

# 实验三

---

## 数据集介绍和地址

[https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/weibo\\_senti\\_100k/intro.ipynb](https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/weibo_senti_100k/intro.ipynb)

## 任务介绍和要求

- 1.分析文本是正面或者是负面
- 2.训练集、验证集、测试集自行划分
- 3.可以使用开源工具进行分词等操作，词向量也采用预训练好的开源资源
- 4.可以调用现成的时序模型

## 说明

可以参考开源资源，但是不可以互相抄袭

## 实验结果提交

- 实验报告(PDF): 需要包含实验过程, 数据处理等
- 实验代码(py文件)
- 以上结果打包为压缩包, 发至[dl2021\\_bigdata@163.com](mailto:dl2021_bigdata@163.com)
- 压缩包命名格式请按: 学号\_姓名\_实验三, 截止日期为11.29晚23:59

——处理数据集的时候大概有1500+条相同的review的label 同时为0和1, 请问是把这样的review当作中性情感来处理吗?

——这类数据是去除掉的, 简单起见, 还是做二分类问题