SENTIMENTAL ANALYSIS

Sentiment analysis, also refers as opinion mining, is a sub machine learning task where we want to determine which is the general sentiment of a given statement. Using python libraries we can extract the information and using that information we can help identify the sentiment of the statement.

In this project I choose to try to classify tweets from Twitter into “positive” or “negative” sentiment by building a model based on probabilities. Because of the usage of Twitter, it is a perfect source of data to determine the current overall opinion about anything.

Scope :

* We can use this project to analyse the stock market, as by taking data from social media sites and analyse the overall opinion of a product.
* This project can analyse the hot topics for any specific statement from the social media websites.
* We can predict the election result by analysing the review through social media.

Requirement:

* Python
* Python libraries – Tkinter, NLTK

Installation :

* (For those who don’t have python installed) First Install Python from this link : <https://www.python.org/downloads/>

And download the python and install it.

* Now setup the environment variable for python.

Path : “C:\Users\Anubhav\AppData\Local\Programs\Python\Python37\python.exe”(Note : Path can be different, so please check it before using).

Now add this path to environment variable

* Now we have to install pip for installation of python packages , for installing pip : <https://bootstrap.pypa.io/get-pip.py> ,then save the file and go to downloaded directory and execute the file.

For executing pip file -> type in command Prompt : “python get-pip.py” (Type the command between “ “ ).

* There are some command we should execute before running project :

pip install nltk

python -m nltk.downloader twitter\_samples

python -m nltk.downloader averaged\_perceptron\_tagger

python -m nltk.downloader punkt

python -m nltk.downloader wordnet

python -m nltk.downloader stopwords

python -m tkinter

* Now we are ready to execute project.

Working:

Let’s start the working of this project

* In this project the project import the NLTK corpus( i.e. dictionary, that contains the data through which the program will train itself).
* Then, we extract data from NLTK and assign positive tweets to positive\_tweet\_tokens and negative tweets to negative\_tweet\_tokens.
* After that, we remove noise ( i.e. @,http:// ,and the words which have different forms but same verbs like “run” , “runs” and “running” ,etc.)
* Then, we use the NaiveBayesClassifier class to build the model. Use the .train() method to train the model .
* After training all the data(Note : Training process will take time around 15sec ), then Tkinter Interface will start



* In this we have to write a tweet and the program will tell us the about the sentiment of the tweet.

Ex -:

Positive Tweet :



Negative Tweet :



Conclusion :

Nowadays, sentiment analysis or opinion mining is a hot topic in machine learning. We are still far to detect the sentiments of s corpus of texts very accurately because of the complexity in the English language and even more if we consider other languages such as Chinese.

In this project we tried to show the basic way of classifying tweets into positive or negative category using Naive Bayes as baseline and how language models are related to the Naive Bayes and can produce better results. We could further improve our classifier by trying to extract more features from the tweets, trying different kinds of features, tuning the parameters of the naive Bayes classifier, or trying another classifier all together.