

Sentiment Recognition in Tweets

Social media monitoring can relate to either gauging thoughts about current events, often known as sentiment analysis, or detecting emotions in created material. The phrase sentiment analysis refers to the automated method of recognising and categorizing text based on the views expressed in it. The technique focuses solely on the examination of the attitude of the individual who gave the material, which is classified as positive, negative, or neutral.

Requirements

Below are a few libraries that are needed to be installed priorly. The general requirements are as follows.

- Numpy
- Scikit-learn
- Scipy
- nltk
- Sklearn with the support of Gaussian Naive Bayes, Logistic Regression, Random Forest
- Pycharm IDE for the development.

Preprocessing and Classification

- Run senti140preprocessor.py to reform sentiment140 data
- Run preprocessor.py to convert test and train files into train-processed and test-processed file
- Run stat.py to create frequency distributions using the processed train and test files.
- Using preprocessed datasets run the files naivebayes.py, randomforest.py, logisticdefault.py to train the models and generate the validation accuracy of each models which are saved in the savedModels folder and can be used to check the testing accuracy
- Run comparator.py file to check the testing accuracy of each stand-alone classifier and compare that with the ensemble voting classifier.

Other files

- positive-words.txt: Directory for positive words.
- negative-words.txt: Directory negative words.
- Dataset: Sentiment140