



**数据挖掘**

**上机实验报告**

|  |  |
| --- | --- |
| 专 业： | 人工智能 |
| 班 级： | 192204 |
| 学 号： | 19220432 |
| 姓 名： | 陆昊宇 |

2024年09月23日

## 一、实验目的

1. 掌握数据挖掘中混合属性的邻近性度量的计算方式。

## 二、实验内容

#### 1. 泰坦尼克号数据集乘客幸存分析实验

（1）分析泰坦尼克号数据集，说明每个属性的类别；

（2）将幸存属性设置为类别标签，使用K近邻算法作为分类算法（无需考虑PassengerID，name，ticket和cabin，去掉带缺失属性的样本），在测试集上预测乘客的幸存情况并计算准确率。

**注1**：数据对象之间的邻近性度量采用混合类型属性的计算方式。

**注2**：从训练集中以合适的比例划分出验证集，利用交叉验证选择合适的K值。

**提交入口：**

<https://send2me.cn/CYaCbSKw/TsCTnNJTxR-KWw>

## 三、实验报告评分标准

1. 完成实验（1）（10分）

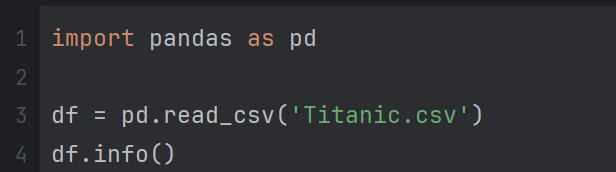
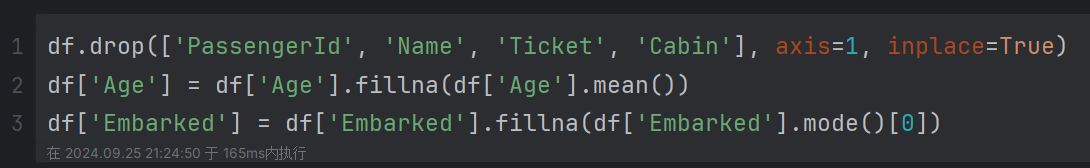
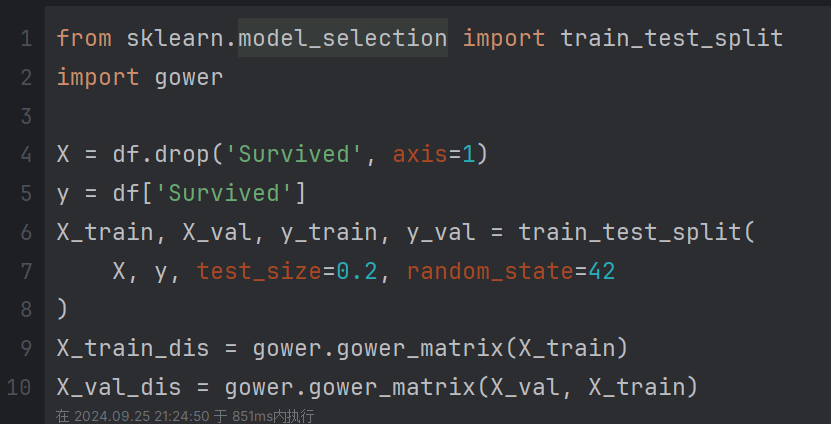
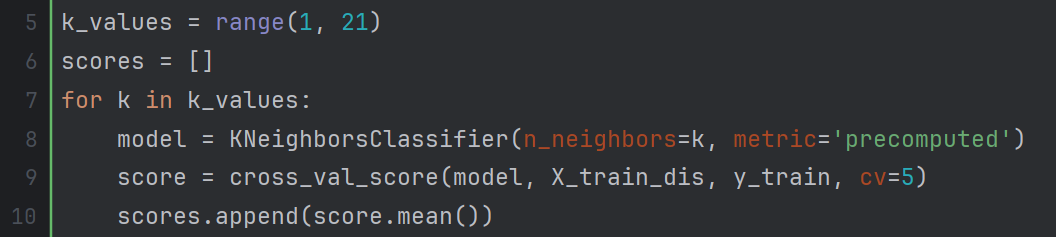
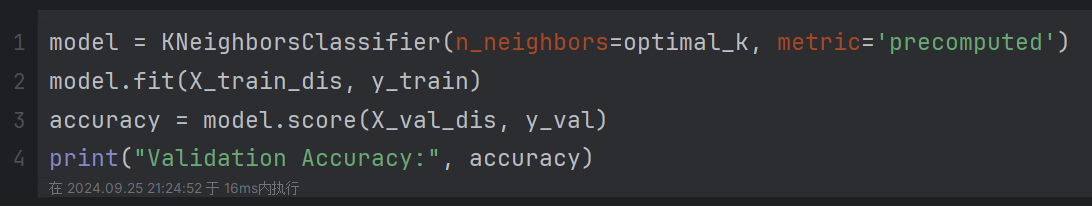
2. 完成实验（2）（60分）

3. 实验过程与实验结论的记录完整（10分）

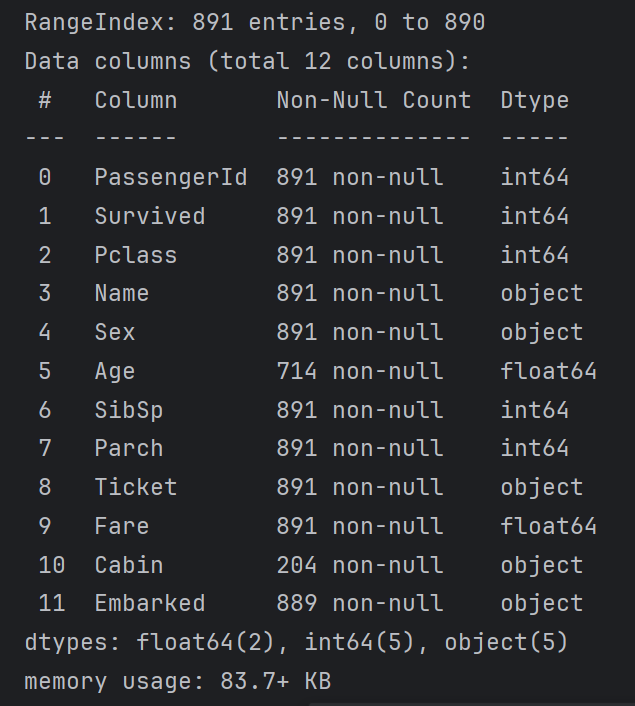
4. 表述逻辑清晰（10分）

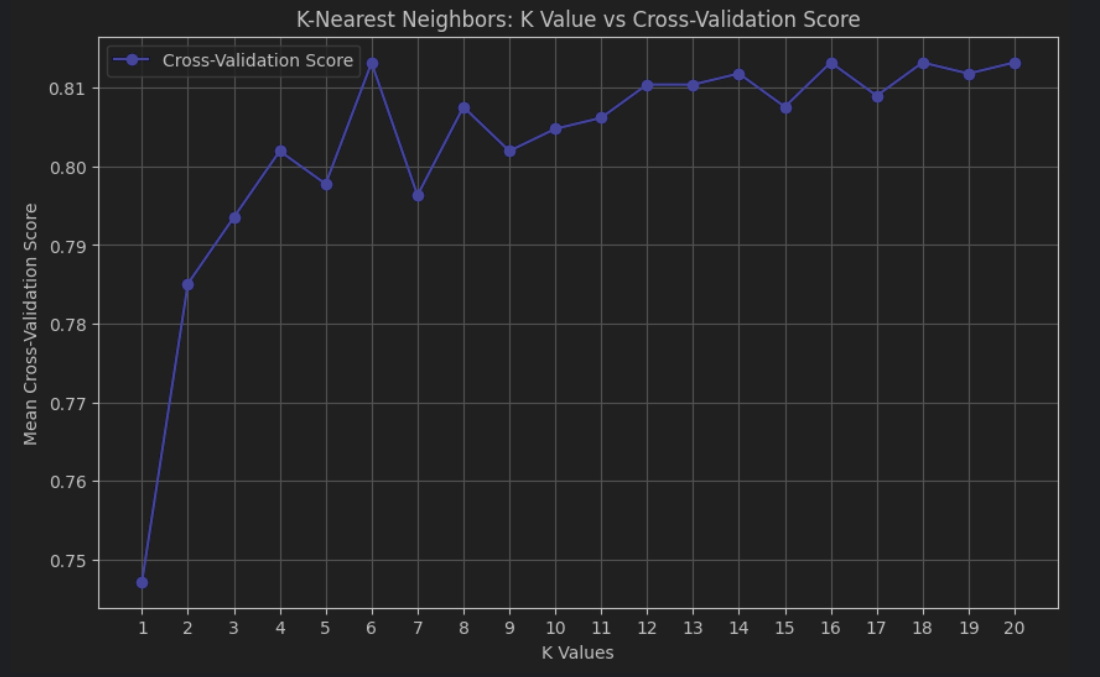
5. 排版工整自洽，图表标题准确（10分）

## 实验过程

1. 数据加载与分析  
   
2. 数据预处理  
   
3. 划分数据集  
   
4. 选择 K 值并训练模型  
   
5. 模型训练与评估  
   

## 实验结论

1. 数据加载与分析  
   
2. 选择 K 值并训练模型



1. 模型训练与评估  
   Validation Accuracy: 0.80