1. 首先创建九个类别的类别标签
2. 然后查看不同类别间的相似度
3. 把数据集划分为训练集和验证集
4. 采用随机森林模型进行预测