

2025-08-25 Meeting Notes

Date

25 Aug 2025

Participants

- RAIL PG-2 project team
- Murtaza (Proxy Client)

Discussion topics

1. Current Progress Report

- Insight Factory Platform: Platform and partial data have been accessed, but only two tables are currently available; the full dataset and schema are still pending.
- GitHub Collaboration: Backlog has been created and divided into five sub-teams.

2. Research Progress

Feature Engineering: At least 5 relevant methods identified:

Temporal data: rolling window statistics, lag features, rate of change, left-right/front-rear differences

Imbalanced data: resampling (undersampling/oversampling), peak/spike counters, combining risk factors

Feature Selection: At least 5 relevant methods identified; Group Lasso has been chosen as the most suitable method at this stage for implementation.

Machine Learning Techniques: At least 5 algorithms identified (LR, ANN, SVM, RF, Transformer); initial coding experiments to follow.

3. Next Steps

- Data Ingestion: Set up and configure ingestion pipeline in Insight Factory.ai, execute the ingestion process with validation.
- Exploratory Data Analysis (EDA): Analyze temporal patterns, class imbalance, and feature relationships; summarize key findings.
- Each sub-team is expected to complete initial training and comparison of at least three models before the next Sprint meeting.

4. **QA**

- Some dataset column descriptions are unclear; the team needs to research independently first, and if not resolved, post clarifications in the channel.
- SQL warehouse access permissions and timeline require further clarification.

5. **Future Plan**

- All teams continue on their assigned research and experimental tasks.
- Complete initial training and comparison of at least three models before the next Sprint meeting, and identify the best-performing approaches.
- During EDA, further investigate and interpret the meaning of features.
- Each user story should include acceptance criteria, and related tasks must be completed within the Sprint. Any expected delays should be communicated in advance.