Document Title: **Train Delay Management**

Intern Name: V Sai Jayesh

Internship Period: 28-07-2025 – 27-10-2025

Department/Team: AIML

Mentor Name: Prasad Krishna & Aditya Raj C

Date of Submission: 10-09-2025

Table of Content

f/1 FRACTION LABS

3 Questions as a Dev

1. How much historical data do we have (number of years/months)? Is it cleaned and labeled (actual vs predicted arrival times)?
2. What's the latency requirement? (e.g., do you want real-time predictions every 1–2 mins or batch-wise every hour?)
3. What external factors should be considered? (weather, maintenance schedules, accidents, congestion, festivals, etc.)

Problem Statement

Passengers rely on the AI delay prediction system to plan travel. However, current predictions are unreliable (false positives and false negatives). This reduces trust in the system.

Solutions

- Improve trust by enhancing prediction accuracy and transparency.
- Provide live updates with confidence scores (so passengers know how reliable the prediction is).
- Easy dashboards for operations team.

Expected Output

1. For Passengers:

- Live train status with reliability score (e.g., "Train expected delay: 10 min, 85% confidence").

2. For Operations Team:

- Dashboard with delay trends, bottleneck stations, reasons for delays.

3. For Railways:

- Reduction in complaints, improved trust, data-driven operations.

Process

1. **Data Collection**: Gather historical train logs and live GPS data from multiple sources.
2. **Data Transformation & Loading**: Standardize formats and store in a unified database.

3. **Relationship Mapping**: Link datasets (e.g., Train IDs, GPS location).
4. **Model Selection:** Choose suitable AI/ML algorithms for regression (delay time and accuracy score).
5. **Model Training:** Train the model on historical data.
6. **Evaluation**: Measure accuracy, precision, recall, and error rates.
7. **Deployment**: Integrate into production with automated notifications (dashboard).
8. Continuous retraining with new data.

Flowchart