《大数据应用开发》课程期末考试方案

1. 题目

根据学号尾号，查询对应考试题目。

|  |  |
| --- | --- |
| 学号尾号 | 题目 |
| 1，9 | 1 |
| 2，8 | 2 |
| 3，7 | 3 |
| 4，6 | 4 |
| 5，0 | 5 |

题目1：文本情感分析

<https://www.pkbigdata.com/common/cmpt/%E6%96%87%E6%9C%AC%E6%83%85%E6%84%9F%E5%88%86%E6%9E%90_%E8%B5%9B%E4%BD%93%E4%B8%8E%E6%95%B0%E6%8D%AE.html>

题目2：幸福指数分析

<https://tianchi.aliyun.com/competition/entrance/231702/information>

题目3：中文医学问答机器人

<https://www.kesci.com/home/dataset/5e4d34b480da780037c0128e>

题目4：学生的学业成绩分析

<https://www.kaggle.com/aljarah/xAPI-Edu-Data?select=xAPI-Edu-Data.csv>

# 题目5：携程房型产品用户行为预测

<https://www.kesci.com/home/dataset/58d4e28c84a25f34b1d94906/files>

自拟题目：

如果题目不满意，也可以自拟题目。从下列网站中寻找想做的比赛题目或者数据集自行进行分析。在线编辑填写到自拟题目.xls，但是每个班重复题目不超过5个，截止到周二定好题目。

kaggle <https://www.kaggle.com/>

AI研习社 <https://god.yanxishe.com/>

天池 <https://tianchi.aliyun.com/competition/gameList/coupleList>

DataFountain：<https://www.datafountain.cn/competitions>

1. 考试时间

2020年6月19日之前截止

1. 评分标准

从数据预处理（20分）、模型建立分析（30分）、参数调优（10分）、预测效果（10分）、数据可视化（10分）、项目报告（20分）给出综合分数。

1. 要求

所有题目可以参考网站的论坛代码，但是不能抄袭。

尽量使用spark ml进行数据分析。并撰写项目报告文档。

五、上交内容

**将代码、数据集、项目报告（见模板）打包命名为学号+姓名，提交作业。**