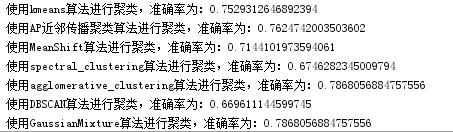
**实验报告**

1. 实验内容

测试sklearn中以下聚类算法在tweets数据集上的聚类效果:

1. kmeans算法
2. AP近邻传播聚类算法
3. MeanShift算法
4. spectral\_clustering算法
5. agglomerative\_clustering算法
6. DBSCAN算法
7. GaussianMixture算法
8. 实验步骤
9. 对tweets数据集进行预处理：分词及形成字典
10. 使用sklearn中各种聚类算法函数处理数据
11. 实验结果



1. 总结

1、聚类准确率不高，需要继续调整参数以及优化程序来提高分类性能。

2、不同的算法对数据集的聚类效果不同，从tweets数据集上的聚类效果看：agglomerative\_clusteri算法、GaussianMixture算法效果较好，DBSCAN算法较差