# Cluster

实验目的：

使用Tweets数据集测试各种聚类算法。

实验步骤：

(1)首先对文本进行处理并将处理的文字和标签分别存入字典中，将其向量化；

(2)统计类的总数量；

(3)调用几种不同的聚类算法；

(4)测试聚类后的效果。

实验结果：

KMeans： 0.721292248386

Affinity： 0.734286617625

Spectral： 0.689574578741

Agglomerative：0.757178787482

Gaussian：0.683066495726

MeanShift：0.690981587324

Dbscan：0.738217267295

实验心得：

第一次使用scikit-learn，虽然有些函数不太了解，过程中也会出现一系列的错误，不过scikit-learn对于聚类确实是很好用，学到了很多新的东西，收获了很多，以后还会加强练习，有更新的认知。