

Big Data Analysis and Mining

大数据分析 & 挖掘 实验指导书

广东财经大学 大数据与人工智能学院

2025 年 9 月

目 录

实验要求.....	1
实验项目二 Python 基础操作	3

实验要求

一、实验意义和目的

本实验课程配合《大数据分析挖掘》理论课同步开设，其中包括验证型、设计型和综合型实验。本实验课程着眼于理论与应用的结合，注重培养学生的实际动手能力，增加学生对大数据背景下数据挖掘中的算法和技术的了解。通过实验教学，使学生掌握基于 Python 语言的数据分析和挖掘算法在实际中的应用。

本实验课程要求学生完成适当的上机实习，并写出相应的实验报告。验证和设计题单独完成，综合题任选一题。验证型题目使学生熟悉大数据分析挖掘平台的搭建和使用。设计型题目使学生掌握大数据分析挖掘中算法的基本理论以及在实际中的应用，培养基本的应用能力。综合型题目在于提高学生分析问题、解决问题的能力，加强学生对数据挖掘的整体理解。

二、实验内容安排

实验内容安排如下：

实验项目编号	实验性质	实验要求	实验项目名称	学时	备注
1	验证	必修	数据分析与挖掘平台的搭建	2	
2	验证	必修	Python 编程实践（函数，复用，组合类型等）	2	
3	设计	必修	数据探索与数据预处理	2	
4	设计	必修	分类与回归实践	4	
5	设计	必修	聚类算法实践	2	

三、实验过程要求

实验前要充分做好准备工作：

1. 复习和掌握与本实验有关的知识内容；
2. 预习、思考实验内容；
3. 对实验内容进行分析 and 设计。

实验过程中，实验者必须服从指导教师和实验室工作人员的安排，遵守纪律与实验制度，爱护设备及卫生。在指定的实验时间内，必须到实验室内做实验。

对于上机过程中出现的问题，尽量先独立思考和解决；对于难以解决的问题可以和同学交流或询问老师；对于同一个实验题目，可以考虑多种方法来实现，然后比较并选择出一种较为有效的方法来实现。

对于设计型和验证型实验，实验时一人一组，独立上机。

四、实验报告要求

实验后，应及时整理出实验报告，实验报告提交电子文档，实验报告具体内容见附录 1：实验报告模板。

实验项目二 Python 基础操作

一、实验类型

实验类型为验证型，2 个学时，必修。

二、实验目的

- 1、 了解 Python 的基本操作。

三、基础知识

- 1、 Python 开发平台的搭建。
- 2、 Python 使用入门。
- 3、 Python 数据分析工具。
- 4、 Vi/Vim 编辑器的基本操作。
- 5、 pip 软件安装工具的使用。

四、实验环境

- 1、 操作系统：Windows XP 以上操作系统，Linux 操作系统、Unix 操作系统或 Mac 操作系统。

五、实验内容

- 1、 元组和列表

- 给定 `lis = [2, 3, 'k', ['qwe', 20, ['k', ['tt', 3, '1']], 89], 'ab', 'adv']`

- (1) 使用两种方法将列表 `lis` 中的 `'tt'` 改变成大写。
- (2) 使用两种方法将列表中的数字 3 变成字符串 `'100'`。
- (3) 使用两种方法将字符串 `'1'` 变成数字 101。

- 给定 `li = ['taibai ', 'alexC', 'Abc ', 'egon', 'Ritian', 'Wusir', 'aqC']`

(1) 查找列表 `li` 中的元素，移除每个元素的空格。

(2) 找出以‘R’或者‘a’开头，并以‘C’结尾的所有元素，将元素分别添加到一个新的列表中，最后循环打印这个列表。

- 给定一段话“数据挖掘利用了来自如下一些领域的思想：(a) 来自统计学的抽样、估计和假设检验，(b)人工智能、模式识别和机器学习的搜索算法、建模技术和学习理论。数据挖掘也迅速地接纳了来自其他领域的思想，这些领域包括最优化、进化计算、信息论、信号处理、可视化和信息检索。一些其他领域也起到重要的支撑作用。特别地，需要数据库系统提供有效的存储、索引和查询处理支持。源于高性能（并行）计算的技术在处理海量数据集方面常常是重要的。分布式技术也能帮助处理海量数据，并且当数据不能集中到一起处理时更是至关重要。”，将这段话中的“数据挖掘”替换成“**data mining**”。

2、字典

- 编写程序，输入两个分别包含若干整数的列表 `lstA` 和 `lstB`，输出一个字典，要求使用列表 `lstA` 中的元素作为键，列表 `lstB` 中的元素作为值，并且最终字典中的元素数量取决于 `lstA` 和 `lstB` 中元素最少的列表的数量。

3、循环

- 使用 **Python** 实现乘法表。
- 编写程序，输入任意大的自然数，输出各位数字之和。
- 编写程序，输入包含若干集合的列表，输出这些集合的并集。要求使用 `reduce()`函数和 `lambda` 表达式完成。

4、 字符串,print 格式化

- 获得输入的一个字符串 s,以字符减号(-)分割 s,将其中首尾两段用加号(+)组合后输出。

例如：输入 Alice-Tom-Eric-Lisa 输出 Alice+Lisa

- 获得用户输入的一个正整数 a，计算 a 的平方根，保留小数点后 3 位，并打印输出。输出结果采用宽度 30 个字符、右对齐输出、多余字符采用加号(+)填充。如果结果超过 30 个字符，则以实际宽度为准。

六、实验要求

1、撰写实验报告（模板参照附录 1），要求报告美观，可读性好。内容包括：

（1）实验目的：要求实验目的明确。

（2）实验环境：要求说明实验的软、硬件配置环境。

（3）实验原理：要求能理解实验的基本原理。

（4）实验步骤和结果：要求实验步骤正确、完整；实验结果清晰、明确。

（5）实验分析和结论：包括实验结果评价，实验中遇到的主要问题的分析与处理，要求评价合理，问题描述清楚具体，分析透彻，处理正确。

2、实验报告交给组长，10 月 9 号前提交。