

**实验(实训)课程报告**

**课程 数据挖掘技术实践**

|  |  |  |  |
| --- | --- | --- | --- |
| 院　　别 | 信息技术与管理 | 专 业 | 计算机科学与技术 |
| 届 别 | 2022 | 班 级 | 计算机科学与技术（职高）一班 |
| 学生姓名 | 邱宇杰 | 学 号 | 202205650121 |
| 指导教师 | 钟敏娟 | 职 称 |  |

湖南财政经济学院

2024年 8 月

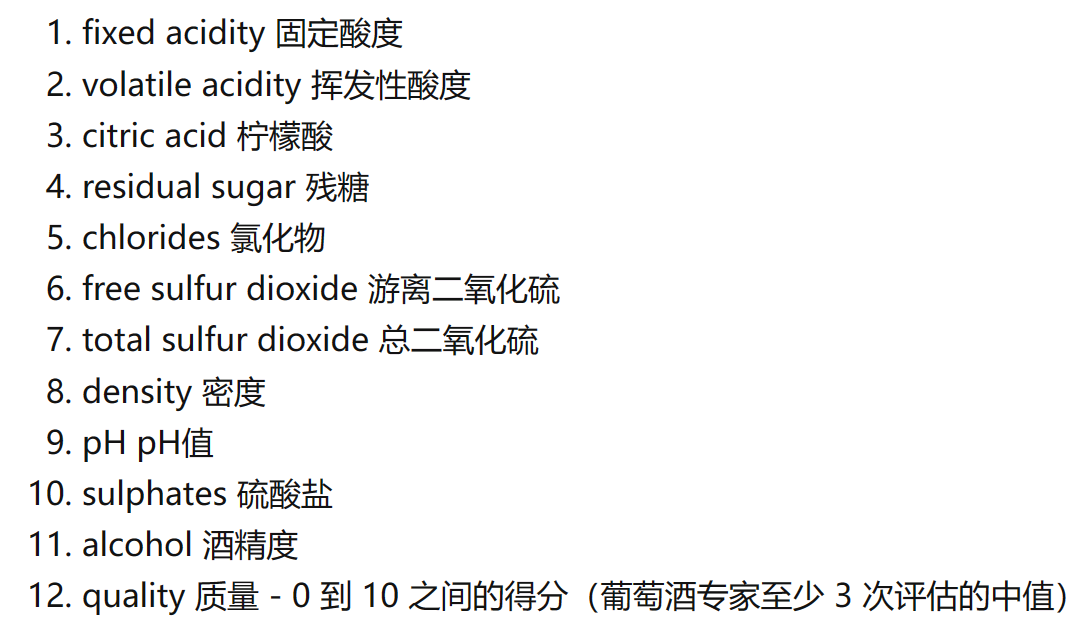
**实验(实训)报告1**

实验(实训)时间: 2024 年 10 月 18 日 指导教师评分：

|  |  |
| --- | --- |
| 实验项目 | 实验一：认识数据—红酒品质数据的可视化 |
| 组员名单 | 独立完成或姓名（学号）（操作性和验证性实验均独立完成，综合性实验1-2人一组） |
| 实验类型 | ■基础性实验 □验证性实验 □综合性实验 |
| 实验(实训)时间 | 2024年 月 日 |

1. 实验目的和要求

今有来自UCI机器学习实验室的葡萄酒质量数据集，分别提供了有关葡萄牙“青酒”的红葡萄酒和白葡萄酒变体的样本信息，每种样本都由专家做了质量评级，并进行了理化指标检验。包含以下12个字段：



要求以红葡萄酒(winequality-red.csv)为对象，对其中数据进行如下初步处理：

(1) 以分号为分隔符进行数据的读入。

(2) 初识数据，给出数据的基本信息，包括字段名称、类型、字段个数、字段是否缺失以及数值数据的统计参数值(均值、个数、最大、最小、标准差等)。

(3) 利用subplot绘制所有字段变量的箱线图、直方图、累计曲线图；绘制变量百分比饼图；分别绘制带注释和对角线矩阵的热力图，并观察各个字段的分布特点。

2．实验步骤（记录程序代码、分析实验结果）

1. 心得体会与建议