**贺州学院课程考核试卷（设计制作）**

**（2021—2022学年第一学期）**

**课程名称：大数据平台实践训练 开课单位：人工智能学院**

**考核年级、专业：19级数据科学与大数据技术** **任课教师：陈醒基**

共享单车大数据分析项目训练，也可以自由选择一个实际的数据分析项目，以分组团队协作的形式，按照大数据分析的生命周期，完成项目需求分析、技术实现、工程管理、撰写大数据研究报告和开发文档等工作，具体要求如下：

1.在该项目中，根据项目特点、规模及团队实际情况选择适当的数据理解、数据准备、模型训练、模型评估、模型优化与重新选择、模型应用的方法和工具。

2.数据理解阶段：首先业务理解，确定业务目标，分析业务现状、趋势及存在的问题；其次数据理解，包括数据与业务之间的映射关系的理解以及采用探索型分析。

3.数据准备阶段：首先完成对数据集进行恰当地划分，如按比例划分训练集、测试集、验证集；其次需要对数据进行适当加工，如数据标准化、特征工程等。

4.模型训练阶段：根据业务需求以及数据特征，选择一种机器学习算法，以训练集为基础训练具体模型。

5.模型评估阶段：选择恰当的方法和工具评估新模型的优度，判断模型是否存在过拟合、欠拟合等现象。

6.模型优化与重新选择阶段：当模型评估结果不满足业务需求时，要进行超级参数的调优以及尝试其他算法重新训练模型。

7.模型应用阶段：将模型部署在应用系统中，并对新数据进行测试。

8.报告撰写阶段：把上述几个阶段整理并形成一份研究报告，其中，研究报告必须包含各个阶段进度安排，每一个阶段说明与相应的代码、测试结果及发现、测试结论。

附件一

**贺州学院人工智能学院**

**研究报告**

|  |  |
| --- | --- |
| 课程名称： | 大数据平台实践训练 |
| 课程论文题目： |  |
| 姓 名： |  |
| 系： | 大数据技术系 |
| 专 业： | 数据科学与大数据技术 |
| 年 级： | 2019级 |
| 学 号： |  |
| 指导教师： | 陈醒基 |

2021年 11 月23日