

Smart Life Clinic Using Data Mining Techniques

Mr. Manoj Shinde

Abstract— 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. Social media becomes much popular from the health care information and Biomedical. This information is commonly shared so healthcare is improves and costs is decrease using opinion which is generated by user. We suggest investigation framework that give attentions on side effects of drugs and also focus on positive and negative response. To Improve health care some Clinical documents are mostly useful because it's are free-text data sources. Clinical documents containing information related to symptoms and valuable medications. To extract a Data from large dataset its become a very popular because users get various ideas from this filtered data. All Data Mining and Knowledge mining become popular because user are process on data and getting information of different area like health, Social, etc. After data processing we focus on users positive and negative opinions. We count this opinions and find out which medication is good, to decide this we also find out the side effects of the 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 users views on various drugs of Disease. This daily updated data helps to pharmaceutical industry, doctors, hospitals, and medical staff, for effective future treatments..

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

I. INTRODUCTION

Social media is providing countless opportunities for patients to give their opinion about particular drugs and devices, many organizations can also acquire feedback on their ser-vices and products [1]-[3]. 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 [4]. Facts of Social media are gathering for bio investigation also mentioned additionally [5]. 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. 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?

Due to the fast increase and improvement of digital textual content statistics made to be had in latest few years, text Mining, Data mining, statistic mining and mining of knowledge becomes more popular to convert such information in to beneficial records and meaningful knowledge. 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. This extracted data is mostly useful to customer. Users Internet struggling is come to be a completely famous with net chat the use of a social websites. 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. This all methods are to be had to get information and system on facts but this information having some limitation due to the fact this all able to paintings on pattern model simplest method we just technique on small amount of statistics and get a small pattern of facts .For preceding strategies there are one-of-a-kind assets from which statistics are collect and developer work in this information, Data source was government fitness monitoring, newspaper articles, on-line buying, and so forth. None of them that impact on network dynamic and Data.

II. REVIEW LITRETURE

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

Many people's are works on the cancer treatments to improve health care, Si Yan and Yanliang Qi worked on cancer research they used the 3 various text mining tools namely MedLEE , HITEx and caTIES. These tools are developed by Columbia University and Harvard University these tools helps to extract medical information and diagnosis from pathology report. After extracting information they made one hypothesis which contains the specific type of cancer and drugs used for same.

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

efficiently handling the underlying subgraph

Alberto Ochoa , Arturo Hernandez [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 [4] 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, up- to-date 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[8] 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 users 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.

III. DRAWBACK OF EXISTING SYSTEM

1. Mining Structural Unstructured Data is complicated.
2. In Existing system Lack of knowledge in Social network analysis.
3. Existing system Implemented only for cancer disease.
4. Existing system cannot uploads prescription paper for patient.

IV. PROPOSED SYSTEM

Social media is presenting countless opportunities for patients to give their opinion about particular drugs and devices, many organizations can also acquire feedback on their services and products [1]-[3]. 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 [4]. Facts of Social media are gathering for bio Investigation also mentioned additionally [5]. Social media permits a VNE (virtuanetworking

isomorphismproblem.

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. Due to the fast increase and improvement of digital textual content statistics made to be had in latest few years, text Mining, Data mining, statistic mining and mining of knowledge becomes more popular to convert such information in to beneficial records and meaningful knowledge. 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. This extracted data is mostly useful to customer. Users Internet struggling is come to be a completely famous with net chat the use of a social websites. 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. Many methods had been working like consisting of link mining [6], type via hyperlinks [7], predictions primarily based on items hyperlinks [8], lifestyles, estimation [9], object [10], institution, and subgroup detection [11], and data mining[12]. By using persons feedback Link prediction, biochemical marketing, online discussion agencies (and rankings) permit for improvement of answers. 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 cancers and associated brought on and medicine. So we need to awareness in these valuable facts as assets. This all methods are to be had to get information and system on facts but this information having some limitation due to the fact this all able to paintings on pattern model simplest method we just technique on small amount of statistics and get a small pattern of facts .For preceding strategies there are one-of-a-kind assets from which statistics are collect and developer work in this information, 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. In the first degree of our modern have a look at, we hire exploratory evaluation the usage of our own dictionary which contains positive and negative words. This dictionary is used to evaluate correlations between user posts and their response on drugs. This response is calculated in the form of positive and negative opinion on drugs. In the second o level we version the users and their posts on the basis of side effects and symptoms. Our Proposed system also capable to generate medication for each disease and also helps to download prescription paper.

V. SYSTEM OVERVIEW

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.

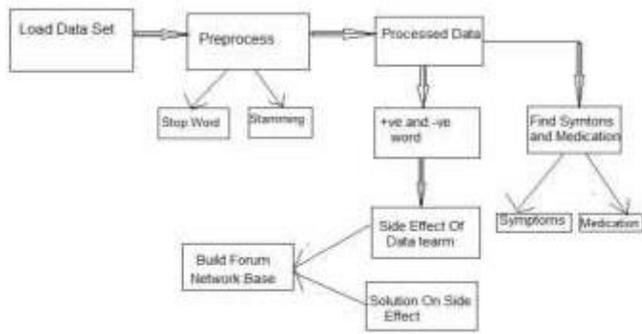


Fig. No 1. System Architecture

A. 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. Also 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 people give positive response. After collecting this dataset, data will send to next process i.e. Text processing.

B. 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.

C. Pattern Taxonomy Process

When the imported document having big size so its 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.

D. 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.

E. 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.

VI. SYSTEM REQUIREMENT

A. Hardware

requirements:

- Hard disk: 128 GB
- RAM: 512 MB
- Processor: Pentium and above
- Input device: Keyboard and Mouse
- Output device: Monitor

B. Software requirements:

- Operating System: Windows 7/Linux
- Back End: MySQL, Oracal 10 g
- UML Design: StarUml

VII. PROPOSED METHODOLOGY

A. Navies Bayes:-

Naive Bayes classifier is a probabilistic machine learning model that's 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.

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

B. Support Vector Machine:-

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).

C. Pattern Taxonomy Process

Knn is a non-parametric supervised learning technique in which we try to classify the data point to a given category with the help of training set. In simple words, it captures information of all training cases and classifies new cases based on a similarity.

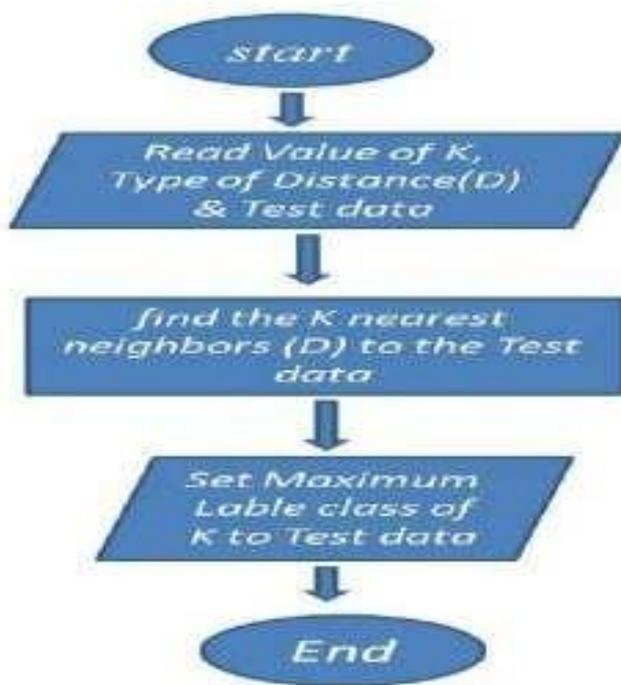


Fig.No 3 Working of KNN Algorithm

VIII. SYATEM ANALYSIS

A. TEXT PROCESSING:-

A mathematical model is a description of a system using mathematical concepts and language. The process of developing a mathematical model is termed as mathematical

- View patient and doctor information.
- Get opinion related medicines.
- Feedback System for Patient.
- Find Symptoms and Medication which is very useful to everyone.

B. SET THEORY:-

LET THE SYSTEM BE DESCRIBED BY S,

$$S = \{ I, P, R, O \}$$

WHERE,

S : IS A SYSTEM.

I: IS INPUT

R: IS SET OF RULES

O: FINAL OUTPUT.

$$I = \{ I1; I2; I3; I4 \}$$

WHERE,

I1 = ENTER PATIENT

I2 = ENTER DOCTOR INFORMATION

I3 = ENTER DISEASE SYMPTOMS.

I4 = FEEDBACK BY PATIENT.

P IS SET OF PROCEDURE OR FUNCTION OR PROCESSES OR METHODS.

$$P = \{ P1, P2, P3 \};$$

WHERE,

P1 = CHECK LOGIN FOR PATIENT.

P2 = SEARCHING FOR DOCTOR.

P3 = READ DISEASE SYMPTOM.

P4 = PREDICTING MEDICATION.

P5 = VIEW PATIENT LIST.

R IS SET OF RULES

$$R = R1, R2;$$

R1 = ENTER VALID INFORMATION.

R2 = MATCH THE DISEASE WITH SYMPTOMS.

$$O = \{ O1, O2, O3 \}$$

WHERE,

O1 = PREDICT MEDICATION.

O2 = DOWNLOAD PRESCRIPTION PAPER.

C. VENN DIAGRAM

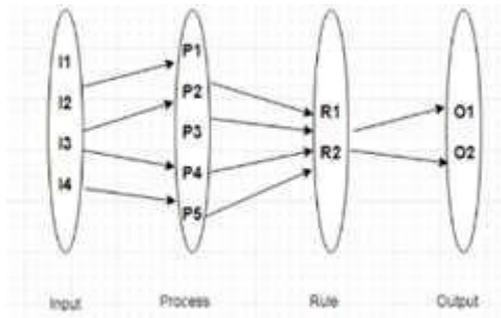


Fig.No 4 Shows Venn diagram

Where,

I1, I2, I3, I4 are inputs,
 P1, P2, P3, P4, P5 are
 process R1, R2 is rules
 And O1, O2, O3 are output.

D. ADVANTAGES OF PROPOSED SYSTEM

- New disease forum that including all disease related information and data flow
- Getting knowledge base information related all disease
- View patient and doctor information.
- Get opinion related medicines.
- Feedback System for Patient.
- Find Symptoms and Medication which is very useful to everyone.

IX. CONCLUSION

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 valuable information which is not easily available, and all information are real time information. Our Proposed system also effectively generates medication for every disease and also allows downloading prescription paper.

ACKNOWLEDGMENT

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, USA, 2003, pp. 549-552.
- [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. 2nd ed. Burlington, MassMA, USA: Morgan Kaufmann, 2006.
- [4] D. Hand, Principles of data mining, Drug Safety, vol. 30, pp. 621-622, Jul. 2007
- [5] Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes,
- [6] 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.
- [7] Jan Noessner, Mathias Niepert, Christian Meilicke, and Heiner Stuckenschmidt Leveraging Terminological Structure for Object Reconciliation
- [8] L. Aroyo et al. (Eds.): ESWC 2010, Part II, LNCS 6089, pp. 334-348, 2010. c Springer-Verlag Berlin Heidelberg 2010
- [9] A. Akay and Bjorn-erik Erlandsson, A novel data mining platform leveraging social media to monitor outcomes of Januvia. In IEEE EMBS Osaka, Japan, 3-7 July, 2013
- [10] Si Yan and Yanliang Qi, Apply Text Mining to Advance Cancer Research, IJPMBS, Vol. 4, No. 2, April 2015. T. Anishal, 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.
- [11] S. Arul Kiruba, IV. Pavithra, IA. Saranya and 2B. Dharani ICON: Improving Care of Human Health using Intelligent Data Mining Advances in Natural and Applied Sciences. 10(2) February 2016
- [12] Sonali More, P. P. Josh Survey on Social Media Data Mining Techniques IJIRCCE DOI: 10.15680/IJIRCCE.2017.0504220
- [13] M. 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 Published: 26.02.2017