

Knowledge Discovery and Analysis in Healthcare Using Data Mining

P. D. Thakre¹, Sheetal A. Tayade²

¹Professor, Department of Computer Science & Engineering, JCET, Yavatmal, India

²M. E. Scholar, Department of Computer Science & Engineering, JCET, Yavatmal, India

Abstract—On the Internet, where the number of choices is overwhelming, there is need to filter, prioritize and efficiently deliver relevant information in order to alleviate the problem of information overload, which has created a potential problem to many Internet users. The Proposed System solve this problem by searching through large volume of dynamically generated information to provide users with personalized content and services. The focus of this paper will be on pharmaceutical institutions and information regarding medicines. Frequent unbalanced distribution of medicines in healthcare institutions may result with serious threats to the health condition of a person. This provides health care institutions, medicine manufacturers and public in general to have insight to the information of the medicine. Thus the proposed system called Alternate Medicine System will be used to create a public awareness about alternative drugs for a particular medicine, availability of that alternative medicine in a locality. This is very useful for public.

Index Terms—Alternative Medicine, Data Mining, Healthcare, Predictive Analytics.

I. INTRODUCTION

Data mining technology provides a user oriented approach to novel and hidden information in the data. Valuable knowledge can be discovered from application of data mining techniques in healthcare system. Data mining in healthcare medicine deals with learning models to predict patient's disease. Data mining applications can greatly benefit all parties involved in the healthcare industry. For example, data mining can help healthcare insurers detect fraud and abuse, healthcare organizations make customer relationship management decisions, physicians identify effective treatments and best practices, and patients receive better and more affordable healthcare services. The huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making.

Alternate Medicine System research has made significant advances over the past decades and has seen wide adoption in electronic commerce. Recently, a variety of types of side information (e.g., social friends, item content) has been incorporated into Alternate Medicine System to further enhance their performance, especially the well-recognized problem of data sparsity. However, most of existing approaches have only investigated the value of a single type of side information at a time, such as social trust, friendship, or

item contents. It is necessary to build new theories, techniques and methods to exploit multi-dimensional (homogeneous and heterogeneous) side information to provide users with better personalized recommendations. At the same time, the large volume and variety of side data and the velocity of incremental updates in live systems provide challenges for the scalable mining and application of user preferences.

The explosive growth in the amount of available digital information and the number of visitors to the Internet have created a potential challenge of information overload which hinders timely access to items of interest on the Internet. Information retrieval systems, such as Google, DevilFinder and Altavista have partially solved this problem but prioritization and personalization (where a system maps available content to user's interests and preferences) of information were absent. This has increased the demand for Alternate Medicine System more than ever before.

II. LITERATURE REVIEW

In this section, an attempt has been made to briefly touch upon some of the mentionable reviews in this domain that has been conducted in last 9-10 years. Paper [12], is a reference work with survey in medical data and biomedicine, and covers elaborate reviews on machine learning models. However, the focus has been more machine learning and data mining oriented and in some cases the categorizations have been not very clear. In another paper [13], authors have discussed few motivating cases of data mining in healthcare along with the results. The paper focuses primarily on classification tasks in healthcare and discusses three models, namely rule induction, decision tree and neural network. In the seminal work presented in [2], the authors have given various cross sectional views (By tasks, By Models, By type of data) of the data mining use cases. However, the discussions have not been from the point of view of the health care entities. [3] Discusses some interesting aspects of data mining in healthcare, however, it focuses more on public health and medicine and coverage of machine learning models are also limited. Some of the surveys have been more specific. As an example [4], focuses on the application of a particular model (Neural Network) in medical science. Some of the surveys have been on narrowed scopes where they have reviewed particular health care problems as example prediction of heart disease [5], [6], and [7].

Different algorithms used in Different papers. The AIS algorithm is the first published algorithm developed to produce all large itemsets in a transaction database [10]. This algorithm

has targeted to discover qualitative rules. This technique is limited to only one item in the consequent. This algorithm makes multiple passes over the entire database. The SETM algorithm is proposed in [8] and motivated by the desire to use SQL to calculate large itemsets [11]. In this algorithm each member of the set large itemsets, L_k , is in the form $\langle TID, itemset \rangle$ where TID is the unique identifier of a transaction. Similarly, each member of the set of candidate itemsets, C_k , is in the form $\langle TID, itemset \rangle$. Similar to [10], the SETM algorithm makes multiple passes over the database. The Apriori algorithm [9] is a great success in the history of mining association rules. It is by far the most well-known association rule algorithm. This technique uses the property that any subset of a large itemset must be a large itemset. Data mining technique is most important technique which is used in Knowledge Discovery in Database (KDD). KDD has different types of steps like Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation, Knowledge presentation etc.

The Proposed System are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to medicines based on their cost analysis and many different parameters. This System are beneficial to both service providers and users. They reduce transaction costs of finding and selecting items in an online shopping environment. Recommendation systems have also proved to improve decision making process and quality. In e-commerce setting, Alternate Medicine System enhance revenues, for the fact that they are effective means of selling more products.

III. PROPOSED METHODOLOGY

The first question that comes to mind while buying medicines is

“How can I check alternatives for medicine brands in India? One medicine costs Rs10 from one company and another company charges Rs.500 for the same dose and same packaging. Why is there such a difference?”

As many Generic brands are there in market the price of medicines should have been lowered down but people don't know about it and that's the reason that still branded companies are selling medicines at high profit rates. To avoid this situation we propose a medicine recommendation system where user can check medicine alternatives with cost based analysis

The medicine recommendation system helps the patients to find an alternative to the medicine which is prescribed by the doctor. This not only helps the patients to find an alternative but also to find an optimum price of the drug required. Doctors are benefited by studying about the different drug composition in a medicine and the effects and side effects of the medicine. Chemist can refer for finding the list of the medicines which are in demand. Pharmaceutical companies can find their competitors and the price of the medicine they are selling at, which will help them in advertising for their respective medicines.

Suppose Medicine X and Y both have a target actions potassium-sparing diuretic, prescribed for hyperaldosteronism, low potassium levels, and for edema (fluid retention) caused by

various conditions. It works by blocking the hormone aldosterone. The drug contained by them is Gliclazide with dosage 10mg/50mg but Medicine X is priced at Rs.56 whereas Medicine Y has dosage 20mg/50mg Rs.64.30. This pinpoints that there isn't much difference between the price but surely dosage difference, this may not be suggested as a cheaper alternative but as a low power alternative to the same Medicine, although it is not advised to consume a higher dosage medicine without doctor's prescription.

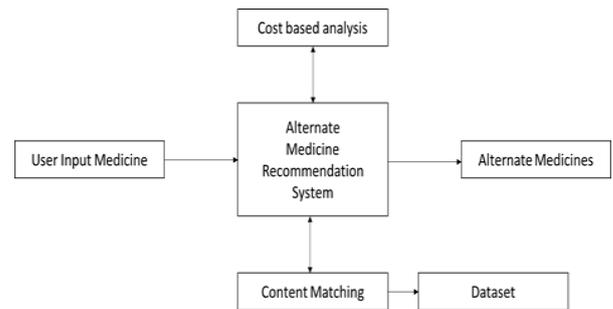


Fig. 1. Model diagram of proposed system

IV. MODULES

There are different modules to find the alternate medicine of prescribed medicine by doctor. This modules includes generation and classification of medicine data, recommending alternate medicine, differentiating medicine on contents and cost factors etc.

A. Dataset Generation

Medicine dataset needs to be generated for multiple medicines having same ratio which can be provided as an alternative to each other. Also for cost analysis it is necessary to know the market cost for such medicines. This dataset will be created for a total of 50 medicines.

B. Dataset Preprocessing

As dataset gets generated it is necessary to preprocess it for any null values if provided and the data should be cleaned and stored into Database for further processing.

C. Data Clustering based on contents and costing

Data clustering needs to be done for grouping similar medicines based on their contents and also it is required for cost based analysis as well.

D. Medicine classification and recommendation

At last ones the user inputs some medicine there is a requirement of finding alternate medicines for users which can only be done using classification of input medicines using some classification algorithm.

V. RESULT

The Fig. 2, is the first page of design for proposed system. In this page user needs to enter the keyword or medicine that's gets searched on database. Using this User will know the alternative medicine for the entered keyword or database. The

next page shows the result of searching this keyword on proposed system.

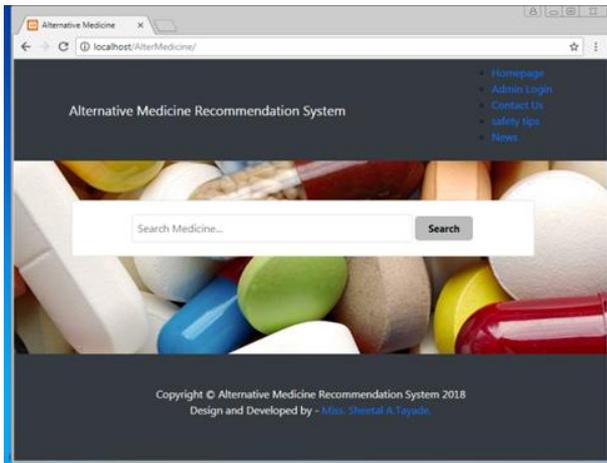


Fig. 2. Home page

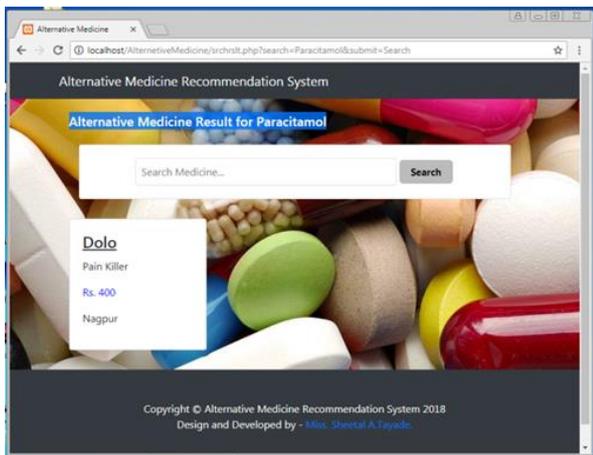


Fig. 3. Result after search alternative medicine for Paracetamol

This page gives result of user's keyword. This will shows the alternative medicine name, price store address etc.

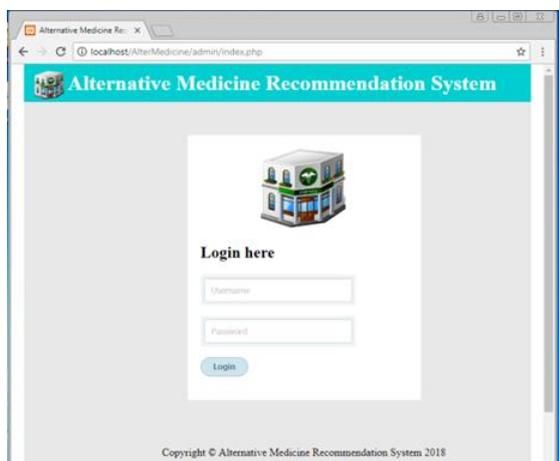


Fig. 4. Admin login

This page shows the admin login form. On this page admin can logged in and make changes such as add medicine details, delete medicine, update medicine etc.

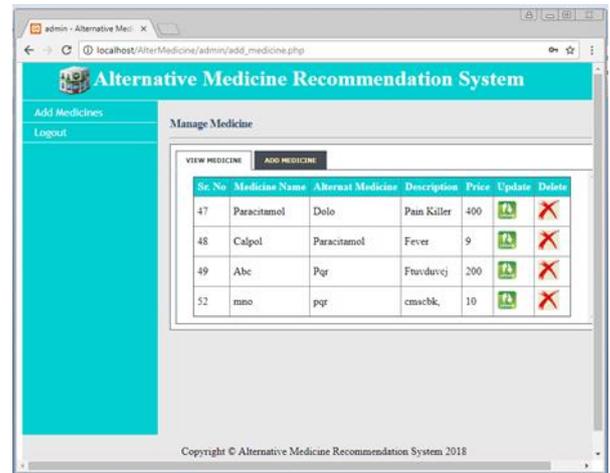


Fig. 5. Admin module homepage

This form shows the admin homepage after admin login. Admin can view all medicine details on its homepage. He/she can add medicine, delete medicine, update medicine etc.

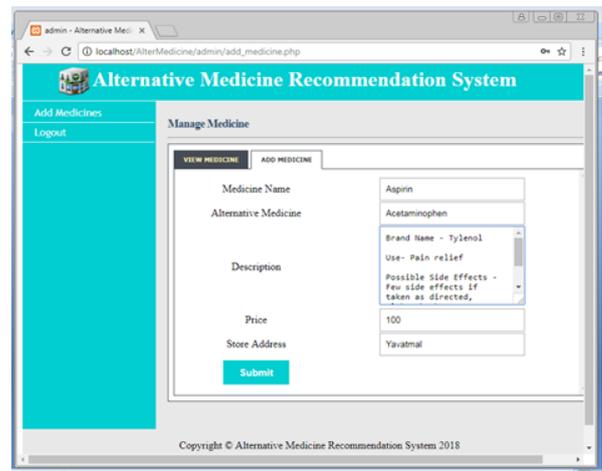


Fig. 6. Admin added medicine

Admin can add the medicine details such as its alternative name, its uses, its price, and store address etc by filling this form.

VI. CONCLUSION

Medicine availability and costing are two major issues in health care system these days. Many time the medicines prescribed by doctor are not available or are very costly based on brand names. But recently many new medicine developers have stepped into market with low cost medicines with high availability. Our proposed system is going to provide users with such medicine information that are less costly as compared to prescribed medicines and at the same time have high availability in market thus helping users in avoid higher cost in health care.

This paper aspects the healthcare domain by providing information to different bodies like Doctors, Patients, Chemists, Pharmaceutical and the Insurance Companies. The Proposed system helps the patients to find an alternative to the medicine which is prescribed by the doctor. This not only helps the patients to find an alternative but also to find an optimum price of the drug required. Doctors are benefited by studying about the different drug composition in a medicine and the effects and side effects of the medicine. Chemist can refer for finding the list of the medicines which are in demand. Pharmaceutical companies can find their competitors and the price of the medicine they are selling at, which will help them in advertising for their respective medicines. The Insurance Companies can find out which area is more prone to diseases and expand their sales out there. The use of data mining not only reduces the human efforts for finding information but also provides the information in fraction of seconds.

REFERENCES

- [1] M. Dey, S. S. Rautaray, "Study and Analysis of Data Mining Algorithms for Healthcare Decision Support System," in *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, 2014, pp. 470-477, 2014.
- [2] J. Iavindrasana, G. Cohen, A. Depeursinge, H. Müller, R. Meyer and A. Geissbuhler, "Clinical Data Mining: A Review," in *IMIA Yearbook of Medical Informatics*, vol. 48, pp. 1-13, 2009.
- [3] R. D. Canlas, "Data Mining in Healthcare: Current Applications and Issues," Master of Science in Information Technology, Carnegie Mellon University, Australia, 2009.
- [4] J. L. Patel and R. K. Goyal, "Applications of Artificial Neural Networks in Medical Science," in *Current Clinical Pharmacology*, vol. 2, no. 3 pp. 217-226, 2007.
- [5] K. Srinivas, B. K. Rani and A. Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks," in *International Journal on Computer Science and Engineering*, vol. 2, no. 2, pp. 250- 255, 2010.
- [6] K. Sudhakar and M. Manimekalai, "Study of Heart Disease Prediction using Data Mining," in *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 1, pp. 1157-1160, January 2014.
- [7] J. Soni, U. Ansari, D. Sharma and S. Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," in *International Journal of Computer Applications*, vol. 17, no. 8, pp. 43-48, March 2011.
- [8] M. Houtsma and A. Swami, "Set-oriented mining for association rules in relational databases," *Proceedings of the Eleventh International Conference on Data Engineering*, Taipei, 1995, pp. 25-33.
- [9] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in *Proceedings of the Twentieth International Conference on Very Large Databases 1994*, Santiago, Chile, pp. 487-499.
- [10] R. Agrawal, T. Imielinski and A. Swami, "Mining Associations between Sets of Items in Massive Databases," in *Proc. of the ACM-SIGMOD Int'l Conference on Management of Data*, 1993, Washington D.C.
- [11] R. Srikant, "Fast Algorithms for Mining Association Rules and Sequential Patterns," Ph. D. Dissertation, University of Wisconsin, Madison, 1996.
- [12] I. Yoo, P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, J. F. Chang and L. Hua, "Data Mining in Healthcare and Biomedicine: A Survey of the Literature," in *Journal of medical systems*, vol. 36, no. 4, pp. 2431-2448, August 2012.
- [13] H. Kaur and S. K. Wasan, "Empirical Study on Applications of Data Mining Techniques in Healthcare," in *Journal of Computer Science*, vol. 2, no. 2, pp. 194-200, 2006.