

NBC

- 1、获取所有的文本类别
- 2、处理数据集数据，获取每一个文本数据及对应的类别标签
- 3、按照 VSM 的方法得到文本的 TFIDF 值
- 4、调用 sklearn 库中的 GaussianNB 方法获取文本的朴素贝叶斯的高斯模型（训练集过大，可用 partial_fit 分批训练模型）
- 5、输入测试集，进行评估获取预测