International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 12
ISSN: 2454-4248
363 – 369

Intelligent Data Monitoring and Controlling System for Health Related Social Networks

[1]N. Mounika
[1]PG Scholar, Department of C.S.E
[1] PDIT, JNTUA University
Andhra Pradesh, India
[1] *nallapanenimounika@gmail.com*

[2] K. Ramesh
[2] Associate Professor, Department of C.S.E
[2] PDIT, JNTUA University
Andhra Pradesh, India
[2] *kumbha.ramesh@gmail.com*

***Abstract*—** Depression is a worldwide wellbeing concern in view of healthcare. Now a days, social media became popular to allow the affected people to share their experience in the form of posts. These kinds of experiences are stored in the database and extracted and analyzed to give the precautions to the other people or to recall the drugs from the side effects, and other service improvements in their treatment regarding to a particular disease. In such cases depression-related social websites are helpful to monitor or get knowledge in various kinds of drugs, side effects and to share the user experiences. In this paper, we proposed a social media website to allow the users to share the experiences of a particular disease i.e. depression and their experience over on it. We used a weighted network model to represent the activities in the social networks. The proposed work has three steps. The first one is to monitor the user activity and followed by network clustering and the module analysis. The persons who likes a particular post comes under a group and those who contrasted belongs to other group. The stop word technique we have implemented in this work is helpful to avoid the misleading communication over the posts and for the efficient user interaction. The statistical analysis of this kind of user interactions are helpful in health networks to gain much knowledge about a specific disease. This approach will enable all the gatherings to take a part and for the future healthcare improvements to the patients suffering from a disease.

Keywords- *Datamining, online fora, depression, stop-words technique.*

*****

I. INTRODUCTION

Over 5 crore Indians suffering from depression, which is a major contributor to the global suicides occurred mainly in the middle to low income countries for example like in India in 2015 according to the study of World Health Organization (WHO) [1]. According to WHO, in its new global health estimates on depression for 2015 said while over five crore Indians suffered from depression, over three crore others suffered from anxiety disorders. The report titled 'Depression and Other Common Mental Disorders — Global Health Estimates' said over two-thirds of global suicides were in low- and middle-income countries like India in 2015. The WHO document said that 322 million people are living with depression worldwide and nearly half of them live in South East Asian and Western Pacific region, reflecting relatively large populations of India and China [3]. "The total number of people living with depression in the world is 322 million. The total estimated number of people living with depression increased by 18.4% between 2005 and 2015, it said. According to WHO figures, the total cases of depressive disorders in 2015 in India were 5 crores which was 4.5% of population in 2015 while total cases of anxiety disorders were 3 crores which was 3% of the population in the same year period. "Suicide occurs throughout the lifespan and was the second leading cause of death among 15-29 year olds globally in 2015," it said. Depression is the leading cause of disability worldwide, and is a major contributor to the overall global burden of disease, WHO said and asserted that more women are affected by depression than men and at its worst, depression can lead to suicide [4, 5]. Forums and social media websites dedicated to depression have recently sprung up for patients and healthcare workers to share their experiences from managing depression in their daily lives to their reactions to antidepressants. Such voluminous information can provide unlimited opportunities for patients, healthcare organizations, and industry to improve solutions through intelligent data mining, extraction, and analysis [2].

A social media network is a virtual networking environment that is composed of nodes and edges. Its contents can be modeled and extracted using computational tools that can map trends, formulate predictions, and assess user relationships. Graphical representation can visually represent the information. A sociomatrix can represent a social network's structure. Topological parameters such as node degrees and network densities can elucidate specific dynamics within a network and specific algorithms can map underlying information-rich structures. Identifying these clusters enables node- (or cluster) centered information mining. Such crucial data can help healthcare organizations, physicians, staff, and patients to improve services based on feedback from "smart" data mining of health-specific social media sites. There exists several methods in the literature that are used in collecting data from social media networks are lexicon-based, supervised classification, and concept extraction [6]. Other methods include the use of graph-based analysis [7], text-based analysis derived from a medical corpus [8], and topic-model statistical analysis [9]. Zhao *et al.* [10] used text-based analysis (length of posts, frequency of certain words) and sentiment analysis to identify influential user's online cancer survival communities. In contrast, our approach combines weighted network models (to represent user activity), module (describing user interaction), and topographical (user activity) analysis with sentiment and text analysis to gain a greater understanding of

 ISSN: 2454-4248
363 – 369

user sentiment on antidepressants and identifying influential users, as well as raise potential flags on drug side effects. This paper is organized as follows. Section II describes the related and literature work and Section III describes the contributions of the proposed work and Section IV discuss about the results and conclusions are drawn in Section V.

II. LITERATURE STUDY

The proposed work is inspired from the work done by [1], in which the authors have contributed the following methods to get the antidepressants. The complete scenario was shown in Figure 1.

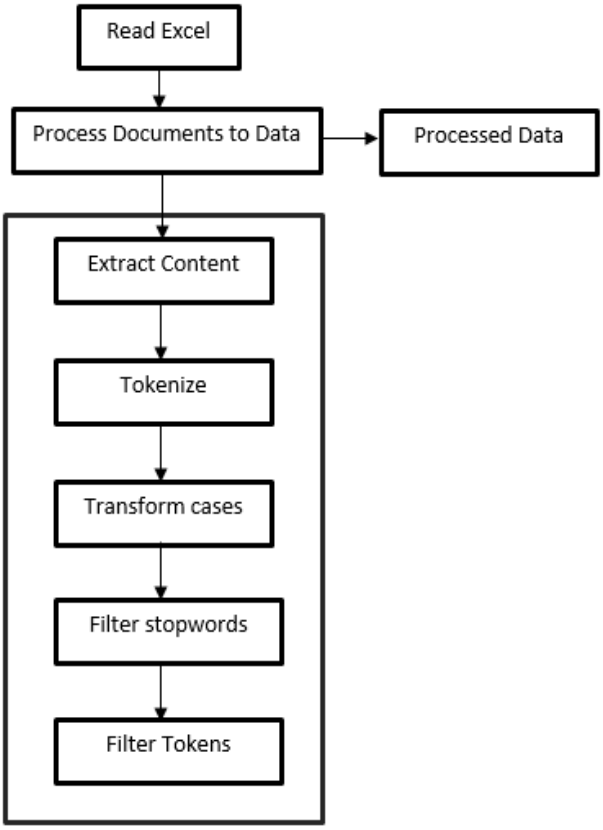

Figure 1: Processing of Rapid miner to get TF-IDF Scores.

The first step was to search for the most sought-after forums dedicated to depression. Our final list, which yielded the following chart, is of descending order. Thus, depressionforums.org was our choice. Secondly, A data collection, analysis, and processing tree was developed in Rapid miner (www.rapidminer.com) to discover the most frequent words (positive, negative, and side effects) to find their term-frequency-inverse document frequency (TF-IDF) scores within each post. The Figure 1 shows the data collection and processing tree. The dataset was uploaded ("Read Excel"), processed ("Process Documents to Data") using subcomponents ("Extract Content," "Tokenize," "Transform Cases," "Filter Stop words," "Filter Tokens," respectively) that filtered excess noise (misspelled words, common stop words, etc.) to ensure measurable variable uniformity. The result ("Processed Data") contained the final word list, with each word containing a specific TF-IDF score. The TF-IDF scores in each post were built based on a representative word set present throughout the forum and reflects the posts' semantic content. Therefore, we viewed a TF-IDF vector as the semantic profile of each post. Consequently, various measures of similarities can be derived to reflect how close the semantic profiles of two posts are, e.g., Euclidian distance or correlation. Additionally, clustering analysis can be performed to identify groups of similar semantic profiles. They have used *k*-means clustering [11] to roughly group the semantic profiles of all posts from our forum, as a preprocessing step necessary for the network-based modeling.

In the next step network based modelling of forum posting is applied. Forum posting activity consisting of threads containing thousands of postings and replies were modeled into a large user centric network. The modeling approach aimed at reflecting user interactions while simultaneously considering the posts' semantic content. The nodes in our network correspond to forum users and connecting directed edges correspond to two different types of interactions: *direct* and *context* interactions. Direct interactions correspond to direct user-to-user replies using the forum's "Reply" option. These interactions were modeled with bidirectional edges connecting the two corresponding nodes. This allowed us to model the mutual exchange of information between a poster and a direct replier. Context interactions reflect users posting within a specific thread (threads are topic-specific, and thread semantic content is homogeneous).The sample network nodes is shown in Figure 2.

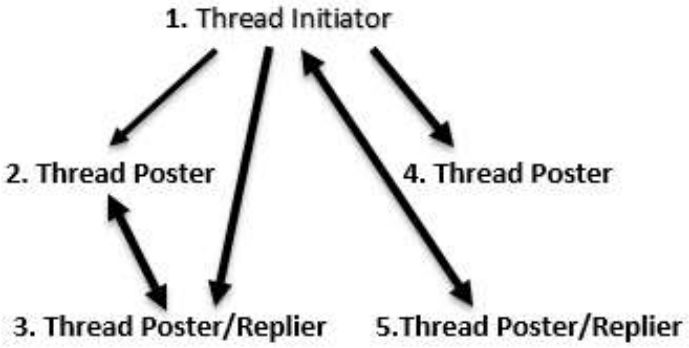

Figure 2: Sample network model with edges.

.

Consequently converted the forum posts into a large directional weighted network containing a number of densely connected units (or modules). Later that are significantly overrepresented within each of the modules found by the network partitioning algorithm. Finally they have proceeded to identify influential users in the retrieved modules. To this goal, we used the HITS, which is a method initially developed for web pages link analysis [12], [13].Information disseminates from authoritative nodes. Hubs link to authoritative nodes, and thus, they broker information flow within the network. This approach for identifying influential users has the advantage of considering both the network's structural properties, and the directionality of information flow.

Lise Getoor et.al. , considered many datasets of interest today are best described as a linked collection of interrelated objects. These may represent homogeneous networks, in which there is

 ISSN: 2454-4248
363 – 369

a single-object type and link type, or richer, heterogeneous networks, in which there may be multiple object and link types (and possibly other semantic information). Examples of homogeneous networks include single mode social networks, such as people connected by friendship links, or the WWW, a collection of linked web pages. Examples of heterogeneous networks include those in medical domains describing patients, diseases, treatments and contacts, or in bibliographic domains describing publications, authors, and venues. Link mining refers to data mining techniques that explicitly consider these links when building predictive or descriptive models of the linked data. Commonly addressed link mining tasks include object ranking, group detection, collective classification, link prediction and subgraph discovery. While network analysis has been studied in depth in particular areas such as social network analysis, hypertext mining, and web analysis, only recently has there been a cross-fertilization of ideas among these different communities. This is an exciting, rapidly expanding area. In this article, we review some of the common emerging themes.

M.E J.Newman et.al. Proposed that there has been considerable recent interest in algorithms for finding communities in networks| groups of vertices within which connections are dense, but between which connections are sparser. Here we review the progress that has been made towards this end. We begin by describing some traditional methods of community detection, such as spectral bisection, the Kernighan Lin algorithm and hierarchical clustering based on similarity measures. None of these methods, however, is ideal for the types of real-world network data with which current research is concerned, such as Internet and web data and biological and social networks.

Andrew Y.Ng,Alice X.Zheng et.al. Proposed that Kleinberg HITS and the Google PageRank algorithms are eigenvector methods for identifying "authoritative" or "influential" articles, given hyperlink or citation information. That such algorithms should give reliable or consistent answers is surely a desideratum, and in [10], we analyzed when they can be expected to give stable rankings under small perturbations to the linkage patterns. In this paper, we extend the analysis and show how it gives insight into ways of designing stable link analysis methods. This in turn motivates two new algorithms, whose performance we study empirically using citation data and web hyperlink data.

David et.al. Have given a snapshot of a social network, can we infer which new interactions among its members are likely to occur in the near future? We formalize this question as the link prediction problem, and develop approaches to link prediction based on measures for analyzing the \proximity" of nodes in a network. Experiments on large co-authorship networks suggest that information about future interactions can be extracted from network topology alone, and that fairly subtle measures for detecting node proximity can outperform more direct measures.

Felix Naumann et.al, considered an abundance of biological data sources contain data on classes of scientific entities, such as genes and sequences. Logical relationships between scientific objects are implemented as URLs and foreign IDs. Query processing typically involves traversing links and paths (concatenation of links) through these sources. We model the data objects in these sources and the links between objects as an object graph. We identify a set of interesting properties for links and paths, such as out degree, image of a link, cardinality of data objects and links, the number of distinct objects reached by some links, etc. Analogous to database cost models, we use statistics from the object graph to develop a framework to estimate the result size for a query on the object graph. Analogous to training and testing, we use sampled data from queries to estimate the result size. We validate our models using data sampled from four NIH/NCBI data sources. Our research provides a foundation for querying and exploring data sources.

Jan Noessner et.al. Have been argued that linked open data is the major benefit of semantic technologies for the web as it provides a huge amount of structured data that can be accessed in a more effective way than web pages. While linked open data avoids many problems connected with the use of expressive ontologies such as the knowledge acquisition bottleneck, data heterogeneity remains a challenging problem. In particular, identical objects may be referred to by different URIs in different data sets. Identifying such representations of the same object is called object reconciliation. In this paper, we propose a novel approach to object reconciliation that is based on an existing semantic similarity measure for linked data. We adapt the measure to the object reconciliation problem, present exact and approximate algorithms that efficiently implement the methods, and provide a systematic experimental evaluation based on a benchmark dataset. As our main result, we show that the use of lightweight ontologies and schema information significantly improves object reconciliation in the context of linked open data.

M.E.J Newman worked on the algorithms for finding communities in networks| groups of vertices within which connections are dense, but between which connections are sparser. Here we review the progress that has been made towards this end. We begin by describing some traditional methods of community detection, such as spectral bisection, the Kernighan {Lin algorithm and hierarchical clustering based on similarity measures. None of these methods, however, is ideal for the types of real-world network data with which current research is concerned, such as Internet and web data and biological and social networks.

III. PROPOSED WORK

This work enabled us to accurately represent user interactions by relying on the data's semantic content. In this paper, enabled us to accurately represent user interactions by relying on the data's semantic content. Analysis to characterize user interactions and extract further knowledge from users posts. The main advantages of our work is that improving health solutions of patients suffering from depression and the affected person to share their experiences.

 ISSN: 2454-4248
363 – 369

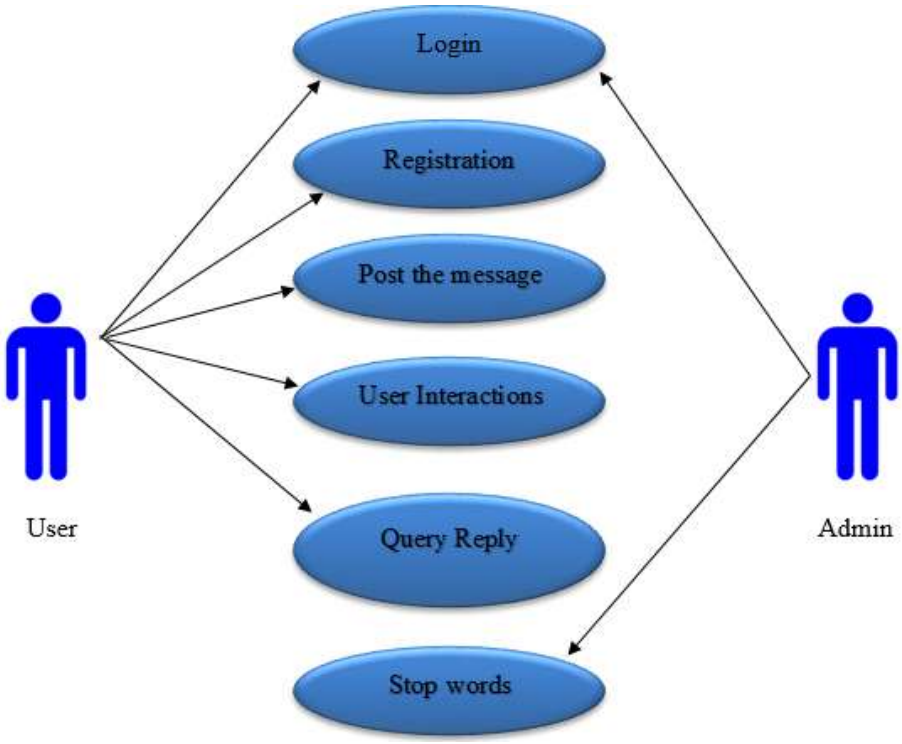

Figure 3: Use case diagram of our work

We have used content network site pages are analyzed and the search engines decide to assign a category to the page. The category is chosen by the software from a finite list. The keywords and ad text of the content campaign are similarly analyzed, and each ad group is assigned a category from the same list. We have used Semantic search seeks to improve search accuracy by understanding the searcher's intent and the contextual meaning of terms as they appear in the searchable data space, whether on the Web or within a closed system, to generate more relevant results. The work is contributed in 5 steps.

1. User Interface Design

The important role for the user is to move login window to social network. This is created for the security purpose. In this login page we have to enter login user id and password. It will check username and password is match or not (valid user id and valid password). If we enter any invalid username or password we can't enter into login window to data owner window it will shows error message. So we are preventing from unauthorized user into the login window to social network. It will provide a good security for our project. So server contain user id and password server also check the authentication of the user. It well improves the security and preventing from unauthorized data owner enters into the network. In this work we used JSP for creating design. Here we validate the login user and server authentication.

2. User Upload Posts

Social media, ranging from personal messaging to live fora's, is providing limitless opportunities for Users to discuss their experiences with their personal and other experiences. It is also providing limitless opportunities for companies to receive feedback on their products and services. Most of the companies are already looking at social network monitoring as a top priority within their IT departments, potentially creating an opportunity for rapid dissemination and feedback of products and services to optimize and enhance delivery, increase turnover and profit, and reduce costs. Information collection, and sharing in the experience.

3. Admin Analyzing Posts

Initial exploratory analysis was performed to determine structures based on user opinions among the posts. The results were a compilation of user's clusters and their correlated opinion of the posts. Subsequent analysis was used to determine influential users among the members. These findings can open new avenues of research into rapid data collection, feedback, and analysis that can enable improved outcomes and solutions for public users and important feedback from admin.

4. Admin Block Posts

A mixed consensus with regard to admin depends on individual patient outcome to solutions. The nature of a social media platform can result in individuals with different outcomes, based on various individual factors and circumstances. Despite such factors, we were able to sift through the data and find positive and negative sentiment, which was later confirmed by the research that emerged regarding Users effectiveness.

5. User View Posted Information

After user login interface the user shares the opinion about in Forum. It may be positive or negative. Many users can login into the forum and the can say their experience. The influential user can say their opinion about the particular topics. It is useful for the awareness of general information. The complete architecture of the proposed work is shown in Figure 4.

 ISSN: 2454-4248
363 – 369

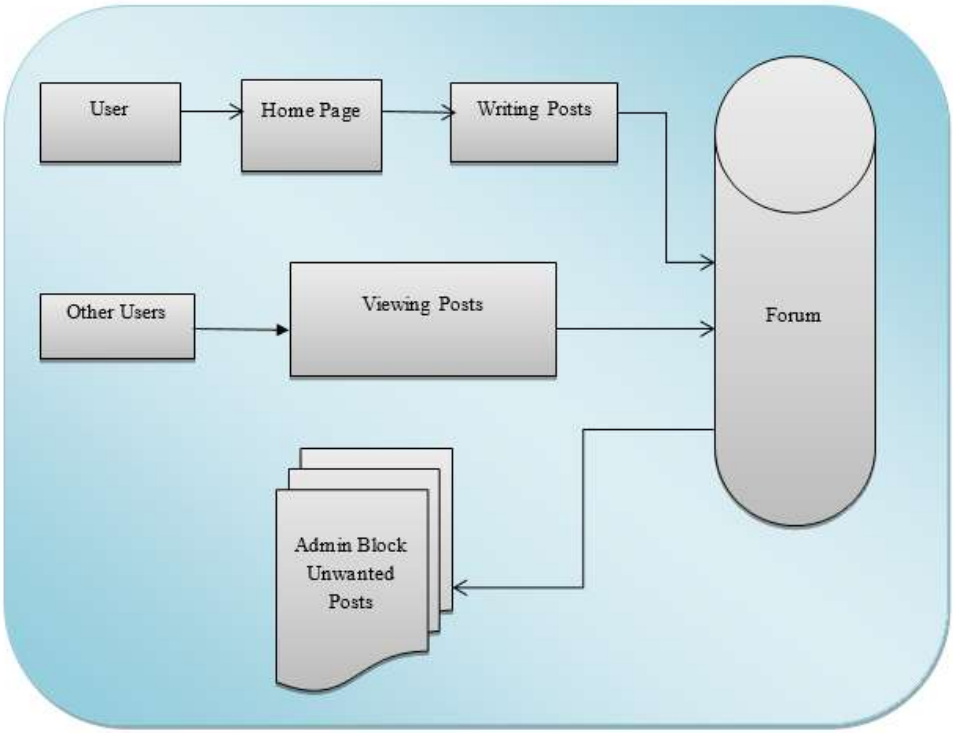

Figure 4: System Architecture

IV. RESULTS & DISCUSSION

The proposed work is inspired from the work done by [1]. In this we have concentrated to minimize the content which is irrelevant to a particular topic. For example usage of abusive words or irrelevant words can be monitored and cannot be used as we have used this as stop words. For example, if the application is fora about the depression related disease, heartache, stomachache are the words considered as irrelevant and the abusive words regarding a particular medicine or a person will be strictly controlled using our approach. Initially the user can login to our fora by the authenticated details. If not he can register providing the details of the person. The home screen contains list of posts that are shared by different community regarding to a particular medicine or their experience over a disease. They can also suggest the symptoms of the disease and the proper treatment to that disease along with the best available hospitality including the address locations or any other related images to this post. The users can give a reply to the post by sharing the other's view on the previous posts. The stop words technique can be used to prevent the misleading communication between the users. The viewers can view the post and they can like the post or dislike the post. The number of likes or votes can be used to identity the group of users that are interested in or faced a similar kind of experience. We have shown few of the screen shots of our work in the following Figures. Figure 5 shows the screen of admin, where he can add the stop words in to the existing list. Once the word can be added in to the database, the word can no long available for any user. I have added the **heartache** as a word in to the list, since we have considered this fora is about depression related.

heartache
python
C++
Java

Figure 5: Sample list of stop words stored in DB.

The following is the home page of our web-site. The post description is stored in the center followed by the person who posted and preceded with the number of views, votes, and any other reply to a particular post. For example, a reply for a first post is given in the Figure 7. The corresponding reply also

 ISSN: 2454-4248
363 – 369

Figure 6: Home Page of our forum

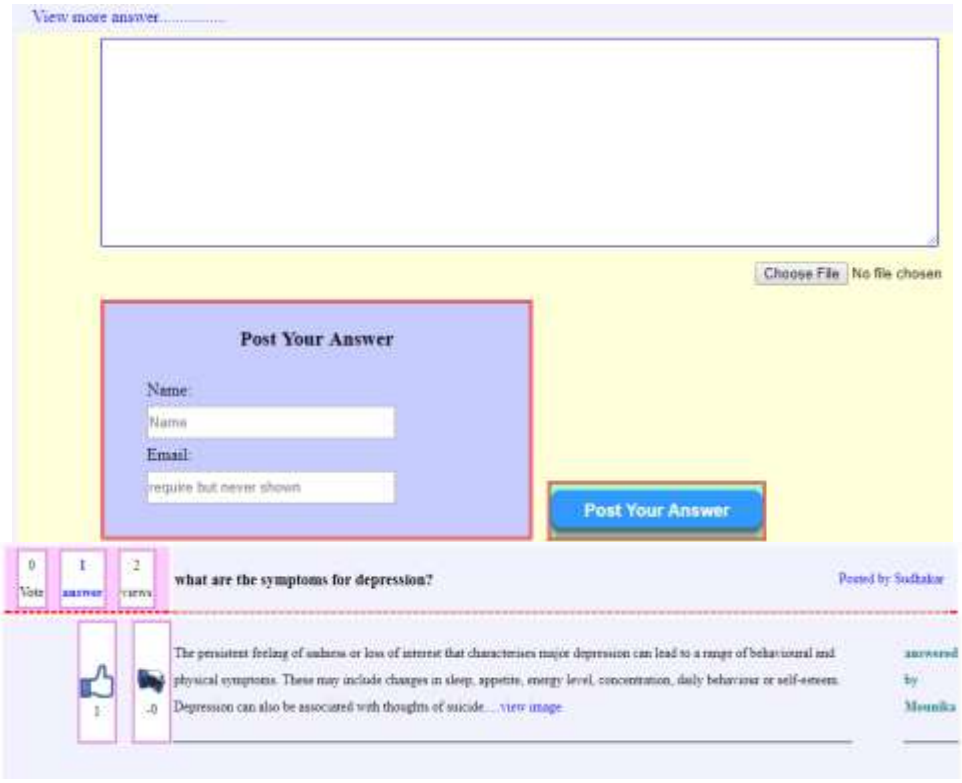

Figure 7: User Interaction for the post

V. CONCLUSION

Depression is the main source of disability, a noteworthy contributor to the world wide burden of several diseases and it is affecting 300 million individuals around the world. Untreated depression has been connected to several issues from stroke to coronary diseases, two of the main illness and cause death in the year 2013. A social network is a virtual system that is a made up of nodes and edges having huge content in it. These contents can be modeled and extracted using numerous tools that can map trends, and formulate predictions and give the relationships between the users. Graphical portrayal give the clear information about these user interactions. In this paper we have contributed 3 things. First one is to monitor the user activity and followed by network-Clustering and the module analysis. The persons who likes a particular post comes under a group and those who contrasted belongs to other group. The stop word technique we have implemented in this work is helpful to avoid the misleading communication over the posts and for the efficient user interaction. The statistical analysis of this kind of user interactions are helpful in health networks to gain much knowledge about a specific disease. This approach will enable all the gatherings to take a part and for the future healthcare improvements to the patients suffering from a disease. To conclude using these kind of intelligent datamining systems can greatly improve the quality of healthcare systems with less time and cost.

 ISSN: 2454-4248
363 – 369

REFERENCES

[1] WorldHealthOrganization.(2015,Jan.5).DepressionFactSheet No.369.[Online].Available:http://www.who.int/mediacentre/factsheets/fs369/en/

[2] Altug Akay et.al. "Assessing Antidepressants Using Intelligent Data Monitoring and Mining of Online For a" in IEEE Journal of Biomedical and Health Informatics, Vol. 20, No. 4, July 2016

[3] A. J. Ferrari et al., "Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010," PLOS Med., Nov. 2013.

[4] Web MED. (2015, Jan. 6). Depression Health Center: Untreated Depression. [Online]. Available: http://www.webmd.com/depression/ guide/untreated-depression-effects

[5] World Health Organizations. (2015, Jan. 7). The Top 10 Causes Death. [Online]. Available: who.int/media Centre/factsheets/fs310/en/

[6] J. L. Wu, L. C. Yu, and P. C. Chang, "Detecting causality from online psychiatric tests using inter-sentential language patterns," BMC Med. In format. Decision Making, vol. 12, 2012.

[7] F. Moradi, A. M. Eklund, D. Kokkinakis, T. Olovsson, and P. Tsifas, "A graph-based analysis of medical queries of a swedish health care portal," in Proc. 5th Int. Workshop Health Text Mining Inf. Analysis, Gothenburg, Sweden, Apr. 26–30, 2014, pp. 2–10.

[8] I. S. Bedmar, R. Revert, and P. Martinez, "Detecting drugs and adverse events from Spanish health social media streams," in Proc. 5th Int. Workshop Health Text Mining Inf. Analysis, Gothenburg, Sweden, Apr. 26–30, 2014, pp. 106–115.

[9] M. Paul and M. Dredze, "Discovering health topics in social media using topic models," PLOS One, vol. 9, pp. e103–e408, 2014.

[10] K.Zhao et.al. "Finding influential users of online health communities: A new metric based on sentiment influence," J. Amer. Med. Inform. Assoc., vol. 21, pp. 212–218, 2014.

[11] T. Kanungo et al., "An efficient k-means clustering algorithm: Analysis and implementation," Pattern Anal. Mach. Intell. Trans., vol. 24, no. 7, pp. 881–892, 2002.

[12] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," J. ACM, vol. 46, no. 5, pp. 604–632, 1999.

[13] J. M. Pujol, R. Sanguesa, and J. Delgado, "Extracting reputation in multi agent systems by means of social network topology," in Proc.1st Int. Joint Conf. Auton. Agents Multi agent Syst., 2002, pp. 467–474.

Author's Profile:

I, N.Mounika is a PG scholar in the school of Computer science and Engineering in PDIT, JNTUA. She received her Bachelor's degree in 2015 from JNTUA University. I am very interested to learn the new technologies and to do various research works and other activities. I have attended various national and international conferences and workshops. My research interests are big data Analytics and Recommendations Systems and Data mining.

Mr. K. Ramesh is currently working as Associate Professor at the school of Computer Science and Engineering in PDIT, JNTUA. He received his master's degree from Sathyabama University in 2007 and Bachelor's degree from JNTUA in 2003. He has 10 years of teaching experience and taught various subjects in computer science stream and organized various national conferences and workshops in the organization. His research interests are cloud computing, computer vision and Computer Graphics and knowledge engineering.