TABLE OF CONTENTS

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place: Jaypee Institute of Information Technology
Date: 11th Dec, 2018

Mahima Varshney (9915103257)
Saksham Raj Seth (9915103270)
Smriti Sidana (9915103100)

CERTIFICATE

This is to certify that the work titled **“Healthcare Twitter Analysis”** submitted by **“Mahima Varshney (9915103257), Saksham Raj Seth (9915103270), Smriti Sidana (9915103100)”** of B.Tech of Jaypee Institute of Information Technology University, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of any other degree or diploma.

Signature of Supervisor	
Name of Supervisor	Ms. Arti Jain
Designation	Assistant Professor
Date	11th Dec, 2018

ACKNOWLEDGEMENT

The completion of any inter-disciplinary project depends upon cooperation, coordination and combined efforts of several sources of knowledge. We are grateful to **Ms. Arti Jain** for her even willingness to give us valuable advice and direction whenever we approached her with a problem. We are thankful to her for providing us with immense guidance for this project.

We are also thankful to Dr. Himani Bansal and Mr. Gaurav Nigam for giving their valuable time for evaluating our project. We would also like to thank our College authorities and Head/Dean for giving us the opportunity to pursue our project in this field and helping us to successfully complete this project.

Signature(s) of Students

Smriti Sidana (9915103100)

Mahima Varshney (9915103257)

Saksham Raj Seth (9915103270)

Date: 11th Dec, 2018

SUMMARY

Health care twitter analysis deals with the sentiment analysis of tweets related to health care by the people on twitter. The application of sentiment analysis has grown enormously. It’s application in health care has great potential to analyze and improve the health of a country. It takes the patients opinions into perspective to make policies and modifications that could directly address their problems. It can help finding a pattern of health disorders in a particular location from the analysis of tweets of users experiences. The tweets can be categorized as positive and negative.

Signature of Student(s)
Smriti Sidana (9915103100)

Mahima Varshney (9915103257)

Saksham Raj Seth (9915103270)

Date: 11th Dec, 2018

Signature of Supervisor
Ms.Arti Jain

Date: 11th Dec, 2018

LIST OF FIGURES

Fig No. Fig. Name **Page No.**

LIST OF TABLES

LIST OF SYMBOLS & ACRONYMS

Symbol	**Symbol Description**
D	Document
W	Word
Q	Query
R	Requirement

IEEE	Institute of Electrical and Electronics Engineers
MNB	Multinomial Naive Bayes
SVR	Support Vector Regression

Chapter 1: INTRODUCTION

A healthy nation they say is a wealthy nation. Healthcare is important to the society because people get ill, accidents and emergencies do arise and the hospitals are needed to diagnose, treat and manage different types of ailments and diseases. Many of people’s aspirations and desires cannot be met without longer, healthier, happy lives.

1.1 MOTIVATION

Human resources are vital to an effective health care system. From an economics viewpoint, health workers’ salaries make up a great share of health budgets in most countries. The health worker is the gatekeeper of the health system. So being the part of the society we should be aware of nearby health because, individuals health can affect the society.

1.2 PURPOSE OF PROJECT

No one plans to get sick or hurt, but most people need medical care at some point. Health insurance covers these costs and offers many other important benefits. Health insurance protects you from unexpected, high medical costs. You pay less for covered in-network health care, even before you meet your deductible. The main motive of the project is to already know your environment, living society which can affect your health.

1.3 DESCRIPTION OF PROJECT

The project provides a systematic and comparative study of methods and techniques for tweets of healthcare and find the keywords such as Depression, Anxiety and help people living in the great environment. Using NLP techniques with the implementation of SVR.

1.4 CURRENT/OPEN PROBLEMS

Health is an important part of well being of a country. A sudden outbreak of a disease can cause panic among people. Early detection of a disease is extremely important to prevent it from spreading in a region.With the rapid growth of usage of social media platform by public to express their opinions, the user posts can be of great help to monitor the spread of diseases in different regions. The collection of health tweets is vast and can be used as a corpus for any machine learning technique. People post their opinions about health care services of a health center or medical products on twitter. Filtering tweets on the basis of geolocation can help health care organizations to provide health care facility in that region to cure and prevent the disease. By the results of analysis, people can learn from others' experiences which can help them make better choices according to their needs. They can learn about the positive and negative experiences of others.

1.5 INTEGRATED SUMMARY OF LITERATURE SURVEY

Research Paper Name	**Author**	**Year**	**Summary**
Mining Social media Streams to Improve public health allergy surveillance	Kathy Lee, Ankit Agrawal, Alok Choudhary	2015	The use of the Bag of words supervised learning approach along with Naive bayes, Naive Bayes Multinomial, Random Forest and SVM with extracting features such as N-grams, POS-tagging, noun-noun bigrams resulting the identification of geolocation with allergy stats.
Using Twitter Data and Sentiment Analysis to study diseases dynamics	Vincenza Carchiolo, Alessandro Longheu, Michele Malgeri	2015	The use of SNOMED-CT was mainly focused. They extracted the geographical location as the feature, using Naive bayes algorithm.
Twitter Sentiment	K Lavanya, C	2017	The bigrams was extracted as a feature and

Analysis using multi-class SVM	Deisy		implemented on multi-class SVM. In additions,Natural Language ToolKit and Pointwise mutual Information and Information Retrieval were the tools and technique to conclude the results with 68% accuracy.
Application of Sentiment Analysis to Language learning	MEI-HUA CHEN, WEI-FAN CHEN, LUN-WEI KU	2018	Language learning implementation with features extracted SentiwordNet, wordnet data cleaning; context-aware emotion synonym suggestion system; with Naive bayes algorithm.
Sentiment Analysis for Health Care	M. Taimoor Khan, Shehzad Khalid	2015	Comparative Study of Naive Bayes, knn, Centroid based, SVM with Bag of words tools and POS-tags as features
Sentiment Analysis Using Fuzzy and Naive Bayes	Ruchi Mehra, Mandeep Kaur Bedi, Gagandeep Singh, Raman Arora, Tannu Bala, Sunny Saxena	2017	SentiStrength, Bag-Of-Words tools were focused and n-grams and POS-tags on tweets with result of Positive:Accuracy-96.634%, recall-100%, precision-93.2692% Neutral: Accuracy-96.15385%, recall-100%, precision-92.3077% Negative: Accuracy-100%, recall-100%, precision-100%
Mining Twitter Data For Influenza Detection and Surveillance	Kenny Byrd, Alisher Mansurov, Olga Baysal	2016	The use of multiple tools such as: Google Maps Developer Tool, Stanford CoreNLP, Apache OpenNLP toolkit, LingPipe toolkit,PostgreSQL, Raspberry Pi 2, OpenLayers 3. With feature extraction of Apache Lucene's Stop Filter, Apache Lucene's PorterStemFilter, Resulting in Stanford CoreNLP-70%
Hybrid Approach	Korawit	2017	The algorithm implemented in the features

Framework for Sentiment Classification on Microblogging	Orkphol, Wu Yang, Wei Wang, Wenlong Zhu		extracted i.e. , n-grams and POS-tags , were SVM, Adapted Lesk Algorithm Resulting in Non WSD mostly has higher F-measure than WSD
Extensive Study of Text Based Methods for Opinion Mining	Prof. D.S. Kulkarni, Dr. S. F. Rodd	2018	Concept-level opinion mining resource with tools like Sentic-Net, WordNet was implemented with Multilayer perception, SVM algorithms on dataset of ISEAR
Summarization and sentiment analysis from user health posts	Vinod L. Mane, Suja S. Panicker, Vidya B. Patil	2015	Lesk Algorithm on features Combination of GSS coefficient and IDF methods with TF , resulting the deadly combination with tools Bag-Of-Words, Apriori, FP-growth Algorithm.

Table 1 Integrated summary of literature survey

1.6 PROBLEM STATEMENT

Sentiment analysis of health care tweets on twitter using keywords such as Depression, Anxiety. To know the public view on health by classifying health tweets using SVR and measuring public satisfaction of health care services by sentiment analysis.

1.7 OVERVIEW OF PROPOSED SOLUTION

First tweets are extracted and data preprocessing is done. Tweets are filtered according Health keywords such as Depression, Anxiety, Mental Illness. Proposed solution consists of feature extraction using NLP techniques to get the statistics. Extracted features were implemented with SVR algorithm and MNB and to use the mapping dictionaries like SentiWordNet Dictionary to get better results.

Chapter 2: ANALYSIS, DESIGN AND MODELING

2.1 OVERALL ARCHITECTURE

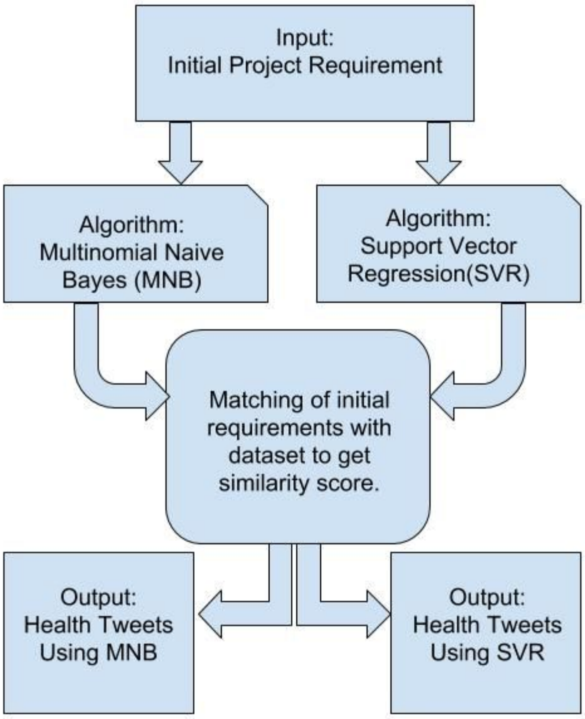

Input:
Initial Project Requirement
Algorithm:
Multinomial Naive Bayes (MNB)
Algorithm:
Support Vector Regression(SVR)
Matching of initial requirements with dataset to get similarity score.
Output:
Health Tweets Using MNB
Output:
Health Tweets Using SVR

Fig. 1 Architecture

The Previous section presented a critical analysis of the existing techniques which can be integrated together to achieve better results from extracting multiple features by texts.
In this section, new proposals will be made for the target system.

Our dataset has been extracted from twitter streaming which consists of tweets as per now, but we will also extract the geo location based on particular tweet, for achieving our goal. Many features have been extracted in order to search the pattern of occurring health disorders in particular local area.

However, we noticed that some of the collected tweets that contain any of these terms are irrelevant to healthcare because of the multiuse of these terms in different domains. For example, in English word ‘virus’ has a potential to exist in healthcare tweets, however it occurs more often in the context of computer virus. Therefore, a more reliable classification is required to distinguish between healthcare relevant and irrelevant tweets. For this purpose, we used more restricted queries to select the tweets among the potentially identified healthcare ones that are very likely to be relevant healthcare. Then we used the narrow identified tweets to train a support vector Regression (SVR) to classify the remaining identified tweets as relevant or irrelevant to healthcare.

SOFTWARE REQUIREMENTS

- Windows 7 and above , or Ubuntu or Centos
- Python Idle, Anaconda

HARDWARE COMPONENTS

- Processor – i3
- Hard Disk – 5 GB
- Memory – 1GB RAM

FUNCTIONAL REQUIREMENTS

- The user should have information of the libraries used.
- The user should be familiar with python.
- The user should be familiar with Anaconda.

NON-FUNCTIONAL REQUIREMENTS

- The program should not have any reliability issues. The program will be thoroughly tested.
- The program should run on any software mentioned above.
- The program should be able to classify videos.

USER REQUIREMENTS

- User should be able to work in python.
- User should be able to work in Anaconda.

2.2 DESIGN DOCUMENTATION

2.2.1 Use Case Diagram

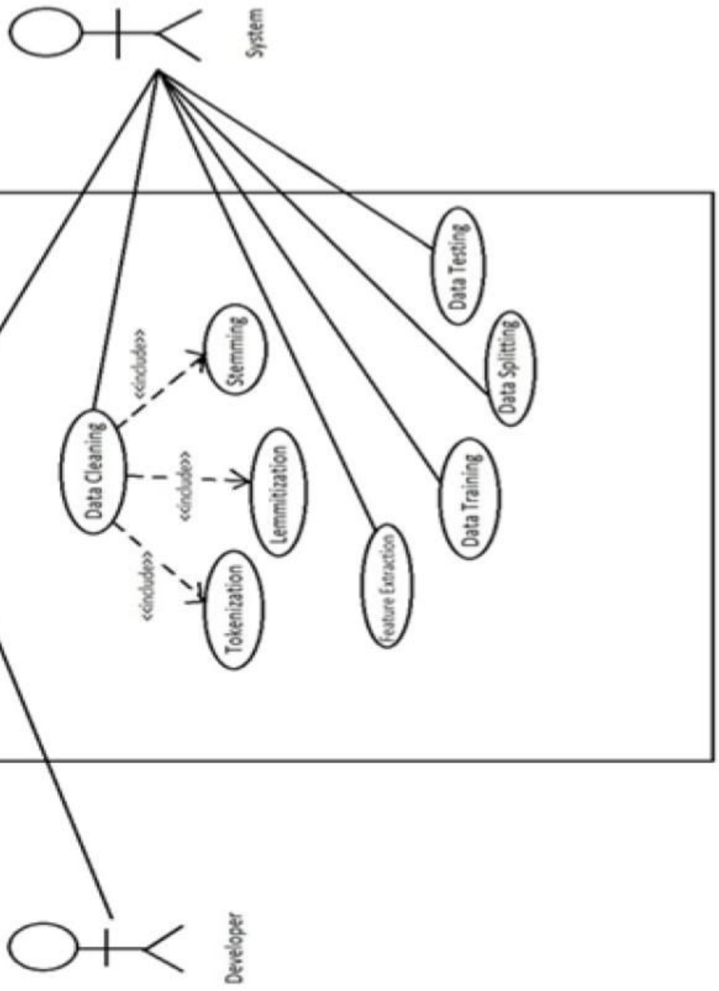


Fig. 2 Use Case Diagram

2.2.2 Control Flow Diagram

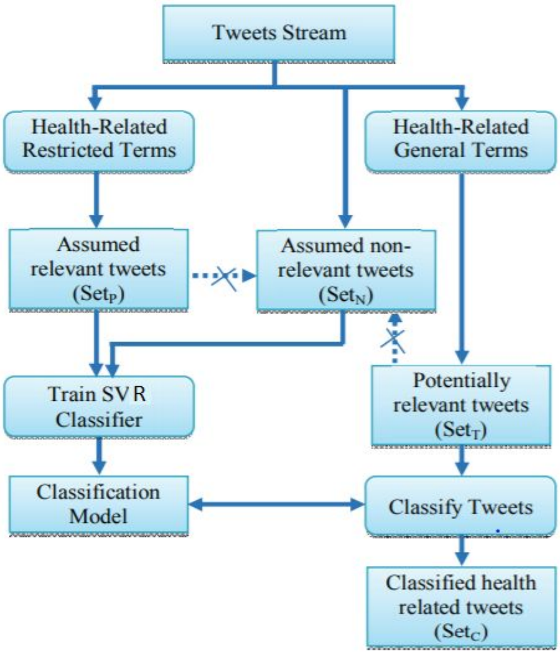

Tweets Stream
Health-Related Restricted Terms
Health-Related General Terms
Assumed relevant tweets (Set$_P$)
Assumed non-relevant tweets (Set$_N$)
Train SVR Classifier
Potentially relevant tweets (Set$_T$)
Classification Model
Classify Tweets
Classified health related tweets (Set$_C$)

Fig. 3 Control Flow

2.2.3 Activity Diagram

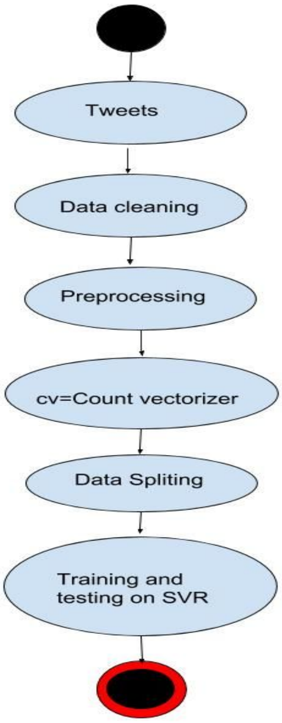


Fig. 4 Activity diagram

2.2.4 Proposed Algorithms

a. Raw data

This involves the dataset collection from twitter.

b. Lexicon construction

This step focuses on the informal language of online social network. For this reason, three type lexicon have been created: lexicon for social acronyms, lexicon for emotions, and lexicon for interjections.

c. Data Pre Processing

The goal behind preprocessing is to clean the dataset by removing words and punctuations that don't have an influence on sentiment classification. This increases the performance of the task. Therefore, the preprocessing is a primordial task. The data preprocessing steps applied are:

1. Case Folding

   All the words in the dataset are converted to lowercase.

2. Due to lack of functionality following have been removed as a part of data preprocessing steps.

   - Unnecessary punctuations
   - Extra blank spaces
   - URL's
   - Hashtags

3. Removal of stopwords

   The stop words are words that do not add meaningful content to the data set (i.e., pronouns, prepositions, conjunctions, etc.). So, removing them reduces the space of the items significantly in the training and testing set.

4. Replacing Emoticons

   Emoticons indicating different sentiments are replaced with keywords ,i.e,
   Positive emoticons- EPOS
   Negative emoticons-ENEG
   Neutral emoticons-ENEUT

5. Replacement of sentiment words

   Words indicating sentiments are replaced with keywords, i.e.,

   Positive words-POS

   Negative words-NEG

6. Replacement of negation and intensity words

   Negation words-NEGATION

   Intensity words-INTENSITY

d. Feature Extraction

- **TOKENIZATION**

  A token is a sequence of characters that we want to treat as a group. Decomposing a text into tokens enables creating counts of tokens, which can be used as features. A token could be a paragraph, a sentence etc. but commonly words are chosen as tokens in text categorization.

- **STEMMING**

  Stemming, together with lemmatization, is a common method in NLP. The goal of the method is to reduce the amount of inflected words by stripping the suffix of the word to retrieve a "base form" of the word. Stemming is built upon the idea that words with the same stem are close in meaning and that NLP will be improved if such different words can be grouped together to one single term, consequently in Machine learning the number of features can be reduced if stems are used instead of the original words. In the example of the words:

  • Argue

  • Argued

  • Argues

  • Arguing

  A stemming algorithm could identify the suffixes "e", "ed", "es", "ing" and strip the words to the stem "argu". As the example demonstrates, a stem does not necessarily has to be a word. The suffix stripping algorithm used for providing this example is the Porter Stemmer, from the python library NLTK (Natural language toolkit), based on the algorithm developed by Martin Porter.

The feature extraction is a process of extracting the main characteristics of the text. For a machine learning algorithm to perform well, it is essential to have features that are descriptive of the text. The total number of occurrences of following features have been taken into account for each sentence:

- **Gram features**

In gram feature extraction calculated the map of the sentence after converting the sentence in tokens and bigrams with inclusive use of porter stemmer for stemming of words. Map contains the key and the value, key is the word extracted after tokenization and bigrams.

- **Sentiment Extractions**

Used TextBlob to extract the polarity and subjectivity of the sentence. Also calculated the positivity and negativity of the sentence using SentiWordNet dictionary. This process is done three times. For complete sentence, first half of sentence and second half of sentence.

- **POS vector**

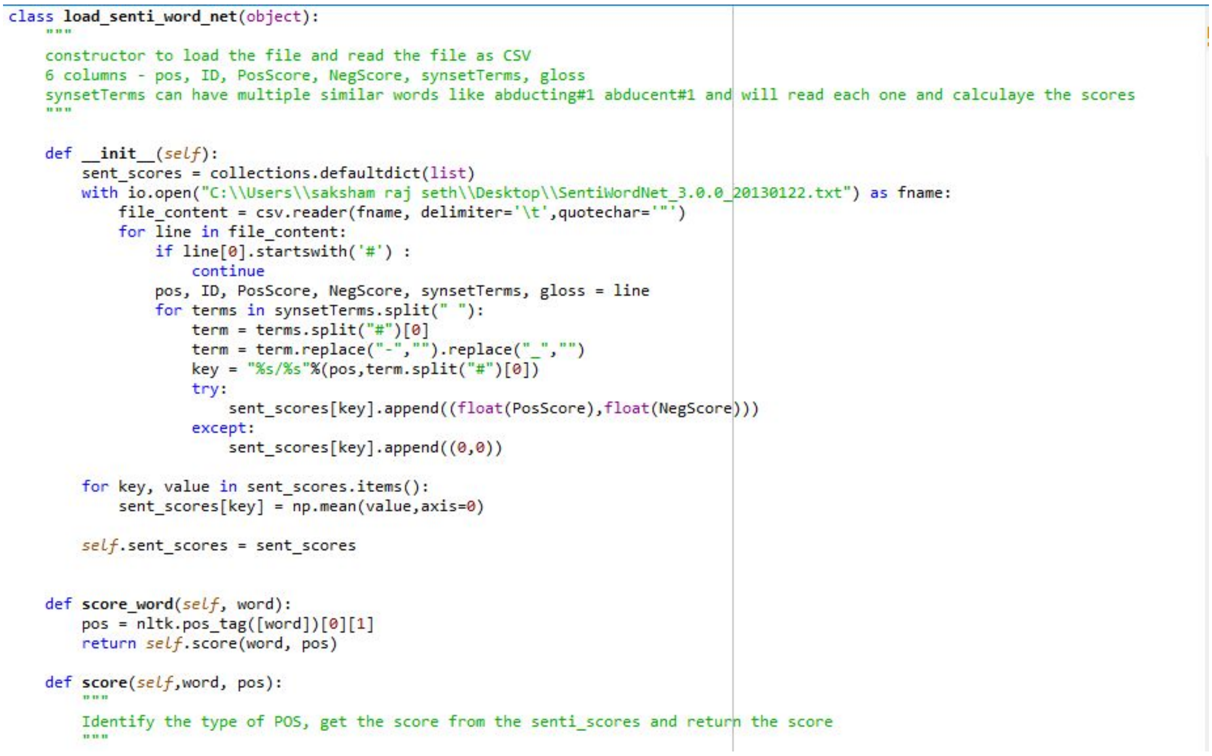
It counts the Part-Of-Speech tags. It uses a numpy array and increases by one when noun, verb, adjective and adverb are found in the pos_tag list.

Chapter 3: IMPLEMENTATION AND TESTING

3.1 IMPLEMENTATION DETAILS

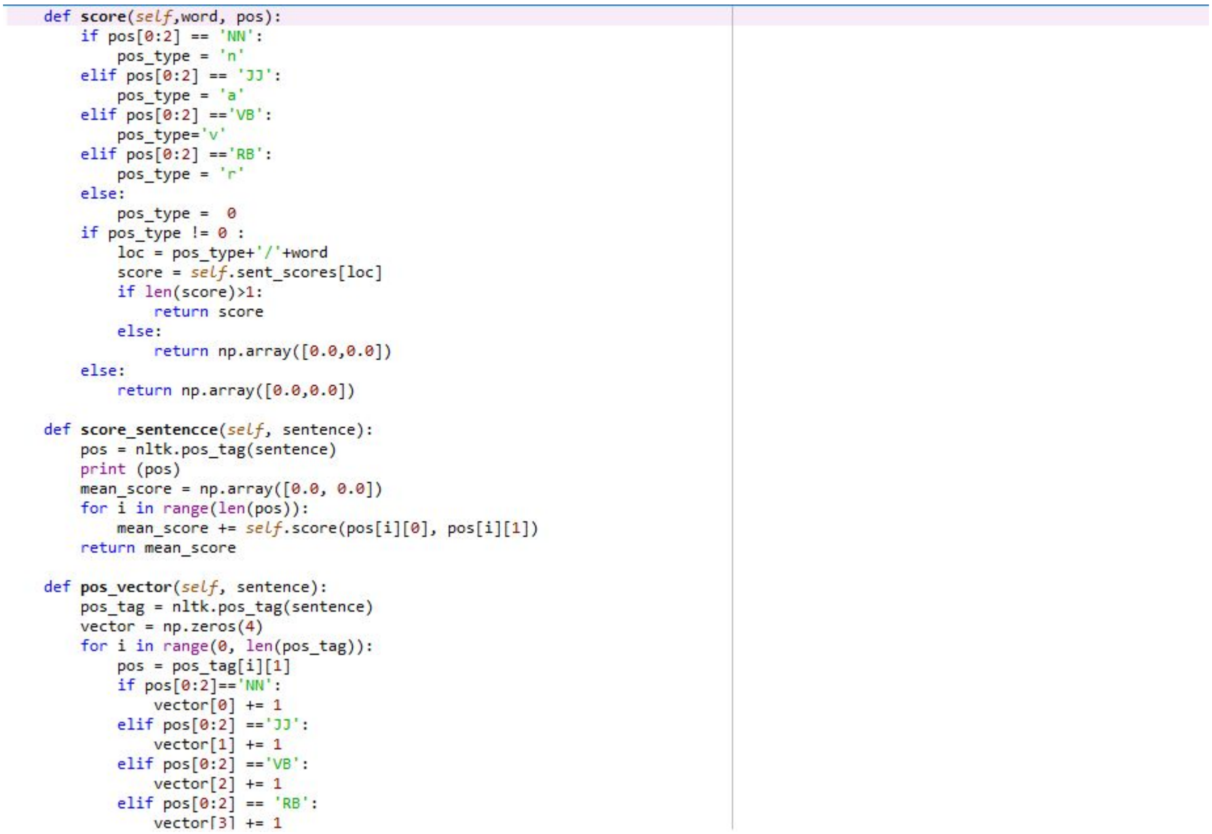
Only a string is required as an input, which is goes through complete process of pre-process which includes Stemming, Lemmatization, also removing Slangs, Emojis, HashTags etc. After which feature extraction took place which are matching columns to calculate the result. In feature extraction we calculated sentiment score using textblob and sentiwordnet dictionary, POS tags, POS vector, POS score and n-grams.

In this process, we have to find the accuracy of the texts. Firstly, we will extract the features of the tweet. Now, we will apply same methodology of complete processing of the provided tweet. Finally, we will compare the features of the provided tweet to our datasets tweets by taking any random tweets for testing to check feature extraction is working fine or not.



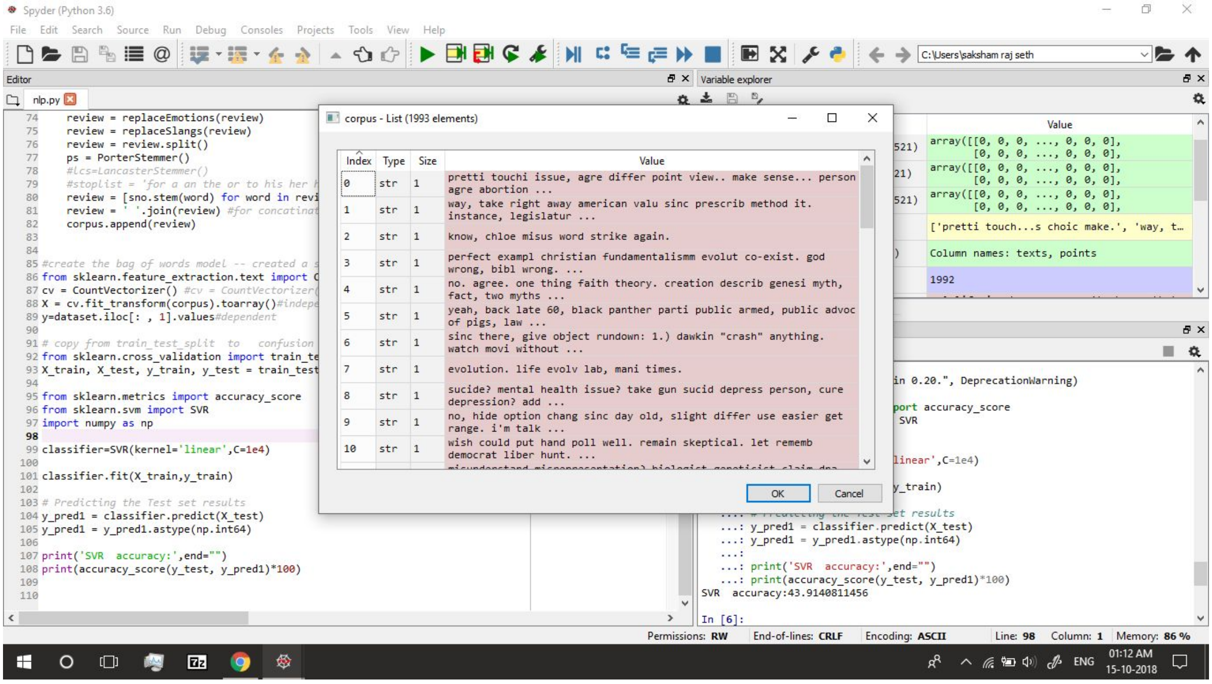
```
class load_senti_word_net(object):
    """
    constructor to load the file and read the file as CSV
    6 columns - pos, ID, PosScore, NegScore, synsetTerms, gloss
    synsetTerms can have multiple similar words like abducting#1 abducent#1 and will read each one and calculaye the scores
    """

    def __init__(self):
        sent_scores = collections.defaultdict(list)
        with io.open("C:\\Users\\saksham raj seth\\Desktop\\SentiWordNet_3.0.0_20130122.txt") as fname:
            file_content = csv.reader(fname, delimiter='\t',quotechar='"')
            for line in file_content:
                if line[0].startswith('#') :
                    continue
                pos, ID, PosScore, NegScore, synsetTerms, gloss = line
                for terms in synsetTerms.split(" "):
                    term = terms.split("#")[0]
                    term = term.replace("-","").replace("_","")
                    key = "%s/%s"%(pos,term.split("#")[0])
                    try:
                        sent_scores[key].append((float(PosScore),float(NegScore)))
                    except:
                        sent_scores[key].append((0,0))

        for key, value in sent_scores.items():
            sent_scores[key] = np.mean(value,axis=0)

        self.sent_scores = sent_scores


    def score_word(self, word):
        pos = nltk.pos_tag([word])[0][1]
        return self.score(word, pos)

    def score(self,word, pos):
        """
        Identify the type of POS, get the score from the senti_scores and return the score
        """
```

Fig. 5 Code Snippet 1



```
    def score(self,word, pos):
        if pos[0:2] == 'NN':
            pos_type = 'n'
        elif pos[0:2] == 'JJ':
            pos_type = 'a'
        elif pos[0:2] =='VB':
            pos_type='v'
        elif pos[0:2] =='RB':
            pos_type = 'r'
        else:
            pos_type =  0
        if pos_type != 0 :
            loc = pos_type+'/'+word
            score = self.sent_scores[loc]
            if len(score)>1:
                return score
            else:
                return np.array([0.0,0.0])
        else:
            return np.array([0.0,0.0])

    def score_sentencce(self, sentence):
        pos = nltk.pos_tag(sentence)
        print (pos)
        mean_score = np.array([0.0, 0.0])
        for i in range(len(pos)):
            mean_score += self.score(pos[i][0], pos[i][1])
        return mean_score

    def pos_vector(self, sentence):
        pos_tag = nltk.pos_tag(sentence)
        vector = np.zeros(4)
        for i in range(0, len(pos_tag)):
            pos = pos_tag[i][1]
            if pos[0:2]=='NN':
                vector[0] += 1
            elif pos[0:2] =='JJ':
                vector[1] += 1
            elif pos[0:2] =='VB':
                vector[2] += 1
            elif pos[0:2] == 'RB':
                vector[3] += 1
```

Fig. 6 Code Snippet 2


Spyder (Python 3.6)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\saksham raj seth
Editor
nlp.py
Variable explorer
corpus - List (1993 elements)
OK
Cancel

```
    review = replaceEmotions(review)
    review = replaceSlangs(review)
    review = review.split()
    ps = PorterStemmer()
    #lcs=LancasterStemmer()
    #stoplist = 'for a an the or to his her h
    review = [sno.stem(word) for word in revi
    review = ' '.join(review) #for concatinat
    corpus.append(review)


#create the bag of words model -- created a s
from sklearn.feature_extraction.text import C
cv = CountVectorizer() #cv = CountVectorizer(
X = cv.fit_transform(corpus).toarray()#indepe
y=dataset.iloc[: , 1].values#dependent

# copy from train_test_split  to   confusion
from sklearn.cross_validation import train_te
X_train, X_test, y_train, y_test = train_test

from sklearn.metrics import accuracy_score
from sklearn.svm import SVR
import numpy as np

classifier=SVR(kernel='linear',C=1e4)

classifier.fit(X_train,y_train)

# Predicting the Test set results
y_pred1 = classifier.predict(X_test)
y_pred1 = y_pred1.astype(np.int64)

print('SVR  accuracy:',end="")
print(accuracy_score(y_test, y_pred1)*100)
```

Index	Type	Size	Value
0	str	1	pretti touchi issue, agre differ point view.. make sense... person agre abortion ...
1	str	1	way, take right away american valu sinc prescrib method it. instance, legislatur ...
2	str	1	know, chloe misus word strike again.
3	str	1	perfect exampl christian fundamentalismm evolut co-exist. god wrong, bibl wrong. ...
4	str	1	no. agree. one thing faith theory. creation describ genesi myth, fact, two myths ...
5	str	1	yeah, back late 60, black panther parti public armed, public advoc of pigs, law ...
6	str	1	sinc there, give object rundown: 1.) dawkin "crash" anything. watch movi without ...
7	str	1	evolution. life evolv lab, mani times.
8	str	1	sucide? mental health issue? take gun sucid depress person, cure depression? add ...
9	str	1	no, hide option chang sinc day old, slight differ use easier get range. i'm talk ...
10	str	1	wish could put hand poll well. remain skeptical. let rememb democrat liber hunt. ...

	Value
521)	array([[0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 0, 0, 0],
21)	array([[0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 0, 0, 0],
521)	array([[0, 0, 0, ..., 0, 0, 0], [0, 0, 0, ..., 0, 0, 0],
	['pretti touch...s choic make.', 'way, t…
)	Column names: texts, points
	1992

```
in 0.20.", DeprecationWarning)

port accuracy_score
 SVR

linear',C=1e4)

y_train)
    ...: # Predicting the Test set results
    ...: y_pred1 = classifier.predict(X_test)
    ...: y_pred1 = y_pred1.astype(np.int64)
    ...:
    ...: print('SVR  accuracy:',end="")
    ...: print(accuracy_score(y_test, y_pred1)*100)
SVR  accuracy:43.9140811456

In [6]:
```

Permissions: RW   End-of-lines: CRLF   Encoding: ASCII   Line: 98   Column: 1   Memory: 86 %

ENG 01:12 AM 15-10-2018

Fig. 7 Corpus

3.2 TESTING OF THE IMPLEMENTED MODULES

3.2.1 Unit Testing

During the implementation, individual components of the module were tested by the developer to reduce the complexity of the overall testing activity and make it easier to detect the components that caused any fault. The test cases were based on the black-box testing method, which considers each component as a black entity and classifies possible inputs into the component. The methodology is only concerned with the input and the result of the tested components.

3.2.2 Integration Testing

This phase involved integrating all the components to increase the reliability of the system. The integration testing detects faults within integration components that cannot be found from unit testing. The technique used was bottom-up testing, which tests components of the lower level first and then integrates them with higher level components. This was decided because this technique allows interface faults to be easily detected. We started with integrating two project components and testing them together. If there is no fault, additional increments are integrated to test. The testing process was repeated several times to ensure that the system worked well after the integration of all parts of the project. Following that, the testing procedure proved all the system increments were working together as they should.

Chapter 4: FINDINGS AND CONCLUSION

4.1 FINDINGS

This Project presents a novel approach of classifying Health tweets for Depression, Anxiety from all mixed tweets which gives us a step ahead to identify health status in a living environment. It also help people to know how to overcome the problems related to Health. It will help you make your living surroundings better.

4.2 CONCLUSION

For efficient retrieval of tweets based on NLP techniques, we had studied different research papers based on tweets classification as well as methods to classify tweets related to health. We tried to extract features and implement algorithms to check the correctness of the method applied.

4.3 FUTURE WORK

In future, we will try to develop an efficient system that can work correctly many times from the current system and also try to expand the project which can help people determine the quality of service of any healthcare center, spread of diseases, and effects of medical products from the analysis of tweets of the others experience. We are trying to expand our system for more than one language, more for languages such as Hindi, Punjabi, Urdu. Also, expansion of project will lead for more health related issues, to make our surrounding better.

REFERENCES

[1] Byrd, Kenny, Alisher Mansurov, and Olga Baysal. "Mining twitter data for influenza detection and surveillance." *Proceedings of the International Workshop on Software Engineering in Healthcare Systems*. ACM, 2016.

[2] Carchiolo, Vincenza, Alessandro Longheu, and Michele Malgeri. "Using twitter data and sentiment analysis to study diseases dynamics." *International Conference on Information Technology in Bio-and Medical Informatics*. Springer, Cham, 2015.

[3] Chen, Mei-Hua, Wei-Fan Chen, and Lun-Wei Ku. "Application of Sentiment Analysis to Language Learning." *IEEE Access* 6 (2018): 24433-24442.

[4] Khan, Muhammad Taimoor, and Shehzad Khalid. "Sentiment analysis for health care." *Big Data: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2016. 676-689.

[5] Kulkarni, D. S., and S. F. Rodd. "Extensive study of text based methods for opinion mining." *2018 2nd International Conference on Inventive Systems and Control (ICISC)*. IEEE, 2018.

[6] Lavanya, K., and C. Deisy. "Twitter sentiment analysis using multi-class SVM." *Intelligent Computing and Control (I2C2), 2017 International Conference on*. IEEE, 2017.

[7] Lee, Kathy, Ankit Agrawal, and Alok Choudhary. "Mining social media streams to improve public health allergy surveillance." *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 2015.

[8] Mane, Vinod L., Suja S. Panicker, and Vidya B. Patil. "Summarization and sentiment analysis from user health posts." *Pervasive Computing (ICPC), 2015 International Conference on*. IEEE, 2015.

[9] Mehra, Ruchi, et al. "Sentimental analysis using fuzzy and naive bayes." *Computing Methodologies and Communication (ICCMC), 2017 International Conference on*. IEEE, 2017.

[10] Orkphol, Korawit, et al. "Hybrid approach framework for sentiment classification on microblogging." *Computing Conference, 2017*. IEEE, 2017.