我们在最初时考虑了使用决策树来进行分类。决策树是一种很简单的算法，他的解释性强，也符合人类的直观思维。这是一种基于if-then-else规则的有监督学习算法，是一种分层树形结构，由根节点、分支、内部节点和叶节点组成。决策树拥有以下的优点和缺点：

优点：

1.易于理解和解释，甚至比线性回归更直观；

2.与人类做决策思考的思维习惯契合；

3.模型可以通过树的形式进行可视化展示；

4.可以直接处理非数值型数据，不需要进行哑变量的转化，甚至可以直接处理含缺失值的数据；

缺点：

1.处理连续变量不好；

2.不好处理变量之间存在许多错综复杂的关系，如金融数据分析；

3.决定分类的因素取决于更多变量的复杂组合时可规模性一般。

我们人为规定了一些决策树进行分支决策的标准，根据该标准，我们制作了如下的决策树，它展示了我们可视化后的部分决策树，同时，为了衡量划分的纯度，我们引入使用gini不纯度来衡量我们分类结果的合理性。

然而决策树分类并不擅长处理具有多个属性的数据。我们划分了训练集和测试集，并统计了测试的结果，最终我们得到的数据如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 准确率 | 召回率 | 精确率 | F1 |
| 训练集 | 0.931 | 0.931 | 0.933 | 0.931 |
| 测试集 | 0.72 | 0.72 | 0.734 | 0.726 |

通过与随机森林作对比，我们很明显地发现决策树在处理此数据集时得到的结果并不理想，低于随机森林的准确率(80.4%)、召回率(80.4%)、精确率(87.4%)以及F1(83.7%)。因此我们最终选择了表现更好的随机森林模型来进行分类。

