

2025 秋《基于 Python 的问题解析》期末大作业

本次作业截止时间为 2025 年 12 月 30 日 23:59

提交作业邮箱: 2024104080@ruc.edu.cn

(注意作业附件名请使用：姓名+学号+期末大作业的格式)

(作业附件仅提交 pdf 格式文件)

期末大作业要求：

1. 自行寻找公开可用的数据集（比如股票数据、房价数据、PM2.5 气象数据、电影票房数据、视频网站数据、基因表达数据、社交网络数据、新冠数据等），并提出相关问题，进行探索性数据分析（可能用到的包：numpy, pandas, matplotlib 等）。利用所学到的《编程基础》的知识来解决提出的问题（**要求包含类的定义、解决异常和调用函数**）。
2. 大作业报告提交格式仅需提供 pdf 文件。
3. 大作业报告应当包括以下内容：
 - 1) 问题背景，问题提出，问题描述。
 - 2) 数据来源（数据源网址或处理后的数据集链接）。
 - 3) 探索性数据分析，包括但不限于数据基本信息、数据清洗过程及数据可视化展示。
 - 4) 结合但不限于本门课程所学习的知识进行分析（比如利用 scikit-learn），并在报告中展现出用到了类的定义、解决异常和调用函数等，注意代码的简洁性和模块化。
 - 5) **至少使用一种深度学习或机器学习算法**，比如神经网络（DNN/CNN/RNN）、树模型（随机森林、boosting）、EM 算法、SVM、逻辑回归等，并且体现**调参过程**。同时，至少包括两种方法/算法/模型解决问题，请在报告中展示不同算法的结果，**对比不同算法的优劣性并尝试分析原因**。
 - 6) 问题的结论及总结。
4. 请把项目上传到 GitHub 上，包括处理后的数据/原始数据/数据链接，可复现代码（使用.ipynb 或.py 格式文件），并在 GitHub 的 Readme 文件中书写项目简介和复现说明（比如运行环境，使用包的版本，复现步骤等），请在 pdf 报告中给出项目的 GitHub 链接。
5. 报告文件不超过 15 页，合理分配各章节内容，条理清晰，字数不限。