ADVANCED ACADEMIC CENTER

A CENTER FOR INTER-DISCIPLINARY RESEARCH
2018-19

SENTIMENTAL ANALYSIS

SUPERVISED BY

J.B.V. PRASAD RAJU
B.S.V.S. ANOOP

GOKARAJU RANGARAJU
INSTITUTE OF ENGINEERING AND TECHNOLOGY
AUTONOMOUS

Advanced Academic Center

(A Center For Inter-Discilinary Research)

This is to certify that the project titled

“ SENTIMENTAL ANALYSIS”

is a bonafide work carried out by the following students in partial fulfilment of the requirements for Advanced Academic Center intern, submitted to the chair, AAC during the academic year 2019-20

NAME	ROLL NO.	BRANCH
Venkata Karthika K	18241A04U0	ECE
Lakshmi Manasa M.V.S	18241A1227	IT
Greeshma		CSE

NAME	ROLL N0	BRANCH
Sanjana M	18241A0428	ECE

This work was not submitted or published earlier for any study

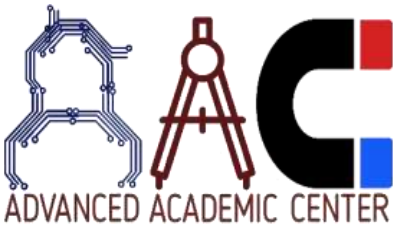
Dr/Ms./Mr.

Project Supervisor

Dr.B.R.K.Reddy
Program Coordinator

Dr.Ramamurthy Suri
Associate Dean,AAC

Table of Contents



ACKNOWLEDGEMENTS

We express our deep sense of gratitude to our respected Director , Gokaraju Rangaraju Institute of Engineering and Technology for the valuable guidance and for permitting us to carry out this project.

With immense pleasure, we record our deep sense of gratitude to our respected principal, for permitting us to carry out this project.

We are thankful to Associate Dean ,Advance Academic Centre for providing us appropriate ecosystem required for the project to complete.

We are thankful to our project supervisor who spared valuable time for us and influence novel ideas to guide us. I am indebted to all the above without whom I would not have concluded the project.

ABSTRACT

The project entitled ***"Sentimental Analysis "*** is the process of computationally identifying and categorizing opinions expressed in a piece of text. It's also known as **opinion mining**, deriving the opinion or attitude of a speaker.

The software is developed using Python and ***Textblob library*** inbuilt in it. Python is now widely used for various projects because it is easy to understand and use. We used python client named tweepy for the official twitter API through which we accessed tweets.

Sentiment Analysis is the process of 'computationally' determining whether a piece of writing is positive, negative or neutral by parsing the tweets fetched from Twitter using Python.

1.INTRODUCTION

1.1 Purpose, scope and limitations

The project aims at analyzing the sentiment of the public for a given topic.

The section Wants to computerize the existing system in order to increase the speed of processing as well as for simplifying the activities , without loosing accuracy.

The system is excepted to computerize the calculations of the sentiment and polarity for the given subject.At present ,this system is restricted to analyze the tweets planning to increse it’s scope in the future.

1.2 Scope of Study and Literature Study

The project is aimed at implementation of concepts of python,Textblob,tweepy.This will help the companies in analyzing the sentiment of their products by providing necessary feedback.

Sentiment analysis is performed by the suggested system on various Chinese micro-blogs. It performs sentiment analysis to determine whether positive/negative or it is an advertisement. The system uses convolutionneural networks for classification. And support vector machine algorithm (SVM algorithm) is used.

It reduces the size of input data by breaking down major data set containing all the information into smaller data set by removing unwanted data like author's name, duplicated data and similar texts. It uses CNN, which has four layers namely input layer, convolution layer, pooling layer and fully connected layer. SVM and lexicon analysis is used as baseline.

2.SYSTEM ANALYSIS

2.1 PRESENT SYSTEM

The present system is manual.The flow of activities is as follows:

Tokenization,Cleaning of data,removing of stop words,Classification and then calculation.

2.2 LIMITATIONS OF THE PRESENT SYSTEM

· The present system is too slow,since generating reports from the reviews is to be done manually.

· Human induced errors are more like to occur in this system.

2.3 USER REQUIREMENTS:

The section wants to computerize the existing system in order to increase the speed of processing as well as for simplifying the activities,without loosing accuracy.

2.4 TECHNICAL FEASIBILITY:

The software can to be developed using the existing technology. Python can be downloaded in free.Social networking platforms **twitter** is user friendly.

3.ANALYZING SENTIMENTS

We have come up with a python code which simplifies the tokenization process.

Textblob is a library for python. Python combined with Textblob provides a fast and easy way to analyze the reviews. We have typed a code in python .

In **Python tokenization** basically refers to splitting up a larger body of text into smaller lines, words or even creating words for a non-English language. The various **tokenization** functions in-built into the nltk module itself and can be used in programs

Once the process of tokenization is done, the next step is cleaning the data.This involves the removal of all the special characters or any other word which donot add any value in analyzing the data.

The next step would be removing the stop words like the ,was,is ,he,she etc.Now we are left with useful adjectives which help us to review a sentence

The next step is classification of the leftover words as positive, negative and neutral.For a positive word we give our sentiment score as +1,for negative -1 and for neutral zero.

we modeled a data with bag of words (lexicons) and tested it on analyzing the statement .more the accuracy score better will be the classification.

Calculating the sentiment score involves combining the polarities .If the polarity is greater than zero it resembles a positive statement and vice-versa .The additon of the polarities of all the statements gives the final sentiment.

All the above mentioned manual processes are combined and put in a python library called textblob. Accessing the review statements is carried out by twitter .

For this , we need to generate the access token and token secret. For this , we need to have a Twitter Developer account and an Application in it.

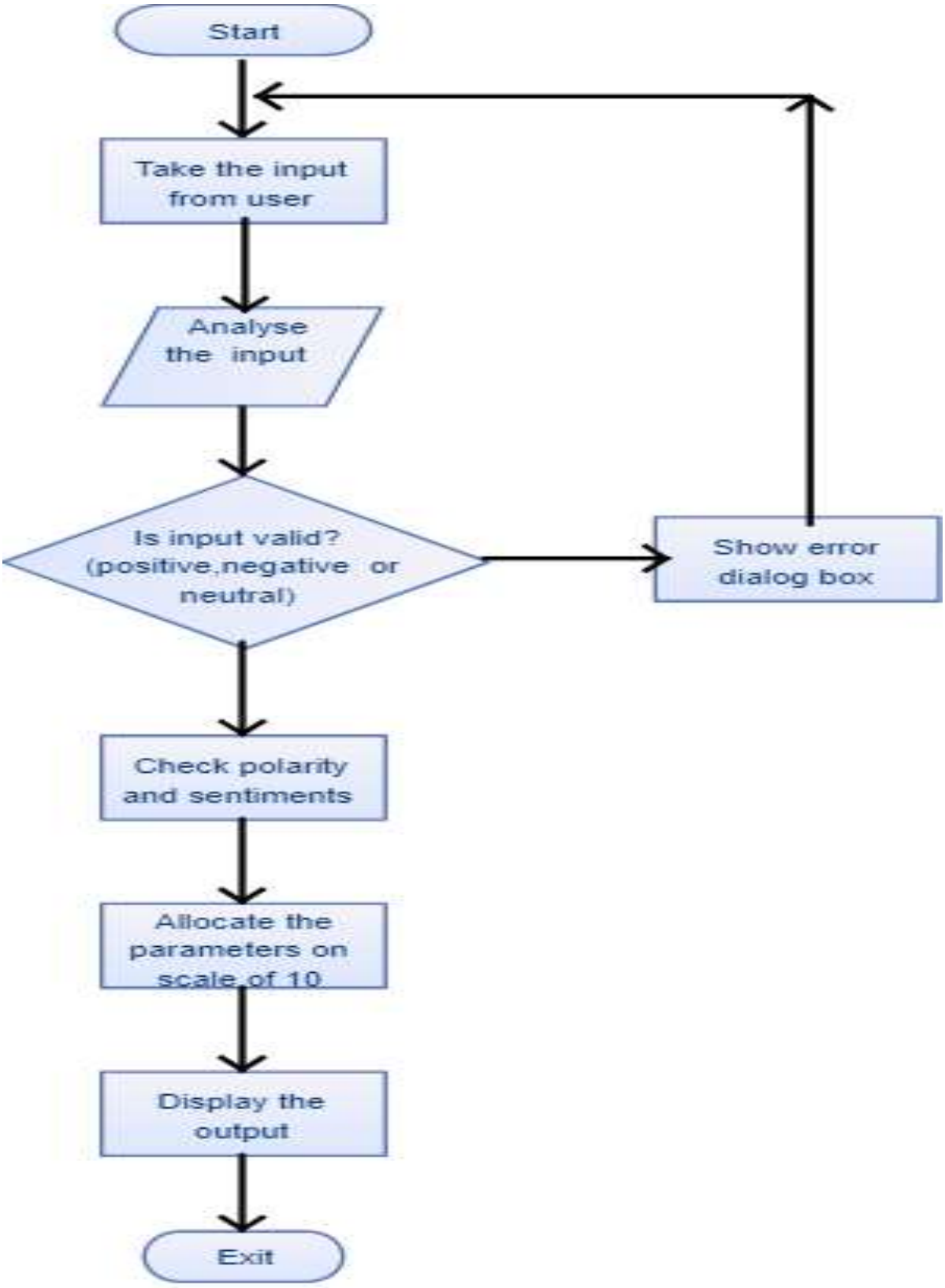
Through this tokens , we have accessed tweets related to a product and analyzed their polarities to calculate the final sentiment using python-textblob code which we have written.

installing: pip install textblob

This is a demonstration of **sentiment analysis** using a `NLTK 2.0.4` powered **text classification** process. It can tell you whether it thinks the text you enter below expresses **positive sentiment**, **negative sentiment**, or if it's **neutral**. Using **hierarchical classification**, *neutrality* is determined first, and *sentiment polarity* is determined second, but only if the text is not neutral.

Installing: `python -m textblob.download_corpora`

FLOWCHART:


Start
Take the input from user
Analyse the input
Is input valid? (positive,negative or neutral)
Show error dialog box
Check polarity and sentiments
Allocate the parameters on scale of 10
Display the output
Exit

FINDINGS:

Sentiment(polarity=0.0, subjectivity=0.0)

Sentiment(polarity=0.29714285714285715, subjectivity=0.6371428571428572)

Sentiment(polarity=-0.11666666666666665, subjectivity=0.3666666666666667)

CONCLUSIONS:

This report discusses in details the various approaches to Sentiment Analysis, mainly Machine Learning and Cognitive approaches. It provides a detailed view of the different applications and potential challenges of Sentiment Analysis that makes it a difficult task.

It achieves the difficult task of performing prediction over a continuum even though trained only on the extreme reviews. Thus machine learning models with a proper kernel that can capture the context will play an important role in SA.

Feature engineering, as in several Machine Learning and Natural Language Processing applications, plays a vital role in SA. It has been seen that Adjectives as word features can capture majority of the sentiment. Use of topic-oriented features and Value Phrases play a significant role to detect sentiment when the domain of application is known.

ATTACHMENTS

IPython console

Console 1/A

```
IPython 7.2.0 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/K V S Achari/Desktop/python/sentimental.py', wdir='C:/Users/K V S Achari/Desktop/python')
Positive tweets percentage: 40.0 %
Negative tweets percentage: 26.153846153846153 %
Neutral tweets percentage: 33.84615384615385 % \



Positive tweets:
RT @SenGillibrand: Remember: Rand Paul was happy to vote for Donald Trump and Mitch McConnell’s trillion-dollar tax handout to millionaires…
RT @DeplorableChoir: There are only 3 people to win the Ellis Island award for their work within the black community...

1. Rosa Parks
2. M…
RT @RawStory: CNN went there -- calling President Donald Trump the #RacistInChief and peddling white supremacist language.

Check out more…
RT @cbouzy: NBC found archived video of Donald Trump and Jeffrey Epstein partying at Mar-a-Lago with a room full of women. Lordy, there are…
RT @RealSaavedra: President Donald Trump on Rep. Ilhan Omar (D-MN): “There's a lot of talk about the fact that she was married to her broth…
RT @PalmerReport: Donald Trump’s day so far:
```

- More racist garbage
- Makes up fake quote from Kevin McCarthy
- House Democrats voting toni…
@realDonaldTrump Yesterday in Greenville was a brilliant indicator of how people respect and love their President-… https://t.co/NBICjf1nte
RT @nytimes: A newly unearthed video from 1992 shows Donald Trump and Jeffrey Epstein watching and commenting on women together at an event…
RT @JohnWDean: Difficult to believe this piece is being written about a sitting POTUS in 2019. We can be sure Donald Trump will live in inf…
RT @Kris_Sacrebleu: “spare me the revolution! It can wait. Win the presidency, hold the House-narrow the spread in the Senate & a lot of go…

Negative tweets:
RT @PalmerReport: Donald Trump's day so far:

- William Barr and Wilbur Ross held in criminal contempt
- Tonight's #TrumpRally is a joke
-…
@ScherieMurray The Republican Party that you may have known before is dead. The new @GOP is Donald Trump's personal… https://t.co/tlAU59bOLv
Trump's 'go back' racism is crude, but may be dangerously effective | By Afua Hirsch https://t.co/T1qTePXRn0
RT @Jimllpaintit: "Can you paint me Theresa May and Donald Trump re-enacting that famous scene from Dirty Dancing in the lake... you know t…
RT @votevets: It looks like Donald "Art of the Deal" Trump tore up Obama's Iran deal ('the worst deal ever") so he could come back to the t…
RT @mehdirhasan: Whether it’s Donald Trump’s Republicans or Ben Shapiro’s Daily Wire team, the anti-black, anti-Muslim racism is across the…

Reference

➢ **Links:**

- https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/
- https://www.youtube.com/watch?v=O_B7XLfx0ic
- https://www.youtube.com/watch?v=1gQ6uG5Ujiw
- https://docs.inboundnow.com/guide/create-twitter-application/#toc-0

- **Books:**

➔ **Applied Text Analysis with Python**
by: Benjamin Bengfort

➔ **Sentimental Analysis and Opinion Mining**
by: Bing Liu

GLOSSARY:

- TextBlob: *TextBlob* is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

- Tweepy: An easy-to-use Python library for accessing the Twitter API.

- OUathHAndler: Tweepy tries to make OAuth as painless as possible for you. To begin the process we need to register our client application with Twitter. Create a new application and once you are done you should have your consumer token and secret.

- API: The API class provides access to the entire twitter RESTful API methods. Each method can accept various parameters and return responses.

- RE library: A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).