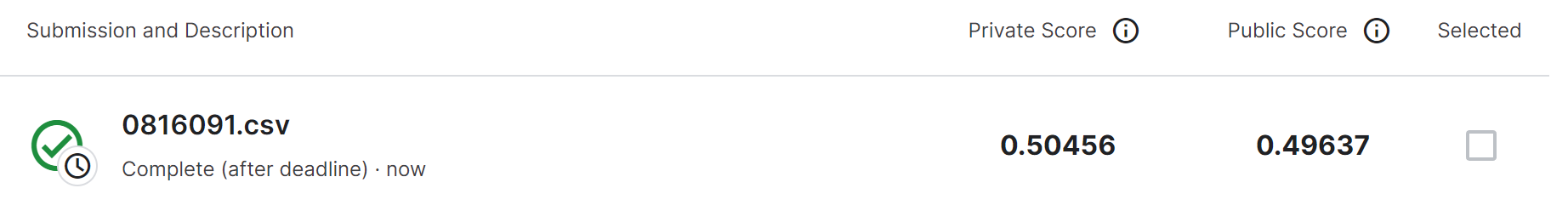
Link of github:

<https://github.com/ginnyching/2022-Fall-Machine-Learning/tree/main/ML_Final_Project>

Screenshot on Kaggle:



參考資料:

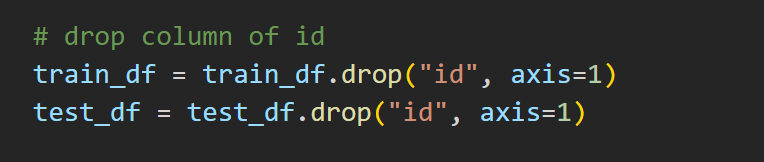
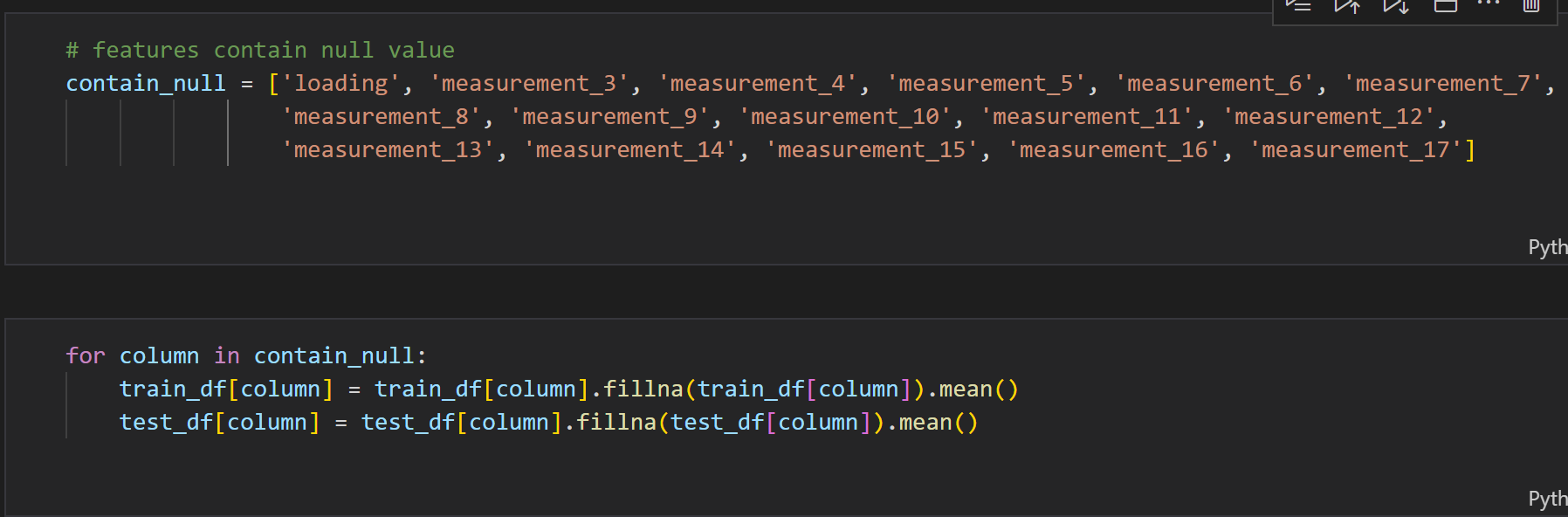
<https://github.com/mohamedali-sc/Tabular-Playground-Series-Aug-2022>

* Inference:

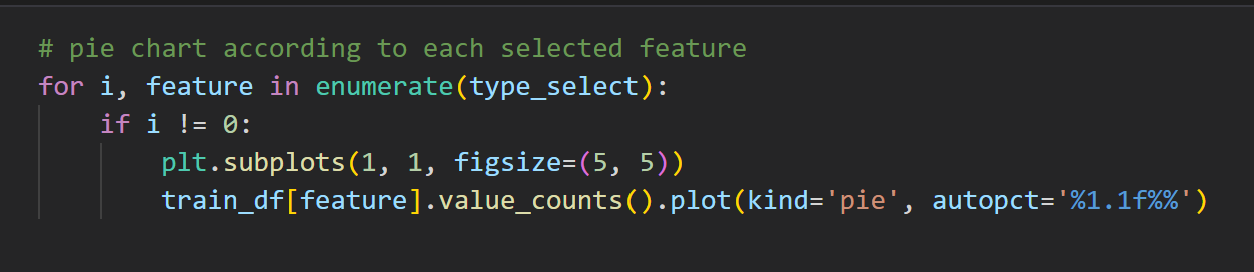
1. Load data from csv file



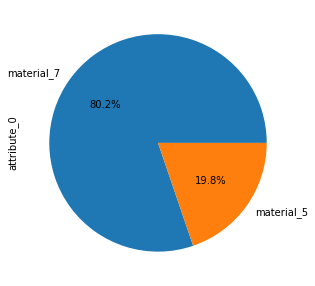
1. Drop unnecessary column and fill null with its mean

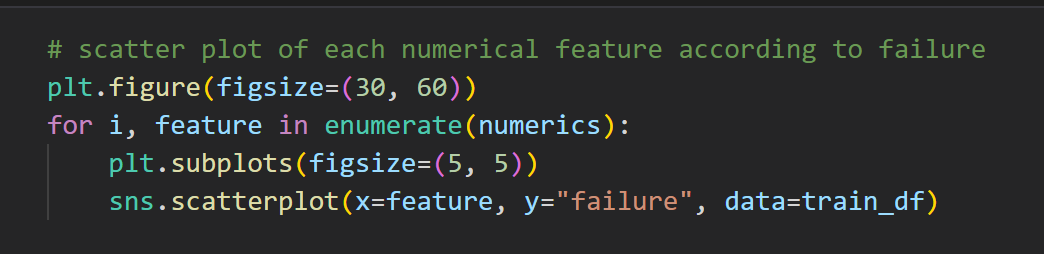
1. plot pie chart to see the ratio of each features



Example of pie chart using attribute\_0:



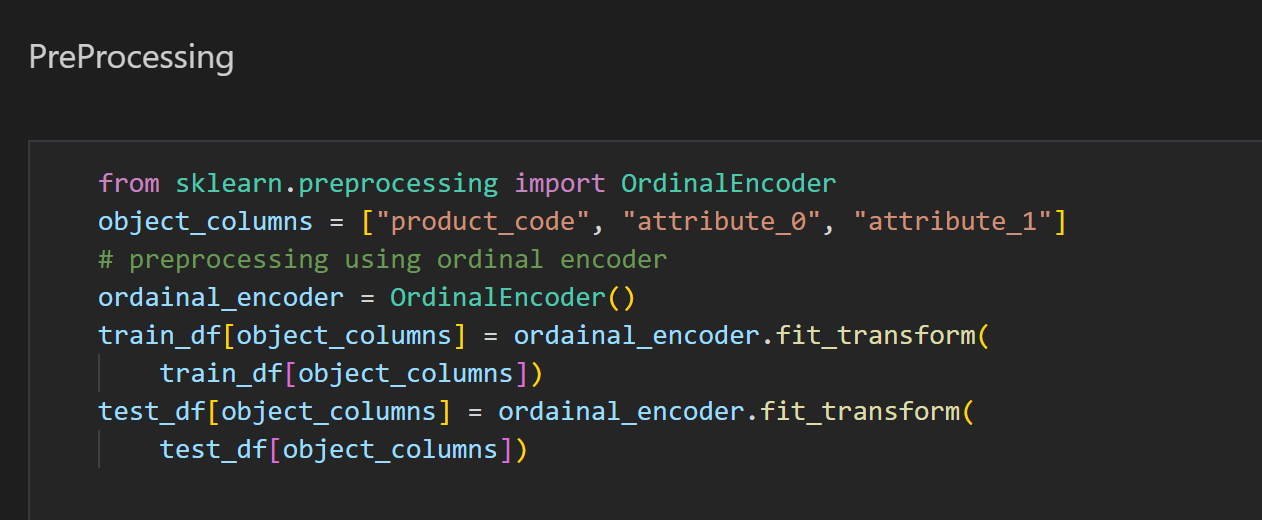
1. plot scatter plot to see how it relates to failure:

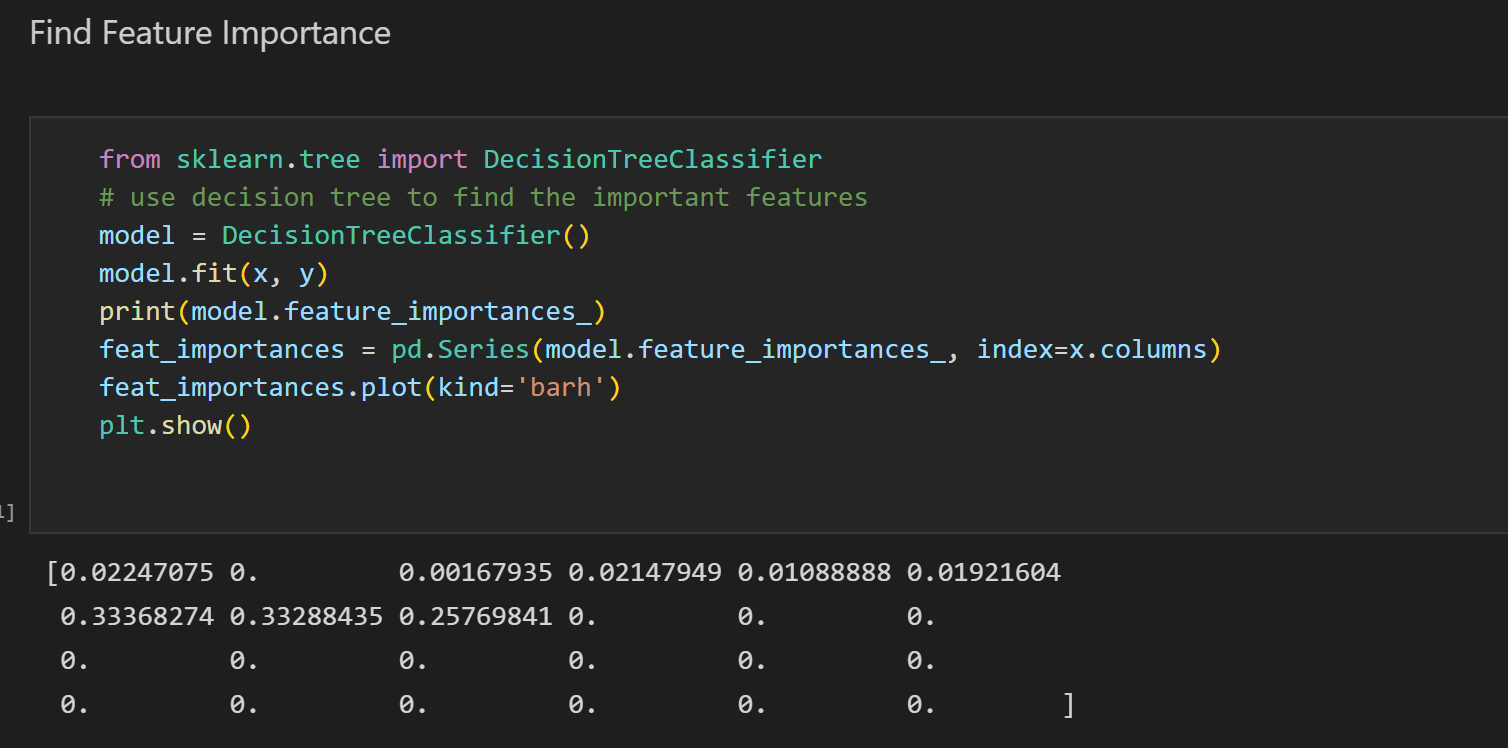


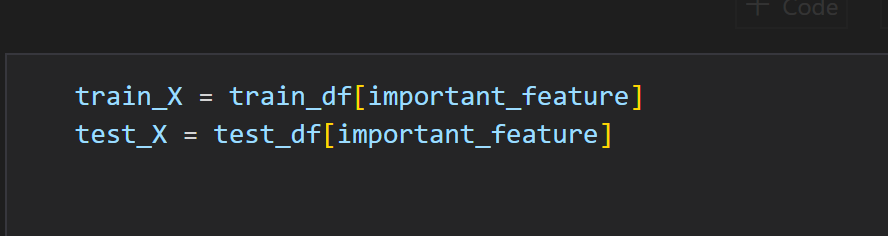
Example of scatter plot:



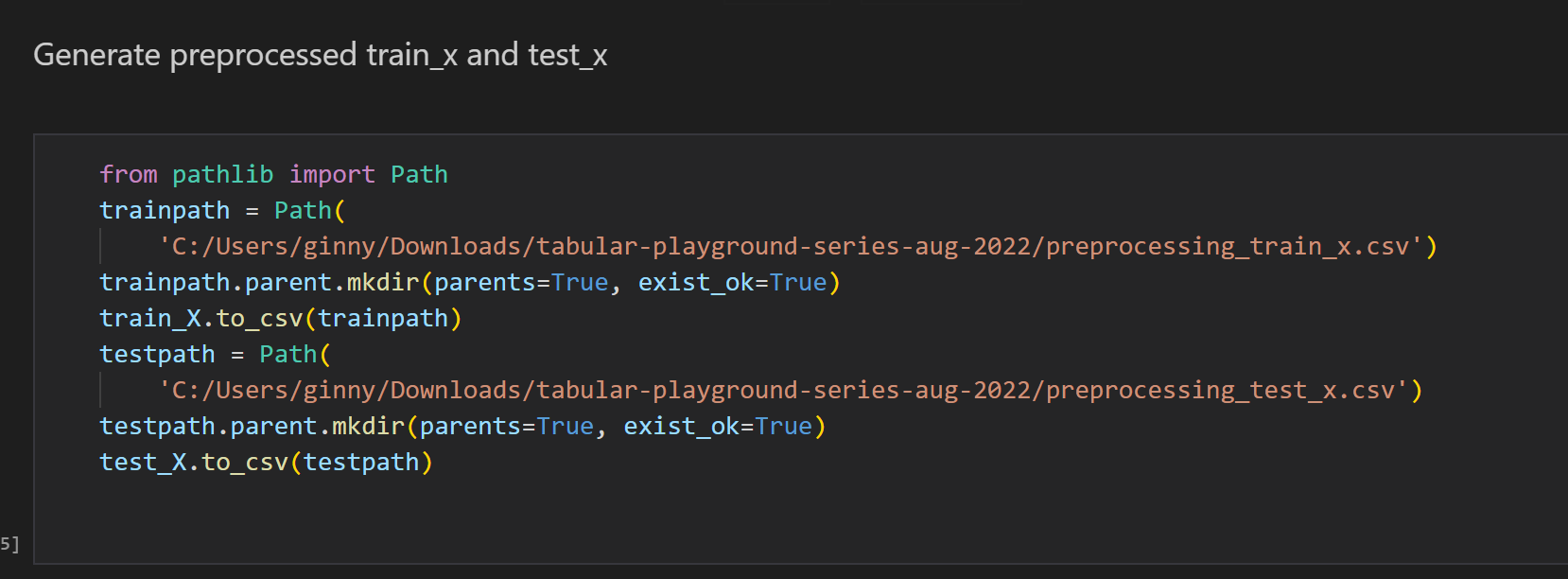
1. PreProcessing using Ordinal Encoder with object type columns:



1. Find feature importance using decision tree: 
2. Generate new train\_X and test\_X using important feature:

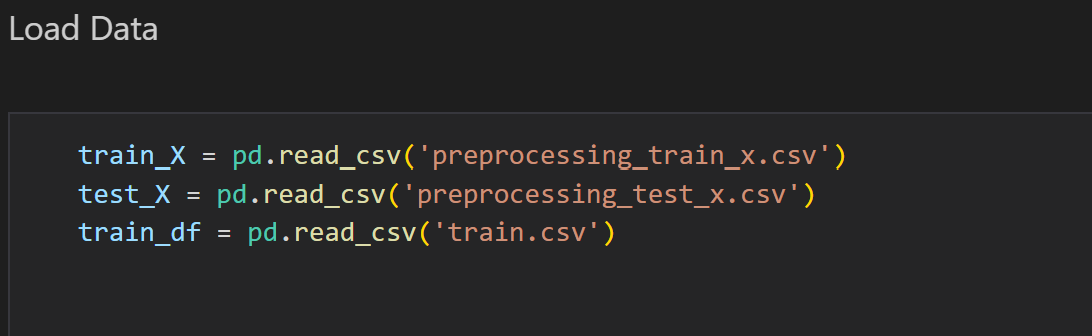


1. Output the new preprocessing.csv



* Training

1. Load data from csv files



1. use different methods to train and find the best method

methods I used: Decision Tree, Random Forest, Histogram-based Gradient Boost, XGBoost, AdaBoost, LGBM, Ensemble(Stacking Classifier)

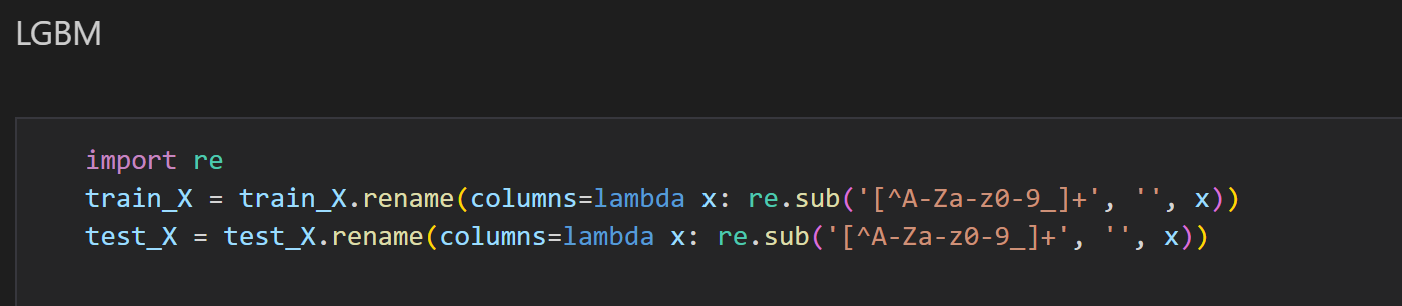
example code of training:



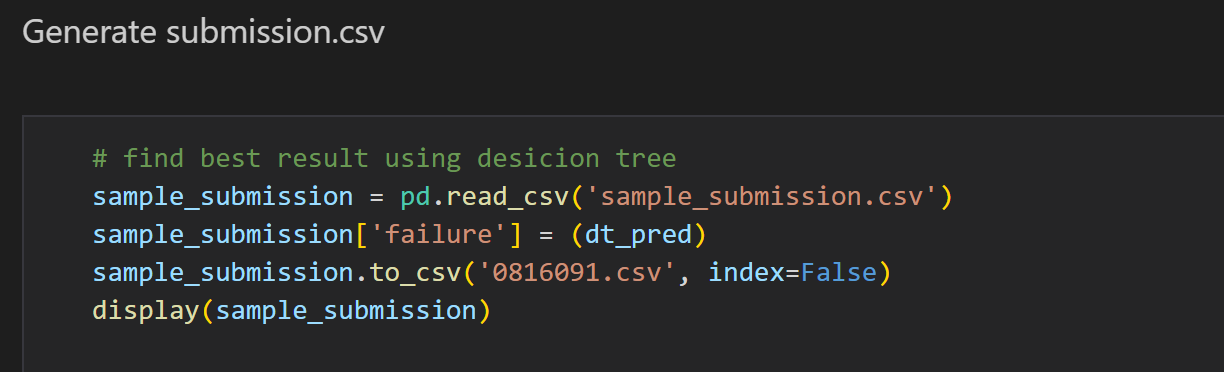
Problem occurred at LGBM:

Do not support special json characters in feature name.

Solution: rename the feature name of train\_X and test\_X.



1. Found best method is Decision Tree, generate submission csv according to its prediction.



1. Submit on Kaggle:

