**课程编号：**A0801051420

**Python**数据分析实验（大作业）指导书

一、实验（大作业）基本信息

Python 数据分析课程的实验总学时为 32 学时，共安排 4次实验，每次实验8学时，这四次实验均为 Python 数据分析类实验。每次实验的内容对应Python数据分析不同知识点，按照数据采集、数据清洗、数据挖掘和数据可视化分配实验任务。要求总体实验报告撰写一份，报告内容包括实验目的、内容、环境、过程与分析、实验创新点、实验总结，详细内容及格式参考报告模板。

实验成果物及报告提交方法：

指定提交截止日前，同学需将项目工程压缩包和实验报告分别上传到QQ发布的大作业连接。压缩包文件名格式：学号-姓名-班级-大作业-工程压缩包.zip；报告文件名格式：学号-姓名-班级-大作业-实验报告.docx。

二、实验（大作业）目的

本课程大作业旨在融会贯通课上所学的基本数据处理和分析技能，综合训练学生使用Python数据分析工具，针对具体应用领域的数据进行相关分析的能力。通过基础知识的学习及开发技能的锻炼，学生将熟练运用Python中常用的数据处理和数据可视化工具，采用常用设计方法，对数据集进行分析建模，进而实现数据分析模型的编码，通过运行结果完成实际分析意义的说明。

三、实验（大作业）内容

实验1. 数据采集

1.1 确定数据分析方向

根据所学内容尝试目标方向数据分析的意义和价值。根据研究方向收集相关数据，并分析其数据特征。可采用数据可视化方法分析数据的特征。

1.2 明确数据分析方向后，寻找合适网站中的数据集。编写爬虫，爬取相关数据集。

1.3 将爬取到的数据集保存到文件与数据库中，以便之后实验开展分析。

实验2. 数据清洗、预处理

2.1 合并、连接、组合爬取数据集。

2.2 对于无效、错误数据进行清洗。

2.3 对数据集中必要的数据项进行标准化、缩放、归一化、特征编码等处理。

2.4 根据需要使用数据挖掘算法的需要，对数据集进行变换操作。

实验3.数据挖掘

3.1 根据分析需要对数据集主要数据项进行统计分析，如：求和、均值、分组等分析。

3.2 选择合适数据挖掘算法，构建分析模型。

3.3 模型中带入清洗后的训练数据集，对模型进行训练。

3.4 使用测试和验证数据集对模型进行评价，并根据评价结果进一步优化模型。

实验4. 数据可视化

4.1 从多个角度出发选择合适的数据可视化展示方式，对数据分析和数据挖掘的结果进行可视化展示，从中获得有价值的知识。

四、实验（大作业）环境

集成开发工具： PyCharm

操作系统：Windows

五、实验（大作业）评分标准

实验成绩50分，占本门总成绩50%；实验报告10分，占本门总成绩的10%。

1 、实验验收成绩（50分）：

* 1. 基本内容（20分）：

1. 完成数据爬取，并保存，5分；
2. 完成数据简单清洗，应包括缺项数据、重复数据、错误数据的处理，5分；
3. 完成简单的数据分析，应包括多角度单项数据统计分析及数据分组统计分析，5分；
4. 完成简单的数据可视化，应包括线图、直方图、饼图、散点图等展示分析结果，5分.
   1. 分析（20分）：
5. 数据集以多种方式进行存储（如：数据库、JSON文件、CSV、Excel等），5分；
6. 数据集进行合并、连接、组合等操作，5分；
7. 分析挖掘的结果采用多角度、多维度的可视化方式展示，5分；
8. 爬取多个数据集，以支撑数据挖掘的需要，5分。

1.3、 创新（10分）：

A．采用人工智能模型，进行回归分析，至少两个，3分；

B．采用人工智能模型，进行分类分析，至少两个，4分；

C、采用人工智能模型，进行聚类分析，至少两个，3分；

2、实验报告成绩（10分）：

报告内容充实、实验内容记录完整，实验结果分析准确，归纳总结充分，实验报告格式正确，文档提交完整、正确。

六、实验（大作业）成绩评定表

|  |  |  |  |
| --- | --- | --- | --- |
| **考察点** | **具 体 要 求** | **满分** | **得分** |
|
| **基本内容** | **能够按照实验要求合理设计并开发出程序，实验态度认真，能够完整记录实验数据、原理及实验结果。** | **20** |  |
| **分析** | **经过分析，除了基本内容以外，设计并实现额外功能，使系统功能更加完善。** | **20** |  |
| **创新** | **实验设计、结果分析具有创新性。** | **10** |  |
| **报告质量** | **报告内容充实、正确，实验内容记录完整，实验报告格式正确，文档提交完整、正确。** | **10** |  |
| **总计** | | |  |