**正负情感词词云&情感预测模型总结**

1. 情感词词云

通过NLP模型对好评和差评中句子的整体情感打分，并筛选出情感分数不为0的句子进行词语拆分和词语情感判断，得出了正向情感词词云和负向情感词词云。

1. 正向情感



正向情感词中出现频率较高的5个为：非常、快、效果、满意和服务。可以得出正向情感的评论（好评）的重点效率、产品功能好、以及服务质量高。

1. 负向情感

文本

描述已自动生成

负向情感词中出现频率较高的五个为：安装、师傅、空调、质量和制冷。可以判断出该空调产品的负向情感评论（差评）多是关于安装问题和产品功能问题。

1. 情感预测模型
2. LinearSVC模型

通过Tfidf在好评和差评中提取词向量，再拟合LinearSVC模型。使用80%的数据作为训练集，剩余20%作为测试集。得到的模型评估参数为：

1. 训练集分数：0.977
2. 测试机分数：0.979
3. 准确率：0.977
4. 召回率：0.975
5. F1值：0.986
6. ROC值：0.981

总体来说，模型质量很好，各项参数都达到了0.97以上。

1. LDA主题模型

图表, 折线图

描述已自动生成

通过拟合LDA模型并可视化不同主题数目得到的一致性，主题数最优为11。建立主题数为11的LDA模型。

图表, 气泡图

描述已自动生成

第一主题占比大于10%（约为15%），其中“非常”出现的频率最高，“安装”其次。可以看出买家很关心该空调产品如何安装。“满意”和“不错”的占比也较高，该产品的正向评价多。