THE UNIVERSITY of ADELAIDE

Software Engineering & Project (COMP SCI 7015)

Snapshot Week 09 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

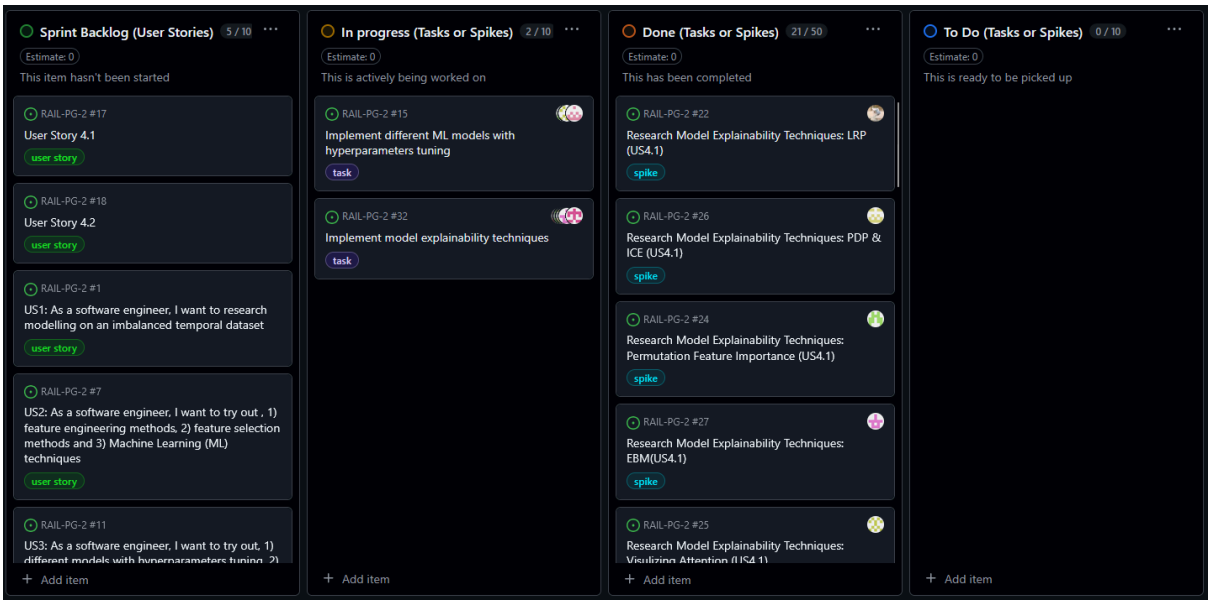
1. Product Backlog and Task Board

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.

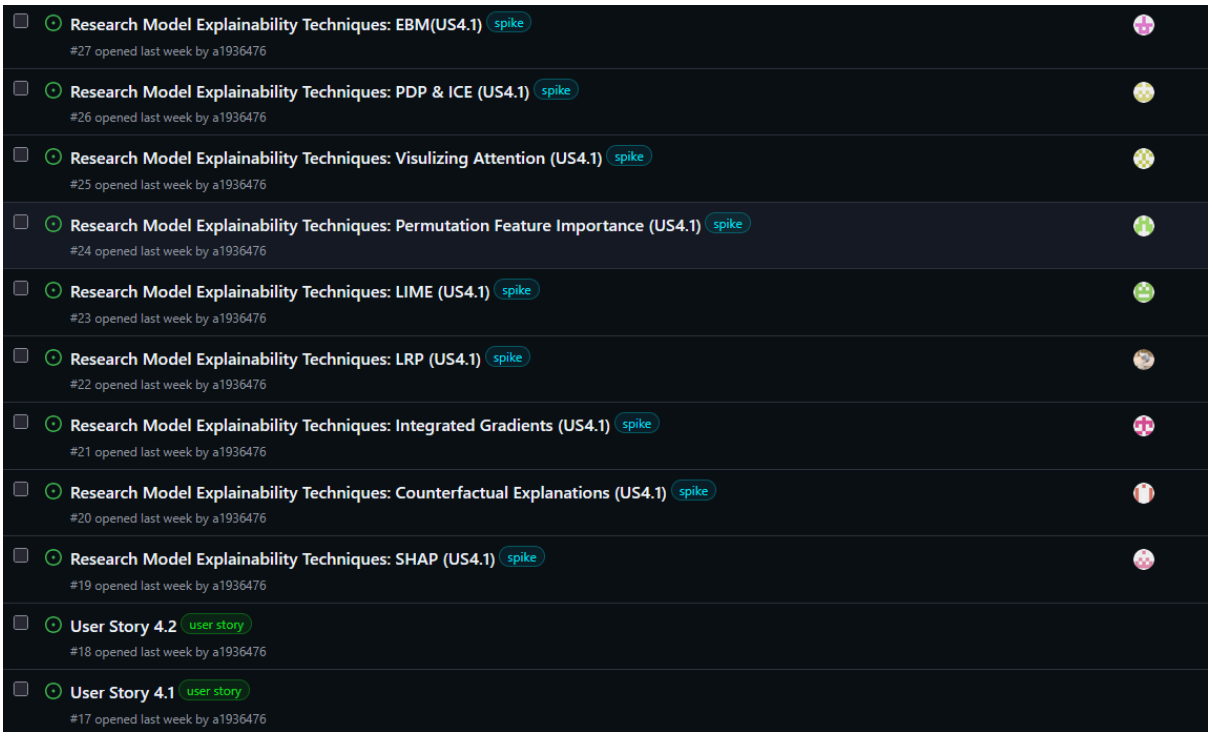
PB16	1	Implement additional feature selection techniques	Try out at least 2 more feature selection techniques.
PB17	1	Implement ML techniques for addressing imbalanced datasets	Try out at least 2 more techniques for handling imbalanced datasets.
PB18	2	Implement different ML models with hyperparameters tuning	Tuning hyperparameter of ● at least 3 different ML models ● feature selection, feature engineering, and imbalanced dataset handling techniques, if they have hyperparameters to tune
PB19	1	Research model explainability techniques	Identify and explore at least 5 model XAi techniques relevant to the project.

1.2. The task board



2. Sprint Backlog and User Stories

2.1. The Sprint backlog



2.2. User stories

User Story 4.1:

As a software engineer, I want to research model explainability techniques (Explainable AI:-XAi) to understand how they can be applied to the project.

Related tasks:

1. Research Model Explainability Techniques: EBM
2. Research Model Explainability Techniques: PDP & ICE
3. Research Model Explainability Techniques: Visualizing Attention
4. Research Model Explainability Techniques: Permutation Feature Importance
5. Research Model Explainability Techniques: LIME
6. Research Model Explainability Techniques: LRP
7. Research Model Explainability Techniques: Integrated Gradients
8. Research Model Explainability Techniques: Counterfactual Explanations (DICE)
9. Research Model Explainability Techniques: SHAP

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.

4. Summary of Changes

This week, each team member is assigned to research one XAI technique and share the result with the team. The research outcomes are added to the wiki page on GitHub.

After the internal meeting of the team, LIME, SHAP, and DICE are chosen to be applied to the machine learning models.

The features in the training dataset will also be refined based on their weight coming from the XAI result.