说明：

1.实验使用了更为原始的 one-hot词向量 表示，准确率可能较低，与其他先进词向量有所差距。

2.请将所需的数据文件：stopwords.txt、train\_ultimate.xlsx、和公共测试集.xlsx（代码中命名为test.xlsx）都放在当前目录下，代码才能正常运行。

代码：

L7 – L27 是数据预处理代码，包含了划分句子和筛去停用词。

L29 – L39 是得到ont-hot词向量的代码。

L42 – L80是文本向量化的代码。

L98 – L102是决策树部分代码。

本实验将词向量作为特征向量进行实验，即每个词都是一个特征。

本实验使用 准确率 和 自定义规则的RMSE 2部分评价，具体数值运行完成代码即可看到。

对公共测试集预测的输出在out.txt中，运行代码后即生成。