暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | 数据处理 | | | | | | |
| 姓 名 | 吴亦珂 | 专业 | 计算机科学与技术（人工智能） | 班级 | 人工智能2班 | 学号 | 09118223 |
| 实验时间 | 2020.8.31-2020.9.23 | | 指导教师 | 孔祥龙 | | 成绩 |  |
| 1. **实验背景和目的**   **近几年来，高考成为了大众越来越关注的一个问题。想考上理想的大学，仅取得高分是不够的，高考志愿的填报也万分重要。如果志愿填报得当，可以考上更好的大学，而志愿填报失误，甚至有可能出现滑档的情况。因此，志愿填报问题就显得愈发重要，如果能较为准确地预测出考生的分数能够上的学校，对于广大学子来说，也是一个福音。**  **我们本次暑期学校实验项目是高考志愿填报助手，目的就在于能够根据考生的分数以及自身其他情况（如竞赛成绩等），为考生推荐志愿填报的大学。这样可以有效避免考到高分，却最终上了一所不满意的学校的惨况。同时，也能让莘莘学子真正收获与自己付出相匹配的结果。** | | | | | | | |
| 1. **小组任务和个人任务**   **小组任务：我们小组的任务主要负责数据的清洗，将一些错误的数据进行更正，冗余的数据进行删除，提供缺失的数据，收集必要的信息从而为其他各组输送可以直接利用的高质量数据。这使得其他小组在进行项目实践的过程中，能够更为便捷地开展自己的工作。**  **个人任务：1. 为部分大学添加省份。**   1. **大学排名查找。** 2. **部分省份双一流大学的整理。** 3. **实现一个查询函数，查询某年某省某类别的一分一段表。** 4. **实现一个查询函数，接收的输入为分数，省份，科目类别，分数上下波动10分，查看去年的这个分数段能有哪些大学。** 5. **实现一个查询函数，接收大学，科类，年份，省份作为输入，返回大学该年份在该省的各专业分数情况及排名的dataframe。** 6. **重新爬取了南京理工大学数据。** 7. **修改了哈尔滨工业大学和西安交通大学的数据。** 8. **检查了重庆以及西藏的数据。** | | | | | | | |
| 1. **个人任务需求分析**   **我个人的任务主要是负责数据清洗，从而为其他组提供优质的数据。数据对于其他组来说非常重要，如果数据有问题，其他组进行实现的过程中就会造成巨大的障碍。并且在测试过程中也会造成极大的干扰，甚至可能导致错误的预测结果，从而造成很多的无用功。对于部分使用深度模型进行训练的小组来说，错误的数据将会是灾难性的，会引入不必要的噪声，极大地影响了最后训练出的模型的结果。**  **因此，我的任务看似非常简单，但实际上也非常重要。对于每个数据都必须本着认真负责的态度，尽量不要出现丝毫的偏差。所以，我在本次实验的过程中也确实对每个数据都认真负责地进行修改或者添加，尽量不要给其他小组增添麻烦。** | | | | | | | |
| 1. **实验过程（需附上关键代码及相关说明）**   **我的编程任务主要集中在查询函数的编写：**   1. **实现一个查询函数，接收大学，科类，年份，省份作为输入，返回大学该年份在该省的各专业分数情况及排名的dataframe。**       **相关说明：**    **查询符合条件大学的相关信息**    **查询对应专业的排名**   1. **查询某年某省某类别的一分一段表**      1. **接收的输入为分数，省份，科目类别，分数上下波动10分，查看去年的这个分数段能有哪些大学。**     **相关说明：**  **读入所需文件**    **波动10分并进行查询** | | | | | | | |
| **五、实验结果与分析**  **1. 实现一个查询函数，接收大学，科类，年份，省份作为输入，返回大学该年份在该省的各专业分数情况及排名的dataframe。**  **查询结果如下：**     1. **查询某年某省某类别的一分一段表**   **查询结果如下：**     1. **接收的输入为分数，省份，科目类别，分数上下波动10分，查看去年的这个分数段能有哪些大学。**   **查询结果如下：**    **分析：这三个查询函数均可以很好地实现各自的功能。所采用的思路基本上都是通过遍历已有的CSV文件，筛选出所需要的数据，并且构建成dataframe的形式返回。** | | | | | | | | |
| **实验总结与心得体会**  **实验总结：这次实验我所在的小组负责数据清洗的相关工作，主要是依靠人工对数据进行调整、修正，也有少量的工作要靠代码实现。虽然这些工作都是较为底层的，主要是为其他小组的工作提供基础的数据来源，以及确保数据的准确性，但是从这些工作中我也的确收获了很多。在这次实验中，我按时按质地完成了组长分配的任务，并且从这些工作当中学会了如何清洗数据，如何进行团队间的合作，为以后的学习和科研打下了一定的基础。**  **心得体会：这次暑期学校的项目，是我第一次参与到这么大的团队的项目之中。除了有一些新奇，更有一丝隐隐的担忧。我之前一直没有机会参与到正式的项目之中，也没有经历过一些较大的团队合作，对于这方面可谓是一片空白，对于这么多人的团队，是否能和小组成员协调好，也是一个棘手的问题。另外我的编程能力也偏弱，能否在有限的时间内完成相应的任务也是一个很大的挑战。但是在做项目的过程中，我的这些焦虑逐渐烟消云散，随着项目的推进，我才意识到团队合作的强大，一个看似浩如烟海的数据筛查工作，经过我们小组的分工，到每个人手中也就并不算多。另外，在与小组成员相处的过程中，我们相互帮忙，共同解决问题，也给我留下了深刻的印象。同时，通过这次项目也暴露出我的不足，如编程能力依然欠缺，交际能力依然不足。**  **在这次项目中，我觉得我学到的最为宝贵的东西就是团队协作。虽然这是100多个人的项目，但是从中我也看出了团队协作的力量和重要性。一个浩大的任务经过各个小组的合理分工，其实也没有太多。由于我们小组是为其他组提供数据，因此我们组内还有一些专门负责对接的成员，他们根据组内的需要，向我们提出一些要求，我们将数据清洗后递交给他们。通过这些往来，使得整个程序进行得井井有条，并且最终也获得了成功。而这个项目的成功与团队协作也绝对是密不可分的。**  **在这次暑期学校的项目中，我获益匪浅。不仅提升了自己的编程、科研能力，最重要的是，我更加清楚地了解到了团队协作的力量以及重要性。这次项目的经历将永远深深地印在我的脑海中，激励着这我以后的学习，同时也时刻鞭策着我向更好的自己迈近。** | | | | | | | | |

2020年9月制