

【课堂汇报部分】

一、第一次汇报（学校排名数据集-后未采用该数据集）：

这一部分我们几乎全程一起讨论方案、分配任务，主要的分工在 PPT 制作部分。而这一部分的分工也与我们原本预计的后期自身工作分配有一定吻合度。

刘琮璟：

- 1、“研究任务和方法——个性化分析清北交复浙 5 所高校”
- 2、“预期成果和时间安排”

徐朱玮：

- 1、“研究方向和预期目标”
- 2、“数据来源和介绍”
- 3、“研究任务和方法——数据可视化、改变权重生成新的大学排名、分析预测 2025 年排名”
- 4、上台汇报（老师随机抽取）

二、第二次汇报（改换数据集为银行信用卡用户数据集-该汇报包含开题和中期）：

徐朱玮：负责全部的 PPT 制作和上台汇报

三、第三次汇报（海报展示和结题答辩）：

刘琮璟：负责全部的海报制作和上台汇报及几乎所有答辩工作

【任务安排和论文撰写部分】

整个项目的过程中，我们全程保持联系和探讨。因此，在完成各自任务时，也会询问对方意见并互相帮助，双方对整个课题的全部进程都是全程参与并且非常熟悉的。当然，我们也做了较为明确的分工。

刘琮璟：

- 1、论文题目、摘要、关键词撰写
- 2、研究方法部分的全部撰写（包括随机森林、神经网络等 6 个模型的原理介绍）
- 3、负责 Lasso 回归、KNN 和神经网络的模型搭建（指代码）
- 4、负责实验过程部分的全部论文撰写（即论文第 5 大块的全部内容撰写）
- 5、负责实验结果分析部分中客户流失预测、模型性能分类、模型综合选择工作和论文撰写
- 6、负责不足与展望中相似矩阵分析的撰写
- 7、负责不足与展望中下采样部分的工作（指代码）
- 8、负责绘制论文第 6、7 大块中的全部图表（详见“大数据引论项目代码(第 6、7 部分图表).ipynb”）

徐朱玮：

- 1、引言部分的全部工作和撰写
- 2、研究背景和意义部分的全部工作和撰写（包括文献查阅的工作）
- 3、数据集介绍和预处理部分的全部工作和撰写（包括格式转换、归一化等）
- 4、性能评价指标部分的全部工作和撰写（包括对混淆矩阵、ROC 曲线的选用等）
- 5、负责决策树、随机森林和 SVC 的模型搭建（指代码）
- 6、负责混淆矩阵和 ROC 曲线图像输出的函数封装（指代码）
- 7、负责随机森林的特征输出（指代码）并撰写实验结果分析的特征分析部分
- 8、负责不足与展望中下采样部分的论文撰写