由于gitlab空间限制，这里将原始数据集以及特征工程生成的数据打包存储到百度网盘上，上传至gitlab中的代码仅包含处理完毕后生成的最终训练测试集，如对原始数据以及特征工程生成的中间文件有兴趣可以去百度网盘下载然后分别解压到data和selected文件中。
（链接: https://pan.baidu.com/s/1miwzqdi 密码: 79km）

代码目录说明:
-data
存放原始数据文件
-feature_engineering
特征工程，包括生成排序特征、组合特征以及模型调参，生成的文件都在selected中
-feature_selected
包括从原始数据集组合训练测试集代码以及数据清洗代码，生成的文件都在selected中
-M1
运用处理完缺失值后并添加完排序特征后生成的训练测试集来训练xgboost模型，得到特征重要性排名，方便接下来组合特征
-M2
包括三个python文件，add_zuhe_feature.py文件是用来组合除特征以及log特征，并生成训练集，xgb_chu.py与xgb_log.py文件是分别使用刚刚生成的除特征与log特征构建xgboost模型，训练完毕后取靠前的特征加入到训练测试集中
-M3
使用处理完毕的训练测试集来训练最终的模型
-selected
存放特征工程生成的文件
train_nofill.csv：才原始数据集组合特征后生成的文件
train_addloss.csv：添加缺失值特征后生成的文件
train_addrank.csv：添加排序特征后生成的文件
train_chufeature.csv：除特征训练文件
train_logfeature.csv：log特征训练文件
train.csv：通过数据清洗以及特征工程对原始数据处理完毕后用了训练最终模型的文件

运行环境要求:
python3.5,scikit-learn,pandas,xgboost,numpy,matplotlib,seaborn,jupyter

代码运行步骤:
1.运行feature_selected中的feature_combination.py文件，运行完毕后生成train_nofill.csv文件；
2.运行feature_selected中的数据清洗.ipynb文件，生成缺失值特征（train_addloss.csv）；
3.运行feature_engineering中的rank.py文件，生成排序特征（train_addrank.csv）
4.运行M1中的xgb.py文件，生成特征重要性排名，方便接下来组合特征
5.运行feature_engineering中的zuhe_tz.py文件，生成除特征（train_chufeature.csv）以及log特征（train_logfeature.csv）
6.运行M2中的xgb_chu.py以及xgb_log.py文件，得到组合特征的特征重要性排名，再运行add_zuhe_feature.py文件，将得出的TOP特征添加至训练集中
运行M3中的xgb.py，训练最终的分类预测模型（最终训练集auc值为0.74001,验证集为0.7070，测试集为0.7077）