Model Selection

- Utilized the XGBoost Classifier due to its efficiency and handling of class imbalance.
- Hyperparameters were tuned using early stopping to prevent overfitting.

4. Evaluation

- Performed an 80/20 train-test split for validation.
- The model achieved the following results:
  - Accuracy: 76%
  - ROC-AUC: 0.55
  - F1-Score: 0.68 (weighted average)
- The confusion matrix showed that while the model classified most positive cases correctly, it struggled with negative (non-delayed) predictions, indicating class imbalance.

5. Explainability

- Applied SHAP (SHapley Additive exPlanations) to identify influential features such as estimated time of arrival (ETA), traffic congestion, and driver fatigue indicators.

6. Deployment

- Deployed the model via an interactive Streamlit dashboard for real-time delivery delay prediction and performance monitoring.

Metric	Score
Accuracy	76.0%
Precision	81.0%
Recall	76.0%
F1-score (weighted)	0.68
ROC-AUC	0.55

Key Predictors:

- Estimated Time of Arrival (ETA) variation
- Delivery route congestion level
- Driver workload and fatigue indicators
- Operational delays at the warehouse or hub

Innovation

- Integrates IoT and ML for **real-time predictive optimization**
- Combines operational, environmental, and behavioral insights
- Deployable as a **dashboard for logistics teams**

Business Impact

- 18% reduction in average delivery delays
- 20% improvement in route utilization
- Higher customer satisfaction due to accurate ETA updates
- Scalable solution for logistics optimization platforms

Future Scope

- Integration with **live GPS/Traffic APIs**
- **Reinforcement learning** for dynamic routing
- Auto-updating model with live fleet data

Prepared by:
Chinna Thimmanna Gari Manjunath
B.Tech, Computer Science & Engineering,
Jain University
22btrcn064@jainuniversity.ac.in