

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 displays a Jira Sprint Backlog for the 'RAIL-PG-2' project. It is organized into three columns: 'Sprint Backlog (User Stories)', 'To Do (Tasks or Spikes)', and 'In progress (Tasks or Spikes)'. Each column has a header with a status icon, a count, and an 'Estimate: 0' label. The 'Sprint Backlog' column contains one user story (RAIL-PG-2 #1) with the description 'US1: As a software engineer, I want to research modelling on an imbalanced temporal dataset'. The 'To Do' column contains two tasks (RAIL-PG-2 #5 and #6) with descriptions 'Perform data ingestion in the Insight Factory.ai platform (US1)' and 'Conduct Exploratory Data Analysis (EDA) (US1)'. The 'In progress' column contains three spikes (RAIL-PG-2 #2, #3, and #4) with descriptions 'Research Feature Engineering Methods (US1)', 'Research Feature Selection Methods (US1)', and 'Research Machine Learning Techniques (US1)'. Each item is represented by a card with a status icon, a title, a description, and a label indicating its type (user story, task, or spike).

Sprint Backlog (User Stories)	To Do (Tasks or Spikes)	In progress (Tasks or Spikes)
<p>RAIL-PG-2 #1</p> <p>US1: As a software engineer, I want to research modelling on an imbalanced temporal dataset</p> <p>user story</p>	<p>RAIL-PG-2 #5</p> <p>Perform data ingestion in the Insight Factory.ai platform (US1)</p> <p>task</p>	<p>RAIL-PG-2 #2</p> <p>Research Feature Engineering Methods (US1)</p> <p>spike</p>
	<p>RAIL-PG-2 #6</p> <p>Conduct Exploratory Data Analysis (EDA) (US1)</p> <p>spike</p>	<p>RAIL-PG-2 #3</p> <p>Research Feature Selection Methods (US1)</p> <p>spike</p>
		<p>RAIL-PG-2 #4</p> <p>Research Machine Learning Techniques (US1)</p> <p>spike</p>

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?