卡通人物

低可信度描述已自动生成

**数据分析与可视化**

**实**

**验**

**指**

**导**

**书**

**田俐**

**智能工程系**

2025年4月

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **课程名称** | | **数据分析与可视化** | | | **实验时间** | | **年 月 日** | |
| **班级** |  | | **姓名** |  | | **学号** | |  |

# 实验一 基于大数据的房产估价

### 一、实验目的与要求

1. 掌握使用numpy和pandas库处理数据的基本方法。
2. 掌握使用Sklearn库对多元线性回归算法的实现及其评价方法。
3. 掌握使用matplotlib结合pandas库对数据分析可视化处理的基本方法。

### 二、实验内容

1. 利用python中pandas等库完成对数据的预处理，最后将处理好的文件进行保存。
2. 利用pandas、matplotlib等库完成对预处理数据的可视化。
3. 结合pandas、matplotlib库对聚类完成的结果进行可视化处理。

### 三、实验步骤

1.数据预处理

2、数据分析

3、数据可视化

|  |  |
| --- | --- |
| **评价项目** | **所得分数** |
| 实验报告书写完整 | □3分 □2分 □1分 其它： 分 |
| 实验代码正确 | □4分 □3分 □2分 其它： 分 |
| 实验结果正确 | □3分 □2分 □1分 其它： 分 |
| **总分：** 分  **日期**： 年 月 日 **签名：** | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **课程名称** | | **数据分析与可视化** | | | **实验时间** | | **年 月 日** | |
| **班级** |  | | **姓名** |  | | **学号** | |  |

# 实验二 客户价值分析

### 一、实验目的与要求

1. 掌握使用numpy和pandas库处理数据的基本方法。
2. 掌握使用RFM分析模型对客户信息进行特征提取的基本方法。
3. 掌握对特征数据进行标准化处理的基本方法。
4. 掌握使用Sklearn库对K-Means聚类算法的实现及其评价方法。
5. 掌握使用matplotlib结合pandas库对数据分析可视化处理的基本方法。

### 二、实验内容

1. 利用python中pandas等库完成对数据的预处理，并计算R、F、M等3个特征指标，最后将处理好的文件进行保存。
2. 利用python中pandas等库完成对数据的标准化处理。
3. 利用Sklearn库和RFM分析方法建立聚类模型，完成对客户价值的聚类分析，并对巨累结果进行评价。
4. 结合pandas、matplotlib库对聚类完成的结果进行可视化处理。

### 三、实验步骤

1、数据预处理。

2、数据分析

3、数据可视化

4、分析评价

|  |  |
| --- | --- |
| **评价项目** | **所得分数** |
| 实验报告书写完整 | □3分 □2分 □1分 其它： 分 |
| 实验代码正确 | □4分 □3分 □2分 其它： 分 |
| 实验结果正确 | □3分 □2分 □1分 其它： 分 |
| **总分：** 分  **日期**： 年 月 日 **签名：** | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **课程名称** | | **数据分析与可视化** | | | **实验时间** | | **年 月 日** | |
| **班级** |  | | **姓名** |  | | **学号** | |  |

# 实验三 智能电网的电能预估及价值分析

### 一、实验目的与要求

1. 掌握使用pandas库处理数据的基本方法。
2. 掌握对时间序列类数据预处理的基本方法。
3. 掌握使用matplotlib结合pandas库对数据分析可视化处理的基本方法。

### 二、实验内容

1. 利用python中pandas等库读取数据，并完成数据的预处理。
2. 利用matplotlib等库完成对数据的可视化。
3. 使用Sklearn库的相关系数建立决策树模型，对模型进行训练，使用测试集测试后对模型的效果进行评价。

### 三、实验步骤

1.数据预处理。打开Jupyter Notebook工具，读取实验目录中所提供的“data3.xlsx”文件，检查文件中时间序列是否完整，有无缺失值，重复值，若在序列中存在缺失值使用前后数据进行拉格朗日插值处理，并新增一列保存日类型属性。

2、数据可视化。将第一步处理预处理好的数据进行可视化，绘制各气象信息的时间序列曲线，及电价和电力负荷的时间序列曲线。

3、相关系数。求出各量与电力负荷之间的相关系数，选择相关系数绝对值前3高的属性作为特征属性，用于下一步进行模型训练。

4、数据分析。使用上一步选择的3个特征属性作为输入属性，电力负荷作为输出属性，合理划分训练集与测试集比例，选择适合的超参数使用Sklearn建立决策树模型，并对模型在测试集上的表现做出评价。

5、模型评价

|  |  |
| --- | --- |
| **评价项目** | **所得分数** |
| 实验报告书写完整 | □3分 □2分 □1分 其它： 分 |
| 实验代码正确 | □4分 □3分 □2分 其它： 分 |
| 实验结果正确 | □3分 □2分 □1分 其它： 分 |
| **总分：** 分  **日期**： 年 月 日 **签名：** | |