A Proposed System on Detecting Stress Based On Social Interactions on Social Networks

[1] Indu Dokare, [2] Gayatri Pawar, [3] Manisha Mirchandani, [4] Shreya Narsapur, [5] Kajal Rajani
[1][2][3][4] Computer Department, VESIT, Mumbai, India.

Abstract – **Stress is essentially humans' response to various types of desires or threats. This response, when working properly, can help us to stay focussed, energized and intellectually active; but if it is out of proportion, it can certainly be harmful leading to depression, anxiety, hypertension and a host of threatening disorders. Cyberspace is a huge soap box for people to post anything and everything that they experience in their day-to-day lives. Subsequently, it can be used as a very effective tool in determining the stress levels of an individual based on the posts and status updates shared by him/her. This is a proposal for a website which takes the Twitter username of the subject as an input, scans and analyses the subject's profile by performing Sentiment Analysis and gives out results. These results suggest the overall stress levels of the subject and give an overview of his/her mental and emotional state. The tool used for analysis of the social media account is Rapidminer. Rapidminer is an environment for various data mining and machine learning procedures with a very effective and simple GUI.[1]**

Keywords- stress; Sentiment Analysis; Rapidminer

1. INTRODUCTION

Social media can be considered as a goldmine for Sentiment Analysis. An active social media user shares the majority of the details of his/her life online. This accounts for a large database on which algorithms can be applied for analysis. There are various social networking sites these days namely, Facebook, Twitter, LinkedIn, Pinterest, Instagram, Tumblr, Reddit, Snapchat, WhatsApp, Quora etc. The fact that there is a need for so many different platforms itself speaks volumes about the number of social media users. The stress detection model is developed by scanning and analyzing the Twitter profile of the subject. The extraction, collection and analysis of these tweets is done using a free software, Rapidminer. The proposed website takes the twitter handle of the client as an input. Subsequent stages of data mining and analysis will be efficiently carried out by Rapidminer, owing to its expedient GUI. In the final stage, the website gives an overview of the stress level of the user and remedial steps, if necessary.
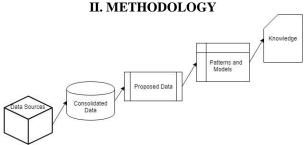
II. METHODOLOGY

Fig 1: Data Mining and Extraction process

Twitter data can be extracted using an SSIS package followed by loading it into an RDBMS and creating a small table to show the garnered tweets. This process creates a saved method to recognize the sentiment of the tweet based on the keywords. Being an elementary approach, it still does the work of analyzing tweets merely by using SQL. Therefore, a more erudite methodology for extraction namely Rapidminer is employed.

A. Data Mining and Extraction

A good amount of tweets have to be extracted to form a database. Tools like Zapier, Google Docs and GDocBackUpCMD are handy for the same. Zapier allows collaboration amongst different web applications. It is thus used to extract tweets in a Google Docs spreadsheet and then copy the data to the local environment for mining. An empty Google sheet is created and three different columns are named as follows: 'twitter_handle' for the twitter handle of the user; 'date_posted' for tracking the date on which the tweet is posted and 'content' for the content of the tweet. Once the framework of the spreadsheet is ready, we proceed to building the zap.

Fig 2: The empty Google Spreadsheet

Creating a Zap is a simple process, it just involves drag and drop with Twitter being the source and Google Docs the target. The contents of the spreadsheet are then mapped to the actual contents on Twitter. After saving the tweets using the GDocBackUp facility (a local copy of the tweets saved to our machine), an SSIS package helps us populate this database.

The next step is to analyse the data for its sentiment, viz. whether the data is to be termed as positive or negative. Following tools have to be used for the same: Rapidminer, two distinct data sets which contain already classified tweets, Microsoft SQL Server. Two distinct tables have to be created which will contain the data. One of them (Table_1) will possess the already analysed data i.e. the reference data while the other table (Table_2) will have the data which is yet to be analysed. The reference tweets will be saved under Table_1 and the tweets that are mined for analysis will be saved under Table_2. Further, cleansing of the data has to be performed as there are many tweets in which people use unnecessary ASCII characters and URLs which are garbage values when it comes to sentiment analysis. This is done by writing an SQL code, in which we create a string splitting function to remove the unnecessary and unwanted characters. On invoking this function, we will be able to remove unwanted characters, reseed identity columns and delete empty SQL records.

Once the data is cleansed, we develop the Rapidminer model.

III. WORKING OF THE SYSTEM

To ensure that the system reaches out to a maximum number of users, a website has been proposed. The website basically takes the users Twitter account details as an input and gives out the stress level of the user as an output. These details will be private to the user and the admin of the website.

The website will have two user types: Admin and Client. Doctors and administrators of the website will operate under the user type Admin. All the users of the system will be under the type Client. Certain functionalities will be different for the two types of users. If the admin logs in to the website, he/she will have access to all the requests made by the Clients. Apart from this, he will also be able to view different feedbacks given by the users and will be able to look after the overall working of the website and will monitor the same. The admin will also have

provisions to answer personalized queries forwarded to them by users.

The user type client will have comparatively less functionalities. The first step will be creating an account on the website and logging in. The clients will enter their Twitter handles as an input, after will they will be given a few choices including the form in which they would like their stress level to be shown (pie chart, graph, histogram, bar chart) and the number of days for which the stress has to be calculated. Once these details are entered, the stress level of the user will be calculated and displayed and reactive measures, if necessary, will be provided.
On the client side, there will be a request form which will be generated.

Fig 3: Request form for the client

Rapidminer will be integrated with the website and the results obtain by the Rapidminer tool are the ones which would be displayed on the website. The Rapidminer model, which is generated, will have workflow as follows:
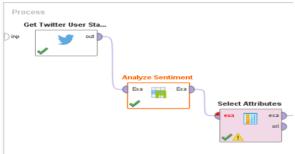

Fig 4(a): Process followed by the Rapidminer tool

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
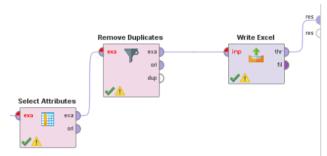**Vol 5, Issue 2, March 2018**

Fig 4(b): Process followed by the Rapidminer tool

IV. EXPECTED RESULTS

The results which will be displayed to the client after analysis of the tweets are as follows:

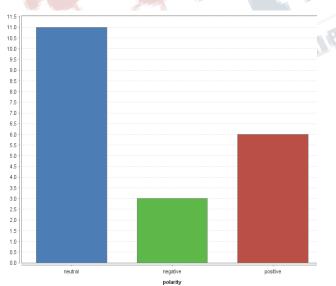| Id	"	"	polarity	subjectivity	Created-At
9679730870...	.	.	neutral	subjective	Feb 26, 2018 ...
9679681173...	.	.	negative	subjective	Feb 26, 2018 ...
9678045482...	.	.	neutral	objective	Feb 25, 2018 ...
9678038564...	.	.	neutral	objective	Feb 25, 2018 ...
9677744335...	.	.	positive	subjective	Feb 25, 2018 ...

Fig 5: The polarity of the tweets

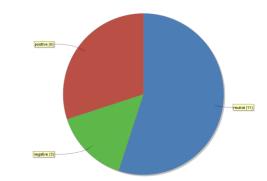*Fig 6: Bar graph output of tweets segregated by polarity*

Fig 7: Pie chart output of tweets segregated by polarity

V. ADVANTAGES AND DISADVANTAGES

A. Advantages:
The system provides easy stress detection based on very little input by the user, while also being non-invasive in nature. The system will also help in detection of victims of cybercrime; e.g.: the recent Blue Whale Game fiasco. Sometimes, people find it difficult to open up to even those closest to them. In such cases, a website would be a simple solution to their problems.

B. Disadvantages:
The user should possess a Twitter account and be an active user of the same; otherwise the results obtained by the system would be inaccurate. Moreover, the application is not reliable in cases of very critical subjects.

VI. CONCLUSION

In today's world, where mainly the youth and almost all of the population is suffering from surmounting stress, be it because of peer pressure, work load or other domestic tensions; it is very crucial to have a reality check about how stressed a person really is.

It is because of this reason that timely detection and prevention of stress is a dire need. We have come up with this project which assists people in scrutinizing the problem of stress.

This project will be very beneficial for those who are not so comfortable in opening up about their problems to others. It will help these people get a reality check and may prompt them to reach out and get medical help, just based on their social interactions.

We have utilized both human as well as machine learning and applied the concepts of Sentiment Analysis. The main characteristic of this system is its non-invasiveness and fast-oriented implementation in detecting stress when compared with the previous approaches.

REFERENCES

[1] Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K.Ratnam, "A Study of Data Mining Tools in Knowledge Discovery Process," International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.

[2] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, "Sentiment Analysis of Twitter Data," unpublished.

[3] "Sentiment Analysis of Twitter Data," Available: https://www.slideshare.net/sumit786raj/sentiment-analysis-of-twitter-data

[4] "Integrating RapidMiner into your application," Available: https://community.rapidminer.com/t5/Original-Rapid-I-Forum/Integrating-RapidMiner-into-your-application/td-p/7071