Software Engineering & Project (COMP SCI 7015)

Snapshot Week 11 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

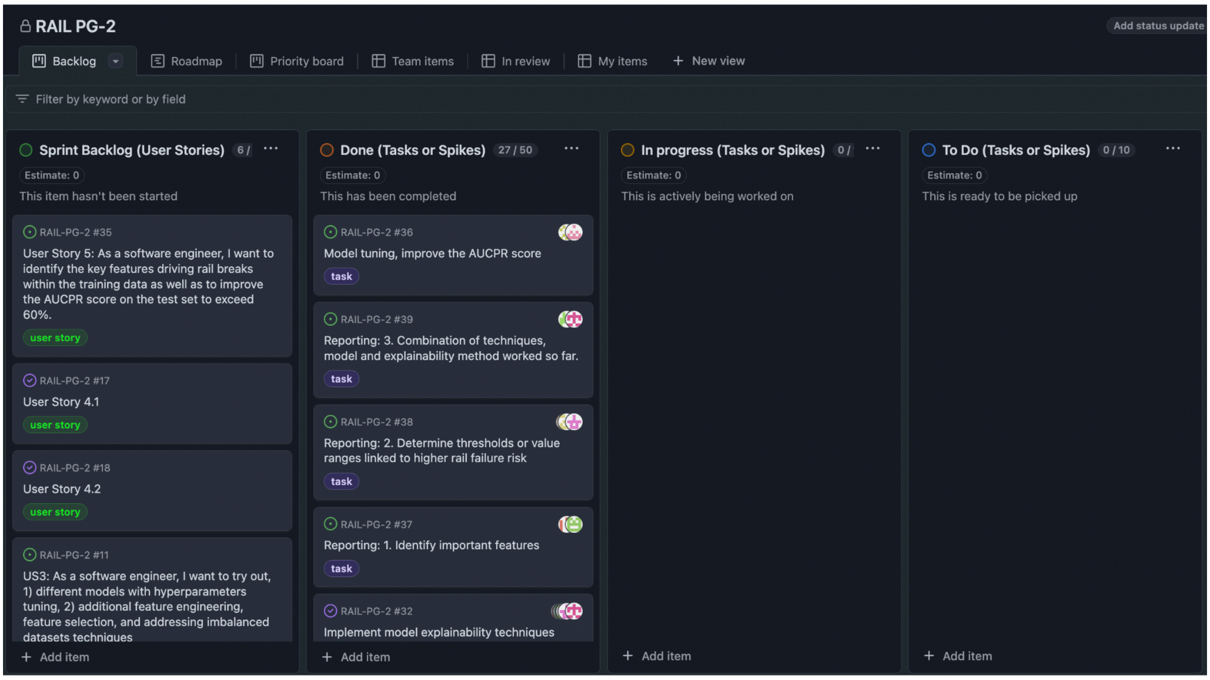
1.1. The product backlog

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.
PB20	2	Implement model explainability techniques	1. Implement and try out at least 3 model explainability techniques. 2. Continue: ● hyperparameter tuning ● experiments on different feature engineering methods, feature selection methods and techniques to approach an imbalanced dataset 3. Report what combination of techniques, model and

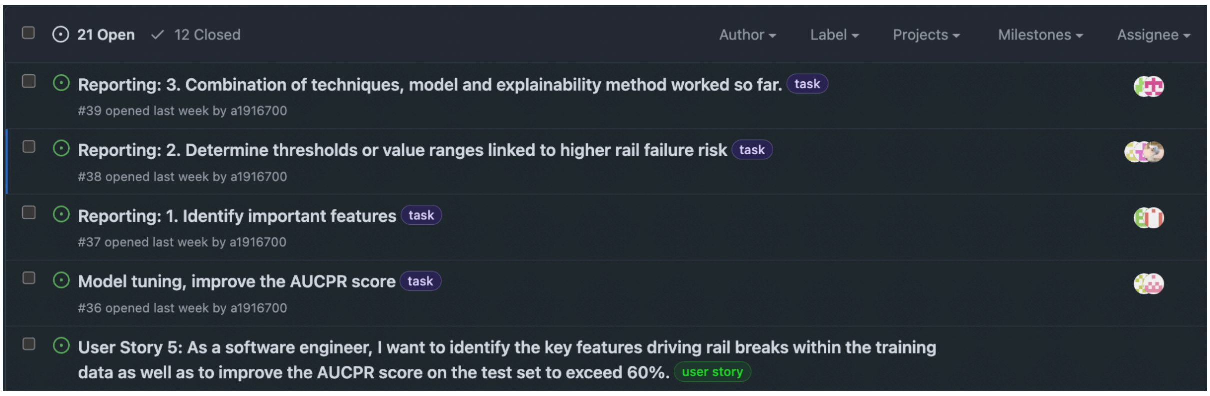
			explainability method worked so far 4. Achieve a minimum InsightFactory leaderboard score of 60%
PB21	1	Model tunning to improve ACCPR score	Improve the AUCPR score on the test set to exceed 60%
PB22	2	Outline the current findings	• Identify the critical features that increase the risk of rail breaks. • Determine the ranges or thresholds of these features that correspond to a higher likelihood of rail failure. • 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:

- Create different branches for different tasks
- Code (including database scripts) is implemented according to acceptance criteria.
- Code has been peer-reviewed and approved if a new table is created in the final schema.
- All relevant tests (unit, integration) have been passed.
- Documentation (code comments, user guides) is updated.
- No major open defects remain.

4. Summary of Changes

In this sprint, we continue to attempt model tuning and try to improve our score based on AUCPR. During this process, it is identified that different models have different important features. The following are the different Best hyperparameters for each model:

1. Model: DNN

Best hyperparameters:

- lr: 0.0017593895286675944
- batch_size: 16
- weight_decay: 0.0016917098161786815
- optimizer: adamw
- use_scheduler: True
- n_units_l1: 118
- n_units_l2: 56
- activation: relu
- dropout_rate1: 0.27981092243499955
- dropout_rate2: 0.3176818552611398
- use_batch_norm: False

c694f15d	RAIL-PG-2	Completed	7 minutes ago	68f6c9434372.csv	Competition 3 - The Defribillator	**Accuracy**: 58.43%, **AUC_PR**: 31.23%, **F1_Score**: 34.90%

2. Model: Transformer

Best hyperparameters:

- d_mode: 128
- nhead: 4
- num_layers: 2
- dropout: 0.4
- lr: 5e-5
- weight_decay: 1e-3
- epochs: 30

f54cd5cb	RAIL-PG-2	Completed	2 minutes ago	68f16f051000.csv	Competition 3 - The Defribillator	**Accuracy**: 57.68%, **AUC_PR**: 45.64%, **F1_Score**: 46.70%

3. Model: SVM

Best hyperparameters:

- C: 1.0
- class_weight: balanced
- dual: False
- penalty: l2
- loss: squared_hinge
- tol: 1e-3
- max_iter: 20000
- fit_intercept: True
- intercept_scaling: 1

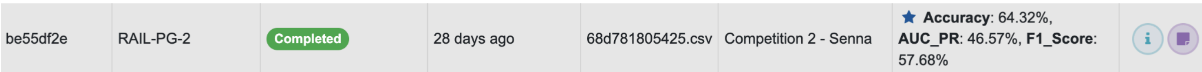
68f1ba0d3013.csv	Competition 3 - The Defribillator	**Accuracy**: 41.57%, **AUC_PR**: 56.55%, **F1_Score**: 47.83%

4. Model: Group Lasso Transformer

Best hyperparameters:

- d_model: 96
- nhead: 2
- num_layers: 2

- dim_feedforward: 256
- dropout: 0.15
- scheduler: LambdaLR
- lr: 2e-4
- optimizer: torch.optim.AdamW (weight_decay: 2e-4)
- L2 regularization: 0.0005
- L1 regularization: 0.00005
- batch_size: 128
- n_epochs: 10
- criterion: nn.BCEWithLogitsLoss (pos_weight = 1)
- sampler: make_balanced_sampler (target_pos_fraction=0.75 for class balancing)
- activation: relu
- gating features: Top 8



In addition, we found out that in the training set, each feature is binned by quantiles, the failure rate of each bin is calculated, and the lift is obtained by comparing it with the overall baseline failure rate. Merge adjacent bins with lift ≥ 1.3 and sample size ≥ 50 into continuous high-risk intervals and output metrics such as (risk_from, risk_to), support, and lift.

1	feature	risk_from	risk_to	support	lift_mean
2	w_row_count	3.6666666666666665	4	172	2.3703308619856434
3	Wagon_Curvature	0.000830912	0.003926437	1829	2.0466915718681147
4	Wagon_IntrainForce	27.446593061847498	143.0808745	3657	1.8192158260110065
5	Tng_Tonnage	11.199999999999998	104.1	7255	1.7433314948217198
6	Wagon_Rail_Pro_R	6.463793879516129	19.027802470781246	1829	1.6819346580698367
7	Wagon_BrakeCylinder	-25.03477878	-14.46400931	1829	1.5907454296202674
8	Wagon_Track_Offset	-703.2058318	-0.226270568	1829	1.5907454296202674
9	Wagon_LP2	-0.343743886	-0.245726018	1828	1.5662714423713568
10	Wagon_Acc4_RMS	-0.056293675	0.014558578995939233	1829	1.565415088384276
11	Wagon_LP1	-0.164947309	-0.121160572	1828	1.5257207254167588
12	Wagon_BodyRockRr	-1.140639945	-0.793626887	1828	1.5105142065587844
13	Wagon_SND_L	-0.132293806	-0.036143398	1828	1.5105142065587844
14	Wagon_Twist2m	-0.191222867	-0.01279519	3656	1.502910947129797
15	Wagon_SND	0.43023351090827383	0.556690612	3656	1.490238848081485
16	Wagon_VACC_R	0.25103104668270837	0.2861872929993333	1828	1.490238848081485
17	Wagon_BounceFrt	-0.214349045	-0.066882476	5484	1.4885492348750435
18	Wagon_Twist14m	-11.94190932	-6.723119089	1828	1.4851700084621604
19	Wagon_Acc2	0	0.12735308424440478	1829	1.4742258599347065
20	Wagon_BounceRr	-1.999244238	-0.223619031	1829	1.4387633822043182
21	Wagon_Rail_Pro_L	-1.70994734	-1.047927332	3656	1.437016032078575
22	Wagon_LP3	-2.578008503	-0.575976314	1829	1.3931687679795335
23	Wagon_SND_R	-0.123991041	-0.030407193	1828	1.3888620556949898
24	Wagon_VACC	0.1163091623158654	0.34148806295166667	3657	1.3884822738914089
25	Wagon_VACC_L	0	0.2512352418223401	5485	1.3801623189262888
26	Wagon_Speed	43.75106245828699	45.29730714285714	1828	1.3736555368370154
27	Wagon_Acc4	0.1844208080403846	0.21784843707193763	1828	1.3483113387403913
28	Wagon_Acc3	0.3849895650414473	0.4184062229717741	1830	1.3468377744357571
29	Wagon_Acc3_RMS	0.041420137826298734	0.046375491924049106	3656	1.3432424991210665
30	Wagon_LP4	-3.450708655	-0.645277697	1828	1.3026917821664683

In conclusion, multiple techniques and training strategies were explored for rail-break risk prediction.

- Base model: Transformer encoder adapted for multivariate rail sensor data
- Regularization: Group Lasso
- Loss functions: Binary Cross-Entropy (BCE), Focal Loss, WeightedRandomSampler
- Feature engineering: value clipping, normalization, zero imputation , removal of low-variance noise features
- Optimization: Cosine-decay scheduling
- XAI: SHAP

The combination techniques finally achieved:

Accuracy: 64.32%, AUC-PR: 46.57%, F1 Score: 57.68%