# Readme

**记录（瞎扯）日期：**2023/03/28

陆胤利

其实是29日凌晨了..🌙

## 模型

单独测试了Xgboost，LightGBM和Catboost，性能尚好。其中，Xgboost的auc较为稳定，LightGBM和Catboos的表现超过0.90，表现优异，但可能存在明显的过拟合。

3个模型最好需要再进行调参。

本来想分别尝试Voting集成，权重集成和Boosting集成3种对于模型的集成学习。但是Voting在预测上有问题，尚未解决；权重集成很简单，但是运算量过于庞大，无法解决（要我有接近1TB的运行内存，我上哪整去啊😭）；Boosting还没测试，看权重集成和Voting集成这样子估计用不了。总之，进一步的集成学习不是必要的，做来玩儿，可以放着不管。

## 数据预处理

选用了生成20维度的PCA，减少了运算压力，这一点可以写进报告里。

不平衡数据集使用的是速度最快的ROA，其他两个需要训练并迭代的模型需要较长的训练时间，洗个澡🛀出来才跑完，比模型还慢（当然，模型用了GPU，肯定快），所以之后再用，选取哪个模型根据报告的写作而定。

## 其他

本次打包上传的每个ipynb和html均有一定的注释，但是大部分是中文，且有些地方没统一。代码是抄Chatgpt 3 的，所以之后可能需要统一，如果看得懂又不用交给老师的话，可以拉倒。😅