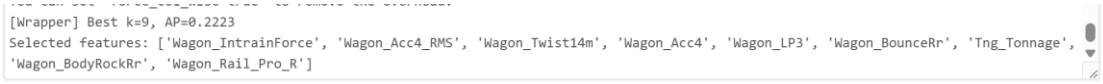
1. Introduction

To explore different feature selection methods and control feature size and training costs, I achieved a new wrapper-style feature subset search process called Sequential Forward Selection (SFS). The goal is to automatically select the top K features with the best performance based on time series cross-validation and the Average Precision metric, without introducing future information leakage.

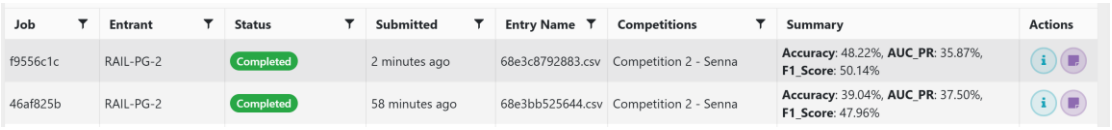
2. Method Overview

The model still uses LightGBM, with lightweight parameters (e.g., n_estimators=50) set to prevent run failures. The TimeSeriesSplit method (n_splits=3) is still used to strictly ensure time order. The "average_precision" metric used reflects the trade-off between recall and precision. The SFS search strategy starts with an empty set and greedily adds the next feature that maximizes the average_precision improvement to the remaining candidates in each round until a limit is reached.

3. Result Comparison



The above picture shows the 9 features selected by SFS.



Job	Entrant	Status	Submitted	Entry Name	Competitions	Summary	Actions
f9556c1c	RAIL-PG-2	Completed	2 minutes ago	68e3c8792883.csv	Competition 2 - Senna	**Accuracy**: 48.22%, **AUC_PR**: 35.87%, **F1_Score**: 50.14%	
46af825b	RAIL-PG-2	Completed	58 minutes ago	68e3bb525644.csv	Competition 2 - Senna	**Accuracy**: 39.04%, **AUC_PR**: 37.50%, **F1_Score**: 47.96%	

Using Sequential Forward Selection (SFS) optimized for Average Precision, the model improved AUC-PR (35.87% to 37.50%) while F1 (50.14% to 47.96%) and Accuracy (48.22% to 39.04%) decreased.