Smart Driver Allocation System

Introduction:

This is a classical use-case of allocating right drivers to the customers based on trip distance, driver distance from the user, event hour, driver gps accuracy & past records of successful rides. In the pipeline we are training a binary classification model with target variable as the which driver has accepted the order request.

This ML pipelines find the drivers in real-time.

Dataset:

booking_log.csv records order-level events: created, driver found/not found, canceled, completed

participant_log.csv records events at the participant (driver) level: created, accepted, rejected, ignored

Implementation:

1. Past record of successful rides has been created through Feature Engineering. The training data after processing, we are trying to find the successfully historical rides group by driver id. This data is also saved in processed.csv for prediction as well.
2. This historical rides data is merged with test data before prediction.
3. Modification in the training module is also done. Gridsearchcv is used for hyperparmeter and hyperparameters modified as list which is passed to gridsearchcv to find best parameters.
4. Since it is a binary classification problem, hence roc threshold is calculated to convert a probabilistic output to label. roc-auc-score is found to be 0.51. another evaluation metric f1 score is also calculated as 62.5% It is good metric to handle imbalanced classes and calculated as harmonic mean of precision & recall.
5. In the prediction also roc threshold is used to get labels from the probabilistic results.

Additional Experiments:

1. Tried with other Models like xgboost and SVM classifier. SVM classifier with rbf & poly kernel is giving better f1 score as ~62% compared to base line randomforest and xgboost.
2. Post hyperparameter tuning of randomforest is also giving approx. 62% f1 score. Hence grid searchcv is implemented in the final submission.