

2025-08-25 Meeting Agenda

Date

25 Aug 2025

Participants

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

Goals

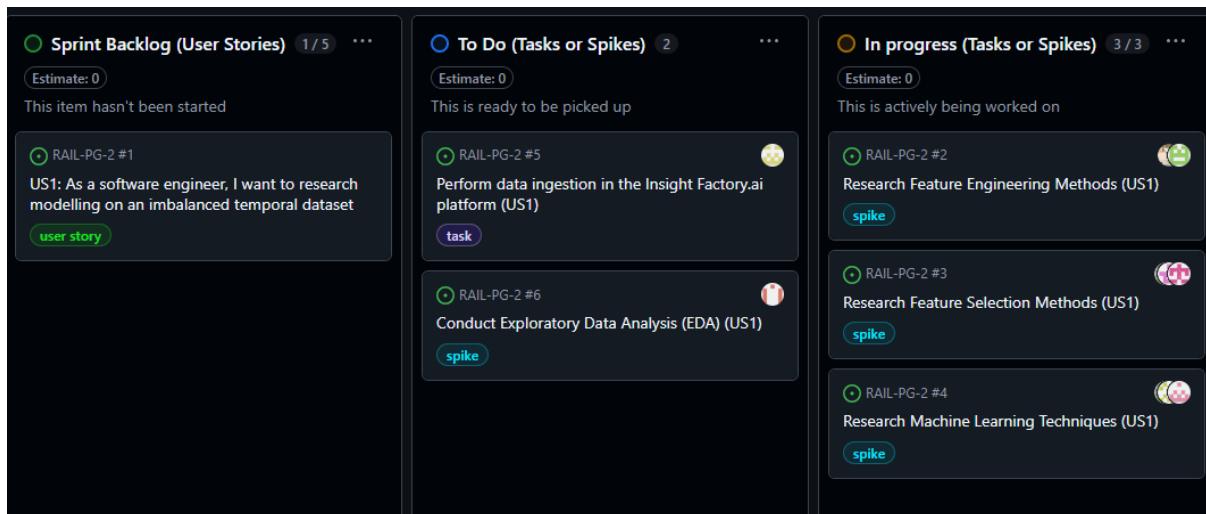
- Report current process
- teamworking
- Future plan
- QA

Discussion topics

1. Report current process:

Insight Factory Platform
Data accessed
Tables checked
Schema reviewed

2. Github teamworking



The screenshot shows a kanban board with three columns: Sprint Backlog (User Stories), To Do (Tasks or Spikes), and In progress (Tasks or Spikes).

- Sprint Backlog (User Stories):** 1 / 5 items. One item is listed: "RAIL-PG-2 #1 US1: As a software engineer, I want to research modelling on an imbalanced temporal dataset". Status: Estimate: 0, Not started.
- To Do (Tasks or Spikes):** 2 items. Two items are listed: "RAIL-PG-2 #5 Perform data ingestion in the Insight Factory.ai platform (US1)" and "RAIL-PG-2 #6 Conduct Exploratory Data Analysis (EDA) (US1)". Status: Estimate: 0, Ready to be picked up.
- In progress (Tasks or Spikes):** 3 / 3 items. Three items are listed: "RAIL-PG-2 #2 Research Feature Engineering Methods (US1)", "RAIL-PG-2 #3 Research Feature Selection Methods (US1)", and "RAIL-PG-2 #4 Research Machine Learning Techniques (US1)". Status: Estimate: 0, Actively being worked on.

3. Research finding:

In progress :

1 Research Feature Engineering Methods:

Find at least 5 relevant feature engineering techniques for temporal and imbalanced datasets.

temporal data:

1. Rolling window statistics
2. Lag features
3. Rate of change
4. Computing left-right/front-rear differences

imbalanced data :

1. Resampling (Undersampling/Oversampling)
2. Peak/spike counters in a window
3. Combining risk factors

2 Research Feature Selection Methods:

Find at least 5 relevant feature selection techniques for temporal and imbalanced datasets.

We plan to adopt group lasso as the most suitable method at the current stage and proceed with code development.

3 Research Machine Learning Techniques:

Find at least 5 machine learning algorithms or approaches suitable for imbalanced temporal datasets.

LR, ANN, SVM, RF, Transformer

Next Steps / To Do:

1 Perform data ingestion in the Insight Factory.ai platform

Set up and configure data ingestion pipeline in Insight Factory.ai.

Execute ingestion process with validation.

2 Conduct Exploratory Data Analysis (EDA)

Analyze temporal patterns, class imbalance, and feature relationships.

Summarize key findings from EDA.

website address: [**Backlog · RAIL PG-2**](#)

4. QA

1. Clarification on future SQL warehouse access – Will the project team have permissions, and what would be the expected timeline?