A sentimental analysis based on mental health calculation for Facebook and Twitter users

[1]L. Praveen Kumar, Asst. Prof, dept of CSE, Malla Reddy Institute of Technology, Kompally, Maisammaguda, Dulapally, Secunderabad, 500100

[2]V. Manikyala Rao, Asst. Prof, dept of CSE, Malla Reddy Institute of Technology, Kompally, Maisammaguda, Dulapally, Secunderabad, 500100

[3]U. Sudhakar Reddy, Asst. Prof, dept of CSE, Malla Reddy Institute of Technology, Kompally, Maisammaguda, Dulapally, Secunderabad, 500100

Abstract

In the entire world, depression is the biggest cause of incapacity. It is important to take the required steps to regulate and lessen the negative impacts of mental illness. To do this, we'll need to gather and analyse social media data, identify those who are afflicted, and provide them with appropriate advice depending on the severity of the issue. We may utilise the sentiment analysis idea to categorise the user's post into positive (alright), neutral (early stages), and negative categories (ill). Following the identification of the people, we need to provide positive, calming or motivating messages to users with psychological disorders in order to prevent them from getting into difficulty in the future.

Keywords: Depression, mental health, sentimental analysis, mental illness.

I.   INTRODUCTION

Depression is a common disorder that impacts an individual ability to perform life activities. Depression also correlates with other mental illnesses, most often debilitating symptoms of anxiety, or drug abuse illnesses. The onset of depression happens earlier in life today than in previous decades. So, It's important to identify and take necessary care as earlier as you can so that one reduce its ill effect significantly. But, the challenge here in identifying is many people who are suffering from depression feels like they are pretty much as the other people around so it makes very difficult to identify the candidates in specific. With the technology now most of the people are involving in social media and expressing their thoughts, views, feelings, opinions etc through which we can able to say how the person is feeling which eventually helps us in recognizing the people who are depressed or stressed or mentally ill for some extent. So, thanks to technology especially in this aspect.

Software Details:

1.   Operating Systems: Windows, Linux
2.   Python Interpreter
3.   MS Excel
4.   COLAB TOOL

Hardware Details:

1.   Processor   -   Pentium-III
2.   Speed   -   1.1 GHz
3.   RAM   -   256 MB (minimum)
4.   Hard disk   -   20 GB
5.   Key Board   -   StandardWindows Keyboard
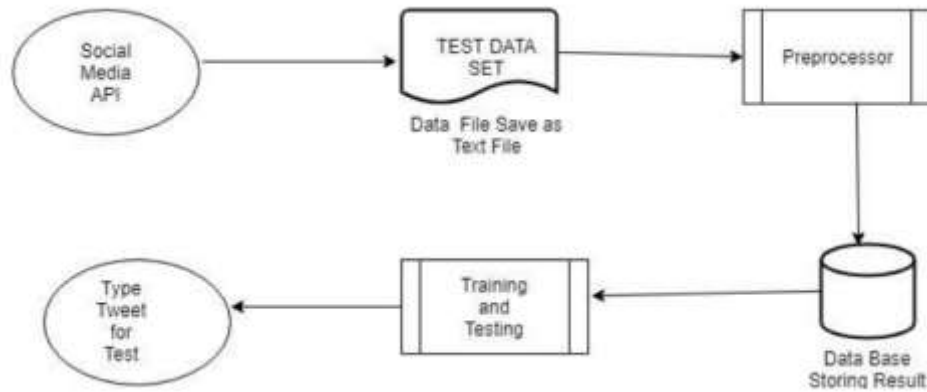6.   Mouse   - Two or Three button mouse
7.   Monitor   -   SVGA

Fig.1: Block diagram for Sentimental analysis

The Existing solutions: The Sentiment Analysis is mainly used for analyzing and knowing pulse of the customers mainly. Generally most of the people use either Python or R language to write the code for sentiment analysis. Generally from collected data sets from customers through surveys or Reviews the data is analyzed and conclusions are drawn with the help of various Algorithms. Generally the pattern is firstly they Tokenize the sentences and then they remove the insignificant portion of words which doesn't contribute any significance to the sentiment analysis and then they perform the analysis to find the polarity. This is how the present Sentiment Analysis work in general.

Our objective is that we can find the people with the depression, mental illness and help them and these days most of the people are expressing their feelings on social media first so it is the best source to find out about how and what people are going through and hence we can save lots of money in the future by knowing the mental health condition of people right now in the present using social media

II.   RELATED WORK

MODULES OF THE SYSTEM

1.   Loading
2.   Data Preprocessing
3.   Visualizing
4.   Training and Testing

Loading: In this we will obtain the Twitter data by streaming the required tweets using keywords and collected into a text file initially created. When data sets are ready you can proceed on to the pre-processing stage.

Pre Processing: In this stage we will go through the data sets and the given dictionary. The dictionary contains words with their corresponding polarity, which is essential to calculate the feeling of each tweet, each word will become tokens and polarity is granted. Each tweet will consist of the sum of all the polarity of each word and will be divided by word count in that tweet

Visualizing: Once the preprocessing is done. You can find the file in the "processed_data / output.xlsx" directory. When you open it, you will find that the ID (tweet) and Sentiment of each tweet are separated into 2 columns. With this output, you now have a set of twitter data and its corresponding sentiment filtered by pressure keywords. (Positive, Neutral and Negative). MS Excel is used to visualize the output data an through python libraries like matlabplot.lib we can visualize in through graphs.

Training and Testing: With output that we got we can use that as reference for training models using other algorithm like Naive Baye Accuracy, Decision Tree, Support Vector Machines, Kneighbors and Random Forest. After that we can calculate accuracy of each model by comparing it to the actual result and we can choose the best of bunch for further doing of sentiment analysis. And Even we can calculate the respective time taken.

Input Design:

The input design is what the user sees. It is the front end. It is designed in such a way that the user understands what is to be done. It comprises the developing specification and procedures for data preparation and those steps are necessary to put data into a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the number of inputs required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy.

Output Design:

A quality output is one, which meets the necessities of the top user and presents the data clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it's determined how the information is to be displaced for immediate need. It's the foremost important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to assist user decision-making. Designing computer output should proceed in an organized, well thought out manner, the correct output must be developed while ensuring that every output element is intended so people will find the system can use easily and effectively. When analyses design computer output, they ought to identify the particular output that's needed to satisfy the necessities.

FEASIBILITY STUDY

we analyse how feasible our project is in all of the aspects like economical, technical and social feasibility in order to study how acceptable and successful the project is. Besides all the theoretical assertions, we study the practicality of the project in this section. Some understanding of the system's main requirements is necessary for feasibility analysis. There are three main criteria to the feasibility study

1.   Technical feasibility
2.   Economic feasibility
3.   Social feasibility

TECHINICAL FEASIBILITY: Technical feasibility is studied to know the technical requirements to the project and how modest these requirements are. Considering the current project, the system merely requires a Wi-Fi connection which is now a very minimal requirement. All the resources we used in the system are simple and minimal and hence, we can do the assertion in which the system is technically feasible.

ECONOMIC FEASIBILITY: we study economic feasibility to check how economically viable the project is. Economic feasibility in general includes the expenditure put in and a justification of how this expenditure is economic.The current project is most likely an economically feasible project as we are capable of controlling the lights and fans at a cost lesser than the cost of a smart bulb. We are also able to protect our home time to time from disasters that could cost us fortunes. Additionally, we save a lot of money by switching off lights timely, which could otherwise add a fortune to our electricity bill.

SOCIAL FEASIBILITY: We study social feasibility to know how acceptable the system is by the user. By testing the social feasibility we can also get constructive criticism that could help us in developing the project in future. The current project involves a simple application that the user has to install. He can control everything in the circuit just by the touch of a button. There's no complexity in the gas leakage system as well, as it works on its own and only notifies the user when the levels of CO are high. Hence, we assert that the system is socially feasible.
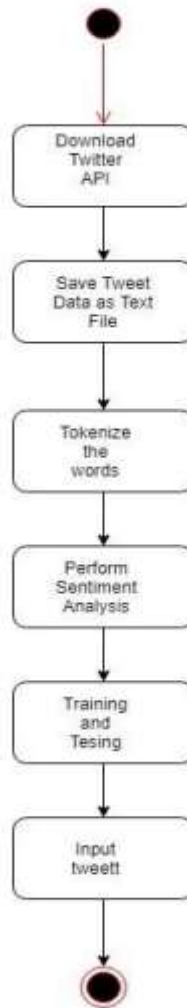
III.   PROCEDURE AND METHODOLOGY

Fig.2: Block diagram for the sentiment analysis process

Packages:

Numpy Package: Numpy is a package in python to perform large scale numerical operations on high dimensional arrays. It can perform scientific computations with large scale data. Numpy provides different functions and attributes that can be applied on the arrays to retrieve necessary information out of the arrays. Numpy is robust and memory efficient. It has several inbuilt optimizations that are written in C which helps in faster execution while dealing with arrays of several hundred dimensions. This library contains wide variety of mathematical and algebraic operations that can be performed on Numpy arrays.

Scikit-Learn Package: This package available in python to deal with machine learning models. It provides pre implemented models of classification, regression and clustering like linear regression, SVM, Decision Trees, Logistic Regression, K-Means, DBSCAN etc. The core part of this library is implemented in C which makes it more efficient and takes less time in execution of models. This package provides a complete pipe line to implement machine learning models right from model compilation to model evaluation. It provides different metrics to evaluate the model. The .fit(), .fit_transform(), .predict() are the common functions used in every model of the scikitlearn package. It provides a generalized functions to deal with different machine learning algorithms as a result it is very easy to implement the models.

Matplotlib package: Matplotlib are open source packages available in python which deals with visualization of data. They help us to draw conclusions about from the plots drawn. They provide different visualization plots to perform univariate and bivariate analysis of data. Matplotlib is a older package but has all the core functionality do deal with the plotting. SeaBorn is the latest package used for visualization which is built on top of Matplotlib contains functions to create plots like scatter plot, violin plot, bar plot, count plot, PDF, CDF etc. Generally in real world to understand the profit or loss of a company by observing or looking upon graphs is easy. In present world data is complex and increasing day by day it leads to so many complex problems. Startups are fail to guess their position in future why because they are unable to know their revenue and expenditure like how much they gain in given period .Matplot is python helpful in efficient way to determined 14 Scattering graphs also we draw below example shows the scatter graph of any company analysis

Tweepy Library: It is a Python library file for getting to the Twitter API. It is well known for basic robotizations and making twitter bots. Tweepy has very numerous highlights. Tweepy is publicly released, and on GitHub and empowers the Python to speak with Twitter stage and utilize its API. For a presentation on the library Twython . At the hour of composing, the present variant of tweepy is 1.13.

Tweepy establishment

 It can be introduced utilizing pip, a Python bundle supervissor. In this article, we're going to utilize a virtual domain (virtualenv) for the ventures to abstain from relying upon framework wide bundles. For more data on virtual conditions and pip, look at Python Virtual Environments: You can begin by making a task called tweepy-bots. The initial step is to make an index and a virtual situation:

```Shell
$ mkdir tweepy-bots
$ cd tweepy-bots
$ python3 -m venv venv
```

The orders above make the virtual condition inside the undertaking index. At that point you can introduce the Tweepy bundle. To begin with, you have to enact the recently made virtual condition and afterward use pip to do the establishment:

```Shell
$ source ./venv/bin/activate
$ pip install tweepy
```

```Shell
$ pip freeze > requirements.txt
```

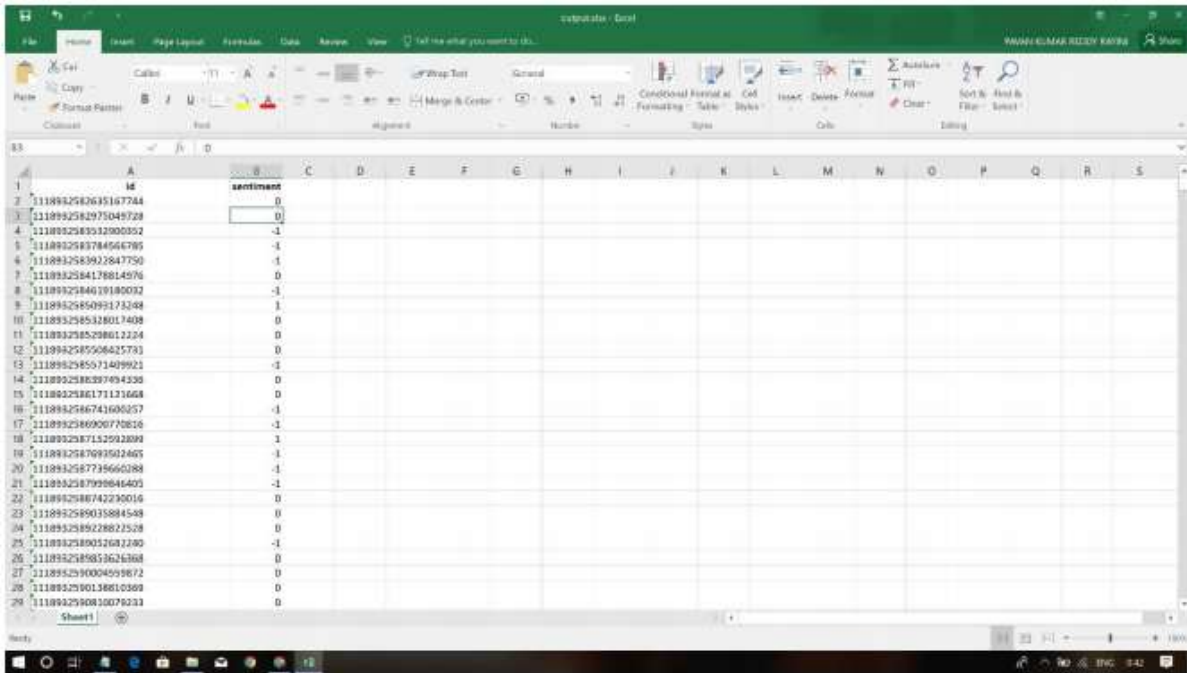You will use requirements.txt while deploying the project. Presently that Tweepy is introduced.

IV.   IMPLEMENTATION AND RESULTS

Output generation

The system's output will be in the form of a number in an excel sheet according to the id that we take. The ouput can be a positive number or a negative number or neutral(zero).

fig.3:Screenshot of input Tweet data in tweetdata.txt

Fig.4:Screenshot of Dictionary File which as polarities of various words

Fig.5: Screenshot of output after executing the preprocessing code

Fig.6: Screenshot of further information after executing the Preprocessing code

Fig.7: Screenshot of Excel File with id and polarity
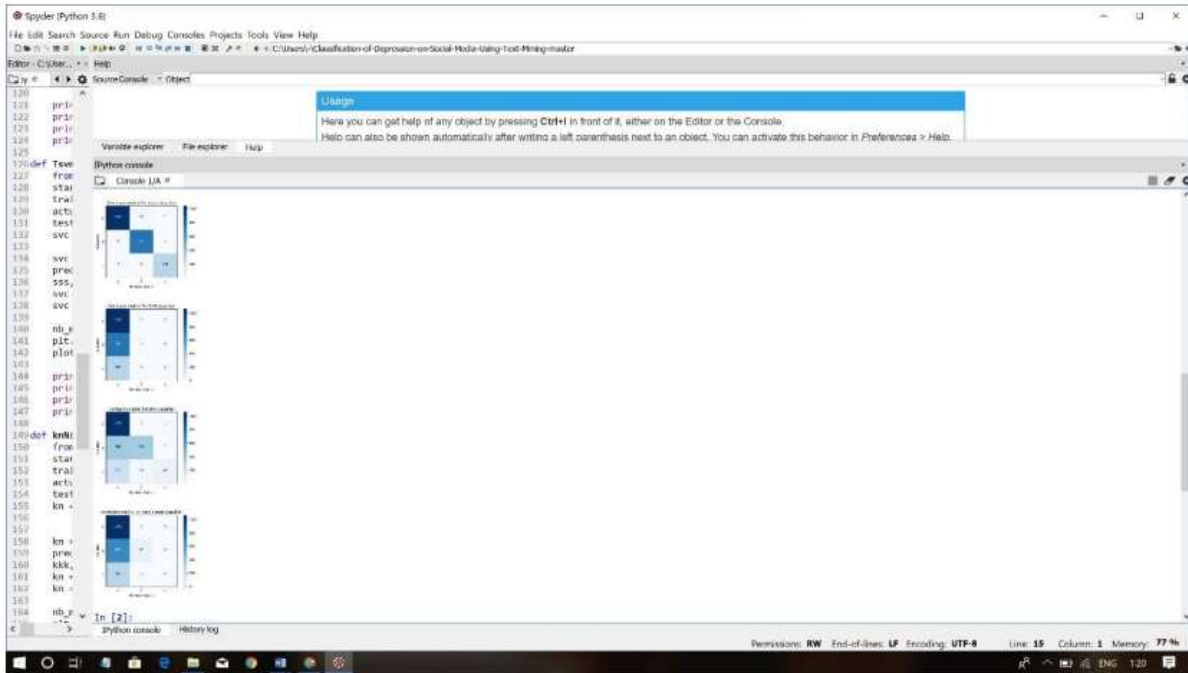
Fig.8: Screenshot of Confusion matrix of various Algorithms

Fig.9: Screenshot of Accuracy and completion time of various Algorithms along with input Tweet

V. CONCLUSION AND FUTURE SCOPE

In these days due to various Reasons people are getting stressed and depressed which is very difficult to identify on their own this project, The DETECTION OF MENTAL HEALTH CONDITION OF SOCIAL MEDIA USER'S USING SENTIMENT ANALYSIS can able to detect the users having depression, Stressed and even who are not Mentally Healthy using the Sentiment Analysis through social media in this project we chosen the Twitter in which we can classify the user into either of the three categories based on their Tweets. Though it is not cent percent accurate it can only get better and better in Future.

FUTURE SCOPE: In this project we used twitter for collecting the data in future one can extend this data to various other social media Applications. In this we haven't included the recommendation as the Twitter actually haven't allowed us to Tweet or Retweet to the users so its further expansion can be a Recommendation System which Actually helps the users to know that they need to consult the doctor incase they are suffering from the stress or depression.

REFERENCES

[1] W. H. Organization. (2017). Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization. [Online]. Available: http://www.who.int/en/newsroom/fact-sheets/detail/depression

[2] M. J. Friedrich, ''Depression is the leading cause of disability around the world,'' JAMA, vol. 317, no. 15, p. 1517, Apr. 2017.

[3] M. Nadeem. (2016). ''Identifying depression on twitter.'' [Online]. Available: https://arxiv.org/abs/1607.07384

[4] S. Paul, S. K. Jandhyala, and T. Basu, ''Early detection of signs of anorexia and depression over social media using effective machine learning frameworks,'' in Proc. CLEF, Aug. 2018, pp. 1–9.

[5] A. Benton, M. Mitchell, and D. Hovy. (2017). ''Multi-task learning for mental health using social media text.'' [Online]. Available: https://arxiv.org/abs/1712.03538

[6] G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, ''From ADHD to SAD: Analyzing the language of mental health on twitter through self-reported diagnoses,'' in Proc. 2nd Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2015, pp. 1–10.

[7]   D. Maupomés and M. Meurs, ''Using topic extraction on social media content for the early detection of depression,'' in Proc. CLEF (Working Notes), vol. 2125, Sep. 2018. [Online]. Available: https://CEUR-WS.org

[8]   P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, and J. Boyd-Graber, ''Beyond LDA: Exploring supervised topic modeling for depression-related language in twitter,'' in Proc. 2nd Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2015, pp. 99–107.

[9]   D. Preotiuc-Pietro et al., ''The role of personality, age, and gender in tweeting about mental illness,'' in Proc. 2nd Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2015, pp. 21–30.

[10]   T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, ''Affective and content analysis of online depression communities,'' IEEE Trans. Affect. Comput., vol. 5, no. 3, pp. 217–226, Jul. 2014.

[11]   H. A. Schwartz et al., ''Towards assessing changes in degree of depression through facebook,'' in Proc. Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2014, pp. 118–125.

[12]   S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, ''Recognizing depression from twitter activity,'' in Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst., Apr. 2015, pp. 3187–3196.

[13]   N.V.KrishnaRao , K Radhika, D Rahul , J Thirupathi and B Madhuravani, " A Survey on Top-K Dominating Queries on Any Incomplete Information", International Journal of Pure and Applied Mathematics, Vol. 119 No. 14 2018, pp. 1807-1811(ISSN: 1314-3395,Scopus Indexed)

[14]   J Thirupathi , D Rahul , K Radhika , N.V.KrishnaRao and B Madhuravani," An Enhanced Novel Based Incentive Framework for Cellular Traffic Offloading", International Journal of Pure and Applied Mathematics, Vol. 119 No. 14 2018, pp. 1813-1817 (ISSN: 1314-3395, Scopus Indexed).

[15]   N V Krishna Rao,YHarikaDevi,NShalini,AHarika,VDivyavani,Dr.N Manga Thayaru"Credit Card Fraud Detection Using Spark and Machine Learning Techniques", ICACECS2020,Machine Learning Technologies and Applications pp 163-172, Algorithms for Intelligent Systems book series (AIS),march 2021.

[16]   J. Wolohan, M. Hiraga, A. Mukherjee, Z. A. Sayyed, and M. Millard, ''Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with nlp,'' in Proc. 1st Int. Workshop Lang. Cognition Comput. Models, 2018, pp. 11–21.

[17]   Y. Tyshchenko, ''Depression and anxiety detection from blog posts data,'' Nature Precis. Sci., Inst. Comput. Sci., Univ. Tartu, Tartu, Estonia, 2018.

[18]   S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, ''Detecting depression and mental illness on social media: An integrative review,'' Current Opinion Behav. Sci., vol. 18, pp. 43–49, Dec. 2017.

[19]   R. A. Calvo, D. N. Milne, M. S. Hussain, and H. Christensen, ''Natural language processing in mental health applications using non-clinical texts,'' Natural Language Eng., vol. 23, no. 5, pp. 649–685, 2017.