**2020-5-31学习阶段总结**

1. 所做事情
2. 课程上课，现在除了计算机体系结构课程和大数据课程报告，其他都已结束。
3. 做疫情可视化交互式网站，将腾讯和百度的数据爬取后用Echarts作图和Flask框架显示。参考的网上的一个视频教程，自己搭的网址<http://175.24.77.110/>。
4. 参加数据挖掘比赛。
5. 两周前参加了老师推荐的一个面试，数据分析实习岗位，二轮面试后没有通过。
6. 遇到问题
7. 做的疫情监控网站在部署到服务器上时遇到了一些问题，在开发者模式可以正常运行，到生产模式后显示不出来，配置nginx时，一直出现端口占用，网上搜索相关解决方法没能找到适用自己这个的。

分析这个问题可能情况：① 真正出问题的地方没有找出来 ②漏掉一些关键步骤自己没有意识到。

1. 数据挖掘竞赛负责写代码：题目是基于网上市民问政平台，做相关文本分析，自己做的过程中大都是参考网上的一些现有的例子进行修改的，很多代码地方是引用的，做的最多的是用pandas处理数据。
   1. 问题一是文本多分类，将市民留言归类到不同的问题解决范畴（如：城乡建设，交通运输，教育文本）。参考网上的案例做的一个多分类，用的LinearSVM模型，精确度有90%。
   2. 问题二是挖掘出市民留言热点问题前五个，采用的LDA模型，将所有留言根据主题模型选出比重最高的前五个。
   3. 问题三是根据政府对留言的回复，做一个答复意见评价的模型。这一题做了留言和对应回复的相似度分析，计算了两者之间的余弦相似度，后面不会没有做了。（但自己做的余弦相似度效果也不好，有一些回复如“收到”，“知道了”，就不能区分出来）。
2. 实习面试，问了一些英语，数学，机器学习算法方面的知识，这方面比较弱，没有什么好的例子来说明自己学的情况比较好。
3. 总结和计划
4. 做的网站还在修改和完善，之后加上世界疫情数据。目前只做了数据显示方面，没有相关分析，老师说能做出数据处理，（如转折点之类的会比较好）。
5. 加强python方面的知识，刷数据结构算法题，看机器学习案例方面的教程和多做相关练习。
6. 还不清楚6月份还能不能去学校，在学校会有规律一点；家里也有好处，补基础的时间会比较多。