Railway Accident Analytics
A Data Driven AI Approach

Yashwanth Sai Kasarabada
220103012
CSE (AI & DS)

Supervised by
Dr. Sanasam Chanu Inunganbi

IIIT MANIPUR

25 April 2025

Contents

Introduction, Objective & Contribution

Introduction:

- Railway accidents pose significant safety and operational challenges.
- Data-driven analytics and AI can help predict, analyze, and reduce accident risks.

Objective:

- To develop a web-based system for analyzing and predicting railway accidents using advanced analytics and AI.

Contribution:

- Integrated multi-section platform for insights, prediction, and reporting.
- Combines statistical analysis, machine learning, and AI assistant features.
- Enables interactive exploration and decision support for railway safety.

Background / Existing Work

What others have done:

- Existing projects use statistical methods and basic dashboards for accident analysis.
- Some systems employ machine learning for accident prediction.
- Examples: Indian Railways accident dashboards, academic studies using regression or classification models.

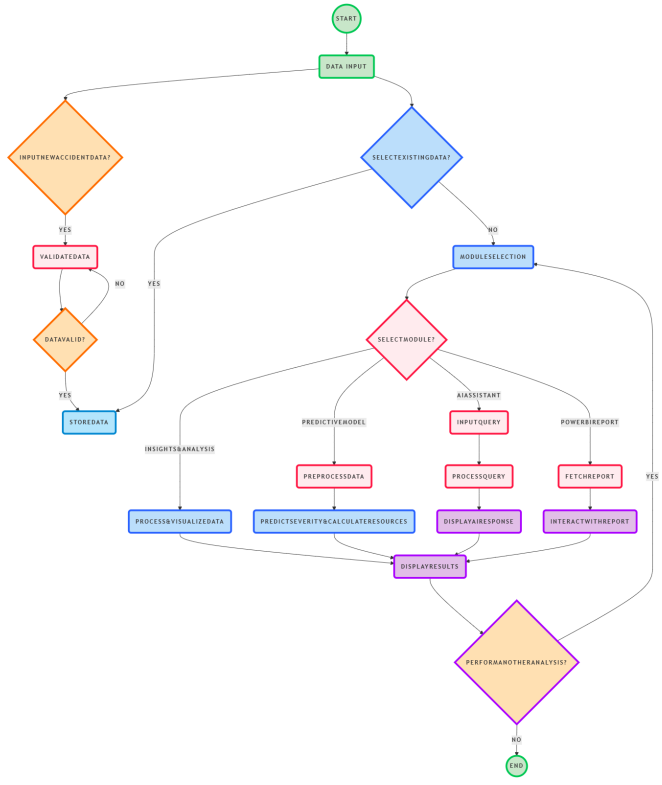
Problems with existing work:

- Limited integration of AI chatbot for queries.
- Lack of interactive, unified platforms combining analysis, prediction, and reporting.
- Insufficient focus on actionable insights and user-friendly interfaces.

Why is your project needed?

- Bridges the gap between data analysis, prediction, and real-time decision support.
- Empowers users with interactive tools and AI-driven recommendations.

System Design



Steps:

- **Input:** Accident datasets (CSV uploads, historical data)
- **Process:** Data cleaning, feature extraction, ML prediction, AI query handling
- **Output:** Visualizations, predictions, AI-driven insights, reports

Technology Stack

Layer	**Technology/Tool**
Frontend	Streamlit (Python-based UI framework)
Visualization	Power BI Embedded, Matplotlib, Seaborn
Data Processing	Python, Pandas, NumPy
Machine Learning	scikit-learn (Random Forest), SciPy
AI Integration	Groq API, LLaMA-3 model
Deployment	Streamlit Cloud (for web hosting)
Security	Streamlit session handling, secure API token management
Data Storage	In-memory upload (CSV files via UI); no external SQL database used

Working

- User uploads or selects datasets through a web interface.
- System processes data, applies ML models for prediction.
- AI assistant answers natural language queries about accidents.
- Results and reports are displayed interactively using Power BI and visualizations.

Result & Analysis

Testing:

- Evaluated on Indian railway accident datasets (1902–2024).
- Used Random Forest Regression for severity prediction.
- Visualizations generated with Matplotlib, Seaborn, and Power BI.

Metric	**Value**
Mean Absolute Error (MAE)	11.39
R^2 Score	0.88

Table: Performance of the Severity Prediction Model

The Mean Absolute Error (MAE) of 11.39 indicates that, on average, the model's predictions deviate from the actual values by about 11.39 units. The R^2 score of 0.88 suggests that the model explains 88% of the variance in the target variable, reflecting strong predictive performance.

Conclusion

Summary:

- Developed an integrated, data-driven AI platform for railway accident analytics.
- Demonstrated effective accident prediction and actionable insights.
- Provided user-friendly, interactive tools for stakeholders.

Future Work:

- Incorporate real-time data streams and alerts.
- Enhance AI assistant's capabilities and language support.
- Integrate with official railway databases and deploy at scale.