Submitted By: Anushka Das (CSE AI ML)

Contact details:

Email: anushkad717@gmail.com

Contact no. 8403942331

LinkedIn: https://www.linkedin.com/in/anushka-das-b1585625b/

Innovation Brief – Intelligent Delivery Delay Prediction System

Keywords:

Machine Learning, Gradient Boosting Classifier, Streamlit Dashboard, Predictive Analytics, Data Preprocessing, Feature Engineering, Delay Risk Prediction, Logistics Optimization, Visualization, Recommendation Engine, Vehicle Selection, Customer Feedback Analysis, Cost Performance Analysis.

Overview

This project presents an AI-powered logistics analytics and delay prediction platform that combines data-driven decision-making with an interactive Streamlit dashboard.
The system analyzes multi-source logistics data, predicts potential delivery delays, and provides actionable operational insights — empowering logistics teams to enhance efficiency, reliability, and customer satisfaction.

Innovation Highlights

1. End-to-End Predictive Pipeline

Unlike static dashboards, this project integrates a complete Machine Learning pipeline — from data preprocessing to model-driven insights.
It not only visualizes logistics performance but also anticipates future risks (e.g., delays, cost overruns) using predictive modeling.

2. Smart Delay Prediction Model

- Implemented Gradient Boosting Classifier (GBC) — chosen over Random Forest for its superior ability to handle imbalanced data and capture non-linear relationships between features.
- Gradient Boosting combines multiple weak learners into a strong ensemble, improving accuracy, interpretability, and robustness even on small or medium datasets.
- Outputs include:
    - Delay probability score
    - Risk level categorization (High / Moderate / Low)

- Dynamic recommendations for delivery optimization

Key Components & Their Purpose

Component	Functionality	Innovation / Impact
Data Preprocessing Module	Loads, cleans, merges, and engineers features from six logistics-related CSV files.	Automated data cleaning ensures consistency; engineered KPIs enable rapid monitoring of operational performance.
Feature Engineering	Creates new variables like “Delayed Flag,” “Weather Impact,” and “Cost per KM.”	Transforms raw operational data into business-ready intelligence.
Model Training (Gradient Boosting)	Predicts delivery delays based on route distance, traffic, weather, and order priority.	Gradient Boosting offers high interpretability and strong performance on structured data.
Recommendation Engine	Suggests next actions based on delay probability (e.g., assign reliable carrier, upgrade to express).	Converts model output into decision support, making predictions actionable.
Vehicle Optimization Module	Recommends vehicle type based on distance and priority.	Adds operational intelligence, aligning logistics resources efficiently.
Visualization Module	Provides KPI cards, carrier comparison charts, correlation heatmaps, and word clouds.	Enables data storytelling — simplifies complex analytics into visual insights.
Streamlit Front-End (app.py)	User-friendly dashboard integrating ML, visual analytics, and prediction in real time.	Offers a complete business application experience, not just an ML demo.

Why Gradient Boosting?

- Performs better on smaller datasets with high feature importance variance.

- Handles non-linear dependencies among logistics factors (e.g., traffic × distance).
- More resilient to noise and class imbalance compared to traditional classifiers.
- Provides feature importance scores that directly inform operational strategy (e.g., focusing on weather-sensitive routes).

Real-World Relevance

In real-world logistics operations, delay prediction plays a crucial role in:

- Reducing customer churn due to missed deadlines
- Optimizing resource allocation (carriers, routes, vehicles)
- Anticipating weather or traffic-based disruptions
- Minimizing operational costs via proactive interventions

This prototype demonstrates the proof-of-concept for an AI-based delay prediction system — which, when trained on large-scale enterprise data, can achieve high accuracy (>85%) and integrate seamlessly with supply chain systems.

Future Enhancements

- Integration with real-time IoT or GPS data for live tracking.
- Model retraining on millions of delivery logs for higher accuracy.
- Use of Explainable AI (SHAP/LIME) to justify predictions to operators.
- Cloud deployment for real-time API-based access to predictions.