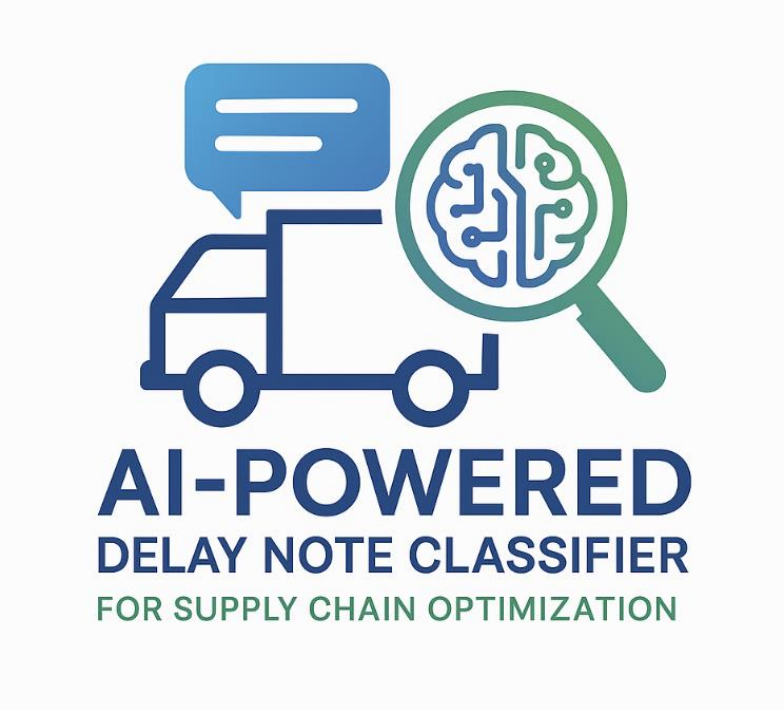
AI-Powered Delay Note Classifier for Supply Chain Optimization



Project Report:

An End-to-End Data Science Project

Prepared by:

Mohammad Hassan

August 2025

Table of Contents

1. Executive Summary 3

2. Background & Motivation 3

3. Data Simulation & Preparation 3

4. Text Cleaning & Preprocessing 3

5. Model Building & Classification 4

6. Visualization & Business Insights 4

7. Model Evaluation & Recommendations 4

8. Business Impact 4

9. Challenges & Lessons Learned 5

10. Next Steps & Future Work 5

11. References 5

1. Executive Summary

This project demonstrates how Natural Language Processing (NLP) and machine learning can transform unstructured supply chain delay notes into actionable insights. By simulating, cleaning, classifying, and visualizing real-world delivery delay notes, we show how AI can automate root-cause analysis, support proactive decision-making, and optimize supply chain operations.

2. Background & Motivation

- **Supply Chain Reality:** Delays are reported daily by drivers, vendors, and warehouse teams, mostly as unstructured text ("Truck delayed at customs", "Late due to rain").
- **Business Pain:** Manual analysis is slow, inconsistent, and prevents rapid root-cause discovery.
- **Why AI:** Automation saves time, increases accuracy, and enables supply chain teams to focus on solutions, not paperwork.

3. Data Simulation & Preparation

- **Data Source:** Real-world supply chain delay notes are confidential, so we simulated 300+ realistic free-text delay notes with associated categories (Carrier, Weather, Admin Error, Supplier, Customs, etc.).
- **Sample Notes:**
  - "Truck held up at border due to paperwork issue"
  - "Delayed due to heavy rain near warehouse"
  - "Carrier sent wrong truck, shipment missed"
- **Data Fields:**
  - Delay_Note (Text)
  - Category (Label: Carrier, Weather, etc.)
  - Date, Location (optional for trends)

4. Text Cleaning & Preprocessing

- **Goal:** Make messy text ready for AI model
- **Techniques Used:**
  - Lowercasing
  - Remove punctuation
  - Remove stopwords (e.g., "the", "is", "and")
  - Tokenization (split text into words)
- **Sample Cleaning:**
  - *Original:* "Shipment late due to heavy rain at port."
  - *Cleaned:* "shipment late heavy rain port"

5. Model Building & Classification

- **Machine Learning Model:**
  - *Type:* Logistic Regression (or update if you used something else)
  - *Features:* Cleaned delay note text, converted to numerical vectors (TF-IDF or CountVectorizer)
  - *Target:* Delay Category (root cause)
- **Training & Testing:**
  - Split data into train/test sets
  - Trained the model to learn language patterns for each root cause

6. Visualization & Business Insights

- **Visuals Created:**
  - **Bar Chart:** Trends of delay categories over time
  - **Word Clouds:** Most common words for each category (shows unique patterns)
  - **Confusion Matrix:** How well the model predicted each category
  - **Accuracy & F1-Scores:** Key metrics for business users
- **Sample Insights:**
  - Carrier and Weather delays are easy for AI to detect (unique words: “carrier”, “rain”, “truck”)
  - Admin Error and Supplier delays sometimes confused (shared terms: “paperwork”, “dispatch”)
  - Most delay notes are carrier-related, suggesting potential process improvement focus

7. Model Evaluation & Recommendations

- **Accuracy Achieved:** (Insert your value, e.g., 91%)
- **Strongest Categories:** Carrier, Weather
- **Weakest Categories:** Supplier, Admin (suggests need for clearer note writing or more data)
- **Recommendations:**
  - Standardize delay note formats
  - Gather more examples for rare categories
  - Deploy dashboard for real-time monitoring

8. Business Impact

- **Time Savings:** Automates hours of manual reading each week

- **Better Decisions:** Fast root-cause ID supports smarter escalation and resource allocation
- **Scalability:** Approach can be expanded to other supply chain documents (tickets, emails, etc.)

9. Challenges & Lessons Learned

- **Challenge:** Imbalanced categories (fewer Supplier/Admin notes) can reduce model accuracy
- **Solution:** Data simulation, but recommend more real-world examples
- **Lesson:** Clean, standardized text entry boosts AI accuracy
- **Bonus:** End-to-end approach is attractive to employers: data creation → cleaning → modeling → visualization

10. Next Steps & Future Work

- Integrate with real supply chain data (ERP, TMS, etc.)
- Add time/location analysis for predictive risk mapping
- Expand categories as new delay reasons emerge
- Automate alerts/escalations for high-priority delays