iMood : Automated Depression Detection Tool

Srividhya Ravichandran

Student, Department of Information Science ,Ramaiah Institute of Technology, Bangalore, India

---

Abstract – *Depression is the leading cause of suicide, with an estimate that up to 60 percent of people who commit suicide have major depression. There isn't a "depression test" that a doctor can use to test its presence. In fact, half the people who are depressed don't realize it themselves, let alone getting it treated. In a country like India, there is a stigma attached to mental health issues that restricts the afflicted from visiting the doctor. iMood is able to use Natural Language Processing Technology and Machine learning algorithms to detect sentiments in a user's social media page. From there, we are able to determine whether the sentiment is uncommonly negative , and if so, we state that the account holder should visit a physician to get help.*

Key Words: Natural Language Processing, Sentiment Analysis, Semantic data, Machine Learning, Tone Analyzer, Data Visualization

1.INTRODUCTION

According to the World Health Organization, around 300 million people in the world are diagnosed with depression. It is also estimated that 15% of the adult population will experience depression at some point in their life time. What's worse is that only one in five people receive treatment for depression. The media portrays the depressed person as someone who sulks around at home and at the workplace , crying a corner alone. While these can definitely be symptoms of depression, this depiction fails to convey the whole story. Some people can go to great lengths to hide these symptoms, which makes it impossible for their loved ones to suspect even a hint of depression. We propose iMood, an automated depression detection tool that detects sentiments from a user's social media account like Twitter. If the sentiment is uncommonly negative, we alert the user about the signs and advise him to visit a doctor. We also display a graph which portrays the user's change in sentiment over time.

1.1 Natural Language Processing

We use Natural Language Processing to detect change in sentiments. Emotional tone is derived from IBM Watson Tone Analyzer ,which is an ensemble framework that infers emotions from text. The tone analyzer is based on Machine learning algorithms that uses features such as n-grams, punctuations, emoticons, greetings and curse words. The tone is calculated from linguistic analysis based on these features. The tone analyzer also leverages lexical features from dictionaries, existence of second person references in a conversation, and certain higher level features such as multiple presence of exclamation marks or question marks. The sentiment is calculated based on these features.

1.2 Sentiment Analysis Framework

To get an understanding of how Sentiment Analysis is performed, I dedicate this section to explaining the framework. Part of Speech Tagging (POS) is performed on the text so that it assigns a label to each word. Thereafter, we can extract patterns in the text.

We then look at sentiment orientation of the patterns we extracted. An example of a pattern may be a noun following an adjective. In this stage, the machine tries to situate the words on an emotive scale. We then compute the averaged sentiment orientation of all the phrases we gathered. Experiments have been done using Naïve Bayes, Maximum Entropy and Support Vector Machines.

An important feature in classification techniques is the determination of the polarity of a word. Thus, the words in a phrase can be grouped together by polarity.

An important process during sentiment analysis is text pre-processing. This involves removing stop words (words that do not contribute to the sentiment score) , lemmatizing the words, reducing the words to their root form and organizing the emotional symbols that people use in texts. Words with a lot of uppercase alphabets generally expresses aggressive text, and a sentence with a lot of exclamations can express joy or surprise.

After the text processing stage, the words are segmented based on their polarity and the part of speech tags are identified. The last step is analysis and scoring. This involves determining the sentiments from the data and scoring them. This process can be carried out in a supervised as well as an unsupervised manner. The former requires labeled data for training classifiers. The latter relies on lexical -based methods that make use of a predefined list of words., where each

word is associated with a particular sentiment. The method that we employ in this paper is based on supervised classification-based sentiment analysis.

2. SYSTEM FLOW

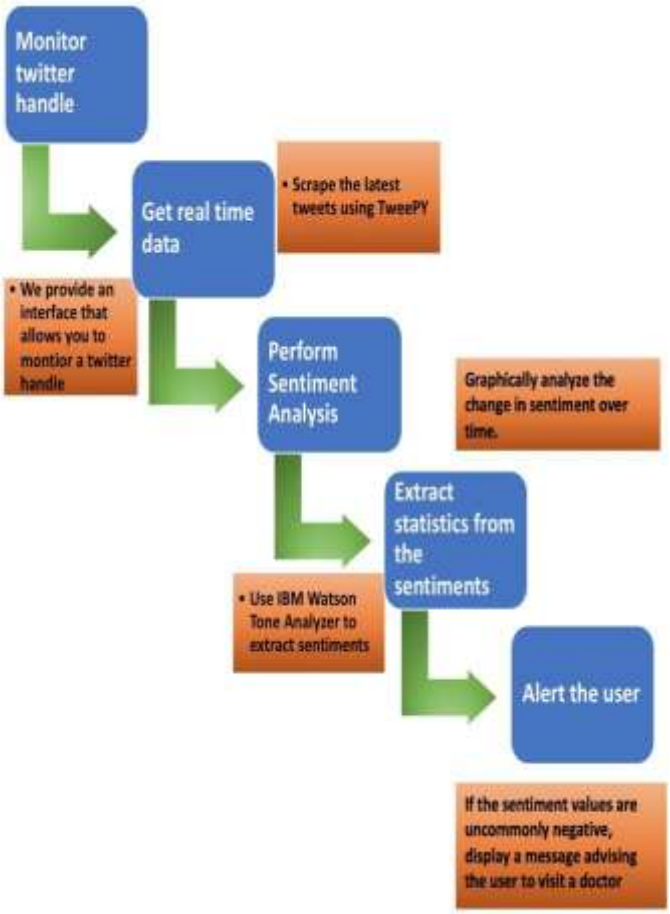

Figure 1: System Flow

3. SOFTWARE SPECIFICATIONS

Requirement	**Software**
Language	Python
Scraping Live Tweets	Requests , TweePy
Sentiment Analysis	IBM Watson Tone Analyzer
User Interface	Flask
Data Visualization	Chart.js

4. RESULTS

After close monitoring of a twitter account for a month, I was able to successfully identify depression.

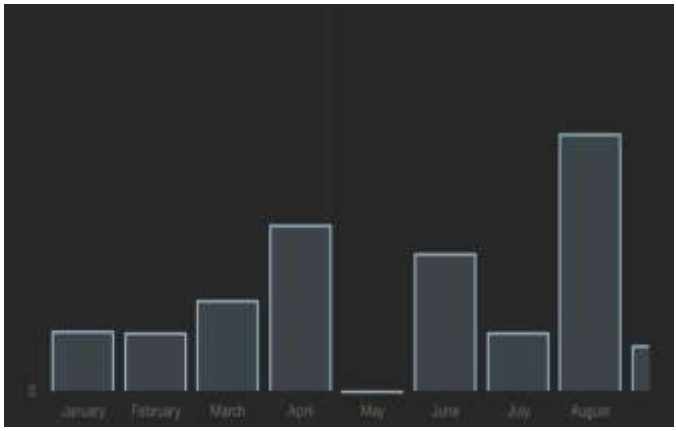

Figure 2: Sentiment Variation over a period of 8 months

REFERENCES

[1] Manasee Godsay, "The process of Sentiment Analysis ", International Journal of Computer Applications, Vol. 126- September 2015

[2] Apoorv Agarwal et.al, "Sentiment Analysis of Twitter Data", Department of Computer Science, Columbia University

[3] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis" , Foundations and Trends in Information Retreival, Vol-2, No. 1

[4] Aaron Lopes Epression, Detecting depression through natural language processing technology and incorporation of machine learning algorithms.