2025-09-08 Meeting Agenda

📅 Date

08 Sep 2025

👥 Participants

- RAIL PG-2 project team

Tao Xu	a1937511
Sheng Wang	a1903948
Jinchao Yuan	a1936476
Zilun Ma	a1915860
Di Zhu	a1919727
Xin Wei	a1912958
Yifan Gu	a1909803
Tianhua Zhang	a1915934
Zihan Luo	a1916700

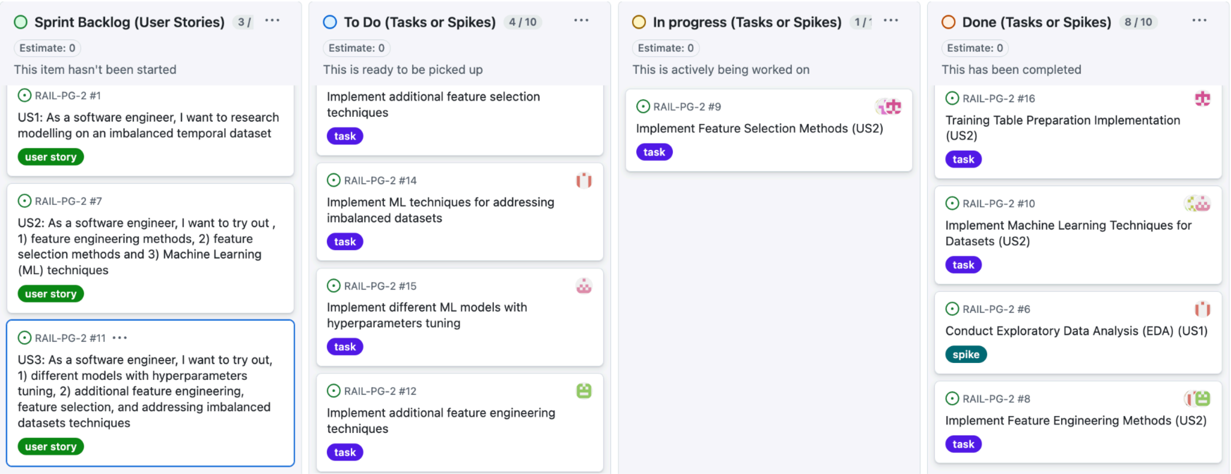
- Murtaza (Proxy Client)

🥅 Goals

- Github backlog overview
- Report progress
- Submission results
- Blockers
- Next steps
- QA

🗣️ Discussion topics

1. GitHub backlog overview



Sprint Backlog (User Stories) 3 / ···
Estimate: 0
This item hasn't been started

RAIL-PG-2 #1
US1: As a software engineer, I want to research modelling on an imbalanced temporal dataset
user story

RAIL-PG-2 #7
US2: As a software engineer, I want to try out , 1) feature engineering methods, 2) feature selection methods and 3) Machine Learning (ML) techniques
user story

RAIL-PG-2 #11 ···
US3: As a software engineer, I want to try out, 1) different models with hyperparameters tuning, 2) additional feature engineering, feature selection, and addressing imbalanced datasets techniques
user story

To Do (Tasks or Spikes) 4 / 10 ···
Estimate: 0
This is ready to be picked up

Implement additional feature selection techniques
task

RAIL-PG-2 #14
Implement ML techniques for addressing imbalanced datasets
task

RAIL-PG-2 #15
Implement different ML models with hyperparameters tuning
task

RAIL-PG-2 #12
Implement additional feature engineering techniques
task

In progress (Tasks or Spikes) 1 / ···
Estimate: 0
This is actively being worked on

RAIL-PG-2 #9
Implement Feature Selection Methods (US2)
task

Done (Tasks or Spikes) 8 / 10 ···
Estimate: 0
This has been completed

RAIL-PG-2 #16
Training Table Preparation Implementation (US2)
task

RAIL-PG-2 #10
Implement Machine Learning Techniques for Datasets (US2)
task

RAIL-PG-2 #6
Conduct Exploratory Data Analysis (EDA) (US1)
spike

RAIL-PG-2 #8
Implement Feature Engineering Methods (US2)
task

2. Report progress

2.1 EDA

- Apply visualization techniques (e.g heapmap, histogram, scatter plot, and time series plot) to analyze data

2.2 Feature selection

- Joining and integrating trainingcontext, wagondata, and tonnagedata table into a unified total_training_table
- Preprocessing the total training table
- Implement a basic version of the Transformer model with Group Lasso and REF using LightGBM, conducting comparative testing to evaluate their performance.

2.3 Feature engineering

- Sensor features and threshold design
- Baseline features, fourier transform, and trend features

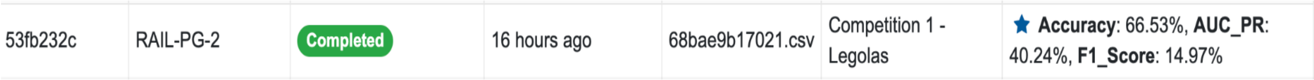
2.4 ML model training

- Training SVM, DNN, and transformer models and submit

2.5 Build production line

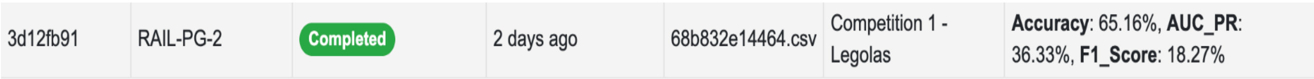
3. Submission results

- **Transformer**



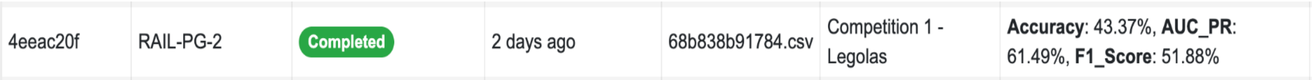
53fb232c	RAIL-PG-2	Completed	16 hours ago	68bae9b17021.csv	Competition 1 - Legolas	★ **Accuracy**: 66.53%, **AUC_PR**: 40.24%, **F1_Score**: 14.97%

- **DNN**



3d12fb91	RAIL-PG-2	Completed	2 days ago	68b832e14464.csv	Competition 1 - Legolas	**Accuracy**: 65.16%, **AUC_PR**: 36.33%, **F1_Score**: 18.27%

- **SVM**



4eeac20f	RAIL-PG-2	Completed	2 days ago	68b838b91784.csv	Competition 1 - Legolas	**Accuracy**: 43.37%, **AUC_PR**: 61.49%, **F1_Score**: 51.88%

4. Blockers

- **Data availability**: At the beginning of the sprint 2, no datasets were provided. Therefore, we started to implement this project a little late.
- **Limited training time:** Team members need time to select features, preprocessing data, and create training tables for training models. Therefore, time for building and training model was quite limited.
- **Unstable platform runtime**: After submitting notebooks, the runtime duration on the IF platform was unstable. For example, sometimes the execution took half an hour, while other times it finished within 4 minutes. The instability increases uncertainty for model training and inference creation.

5. Next steps

Goal: In the sprint 3, we will improve models overall performance and achieve an F1 score over 55%.

- Implement at least 2 methods of feature engineering, feature selection, and handling imbalance datasets.
- Tuning hyperparameters of 3 models, feature selection, feature engineering, and imbalanced dataset handling technique.

6. QA

Q: For addressing imbalanced datasets, do we need to create a balanced table by preprocessing techniques, or can we directly apply handling imbalanced datasets techniques when training model (e.g. class weights)?