International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Research on Smart Health Prediction System using Data Mining Technique

Mr. Shinde Manoj Bapurao[1], Prof. Kumbharde M. V[2]

[1]Student of Computer Engineering, SND College of Engg. Yeola, India
[2]Professor, Dept. of Computer Engineering, SND College of Engg. Yeola, India

---

Abstract – The Data offers many facilities to the end users such as software, organization and platform go on. In this proposed system, we study about the wisely mining knowledge of social media data. This information is commonly shared so healthcare is improves and costs is decrease using opinion which is generated by user. We suggest the investigation framework that to give attentions on side effects of drugs and focus on feedback of user. For the Improve health care some Clinical documents are mostly useful because it has are free-text data sources. The Clinical documents containing information related to symptoms and valuable medications. Further, we focus on the symptoms of the disease of patient. By taking the expert doctors suggestion, we list out the medication of the any disease according to the symptoms and we provide this medication or treatment to the user on our forum.

We can expand our research into Data and Knowledge mining of social media and takes the user's views on various drugs of Disease. This daily updated data helps to pharmaceutical industry, doctors, hospitals, and medical staff, and for patients for effective future treatments.

***Keywords*- Knowledge mining, Complex networks, social computing, Data mining, semantic Web.**

2. INTRODUCTION:-

In the last decade many advances have been made in Social media is presenting countless opportunities for patients to give their opinion about particular drugs and devices, many organizations can acquire feedback on their services and products. Medical organizations are giving first priority to social network tracking in their Information Technology departments, creating possibility for speedy distribution and products related comments. It offers to beautify and optimize shipping. This optimization helps to decrease charges and finally it increases profit and turnover. The Facts of Social media are gathering for bio investigation also mentioned additionally. Social media permits a VNE (virtual networking environment). We can extract Knowledge by using various computational tools. By using available networks, making a group of social media is also a one of the way of knowledge extracting. Social network is just like a structure, which is a collection of edges, and nodes these nodes and edges are connected with each other in numerous relationships.

The Data mining can be considered as the extraction of raw or useless data from huge databases, many applications like healthcare systems, market analysis get advantages by such mined data and also they came to know how to extract useful data from a big amount of data. Now humans will provide specific critiques on net website online, like a massive U. S. India range of people are using social web sites to tweet, chat. So this internet records become a totally critical element to peoples who want to get a few nice and negative statistics of very own field. Very Important Data are launch with discussion board method we have a completely treasured information related to fitness care because all subject matter going associated with most Disease and associated brought on and medicine. So we need to awareness in these valuable facts as assets. Data source was government fitness monitoring, newspaper articles, on-line buying, and so forth. None of them diagnosed a social forum that impact on network dynamic and Data.

2. RELATED WORKS:-

Jun Huan, Wei Wang, Jan Prins [2] algorithm FFSM for the frequent subgraph-mining problem. Comparing to existing algorithms, FFSM achieves substantial performance gain by efficiently handling the underlying subgraph isomorphism problem.

Alberto Ochoa, Arturo Hernndez [3] shows study about artificial Societies and Social Simulation using Ant Colony, Particle Swarm Optimization and Cultural Algorithms. Jan Noessner, Mathias Niepert, Christian Meilicke, and Heiner Stuckenschmidt. 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.

Mr. Pramod B. Deshmukh , Mrs. Aditi A. Kalia Mrs. Vrushali U. Utterwar , Mrs. Dipali M. Patil [5] shows study on Intelligently extracting knowledge from social media has newly attracted great interest from the Biomedical and Health Informatics community to simultaneously improve healthcare result and moderate costs using consumer-generated viewpoint that is from opinion mining.

T. Anisha , Mr. N. Thulasi [7] This approach can expand research into intelligently mining social media data for consumer opinion of various treatments to provide rapid, update information for the pharmaceutical industry, hospitals, and medical staff, on the effectiveness (or ineffectiveness) of future treatments.

S. Arul Kiruba, V. Pavithra, A. Saranya and B. Dharani. To improve the care of human health by consumers opinion from the forum posts is our aim. He propose a system for discovering and extracting a positive and negative symptoms and side effects of different drugs for lung cancer disease from influential user's forum posts. Based on these outcomes the drugs are rated and ranked based on TF-IDF.

Sonali More, P. P. Joshi [9] Survey on Social Media Data Mining Techniques for detecting useful knowledge from massive datasets like trends, patterns and rules. This survey discusses different data mining techniques used in mining social media.

3. PROPOSED SYSTEM:-

The system architecture of Smart Health Care system by Social Media Using Data Mining Technique which shows how the data can be accessed from user and how does process step by stem and also shows how system generate medication for user with also considering opinion mining by user.

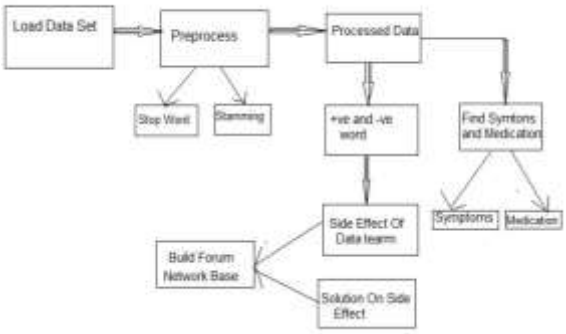

Fig 3.1 Proposed System Architecture

The system is divided as two phases or modules Pre-processing and Feature extraction. The figure gives the detail flow of system.

1) Get Dataset:-
2) Text Processing:-
3) Pattern taxonomy process:-
4) Find out positive and negative words:-
5) Calculate Symptoms and Medication:-

1) Get Dataset:-We are taking data from Disease forum site. This data is in the form of users tweets related to all Disease types and its treatments. In addition, this module provides the facility to live tweet and these tweets are taken as an input dataset for processing. These data is related to treatments and experience of particular drugs on a cancer so it is a raw data from this we have to find out the how many peoples give positive response. After collecting this dataset, data will send to next process i.e. Text processing.
2) Text Processing:-Text processing is a process in which we remove the stop words and text stemming
   - Stop words removal:-The words rather than natural language words are the stop words. In short, stop words are words, which are meaningless.
   - Text stemming:- Inflected and derived words are removed in stemming process.

These are removing on their stem base or root form. It generally a written word forms.

3) Pattern taxonomy process:- When the imported document having big size so it's become difficult to the text processor to process dataset. So to overcome this problem the document or dataset is break in to the small paragraphs. Now each paragraph is considered as separate document. Some terms are extracting from positive document and these terms are extracting in to each document. Formation. Along with this, social and group information is more valuable to cold users than to heavy users.
4) Find out positive and negative words:- Processed data is useful to find out positive and negative opinion as comments of user. To find this we used our own predefined dictionary in which we add the positive and negative words. To avoid the repeated words we used TF-IDF algorithm. Term frequency (TF) count the frequency of words found in the document it means it how many times occurs a particular word.IDF (Inverse document frequency) calculate the percentage of term occurs in main document.
5) Calculate Symptoms and Medication:- Calculate Symptoms and Medication By using Our predefined dictionary of symptoms words we can find out the symptoms from users tweets. By taking expert doctors opinion or using online Medical dictionary, we can list out the medication according to symptoms. We can add this information in our database for best result. So that the user can select their symptoms and came to know medication for cancer so that the health care is improve by using our forum because its update regularly by users and admin. Experimental results show that multi-view NMF is a preferable method for clinical document clustering. Moreover, we find that using extracted medication/symptom names to cluster clinical documents outperforms just using words.

4. SYSTEM OVERVIEW:-

A. Methodology used:-

1.1 Naives Bayes Algorithm:-

Naive Bayes classifier is probabilistic machine learning model that is used for classification task. The crux of the classifier is based on the Bayes theorem. Using Bayes theorem, we can find the probability of a happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence, it is called naive.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

1.2 SVM Algorithm:-

Support Vector Machine (SVM) is a supervised machine-learning algorithm, which can be, used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).

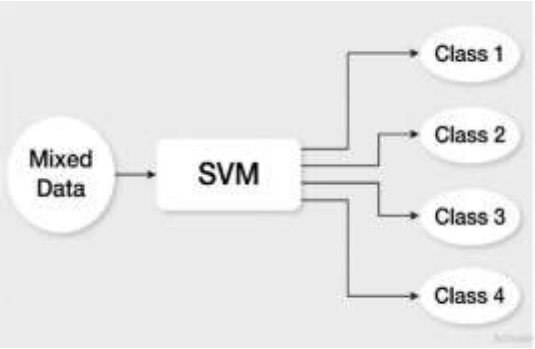

Fig 4.1 SVM algorithm

1.3 ANN Algorithm:-

ANN stands for *Artificial Neural Networks*. It is a computational model. That is based on structures

and functions of biological neural networks. Although, the structure of the ANN affected by a flow of information. Hence, neural network changes were based on input and output basically, we can consider ANN as nonlinear statistical data. That means complex relationship defines between input and output. As a result, we found different patterns. In addition, we call the ANN as a neural network.

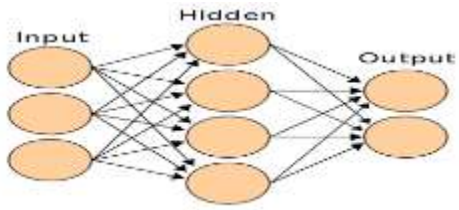

Fig 4.2 ANN algorithm

B. Requirement of Proposed system

1) Hardware Requirement:-

Hardware requirement for the system are given below:

1.Hard disk: 128 GB

2.RAM: 512 MB

3.Processor: Pentium and above

4.Input device: Keyboard and Mouse

5.Output device: Monitor

2) Software Requirements

Software requirement for the system are given below:

1. Operating System: Windows 7/Linux
2. Front End: HTML, JS, Bootstrap
3. Back End: MySQL, Oracle 10 g
4. UML Design: StarUml

C. Result Analysis

Fig 4.3 System Snapshots

The Results of this system Smart Health Prediction System Using Data Mining Technique contains patient treatment, positive Negative effect, Medication name, Symptoms detail, and as size of information present on the internet has taken a shape of the giant, it has become a necessity to increase the efficiency of the search. Data.in this paper we also adds some snapshots of our system as showing system result. The Total 100 post were found on metformin type 2 medicine .this post are further used for processing .raw data is proposed and word with tf-idf score are obtained these words are assigned with weight .the positive and negative opinion are obtained by using patient feedback. Obtained words are feed for clustering. Similar weighed words and grouped in one cluster. This shows the correlation between positive and negative words.

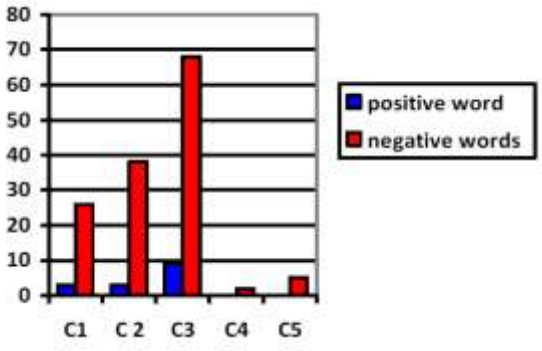

Fig 4.4 Positive and Negative words

Fig: chart of positive and negative word correlation.

After analysis network is built by using post threads and replies and sub graph are identified. The side effect list is generated of the overall post collected. From the modules obtained, the most discussed side effects are found by comparing their overall tf-idf score to the list obtained within the module. The t-test is applied to analyze the result according to the t-test performed weakness; diarrhea and tiredness are the most discussed side effects within the module.

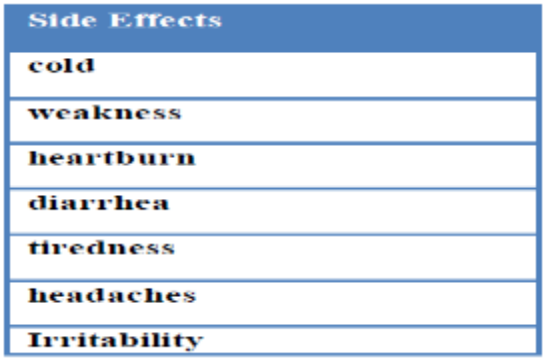

Fig of side effect of medication.

From the modules obtained, the most discussed side effects are found by comparing their overall tf-idf score to the list obtained within the module. The t-test is applied to analyze the result according to the t-test performed weakness; diarrhea and tiredness are the most discussed side effects within the module.

Modules	Side effect	p-value
Module 1	weakness	P< 0.05
	Diarrhea	P< 0.01
	Tiredness	P< 0.01
Module 2	Diarrhea	P< 0.01
	Tiredness	P< 0.01

Table: Side-Effect Frequency and Location in Selected Modules

We converted a forum focused on oncology into weighted vectors to measure consumer thoughts on the drug using positive and negative terms parallel another list containing the side effects. Social media can open the door for the health care sector in address cost reduction, product and service optimization, and patient care. These obtained results could be used as feedback loop for medicine manufacture companies. In future studies the post can be categorized on basis of their rankings or likes of post.

5. CONCLUSIONS

In this Proposed System, we build an integrating system to extract treatment, side effect symptom/medication names from unstructured/semi-structured Data from Disease forum. The overall system contains patient treatment, positive Negative effect, Medication name, Symptoms detail, As size of information present on the internet has taken a shape of the giant it has become a necessity to increase the efficiency of the search Data. Data and knowledge mining on data is very important because we are getting a valuable information which is not easily available, and all information are real time information.

The Next work will be research of We have focused on the Use the Document Clustering Results to Improve Medication Recommendation. Also Patient's Age/Gender/Demographical Information. In addition, our future work would be to extend the Statistical Analysis of Patient, inquiries, Medicine Records.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my profound gratitude and deep regard to my Project Guide Prof. M. V. Kumbharde, for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. Also the valuable help of Prof. I. R. Shaikh (HOD Comp. Dept.) and Prof. V. N. Dhakane (PG coordinator) who provided facilities to explore the subject with more enthusiasm. I express my immense pleasure and thankfulness to all the teachers and staff of the Department of Computer Engineering, S.N.D. College of Engineering and Research Center, Yeola, Nasik for their co-operation and support.

REFERENCES

1. J. Huan and J. Prins, Efficient mining of frequent subgraphs in the presence of isomorphism, in Proc. 3rd IEEE Int. Conf. Data Mining, Melbourne, Florida. FL,
2. June Almenoff, Joseph M. Tonning, Perspectives on the Use of Data Mining in Pharmacovigilance Leading Article Drug Safety 2005; 28 (11): 981-1007 0114-5916/05/0011-098.
3. J. Hans and M. Kamber, Data Mining Concepts and Techniques. Second ed.Burlington, MassMA, USA: Morgan Kaufmann, 2006.
4. Hand, Principles of data mining, Drug Safety, vol. 30,pp. 621622, Jul. 2007 , Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes, L. Li, and L. Janacek. Artificial societies and social simulation using ant colony, particle swarm optimization and cultural algorithms, Source: New Achievements in Evolutionary Computation, Book edited by: Peter Korosec, ISBN 978- 953-307-053-7, pp. 318, February 2010.
5. Jan Noessner, Mathias Niepert, Christian Meilicke, and Heiner Stuckenschmidt Leveraging Terminological Structure for Object Reconciliation L. Aroyo et al. (Eds.): ESWC 2010, Part II, LNCS 6089, pp. 334348, 2010. c Springer-Verlag Berlin Heidelberg 2010
6. A.Akay and Bjorn-Erik Erlandsoon,A novel data mining platform leveraging social media to monitor outcomes of Januvia. In IEEE EMBS Osaka, Japan,3-7 july,2013.
7. Si Yan and Yanliang Qi, Apply Text Mining to Advance Cancer Research, IJPMBS, Vol. 4, No. 2, April 2015.
8. T.Anisha1 , Mr.N.Thulasi2 Improving Health Care Based on opinion Mining Using KNN International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol. 5, Issue 4, April 2016. S.
9. Arul Kiruba, 1V. Pavithra, 1A. Saranya and 2B. Dharani ICON: Improving Care of Human Health using Intelligent Data Mining Advances in Natural and Applied Sciences 10(2) February 2016.
10. Sonali More, P. P. Josh Survey on Social Media Data Mining Techniques IJIRCCE DOI: 10.15680/IJIRCCE.2017. 0504220M.
11. Kiruthika1 , M.Kokilavani2 , N.Mouniga3 , R.Anitha Nithya4 Social Network Based Health Analysis Using User Trust Behavior Model IRJAET E - ISSN: 2454-4752 P - ISSN : 2454-4744 Vol 3 Issue 1 (2017) Pages 1660 - 1665 Received : 15.02.2017
12. Published: 26.02.2017 https://www.nih.gov/health information/nih-clinical-research-trialsyou/ list-registries.
13. https://www.who.int/news-room/fact-sheets/detail/disability-and-heal.