2025-09-20 Meeting Notes

🗓️ Date

20 Sep 2025

👥 Participants

- RAIL PG-2 project team
    - Tao Xu a1937511
    - Jinchao Yuan a1936476
    - Zilun Ma a1915860
    - Di Zhu a1919727
    - Xin Wei a1912958
    - Yifan Gu a1909803
    - Tianhua Zhang a1915934
    - Zihan Luo a1916700

🗣️ Discussion topics

1. **Share team progress in the end of Sprint 3**
    - **Feature engineering team:**
        - Applied resampling techniques on three training tables, including stratified and random oversampling/undersampling. In total, there are 12 training datasets for model evaluation.
        - Refactored the feature engineering script, facilitating the addition of new features.
        - Added lag features for all numeric columns to observe their impact on model performance
        - Introduced interaction features (e.g., curvature × speed, speed × twist) to assess potential performance improvements.
    - **Feature selection team:**

    - Explored feature selection strategies with varying time windows, Group Lasso settings, and parameter tuning in Focal Loss for imbalance handling.
    - Applied a new method, gating, achieving an F1 score of 53.64%.
- **ML model training team:**
    - Combined models with datasets that were handled by different feature selection and feature engineering techniques.
    - Reported which feature selection and feature engineering techniques are suitable and facilitate us to achieve a higher F1 score.
    - Tuning hyperparameters for three models, including SVM, DNN, and transformer.
    - Report the highest F1 score for every model. The highest F1 score for SVM is 52.51%, for DNN is 30.00%, and for Transformer is 53.11%.

2. **Discuss problems that were faced in the end of Sprint 3**
    - **Overfitting:** Trained models have good overall performance on validation and testing datasets. However, the leaderboard scores are very low.
    - **Model overall performance:** If the F1 score increased, the accuracy score decreased. Every trained model has this problem.
3. **Discuss the snapshot 3.2**
    - Check the content of the snapshot 3.2.
    - Every team summarized changes this week.
4. **Next steps**
    - The overfitting problem will be mitigated to improve overall performance.
    - The feature selection and feature engineering team will optimize useful techniques based on the feedback from the model training team.
    - The model training team will continue to experiment by combining new techniques or improved existing techniques with trained models.