Sprint 4 Plans and Results

1. Introduction

Based on the sprint 3 result and user story 4, this sprint will apply a SHAP technique to select important features to improve model performance. Furthermore, hyperparameters for transformer model will be tuned based on the auc metric in this sprint.

2. Plans

In this sprint, we will tune hyparameters based on auc scores. The SHAP technique will also be used to find the importance and contribution of input features. This technique is used to help us select effective features.

2.1 Tuning hyperparater

	input_size	seed	d_model	n_head	num_layers	dropout	lr	weight_ decay	epochs
1	20000	42	32	4	2	[0.3, 0.4]	[1e-4, 5e-5]	[1e-3, 3e-3]	30
2	20000	42	64	4	2	[0.3, 0.4]	[1e-4, 5e-5]	[1e-3, 3e-3]	30
3	20000	42	96	4	2	[0.3, 0.4]	[1e-4, 5e-5]	[1e-3, 3e-3]	30
4	20000	42	128	4	2	[0.3, 0.4]	[1e-4, 5e-5]	[1e-3, 3e-3]	30

Table 1: Tuning hyperparameter based on auc

2.2 XAI technique: SHAP

According to Sprint 3 results, using the SHAP technique after training to find which features have high importance. The GradientExplainer of SHAP will be used to explain feature importance and contribution. We will input 36 features and virtualize the top 20 and top 10 feature importance by bar plot and contribution by beewarm plot. We will also select the top 20 and 10 important feature to retrain the model.

3. Results

3.2 Tuning hyperparamter

- Group 1: The best hyperparameters are {'d_model': 32, 'nhead': 4, 'num_layers': 2, 'dropout': 0.3, 'lr': 0.0001, 'weight_decay': 0.003, 'epochs': 30}. The leaderboard result

is Accuracy: 42.88%, AUC_PR: 38.44%, F1_Score: 48.39%.

01563cfa	RAIL-PG-2	Completed	5 days ago	68e834cd6164.csv	Competition 3 - The Defribillator	**Accuracy**: 42.88%, **AUC_PR**: 38.44%, **F1_Score**: 48.39%

- Group 2: The best hyperparameters are {'d_model': 64, 'nhead': 4, 'num_layers': 2, 'dropout': 0.3, 'lr': 0.0001, 'weight_decay': 0.003, 'epochs': 30}. The leaderboard result is Accuracy: 42.88%, AUC_PR: 27.04%, F1_Score: 48.39%.

6d7a4895	RAIL-PG-2	Completed	5 days ago	68e85ed23176.csv	Competition 3 - The Defribillator	**Accuracy**: 42.88%, **AUC_PR**: 27.04%, **F1_Score**: 48.39%

- Group 3: The best hyperparameters are {'d_model': 96, 'nhead': 4, 'num_layers': 2, 'dropout': 0.3, 'lr': 0.0001, 'weight_decay': 0.001, 'epochs': 30}. The leaderboard result is Accuracy: 42.88%, AUC_PR: 35.17%, F1_Score: 48.39%.

5cabd6e9	RAIL-PG-2	Completed	6 days ago	68e71cd65581.csv	Competition 3 - The Defribillator	**Accuracy**: 42.42%, **AUC_PR**: 31.50%, **F1_Score**: 47.75%

- Group 4: The best hyperparameters are {'d_model': 128, 'nhead': 4, 'num_layers': 2, 'dropout': 0.3, 'lr': 0.0001, 'weight_decay': 0.003, 'epochs': 30}. The leaderboard result is Accuracy: 42.70%, AUC_PR: 32.23%, F1_Score: 48.14%.

8ab681f4	RAIL-PG-2	Completed	4 days ago	68e9847a7475.csv	Competition 3 - The Defribillator	**Accuracy**: 42.70%, **AUC_PR**: 32.23%, **F1_Score**: 48.14%

Therefore, based on AUC score, the best group is Group 1. However, its Accuracy, AUC_PR, and F1 scores are lower than sprint 3 scores. We will use sprint 3 tunned hyperparameters based on F1 metric.

3.1 XAI technique: SHAP

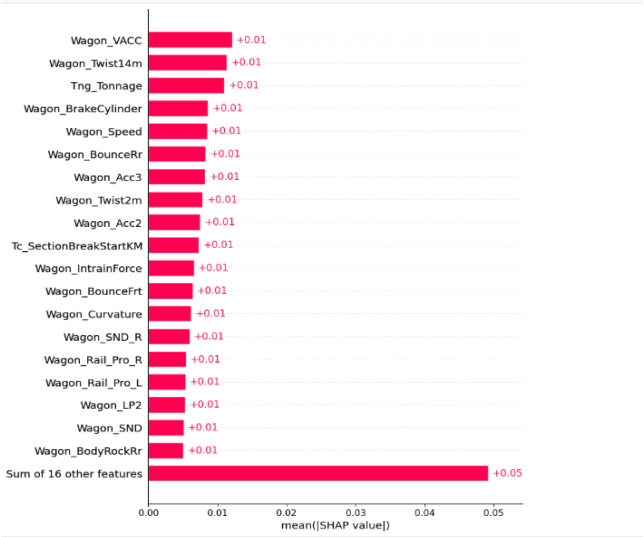


Figure1: Top 20 Feature Importance

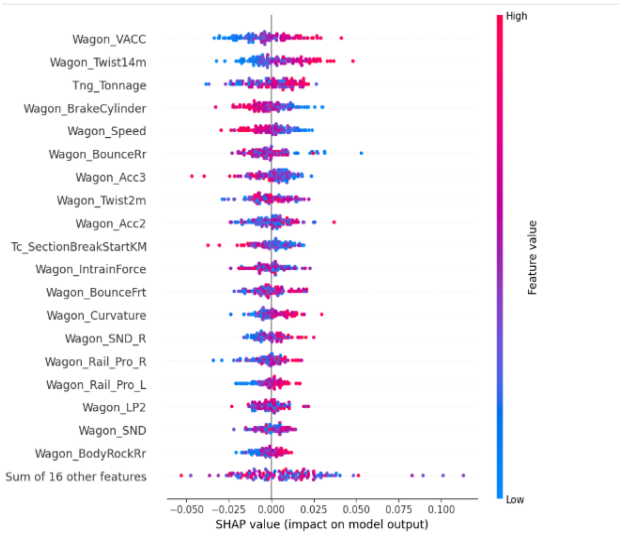


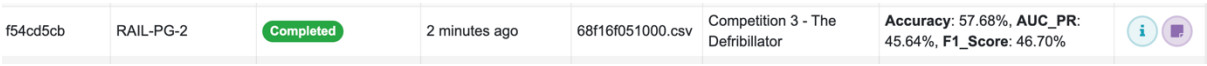
Figure 2: Top 20 Feature Contribution

We virtualize the top 20 feature importance with its contribution. We applied these features to retrain model and submit. We also compared the results of all input featurs (36 features) between the top 20 and 10 important features.

36 features:

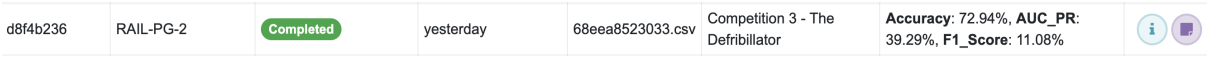
ba6978cc	RAIL-PG-2	Completed	yesterday	68eec2a72732.csv	Competition 3 - The Defribillator	**Accuracy**: 64.42%, **AUC_PR**: 34.68%, **F1_Score**: 32.14%

Top 20 important features:



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%

Top 10 important features:



d8f4b236	RAIL-PG-2	Completed	yesterday	68eea8523033.csv	Competition 3 - The Defribillator	**Accuracy**: 72.94%, **AUC_PR**: 39.29%, **F1_Score**: 11.08%

8 features based on EDA (Sprint 3 result):

33b2dc2d	RAIL-PG-2	Completed	15 days ago	68cccf8b6106.csv	Competition 2 - Senna	**Accuracy**: 45.16%, **AUC_PR**: 44.17%, **F1_Score**: 53.11%

Therefore, selecting the top 20 important features is suitable. This group gets high and similar accuracy, auc, and f1 scores among three groups. Also, compared this group with the Sprint 3 results, the accuracy and auc score are higher, but the f1 score reduces slightly.

Conclusion

In Sprint 4, we tuned hyperparameters based on the auc score. The scores are lower than the scores in Sprint 3. Therefore, we will continue to use tuned hyperparameters from the last sprint. Furthermore, we combine the SHAP technique with the Sprint 3 transformer model. This method explains the feature importance and contribution to the transformer model, which is more suitable and reliable for selecting features compared with EDA. Also, the trained model through SHAP has a higher accuracy and auc scores.