# 期末作业 数据挖掘

## 题目描述

在课后作业.csv中，第一列group是我们的目标值，其他列都是用来预测group的属性值，请使用数据挖掘方法建模通过其他属性预测group并提交结果报告。

1. 可以使用各种数据挖掘或机器学习的算法，包括课上讲的或者没讲过的。验证必须采用5折验证。
2. 可以使用数据预处理。
3. 也可以不做分类，采用聚类方法通过其他属性做聚类，但结果必须有意义，例如和现有的group对比结果很相关。
4. 做一定的可视化工作和后续工作，包括但不限于重要特征分析等。
5. 也可以采用其他数据挖掘方法做任何你感兴趣的方向，但必须保证结果有意义。
6. 凡是上述做的工作都要在报告中写明，并说明理由，例如解释做的数据预处理步骤，验证步骤等。
7. 提交一个完整的可验证的python 3.x程序。
8. 成绩由程序结果和报告综合判定。不做预处理、采用sklearn（或其他工具包）的default参数的某种机器学习0/数据挖掘方法，给出重要特征的视为baseline，40分，之后每一个改进视情况加分，每个改进须在报告中明确写出。注：不调参数的xgboost方法给出的5-折验证accuracy约在75%左右，能提升accuracy的可以加分。其他未使用分类算法的视情形相应给分。
9. 最后结果提交截止时间为：**2024-12-22日中午12点**。