

《数据挖掘基础 A》线上实验任务书

一. 目的

基于本学期的特殊原因,综合考虑集中风险和课程要求,以及学生仍需完成17-19周的集中实验等原因,提出本任务书方案。《数据挖掘基础 A》是计算机科学与技术院智能科学技术的必修课,是掌握数据分析能力的一门重要基础课程。本课程首先讲授了数据分析的基本知识概念、数据分析预处理的手段,接着从数据分析方法的角度,介绍了数据挖掘关联分析、分类以及聚类三大类算法的基本知识、必要理论基础以及一些经典的数据挖掘算法。通过对本门课程的学习,学生能够系统地获得数据分析方法的基本概念和理论技术,掌握关联规则分析、分类和聚类数据挖掘算法,从而使学生会利用数据预处理和数据挖掘的技术去分析和解决不同行业应用领域中对数据进行处理和获取知识的问题,对培养学生形成良好的计算机科学技术和人工智能领域知识的运用能力有很大的帮助。

二. 要求:

1. 1~3 人一组(按课表上划分的实验班内自由组合,不可跨班实验)。
2. 通过互联网、图书馆搜集相关资料,认真分析题目。
3. 掌握数据挖掘不同算法的原理及编程应用,准确把握所要完成任务的功能需求,认真完成任务前期的需求分析。
4. 模块的结构设计完成后,进行功能模块的界面风格和要求规范的设计,实现变量命名规范和模块命名规范的设计,根据设计结果撰写各模块说明书
5. 进行检查,并提交设计报告。

三. 报告要求:(格式见报告册模版)

1. 实验题目
2. 需求分析
3. 任务总体设计:包括任务说明与描述,用数据分析处理流程的说明与描述等内容。
4. 任务详细设计:包括详细的数据预处理过程、算法实现的设计和实验结果展示图,对实验结果的说明与描述等。

5. 心得体会

四. 设计内容

1. 利用 **Kaggle** 数据集（股票、新闻、金融、商业皆可），要求：自行找到公开数据集，进行基本的数据存储、数据加载、数据清洗、数据挖掘算法建模，报告中必须包含数据的详细信息和建模结果。
2. 利用 **UCI** 公开数据集进行数据挖掘（不能使用其中 **iris** 数据）。要求：选择公开数据集，进行基本的数据存储、数据加载、数据清洗、数据挖掘算法建模，报告中必须包含数据的详细信息、建模过程和建模结果。
3. 利用相关算法解决 **MINIST** 手写数据集分类问题，要求：进行基本的数据清洗、数据分析，选择传统机器学习模型的在报告中必须包含数据的详细信息、建模过程和建模结果；选择深度学习模型的团队在报告中必须清楚注明模型内容。

五. 报告提交时间及方式

本次报告仅提交实验报告电子档，电子档文件名命名为全组同学学号.doc 或学号.docx，多人完成的学号用逗号隔开。

（如 **2017211100, 2017211101, 2017211102.docx**）

报告要求在 **19 周周日（6 月 28 日）**前由自然班学习委员收齐后，打包发给实验课任课教师邮箱（任课教师、所属实验班见教务在线课表）。**选修、重修、补修**的同学直接发送到**所属实验教师邮箱**。任课教师邮箱如下：

刘 群：liuqun@cqupt.edu.cn

曾宪华：zengxh@cqupt.edu.cn

戴大伟：daidw@cqupt.edu.cn