GraphicEra
Deemed to be University
Accredited by NAAC with Grade A
NBA Accredited Program in CSE, ECE & ME
Approved by AICTE, Ministry of HRD, Govt. of India

MINI-PROJECT REPORT

TOPIC: "SENTIMENT ANALYSIS ON TWITTER DATA"

Shivanshu Tiwari

2013507

BTech CSE | 6th Semester

Roll No. 56 | (Sec: C)

GitHub: https://github.com/ShivanshuTiwari-github/Sentiment-Analysis-of-Twitter-data-using-Python

ABSTRACT

In this report, we address the problem of Sentiment analysis in social media. We use a python language and methods to perform the Sentiment analysis on twitter to classify tweets as positive, negative or neutral.

PROBLEM STATEMENT

SR NO 16. Sentiment Analysis on Twitter data:

Sentiment analysis is a kind of data mining where you measure the inclination of people's opinions by using NLP, text analysis, and computational linguistics. We perform sentiment analysis mostly on public reviews, social media platforms, and similar sites.

INTRODUCTION

Social media is a very popular way for people to express their opinions publicly and to interact with others online. In aggregation, social media can provide a reflection of public sentiment on various events. Unfortunately, many users engaging online, either on social media, forums or blogs, will often have the risk of being targeted or harassed via abusive language, which may severely impact their online experience and the community in general.

Twitter is a popular social networking website where members create and interact with messages known as "tweets". This serves as a mean for individuals to express their thoughts or feelings about different subjects.

PREREQUISITES

- Tweepy, the official Python library for accessing the Twitter API
- TextBlob, a Python library for processing textual data
- NLTK dataset, to help better natural language processing
- Keys from the Twitter Developer Application Management site
    - Consumer key
    - Consumer secret
    - Access token
    - Access token secret

TWITTER SENTIMENT ANALYSIS USING PYTHON

This article covers the sentiment analysis of any topic by parsing the tweets fetched from Twitter using Python.

Sentiment Analysis is a technique widely used in text mining. Twitter Sentiment Analysis, therefore means, using advanced text mining techniques to analyse the sentiment of the text (here, tweet) in the form of positive, negative and neutral. It is also known as Opinion Mining, is primarily for analysing conversations, opinions, and sharing of views (all in the form of tweets) for deciding business strategy, political analysis, and also for assessing public actions.

The project requires authentication via the Twitter API. A new application needs to be created to get the necessary keys. A few libraries also need to be installed for the script to run properly.

METHODOLOGY

1. **Authorize twitter API client.**
2. **Make a GET request to Twitter API to fetch tweets for a particular query.**
3. **Parse the tweets. Classify each tweet as positive, negative or neutral.**

Tweepy: Tweepy, the Python client for the official Twitter API supports accessing Twitter via Basic Authentication and the newer method, OAuth. Twitter has stopped accepting Basic Authentication so OAuth is now the only way to use the Twitter API.

Tweepy gives access to the well documented Twitter API. Tweepy makes it possible to get an object and use any method that the official Twitter API offers. The main Model classes in the Twitter API are Tweets, Users, Entities, and Places. Access to each returns a JSON-formatted response and traversing through information is very easy in Python.

TextBlob: TextBlob, one of the popular Python libraries for processing textual

data, stands on the NLTK. It works as a framework for almost all necessary tasks, we need in Basic NLP (Natural Language Processing). TextBlob has some advanced features like –

Sentiment Extraction

Spelling Correction

Matplotlib: Matploitlib is a Python Library used for plotting, this python library provides and objected-oriented APIs for integrating plots into applications.

matplotlib.pyplot is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.

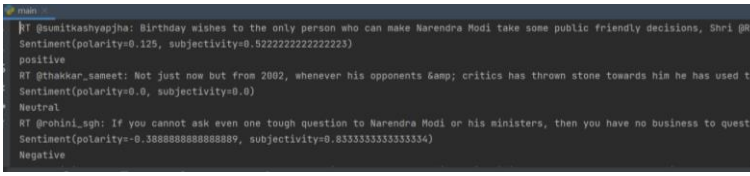
SENTIMENT ANALYSIS USING TEXTBLOB

TextBlob is a python library for Natural Language Processing (NLP).

TextBlob returns polarity and subjectivity of a sentence.

Polarity lies between [-1,1], -1 defines a negative sentiment and 1 defines a positive sentiment. Negation words reverse the polarity. TextBlob has semantic labels that help with fine-grained analysis. For example — emoticons, exclamation mark, emojis, etc.

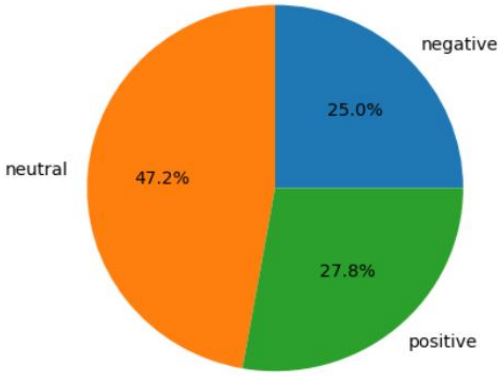
Subjectivity lies between [0,1]. Subjectivity quantifies the amount of personal opinion and factual information contained in the text. The higher subjectivity means that the text contains personal opinion rather than factual information.



PLOTTING THE DATA

Now, after performing sentiment analysis on twitter data:

For all calculated parameters (polarity, subjectivity, sentiment, negative, positive, neutral and compound parameters), I create new features to my data frame.



CONCLUSIONS AND FUTURE WORK

Finally, parsed tweets are returned. Then, we can do various type of statistical analysis on the tweets. For example, in above program, we tried to find the percentage of positive, negative and neutral tweets about a query.

I believe that deep learning models have a high potential w.r.t. classifying text or analysing the sentiment in general. In my opinion there is still space for further improving the classification algorithms. For future work I plan to investigate other sources of information that can be utilized

to detect hateful messages. In addition, to generalize the output received in the current experiment, with evaluation over other datasets, including analysing texts written in different languages.

Thank You.