

QG 工作室暑期实习生两日结

姓名： 许继元 组别： 数据挖掘 年级： 19 级 日期： 2020.07.28

生活小记	<p>7.27 日，由于之前没有系统学习 Matplotlib 和 Seaborn 等绘图库，上午花时间去查阅资料寻找多维数据的可视化方法，一共找到了平行坐标、RadViz 雷达图、Andrews 曲线这三种可视化方法。不足之处是这三种方法只能处理输出变量为离散值的情况，也就是分类问题，而对于回归问题，这三种方法不可行。下午完成了数学建模校赛的第三题，也就是电影票房模型的构建，但是找了很久也找不到存储票房信息的网站，后面想到了一个方法：首先用爬虫爬取电影数据，然后提取电影名称，接着通过谷歌搜索获取每一部电影的票房信息，然后清洗数据，用多元线性回归模型进行训练，最后得出的结果还不错，预测两部电影的票房也在正常范围内，模型平均误差达到 3.04 左右。晚上还是日常抽出时间练琴。</p> <p>7.28 日，上午继续学习 Web 开发的教程，学习了用户名和密码的数据库检索查重和利用 Flask-login 进行用户登录，下午和晚上完成了中期考核项目的 iris 数据集的原始分布和逻辑回归算法模型训练过的数据挖掘结果可视化，第一次提交开发内容到 GitHub 的时候出了点差错，代码全部不见了，冷静分析并在网上搜索解决方案之后，顺利合并到了 master 分支。睡觉前看了一下最近的乐队节目的视频，花了点时间练琴。</p>
学习开发比赛概要	<p>学习上，学习了平行坐标、RadViz 雷达图、Andrews 曲线这三种基于分类问题多维数据集的可视化方法。在 Web 开发的学习上，学会了用户名和密码的数据库检索查重和利用 Flask-login 进行用户登录。</p> <p>开发上，完成了中期考核项目逻辑回归算法部分的原始数据多维可视化分布和数据挖掘结果分布图。</p> <p>比赛上，完成了数学建模校赛的电影票房模型搭建，模型平均误差达到 3.04 左右。</p>

感想收获	<p>生活上，虽然每天的任务都比较繁重，但是只要每天抽出时间练习吉他，演奏水平还是有在慢慢进步，也可以慢慢学习新的歌曲。</p> <p>学习上，随着Web开发知识的学习，感觉这个新的领域也蛮有意思的，想再继续了解下去。Python的一些绘图库也想系统地过一遍，因为这是数据可视化的一个重要环节。</p> <p>开发上，和小伙伴们一起讨论问题、开发项目，是一件很有趣的事情，也很期待线下的多人协作。</p>
存在问题 (备注)	