THE UNIVERSITY of ADELAIDE

Software Engineering & Project (COMP SCI 7015)

Snapshot Week 03 of Group RAIL PG-2

Rail Break Prediction ML

Tao Xu a1937511

Sheng Wang a1903948

Jinchao Yuan a1936476

Zi Lun Ma a1915860

Di Zhu a1919727

Xin Wei a1912958

Yifan Gu a1909803

Tianhua Zhang a1915934

Zihan Luo a1916700

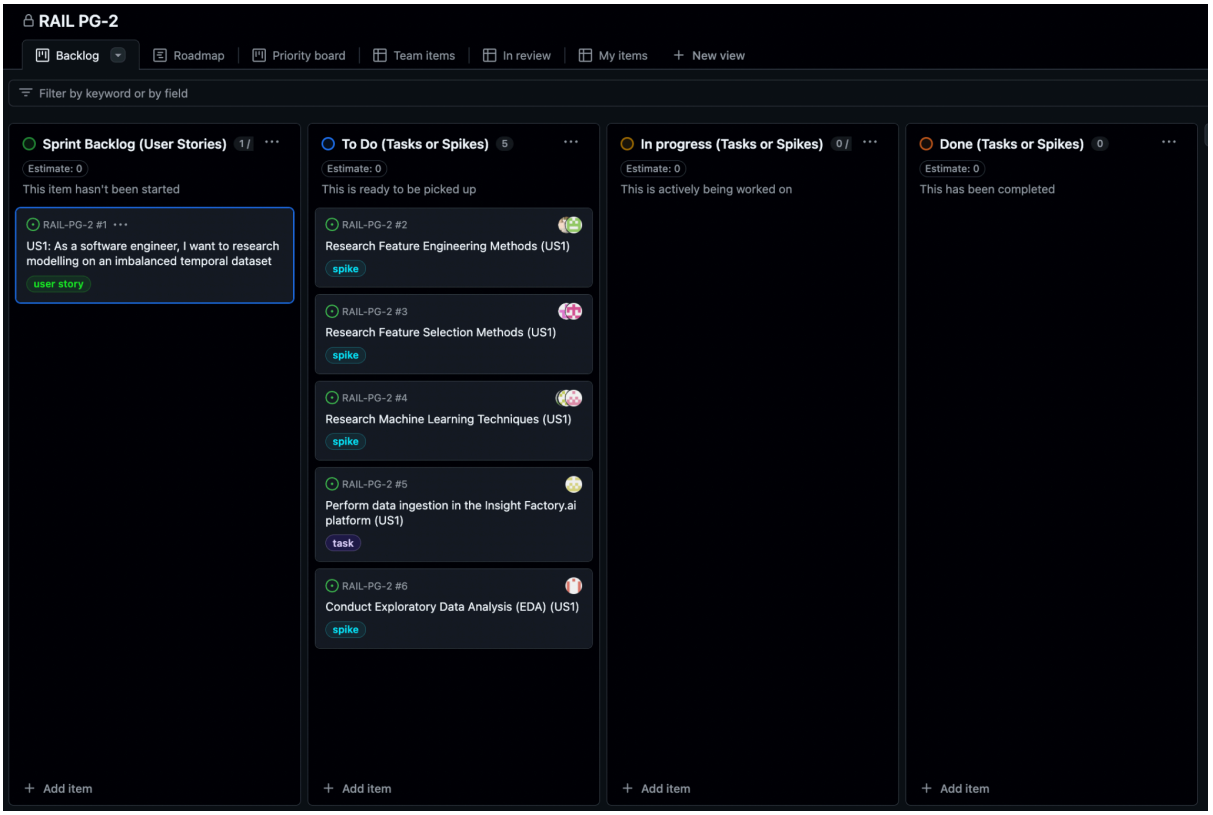
Supervisor : Murtaza Bootwala

1. Product Backlog and Task Board

1.1. The product backlog

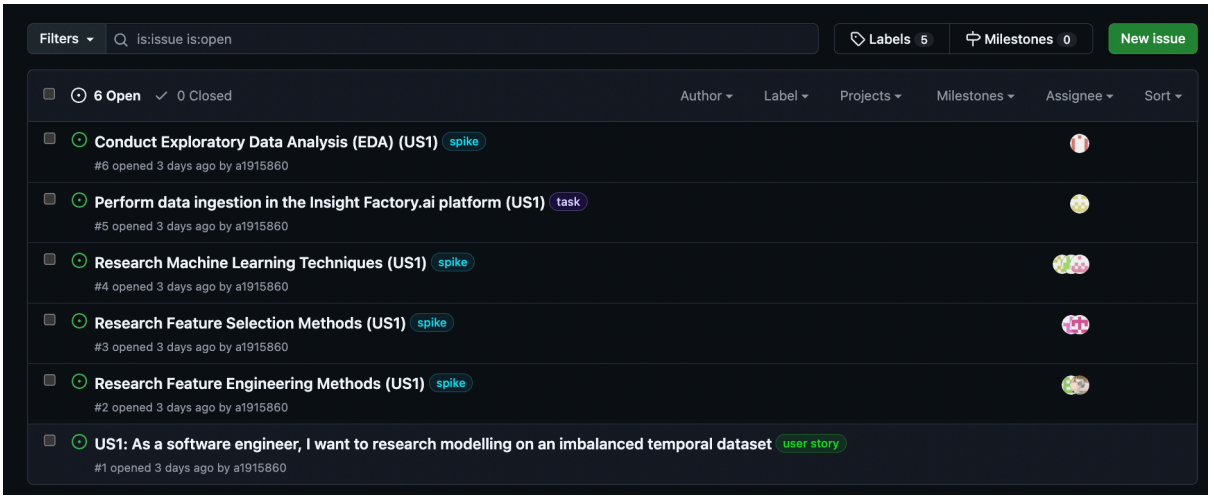
ID	Priority	User Story/Task/Spike	Description
PB1	1	Feature Engineering	Create new features based on domain knowledge and data patterns to improve model performance.
PB2	1	Feature Selection	Identify and retain the most relevant features to reduce noise and improve efficiency.
PB3	1	Model Research & Selection	Investigate suitable machine learning techniques for imbalance temporal datasets
PB4	2	Data Ingestion into InsightFactory.ai	Import the provided real-world production dataset into the InsightFactory platform.
PB5	2	Data Cleaning & Preprocessing	Handle missing values, outliers, and inconsistencies in the dataset.
PB6	2	Exploratory Data Analysis (EDA)	Analyze data distributions, trends, and anomalies to understand key characteristics.
PB7	3	Model Training	Train predictive models using the processed and engineered dataset.
PB8	3	Model Evaluation	Assess models using Accuracy, F1 Score, and AUCPR metrics.
PB9	3	Benchmark Comparison	Compare the model's performance against the InsightFactory bench mark model for potential bonus marks.
PB10	4	Model Optimization & FInalization	Fine-tune model parameters, optimize features, and prepare the final deliverable.

1.2. The task board



2. Sprint Backlog and User Stories

2.1. The Sprint backlog



2.2. User stories

Research techniques for modelling an imbalanced temporal dataset, including feature engineering, feature selection, and suitable machine learning methods. Ingest the data into the Insight Factory.ai platform. Conduct exploratory data analysis (EDA) to develop a plan to approach the project.

Related tasks:

1. Research Feature Engineering Methods
2. Research Feature Selection Methods
3. Research Machine Learning Techniques
4. Perform data ingestion in the Insight Factory.ai platform
5. Conduct Exploratory Data Analysis (EDA)

3. Definition of Done

A backlog item is considered “Done” when:

Spike:

- The research is complete, including findings, identified risks and challenges, and any recommendations.
- All relevant documentation is shared with the team.

Task*:

- Code (including database scripts) is implemented according to acceptance criteria.
- Code has been peer-reviewed and approved.
- All relevant tests (unit, integration) have been passed.
- Documentation (code comments, user guides) is updated.
- No major open defects remain.

* The current sprint is research-based, so the DOD for tasks might not be applicable. However, it's better to decide the expectation for tasks earlier.

4. Summary of Changes:

Since the start of the project, our team has analysed the provided user stories. Based on these, we have defined the key tasks for the first sprint.

Some of the team members will focus on researching feature engineering methods, feature selection methods, and machine learning techniques. After these are done, a report on findings and recommendations for implementation will be submitted.

The other members will perform data ingestion into the InsightFactory.ai platform, and then conduct exploratory data analysis (EDA) to understand the dataset's structure, distributions, and potential data quality issues.

In our opinion, this work will provide a solid foundation for developing the modelling approach in subsequent sprints.