Logistics Innovation Challenge – AI Internship Case Study

Predicting Delivery Delay Risk using Machine Learning

1. Problem Statement

NexGen Logistics Pvt. Ltd. operates a multi-city logistics network handling diverse product categories and delivery priorities. Despite steady growth, the company faces frequent delivery delays that negatively impact customer satisfaction, operational efficiency, and costs.

Currently, delivery issues are identified **after** delays occur, leading to reactive decision-making. There is a need for a **predictive, data-driven system** that can identify delay risks early and enable proactive intervention.

2. Objective

The objective of this project is to:

- Predict whether an order is likely to be delayed before delivery

- Identify high-risk deliveries using historical data

- Enable proactive operational and customer-facing decisions

- Demonstrate how AI can transform logistics from reactive to predictive operations

3. Data Used

The solution uses multiple interconnected datasets provided in the case study:

- **Orders Data**: Order details, priority, product category, and order value

- **Delivery Performance Data**: Promised vs actual delivery days, delivery status, costs

- **Routes Data**: Distance, traffic delays, fuel consumption, toll charges

- **Customer Feedback Data**: Ratings and customer sentiment indicators

- **Cost Breakdown Data**: Fuel, labor, maintenance, and overhead costs

All datasets were merged using a common key: **Order_ID**.

4. Approach & Methodology

4.1 Feature Engineering

Key features selected for prediction include:

- Delivery priority

- Product category

- Order value

- Route distance

- Traffic delay

- Fuel and labor costs

- Delivery cost

- Customer rating

Missing values were handled using median imputation for numeric features.

4.2 Machine Learning Model

- **Model Used**: Logistic Regression

- **Problem Type**: Binary classification (Delayed / Not Delayed)

Logistic Regression was chosen because:

- It is interpretable and explainable

- It performs well on structured tabular data

- It is suitable for fast deployment in business environments

Categorical variables were label-encoded, and numerical features were standardized using feature scaling.

5. Solution Prototype

A **Streamlit-based interactive dashboard** was developed to demonstrate the solution.

Key functionalities:

- User inputs order and route details

- The system predicts delay risk probability

- Risk is categorized as Low, Medium, or High

- Business-friendly recommendations are displayed

This prototype demonstrates how AI predictions can be directly consumed by business users.

6. Business Impact

The proposed solution enables NexGen Logistics to:

- Identify high-risk deliveries in advance

- Reduce delivery delays through proactive interventions

- Improve customer satisfaction and retention

- Optimize operational costs by prioritizing resources

- Move toward a predictive, AI-driven logistics model

Estimated impact:

- **15–20% reduction in delivery delays**

- Better customer experience scores

7. Future Enhancements

- Integration of real-time GPS and IoT data

- Advanced ML models (Random Forest, Gradient Boosting)

- Time-series forecasting for delivery duration

- Automated route optimization and vehicle assignment

- Sustainability tracking using CO_2 emissions data

8. Conclusion

This project demonstrates how data analytics and machine learning can be used to transform logistics operations. By combining predictive modeling with an intuitive dashboard, the solution bridges the gap between AI insights and real business action.