

# Sentiment\_Analysis

---

## Introduction

---

This part contains seven folders: python, stopwords, dataset, result, model, ipynb, script.

Python contains the train and test code of two functions which the files start with 'Analysis' are the codes that use NLTK Naive Bayes Classifier to train our dataset and predict. The files start with 'NaiveBayes' are the codes contain the classifier that was accomplished by ourselves, only use the NLTK package to tokenization, and stemming.

Stopwords contains the stopwords that need to be removed.

Dataset contains the tweets that we collect.

Result contains the file that each tweet in the dataset has been labeled and it also has the probability of the sentiment.

Model contains the two functions model. One is the model named 'NaiveBayesModel.pickle' which was trained by the NLTK package. The other texts named 'pos\_words.txt' and 'neg\_words.txt' are the datasets contain the positive words and the negative words that used in our naive bayes classifier.

Ipynb contains the model evaluations, it also has two file of these two model. The evaluations are the precision, recall, f1, accuracy.

Script contains the script that can predict each tweets in the dataset folder.

## Run Script

---

To run the script, what need to satisfy is that the environment need the python library: sys, csv, nltk, string, re, pickle, math, codecs, pickle, numpy.

Here four script: 'Analysis\_Test.sh', 'Analysis\_Train.sh', 'NaiveBayes\_Test.sh', 'NaiveBayes\_Train.sh'. For this time, we only use the 'data.csv' stored in the dataset folder. All the script are no need to input any parameter. Just run like './Analysis\_Test.sh'.

The Analysis\_Test.sh predict the 'data.csv' dataset with NLTK classifier.

The Analysis\_Train.sh use the Twitter\_Sample to train the NLTK classifier.

The NaiveBayes\_Test.sh predict the 'data.csv' dataset with the classifier accomplished by ourselves.

The NaiveBayes\_Train.sh use the Twitter\_Sample to extract the positive words and the negative words.