# 实验报告

201844911 杨子玉

**1.相关资料**：[https://scikit-learn.org/stable/modules/clustering.html#](https://scikit-learn.org/stable/modules/clustering.html)

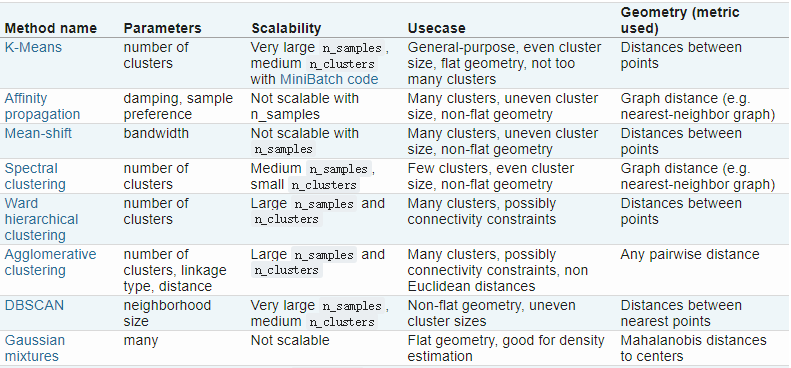
**实验任务：**测试sklearn中以下聚类算法在tweets数据集上的聚类效果。

使用NMI(Normalized Mutual Information)作为评价指标。

**2.相关方法：**

**scikit-learn**简称sklearn，支持包括分类、回归、降维和聚类四大机器学习算法。还包含了特征提取、数据处理和模型评估三大模块。

此次作业主要使用以下几种聚类方法：



**3.处理文本数据集:**

1)将实验数据的文本和应属于的类别放入两个向量中

2)调用库函数计算每个文本的tf-idf值

**4.进行聚类：**

调用函数聚类，同时采用NMI(Normalized Mutual Information) 标准化互信息 评价效果

**5.实验结果如下图所示**

# 可以看到大多集中在0.7左右范围，AffinityPropagation的效果最好。

