|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ---------------------------------------------装---------------------------------------------------订-----------------------------------------------线----------------------------------------------  班级 姓名 学号 教学班名称 序号 任课教师 - ---------------------------------------------------密----------------------------------------封-------------------------------------------线----------------------------------------------------- | | **广东财经大学线上考试答题纸（格式二）**  课程 大数据分析与挖掘 考试时间  成绩 评阅人 胡苏   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 评语：   |  |  |  | | --- | --- | --- | | **检查项目** | **权重** | **得 分** | | **(1)** **选题合理性：**选题与本课程的关联性。 | 10 |  | | **(2)** **材料完整性**：包括数据获取及清洗，特征工程，模型选择及对比，模型调优，模型评估及结果可视化展示等；材料真实，内容翔实。 | 35 |  | | **(3)** **模型设计的科学性和合理性**：主要考虑特征工程阶段及模型选择阶段的分析与设计过程的科学性及合理性。 | 35 |  | | **(4)** **文档规范性**：项目报告规范，内容完整，术语表达准确，描述清晰，结果可视化直观。 | 20 |  | | **合计** | **100** |  | |   ＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝  题目：综合应用数据挖掘技术解决实际问题，问题从以下项目选题中任选一道完成。  设计要求：基于Python语言，操作系统不限。  提交文档要求：设计报告格式参照设计模板（在正文处给出）。  项目选题：  **【选题1】** 根据2007-2020年俄罗斯各地机场的货运量及客运量数据，分析新冠疫情对这些机场的影响。  **数据获取**：链接: https://pan.baidu.com/s/1Ii\_QntUo0XSSB9BgHC94QA 提取码: vf9p  **数据描述**：有关2007年至2020年俄罗斯机场客运量及货运量信息。取自俄罗斯联邦运输部的官方网站。文件按月包含机场ecah中的abaut passer航空交通信息，机场也有地理坐标。  **【选题2】**[真实或伪造]伪造职位发布预测  **数据获取**：链接: https://pan.baidu.com/s/1YzDakK-0b98B\_XqlxjHTnA 提取码: rdev  **数据描述**：该数据集包含18K职位描述，其中约800个是伪造的。数据包括文本信息和有关作业的元信息。该数据集可用于创建分类模型，该分类模型可学习欺诈的职位描述。  **【选题3】** 上海车牌拍卖价格  **数据获取**：链接: https://pan.baidu.com/s/1MiEUnPYvitDOv7K49JrpJg 提取码: f78n  **数据描述**：上海每个月都会使用拍卖系统向化石燃料汽车购买者出售数量有限的车牌。这个牌照的平均价格约为13,000美元，通常被称为“世界上最昂贵的金属片”。因此，我们的目标是预测下个月的平均价格或最低价格。  **【选题4】**电信客户流失  **数据获取**：链接: https://pan.baidu.com/s/1S8EmIsA67HMgAQywwL\_2nw 提取码: gpwu 复制这段内容后打开百度网盘手机App，操作更方便哦  **数据描述**：分析所有相关的客户数据并制定有针对性的客户保留计划。  **属性描述：**如下图 | |
| （正文）  **《×××》项目报告**  **一、项目概况**  （五号宋体）  **二、数据获取及分析**  （五号宋体）  **三、特征工程**  （五号宋体）  **四、模型设计、选择**  （五号宋体）  **五、模型评估**  （五号宋体）  **六、实验结果对比及结果可视化**  （五号宋体）  **附录 A主要实现代码** | |
|  | |

|  |  |
| --- | --- |
|  | |
|  | |
|  | |