THE UNIVERSITY of ADELAIDE

Software Engineering & Project (COMP SCI 7015)

# Snapshot Week 12 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

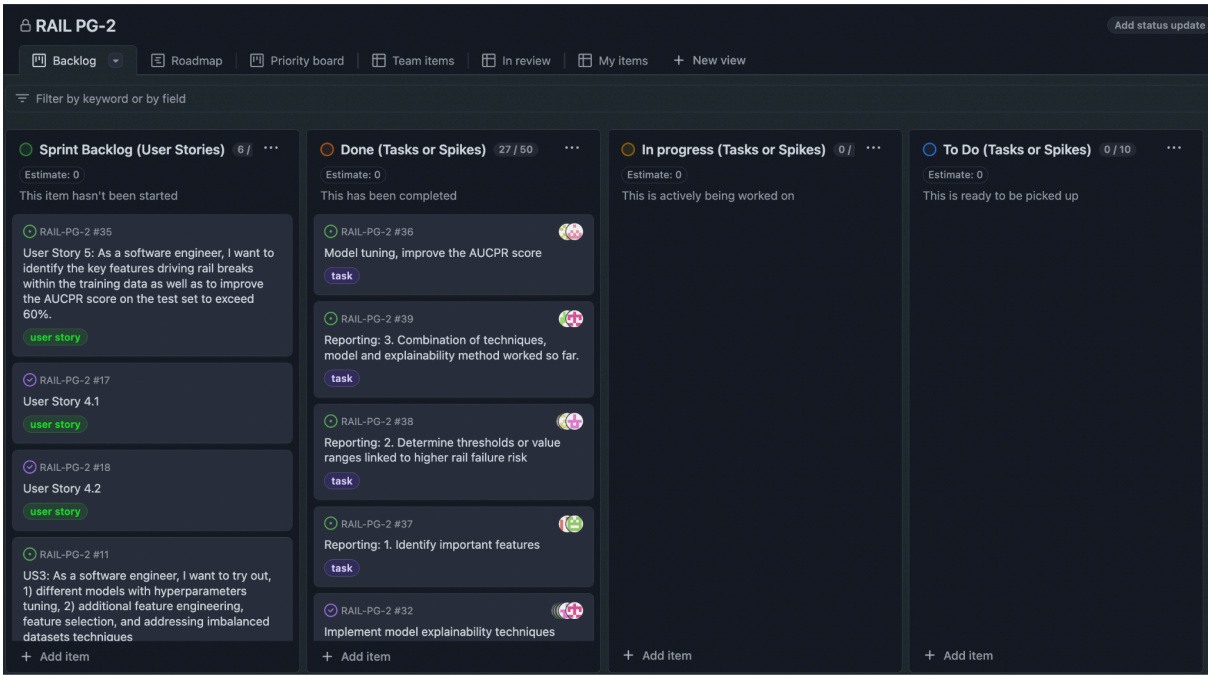
## 1.1. The product backlog

| 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. |

| | | | |
|---|---|---|---|
| PB11 | 1 | Implement Feature Engineering Methods | exploring and testing different feature transformation and construction approaches to enhance the predictive power of the dataset. |
| PB12 | 1 | Implement Feature Selection Methods | applying statistical and algorithmic techniques to identify the most relevant features and reduce dimensionality for improved model efficiency. |
| PB13 | 2 | Implement Machine Learning Techniques for Datasets | investigating specialized algorithms and resampling strategies to handle class imbalance effectively. |
| PB14 | 2 | Training Table Preparation Implementation | Implement Training table preparation scripts, these scripts provide the fundamental data integration for the overall project pipeline. |
| PB15 | 1 | Implement additional feature engineering techniques | Try out at least 2 more feature engineering techniques. |
| PB16 | 1 | Implement additional feature selection techniques | Try out at least 2 more feature selection techniques. |
| PB17 | 1 | Implement ML techniques for addressing imbalanced datasets | Try out at least 2 more techniques for handling imbalanced datasets. |
| PB18 | 2 | Implement different ML models with hyperparameters tuning | Tuning hyperparameter of<br>● at least 3 different ML models<br>● feature selection, feature engineering, and imbalanced dataset handling techniques, if they have hyperparameters to tune |
| PB1 | 1 | Research model | Identify and explore at least 5 model XAi techniques |

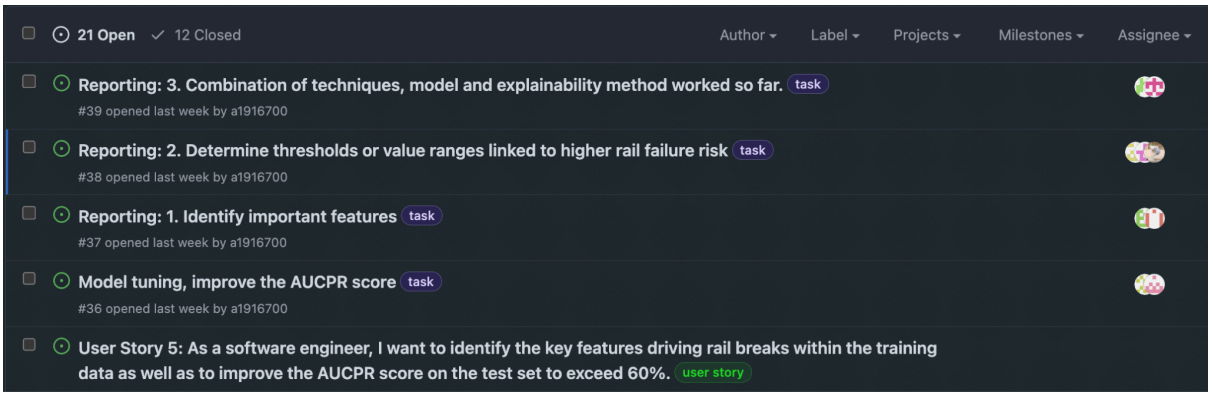
| | | | |
|---|---|---|---|
| 9 | | explainability techniques | relevant to the project. |
| PB20 | 2 | Implement model explainability techniques | 1. Implement and try out at least 3 model explainability techniques.<br>2. Continue:<br>● hyperparameter tuning<br>● experiments on different feature engineering methods, feature selection methods and techniques to approach an imbalanced dataset<br>3. Report what combination of techniques, model and explainability method worked so far<br>4. Achieve a minimum InsightFactory leaderboard score of 60% |
| PB21 | 1 | Model tunning to improve ACCPR score | Improve the AUCPR score on the test set to exceed 60% |
| PB22 | 2 | Outline the current findings | ● Identify the critical features that increase the risk of rail breaks.<br>● Determine the ranges or thresholds of these features that correspond to a higher likelihood of rail failure.<br>● Report what combination of techniques, model and explainability method worked so far. |

### 1.2. The task board




# 2. Sprint Backlog and User Stories

## 2.1. The Sprint backlog




## 2.2. User stories

User Story 5:

As a software engineer, I want to identify the key features driving rail breaks within the training data as well as to improve the AUCPR score on the test set to exceed 60%.

**Related tasks:**

1. Model tuning, improve the AUCPR score #36
2. Reporting: 1. Identify important features #37
3. Reporting: 2. Determine thresholds or value ranges linked to higher rail failure risk #38
4. Reporting: 3. Combination of techniques, model and explainability method worked so far #39

## 3. Definition of Done

A backlog item is considered “Done” when:

Task:

- Documentation (code comments, user guides) is updated.
- No major open defects remain.

## 4. Summary of Changes

In the final week of this project, we did not continue to do model tuning work. All the conclusions belonging to this sprint could be found in Snapshot_week11_RAIL PG-2, we sorted out the previous achievements, shared learning experiences within the team, discussed the content of the final report, and prepared the presentation slide to ensure the smooth completion of the final project deliverables.