# 软件设计说明书

1. 软件名称

基于决策树的葡萄酒的质量预测

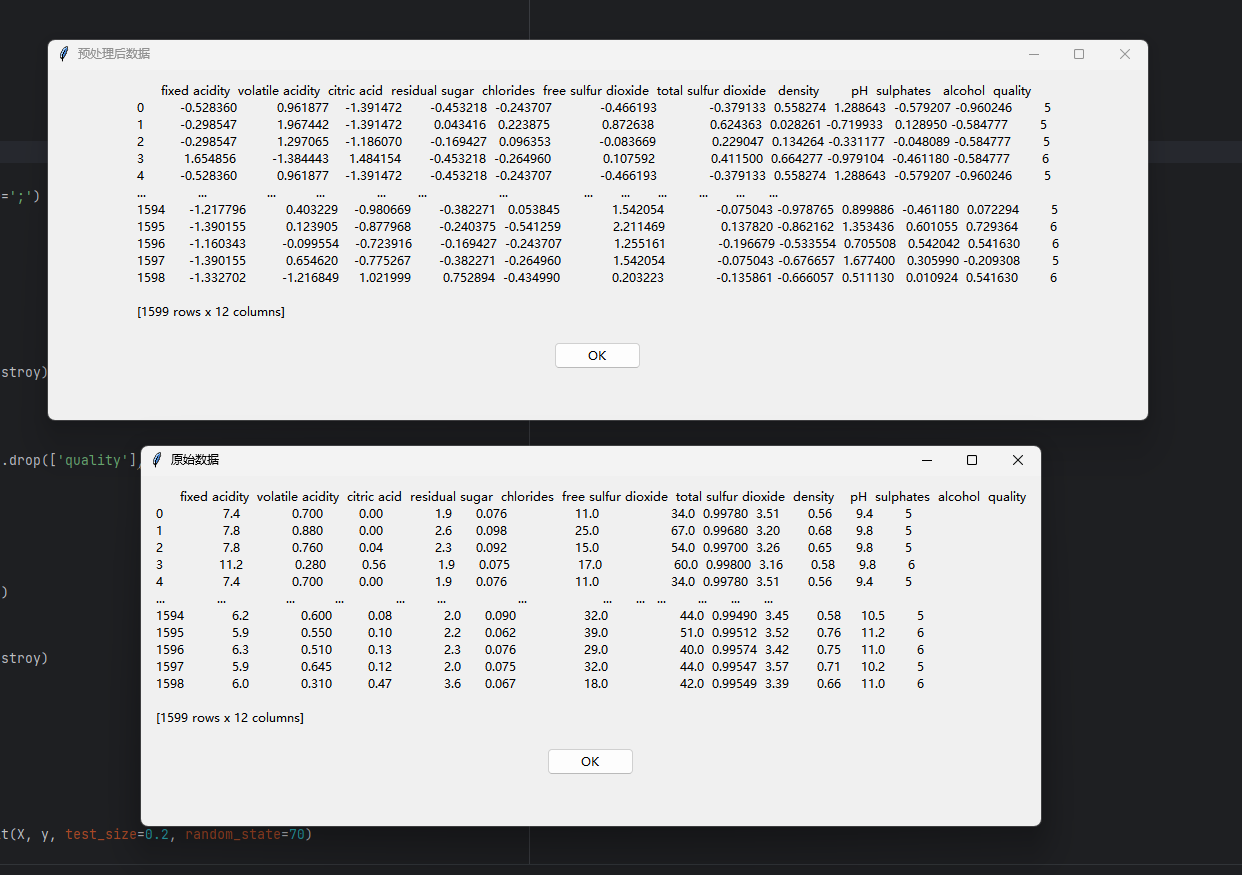
2. 软件功能介绍

这个软件是一个基于Python的Tkinter GUI应用，旨在为用户提供一个简单易用的界面来处理葡萄酒质量数据集，训练模型，并进行质量预测和模型评估。

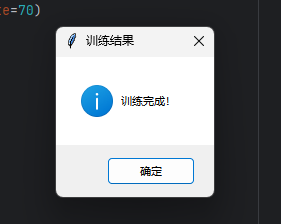
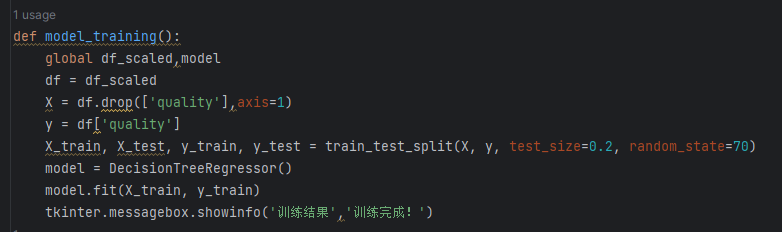
软件主要包含以下四个功能：

数据预处理： 通过点击"数据预处理"按钮，软件会自动读取葡萄酒红酒数据集，对除了需要预测的质量列以外的所有列进行标准化，目的是为了更好地训练决策树模型。标准化后，屏幕会显示处理前和处理后的数据表格。



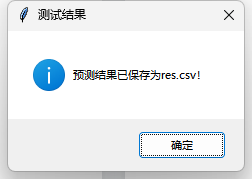


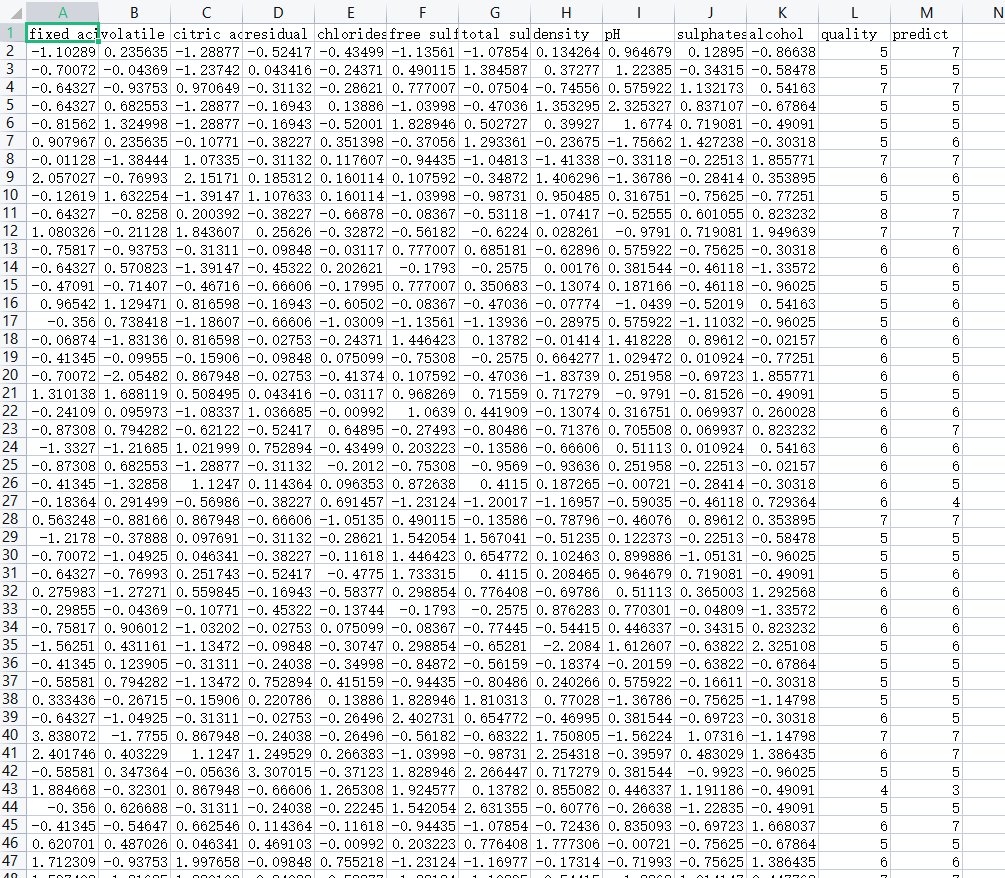
模型训练： 通过点击"模型训练"按钮，软件会使用预处理过的数据，采用决策树算法进行模型训练。



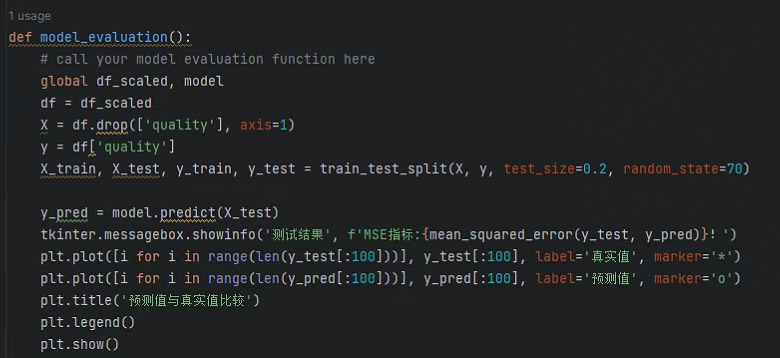
模型预测： 通过点击"模型预测"按钮，软件将对测试集数据进行预测，并将预测结果存储为excel文件，方便用户查看以及用于后续模型评估。

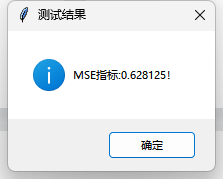


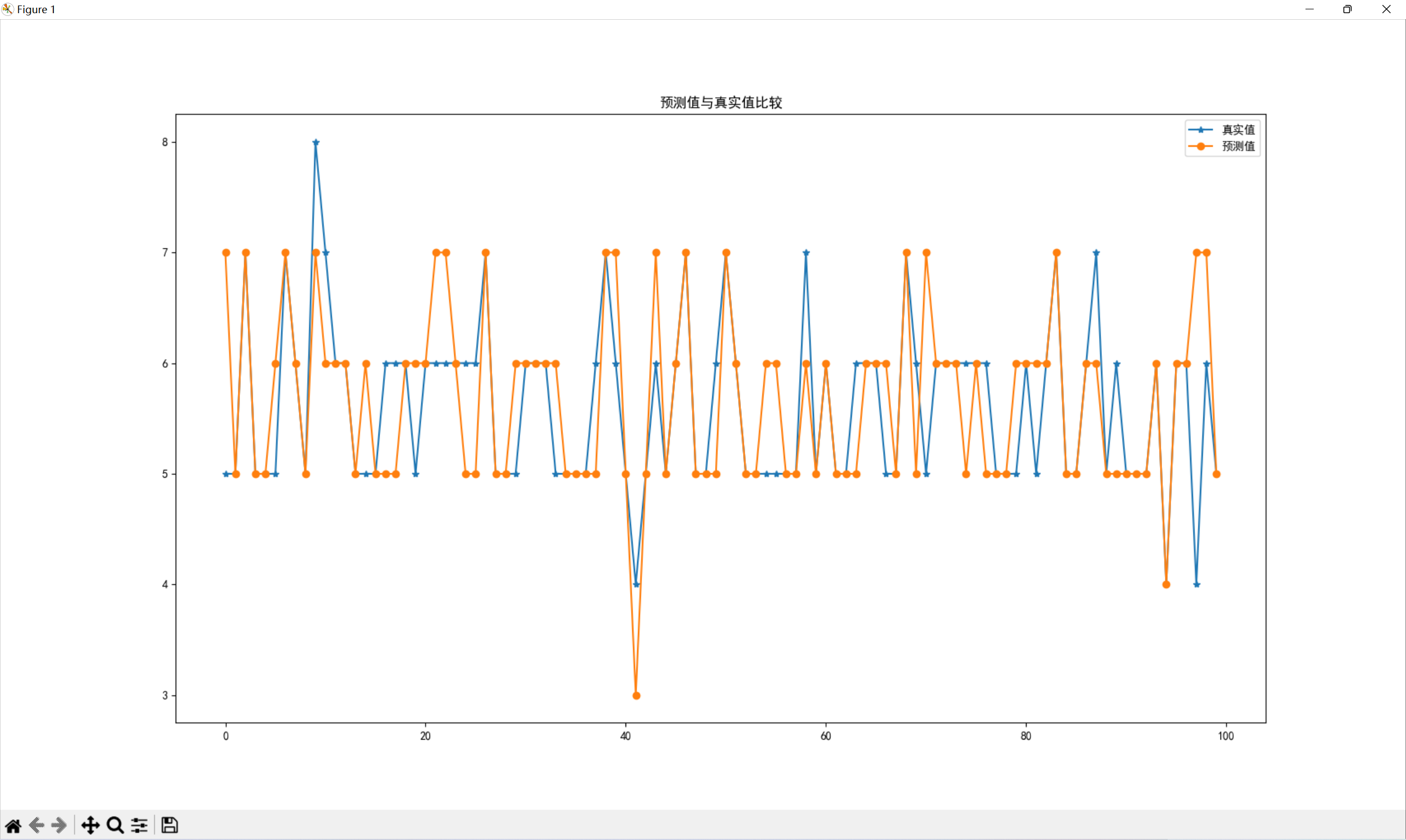




模型评估： 通过点击"模型评估"按钮，软件会计算模型的预测准确率，并显示在弹出的消息框中。之后会抽抽取前100个测试样本进行真实值和预测值的可视化对比。







3. 软件使用说明

下载和启动软件： 下载提供的Python脚本，然后在Python环境中运行脚本，将打开软件的GUI界面。

执行数据预处理： 首先点击"数据预处理"按钮，开始处理数据。处理完毕后，会弹出一个消息框，提示"数据预处理完成"。

进行模型训练： 数据预处理完成后，点击"模型训练"按钮，开始训练模型。训练完毕后，会弹出一个消息框，提示"模型训练完成"。

进行模型预测： 模型训练完成后，点击"模型预测"按钮，开始对测试集数据进行预测。预测完成后，会弹出一个消息框，提示"模型预测完成"。

进行模型评估： 完成模型预测后，点击"模型评估"按钮，开始计算模型的预测准确率。完成计算后，会弹出一个消息框，显示模型的准确率。

1. 算法原理

决策树原理：

决策树（Decision Tree）是一种常见的机器学习方法，用于分类和回归问题。决策树的思想来自于人类的决策过程。我们在做决策时，通常会根据各种情况和可能的结果来逐步缩小选择的范围，直到得出决策。决策树的每个节点都表示一个特征（或属性），每条边代表一个决策规则，叶子节点表示一个决策结果。

决策树的构建过程是一个递归过程：

(1)选择最佳特征进行分割： 在每个节点，选择最好的特征来分割数据。选择哪个特征作为最佳特征取决于各个特征对于目标变量的影响程度。这个度量标准可以是信息增益（在ID3算法中），信息增益比（在C4.5算法中），或者基尼指数（在CART算法中）。

(2)创建决策节点： 创建一个决策节点，将选中的最佳特征放在该节点。

(3)分割数据集： 根据选中的特征和特征的取值，将数据集划分为几个子集。

(4)构建子树： 对每个子集递归地重复上述步骤，直到所有特征已经用完，或者数据集在当前特征下已经没有更多的分割可能（也就是说，该子集中的所有实例都属于同一个类别，或者所有实例在所有未使用的特征上都有相同的取值），此时创建一个叶节点，将当前子集中出现最多的类别标记为该节点的类别。

MSE原理：

MSE是均方误差（Mean Squared Error）的缩写，是一种常见的用于衡量回归模型预测精度的指标。具体来说，它计算的是预测值和真实值之间差值的平方的平均值。MSE越小，说明模型的预测精度越高。特别地，如果MSE为0，那么模型的预测结果完全正确，没有任何误差。