本周在上周的学习基础上，进一步的学习了LDA主题模型的基本建立方法，通过之前处理过的文本评论数据建立了词典和语料库，并且学习了CountVectorizer类的基本使用方法。得知其是属于常见的特征数值计算类，是一个文本特征提取方法。对于每一个训练文本，它只考虑每种词汇在该训练文本中出现的频率。它能够将文本中的词语转换为词频矩阵，它可以通过fit\_transform函数计算各个词语出现的次数。

相关代码截图如下：



