

Balanced Health Monitoring System using Stop Words for Social Big Data Applications

Chatakunta Praveen kumar¹, Dr B.V.Ramnaresh yadav²,

1.Department of Computer Science and Engineering, Institute of Aeronautical Engineering, Autonomous, Dundigal, Hyderabad-500043 T.S, 526.praveen@gmail.com¹,

2.Department of Computer Science and Engineering, JNTUH College of Engineering, Hyderabad, TS, bvramnaresh@gmail.com

Abstract

As per health care perception depression is a wellbeing concern at global range. Today social media allows the people those who affected can share their experiences through posts. Such experiences are stored in database and can be extracted, analyzing to assist the precautions for drugs from side effects, and other service improvements in their particular disease treatment. In those aspects, social websites related to depression are helpful to extract knowledge or monitoring various types of drugs and its side effects and also for sharing their experiences on depression. We have taken a weighted edge network model for representing social networks activities. The proposed work undergoes with the three steps. The first step is user activity monitoring, followed by network clustering and module analysis. Whoever the person like a specific posts belongs to a group and those who are not are belong to other group. We implemented the stop word technique here which is helpful in avoiding the misleading communication on the posts and efficient interaction of user. The statistical analysis of such user interactions are beneficial for health networks to acquire more knowledge on particular disease. This approach enables us all the gatherings took a part and for healthcare improvements in future to the patients of that disease.

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

I. Introduction

Depression is a disease and said as a major contributor to the world wide suicides that happened especially in middle to low income countries. For instance, like India based on the study of World Health Organization (WHO) [1]. As per WHO basis of 2015 in its recent global health estimation about discouragement in 2015 expressed that around 5 crore Indians are gloom enduring one's, though three crore individuals are experienced uneasiness disorders. The report is entitled as "Misery and Other Common Mental Disorders — Global Health Estimates" said in regards to 66% of worldwide suicides are occurring particularly in low and center wage nations like India in 2015. The archive of WHO demonstrates that 322 million individuals are with despondency and roughly half individuals from them live in South East Asian and Western Pacific locale, reflecting generally vast populaces of India and China [3]. Add up to individuals with gloom in world are 322 million. Among them there assessed individuals with gloom expanded by 18.4% in the vicinity of 2005 and 2015. As for each WHO figures in 2015 the quantity of depressive issue cases are 3 crores of populace while 3 crore of populace is with tension issue. It likewise said that "Suicide happens all through the life expectancy and was the second driving reason for death among 15-29 years old internationally in 2015". Sadness is the primary factor for causing incapacity in worldwide and is prosperity worry to the general worldwide weight of infection, WHO said and attested ladies are influenced by wretchedness than men and is prompts suicides [4, 5]

Gatherings and web-based social networking sites are for discouragement by sharing the encounters of human services laborers and patients to oversee in their standard lives and reacts to antidepressants. Such enormous information offers more noteworthy possibilities for patients, medicinal services associations, and industry to improve arrangements through shrewd information mining, extraction and analysis [2]. A virtual web-based social networking organizing condition comprises of hubs and edges. Its substance are demonstrated and separated with

the help of computational tools are trendy which formulate expectations and buildup the user relationships. The information will be represented visually by graphical representation. A informal community's structure is spoken to by a socio framework. Topological parameters like hub degrees and system densities clarify specific progression inside a system and a specific calculation watches out for fundamental data arranged structures. Finding those bunches empowers hub (or group) focused information mining. Such vital information helps the group for enhancing administrations that depends on input from "brilliant" information mining of wellbeing related web-based social networking sites. There are different strategies in writing that aides in information gathering from web-based social networking systems are lexion-based, directed order, and idea extraction [6]. The rest of techniques utilize diagram based investigation [7], content based examination got from a medicinal corpus [8], and a theme display factual examination [9]. Zhao et al. [10] utilized content based analysis (posts length, certain words recurrence) and conclusion investigation to recognize compelling client's online malignancy survival groups. Conversely our approach incorporates weighted system models (for client action portrayal), module (describing client cooperation) and topological (user movement) examination with assumption and content investigation to get great comprehension of client supposition on antidepressants and finding persuasive clients, and reacts on drug side effects. Our work is organized as follows. The related and literature work depicted under Section II and the proposed work and its contribution described under Section III and results and discussions under Section IV and conclusions is depicted under Section V.

II. Literature Study

The proposed work is inspired by the work that is done [1], where the authors contributed the below methods to find antidepressants. The overall scenario is shown in Figure 1:

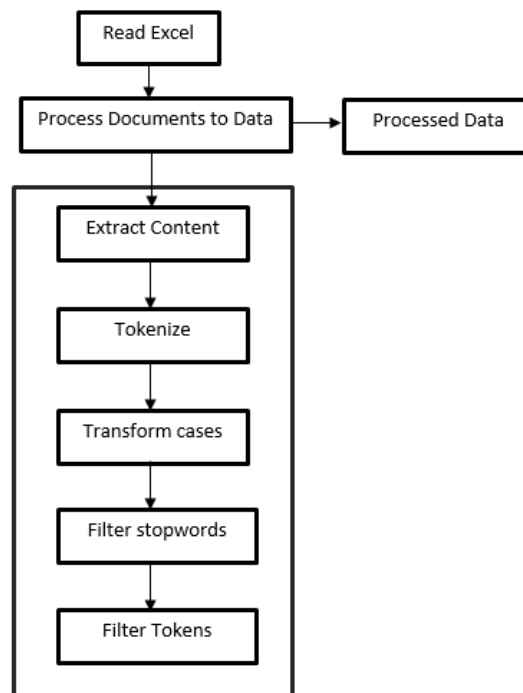


Figure 1: Processing of Rapid miner to get TF-IDF Scores.

The initial step was a search for depression dedicated forums. Our last list, which produces the below chart of descending order. Hence we choose depressionforums.org. After that collecting, analyzing and processing tree of data

was generated in Rapid miner(www.rapidminer.com) to recognize the most happening words(positive,negative and reactions) to get their Term-Frequency-Inverse Document Frequency) scores inside each post. The Figure 1 shows the accumulation and preparing of a tree. The dataset was transferred ("Read Excel"), handled ("Process Document to information") with the utilization of subcomponents("Extract Content", "Tokenize", "Change Cases","Filter Stop words", "Channel Tokens" individually) that channels additional commotion (incorrectly spelled words, basic stop words and so on.) to have variable measures consistency. The yield ("Processed Data") has the last word list; with each word has a specific TF-IDF score. The TF-IDF scores in each post that constructed was subject to an agent word set in whole discussion and mirrors the semantic posts content. Subsequently, we demonstrated a TF-IDF vector as each post's semantic profile.

Significantly, much similarity measure can be inferred to demonstrate how shut the two post's semantic profiles are, as Euclidian separation or correlation. In expansion to this cluster investigation will performs to discover the gatherings of same semantic profiles. They utilized k-implies clustering[11] to amass every one of the posts semantic profiles of our discussion as a required preprocessing venture for organize based displaying. Later the further system based demonstrating venture of posting gathering is connected. The movement of gathering posting having of strings with a large number of postings and reactions were outlined as a high client driven system. This demonstrating approach going for showing client communications by taking posts semantic substance into consideration. Our organize hubs compare to gathering clients and associates coordinated edges relate to direct and setting interactions. The client to-client answers with the utilization of discussion's "Answer" alternative. Those collaborations are said to be immediate cooperations and are demonstrated with the edges of bidirectional interfaces the comparing nodes. This influenced us to shared information to trade between a publication and a direct replier. The clients posting inside a specific thread(threads might be theme related and string semantic substance is same).are reflected by context interactions. The network nodes sample is shown in figure 2.

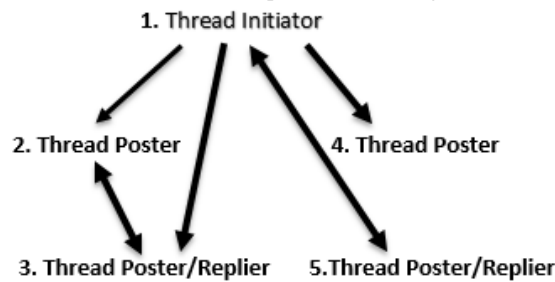


Figure 2: Sample network model with edges.

Significantly, the forum posts are converted into a huge directional weighted network with multiple densely connected units(or modules). The network partitioning algorithm helps in detecting the nodes' over representation within every module. For this objective we utilized HITS, which is an at first created technique for website pages interface analysis[12][13]. Data disseminates from definitive hubs. There is a connection from center points to legitimate hubs thus they broke stream of data inside the network. This approach helps in finding influential users and benefit in considering the structural properties of both networks' and the direction of data flow.

LiseGetoor et.al took more datasets which are portrayed as connected accumulations of items which are interrelated. These speak to homogeneous sort systems with single protest write and interface compose and heterogeneous systems with different question and connection writes (other semantic information if conceivable). Single mode informal organizations like individuals interface by fellowship connections or WWW, a connected pages accumulation are the cases of homogeneous systems while restorative spaces delineating patients, maladies, medications and contacts, authors and so on are the instances of heterogeneous networks.

Link mining is one of information mining systems that remotely take these connections when creating prescient or enlightening models of connected data. In general connection mining exercises incorporate positioning items, discovering gatherings, aggregate order, interface forecast and revelation of sub chart. Where organize examination was considered in indicated regions like investigation of informal organization, hypertext mining and web examination just in later there have been a thoughts cross-treatment among those changed groups. This is a developing region. In this paper we inspected a few common emerging themes. Therecent considerable interests in algorithms were proposed by M.E Newman et.al for detecting the communities in networks. The connections within

the groups of vertices are denser while the associations between them are sparser. The advances checked on here which is tending towards this end. We begin by portraying some ordinary techniques in discovering group like unearthly division, the Kernighan Lin calculation and various leveled bunching on comparable measures premise. No strategy is same for true system information sorts with which introduce look into concerns like Internet and web information and natural and interpersonal organizations.

Kleinberg HITS and the Google PageRank algorithms was proposed by Andrew Y.Ng,AliceX.Zheng et.al. Which are eigenvector methods for detecting “authoritative” or “influential” articles, given information on hyperlink or citation? Those algorithms must give reliable answers is exactly a desideratum and in [10], we analyzed when they are expecting stable rankings under small perturbations to patterns linkage. In this paper we extend the analysis and shows how it gives insight into ways of stable links designing analysis methods which in turns motivates two more new algorithms using citation and web hyperlink data.

The social network’s snapshot was given by David et.al ,could we infer which recent interactions are likely going to occur in future? We formalize the above question as the link expected issue, and generate approaches to link expected that measure for network nodes proximity.Experiments on large co-authorship network assist that data regarding future interactions can be gathered from network topology alone and that fairly measures the proximity of node can outperform more direct measures.

Felix Naumann et.al, took more natural information sources that has information on logical substances classes like qualities and successions. The logical protest's intelligent connections are actualized like URL's and remote ID's. To navigate connections and ways (joins link) through these sources Query preparing is performed. We outline information questions in these sources and a protest chart is the connections between objects. We identify a gathering of fascinating properties for connections and ways like out degree, interface picture, information protest's cardinality and connections, the quantity of particular items came to by a few connections et cetera. Similar to database cost models; To build up a structure from question chart we utilized measurements for evaluating inquiry result estimate on protest diagram. Comparable to preparing and testing, to gauge the outcome measure we utilized examined information from inquiries. Our models are approved with the utilization of test information from NIH/NCBI information sources. Our exploration gives an establishment information sources questioning and investigating.

Jan Noessner et.al was argued that linked open data is the major advantage of semantic technologies for web since it gives more structured information with effective way of access than web pages. In this paper, we proposed another approach for object recognition that relies on prevailing semantic similarity measure for linked data. We choose a measure to the problem in object recognition, and presented precise and relevant algorithms that implement the methods and give a systematic experimental evaluation on benchmark dataset basis. As our result, we shown that the use of lightweight ontologies and schema information mainly enhances object recognition in the terms of linked open data.

III. Proposed Work

This empowered us to speak to correct client connections by relying upon information's semantic substance. In this paper, it empowered us for exact portrayals of client cooperations with the reliance on information's semantic content.Analyze the client collaborations and concentrate additional learning of client's posts.Our work key advantage is increasing health solutions for patients with depression and the person who is affected with depression can also share their experiences.

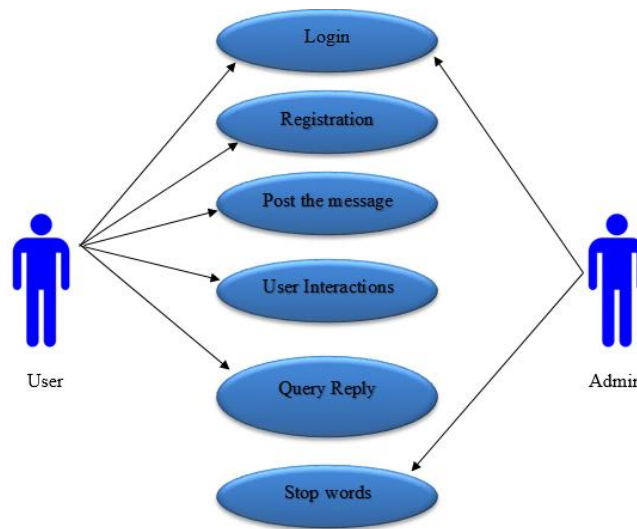


Figure 3: Use case diagram of our work

We analyzed content network site pages and search engines commit to allot a category for page. A Software chooses the category from finite list. Likewise the keywords and ad text of content campaign are analyzed and every ad group will be given a category of the same list. We used Semantic search seeks to enhance accuracy of search by knowing the searcher’s intension and the terms context as they are shown in data space that is searchable, either on the Web or inside a closed system, for generating some more related outcomes. The work contributes in 5 steps:

1. User Interface Design

The user’s important role is moving login window to social network. It is done for the purpose of security. In login page the login user id and password are entered. It checks for username and password are matched or not (valid user id and password). If user name and password are not valid we are unable to enter login window and it will displays error message to data owner window. Thus we prevent the login window to social network from unauthorized access. It made our project more secure. So server has userid and password and server also performs user authentication. It enhances the security and prevents illegal access by the user to enter the network. For our work JSP is used for designing. We validated the user’s login and authentication of server.

2. User Upload Posts

Social media, from personal texts to live fora’s providing boundless opportunities to clients to share their encounters. Moreover it is giving more opportunities to organizations for getting input on their items and services. Many organizations are currently focusing on interpersonal organization observing as their first need in their IT offices, and furthermore making a shot for getting fast criticism on their items and administrations to diminish and enhance conveyance, increment turnover and benefits and limit costs.

3. Admin Analyzing Posts

On the basis of user opinions on the posts the structures are determined by the initial exploratory analysis. The outputs are the user’s clusters compilation and their opinion on the posts. To determine the users who are influenced among the members subsequent analysis was used.

4. Admin Block Posts

A multi consensus relates to admin relies on every patient solutions. The platform of social media results in individuals with varied outputs depending on different individual factors and circumstances. Apart from those

factors we are able to move through the data and can collect favorable and unfavorable assumption, which was then chosen by inquire about that rose on client's adequacy.

5. User View Posted Information

Whenever the user login the interface and they share the opinion in Forum either positive or negative. Most of the users may login the forum and can share their experience. The user who is influenced can reveal their opinion about specific topics. It is beneficial for being aware about general information. The overall architecture of our work is shown in Figure 4.

IV. Results and Discussion

We have concentrated to reduce the content which is not matched with the topic. Like using abuse words or words those are irrelevant are monitored and can't be used because we considered this as stop word. For instance, let fora be the application regarding depression disease. Heartache, stomach pains are the words treated as irrelevant and abusive so that specific medicine or a person will be strictly controlled by this approach. Firstly the user will login to fora with authenticated details. If won't he can provide details for registration. The home screen has the posts lists which are shared by different people on specific drug or on disease. They are also able to assist the disease symptoms and better treatment for that disease, good available hospitality with the address locations or any post related images. The users can reply to the post by sharing the other's view on the before posts. To avoid misleading communication among users stop words technique will be used. Any viewer can view the post and can like or dislike the posts. The number of likes or votes is helpful in the users' group identification that those are interested in or faced the same experience. We show some of our work related screenshots in the following figures. Figure 5 shows the admin screen, where user is able for adding stop words into prevailing list.

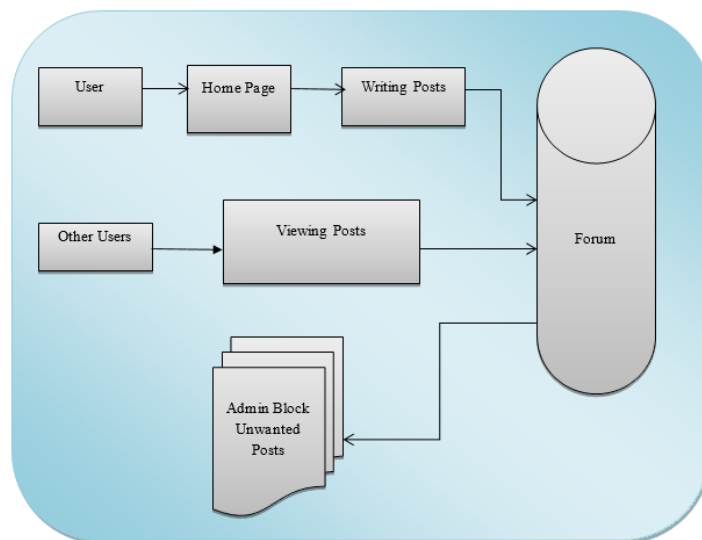


Figure 4: System Architecture

Once the user added the word into database, the word is unavailable for any user. Here I was added the word heartache into the list, because we took this fora related to depression.

heartache
Python
C++
Java

Figure 5: Sample list of stop words.

The below figure is our website’s home page. The description of the post is stored in the centered followed by the posted person and with the number of views; votes and any other reply proceeded to a particular post. For instance, a reply for first post is shown in figure 7 with the corresponding replies.



Figure 6: Home Page of our forum

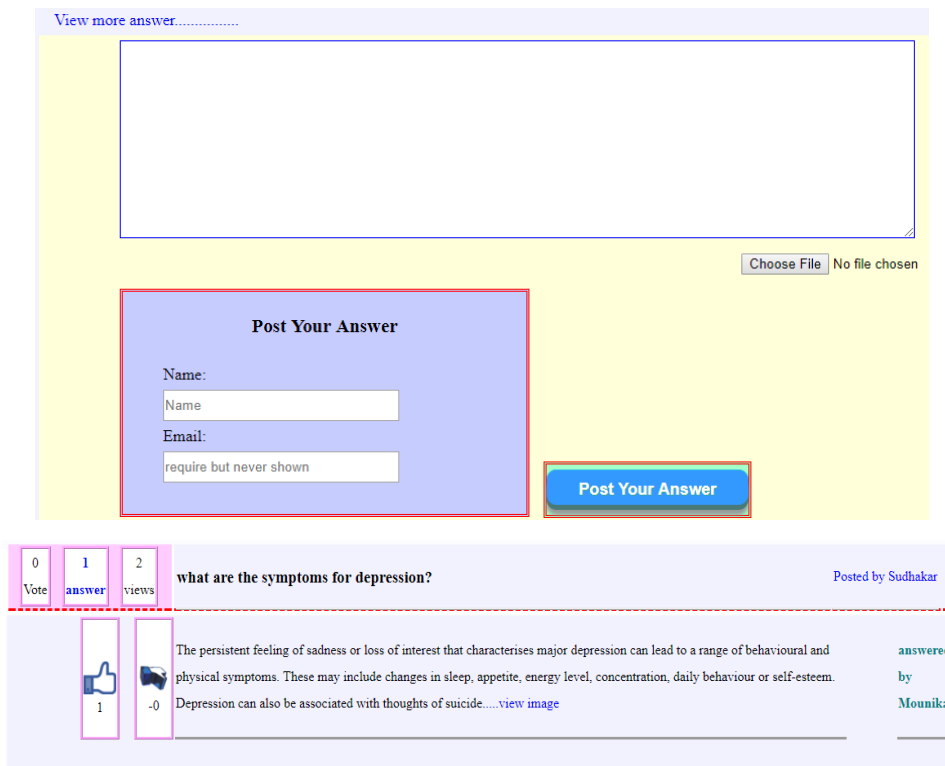


Figure 7: User Interaction for the post**V. Conclusion**

Depression is said to be the main reason for disability, a wellbeing and burden of various diseases in global range and 300 million people are affecting in world wide. The depression that can't be solved will leads to various issues from stroke to coronary ailments, both of primary sickness and cause demise in 2013. A virtual framework as informal community that is formed with hubs and edges has more substance. Those substance can be planned and assembled utilizing different apparatuses that are popular and define desires and builds the relationships among users. Graphical portrayal provides the information clearly on these user interactions. In this paper we contributed 3 aspects. One is the user activity monitoring and followed by network clustering ad module analysis. An individual who likes a particular post will be taken as a group while the other belongs to other group. To avoid misleading communication among users stop words technique is used and also for efficient interaction of user. The statistical analysis of such user interactions is beneficial in health networks to be aware of particular diseases. It enables us all the gatherings took part and for enhancements in healthcare to suffering people from disease in future. Eventually concluding that using these types of data mining systems can widely enhance the healthcare system's quality at cheaper cost with in time.

REFERENCES

- [1] World Health Organization. (2015, Jan. 5). Depression Fact Sheet No. 369. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs369/en/>
- [2] Altug Akay et al. "Surveying Antidepressants Using Intelligent Data Monitoring and Mining of Online For an" in IEEE Journal of Biomedical and Health Informatics, Vol. 20, No. 4, July 2016
- [3] A. J. Ferrari et al., "Weight of depressive issue by nation, sex, age, and year: Findings from the worldwide weight of sickness think about 2010," PLOS Med., Nov. 2013.
- [4] Web MED. (2015, Jan. 6). Sadness Health Center: Untreated Depression. [Online]. Accessible: <http://www.webmd.com/sadness/control/untreated-misery-impacts>
- [5] World Health Organizations. (2015, Jan. 7). The Top 10 Causes Death. [Online]. Accessible: [who.int/media-center/factsheets/fs310/en/](http://www.who.int/media-center/factsheets/fs310/en/)
- [6] J. L. Wu, L. C. Yu, and P. C. Chang, "Recognizing causality from online mental tests utilizing between sentential dialect designs," BMC Med. Informat. Basic leadership, vol. 12, 2012.
- [7] F. Moradi, A. M. Eklund, D. Kokkinakis, T. Olovsson, and P. Tsifas, "A chart based examination of therapeutic questions of a swedish medicinal services entrance," in Proc. fifth Int. Workshop Health Text Mining Inf. Examination, Gothenburg, Sweden, Apr. 26– 30, 2014, pp. 2– 10.
- [8] I. S. Bedmar, R. Return, and P. Martinez, "Identifying drugs and antagonistic occasions from Spanish wellbeing online networking streams," in Proc. fifth Int. Workshop Health Text Mining Inf. Examination, Gothenburg, Sweden, Apr. 26– 30, 2014, pp. 106– 115.
- [9] M. Paul and M. Dredze, "Finding wellbeing points in online networking utilizing subject models," PLOS One, vol. 9, pp. e103– e408, 2014.
- [10] K. Zhao et al. "Finding influential users of online health communities: A new metric in view of supposition influence," J. Amer. Med. Advise. Assoc., vol. 21, pp. 212– 218, 2014.

- [11] T. Kanungo et al., "An efficient k-implies bunching calculation: Analysis and usage," *Pattern Anal. Mach. Intell. Trans.*, vol. 24, no. 7, pp. 881– 892, 2002.
- [12] J. M. Kleinberg, "Legitimate sources in a hyperlinked domain," *J. ACM*, vol. 46, no. 5, pp. 604– 632, 1999.
- [13] J. M. Pujol, R. Sanguesa, and J. Delgado, "Removing notoriety in multi agentsystemsbymeansofsocialnetworktopology," in *Proc. 1st Int. Joint Conf. Auton. Operators Multiagent Syst.*, 2002, pp. 467– 474.
- [14] Chatakunta Praveen Kumar, Pole Anjaiah, "Mining Association Rules from No-SQL data bases using Map-Reduce Fuzzy Association Rule Mining Algorithm", *International Journal of Applied Engineering Research*, Volume 12, Number 21, pp. 10472-10476, Dec. 2017. (ISSN 0973-4562, Scopus Indexed)
- [15] K. Rajendra Prasad and C. Raghavendra, "Effective Mammogram Classification Using Various Texture Features", *Journal of Advanced Research in Dynamical and Control Systems*, Vol. 9, Sp– 12, pp. 2309- 2320, Dec. 2017. (Scopus Indexed)
- [16] K. Rajendra Prasad, I Surya Prabha, N Rajasekhar, M Rajasekhar Reddy, "Social Data Analytics by Visualized Clustering Approach for Healthcare", *International Conference on Advanced Computing and Intelligent Engineering*, C.V. Raman College of Engineering, Bhubaneswar, India, 2016. (Springer, SCOPUS indexed)

