DATA Mining Application towards Adverse Effects of Anti-Diabetic Drugs

Gopishetti Venkatesh, M. Lawanyashri, Sai Saraswathi V

Abstract: *Data mining applications in health care information system is the trend set and shaping the IT industries. Data mining techniques for Anti-Diabetic Drugs is a new side of application which plays a major role in health care information system. We all wanted to know how a person is been affected by the particular medicine. Usually, many researchers are trying to predict the diseases using data mining algorithms but at the same time we have to take into consideration how the drugs are affecting a particular person. It helps a doctor to analyse these results to treat them in a proper way. The implementation can be done through the predictive data mining techniques like Decision Tree Induction, J48 classifier and Naive Bayes classifier algorithms. It predicts whether there is any change that has occurred or not after having the medicine i.e., impact, it also predicts the percentage to which, it has been cured.*

Index Terms: *Classification, diseases Diagnosis, predictive data mining, Naïve Bayes Classifier, impact.*

I. INTRODUCTION

As per the WHO and International Diabetes Federation (IDF) Technical Advisory Group, Geneva updated the guidelines for Diabetes Mellitus [1], the current criteria for identifying diabetes should be maintained - fasting plasma glucose- 26 mg/dl and 2 h plasma glucose- 200 mg/dl. Nearly 171 million people around the globe are affected by diabetes mellitus and it is expected to increase 366 million by 2030. (1, 2). IDF has given the series of guidelines for management of Diabetes, Prevention and care[2]. IDF classifies diabetes disease like Type 1 diabetes, where beta cells of pancreas produces insulin- Juvenile Onset Diabetes, Type 2 is called as Non- Insulin Dependent Diabetes or adult-onset diabetes and nearly 90% of population is affected by Type 2 and Gestational Diabetes (GDM) is a form of diabetes consisting of elevated blood glucose level noticed during the pregnancy. Sulfonyl Ureas, biguanides[3-6], alpha-glucosidase inhibitors, Thiazolidinediones, meglitinides are the classification of drugs used to treat diabetes mellitus[7]. The potential side effects of common oral diabetes drugs are stomach upset, skin rash, weight gain, kidney complications, dizziness, bloating, diarrhoea, the risk of liver disease, anaemia, swelling of legs, low blood sugar etc. [8]

The main purpose of this paper is to identify how diabetic drugs are affecting a particular person. In this paper we have used data mining techniques like naïve Bayes classification to identify the effect of diabetic drugs on individuals.so that it becomes easy to prescribe the drugs[9]. Now a day's, a most common disease is Diabetes. Doctors who are having less experience, they don't know correctly what medicines to be prescribed, how it's going to affect the body and to what extent it's going to be cured, we can take the old records of the admitted patients then, we apply some data mining algorithms like Naïve Bayes Classifier and Decision Tree induction to know how it's affecting and its now easy to prescribe medicines based on results[10-15].

Diabetes has been major problem in most of the population in recent times; even people at younger age are affected by this disease. Now a day's we have lot of improvised drugs to control and cure diabetes. people with diabetes has to go for long term medications[16]. In recent times good research is done on the side effects of diabetic drugs, the side effects of diabetic drugs depend on individuals and their health factors. The data available in hospitals can be used to identify the side effects easily by performing predictive data mining techniques[17].

II. PROPOSED WORK

In this paper data collected from the hospitals is used to apply the predictive data mining techniques. Data pre- processing techniques like removing outliers, data cleaning, data normalization and data reduction are done. we apply Naive Bayes classification algorithm on the data, that is obtained after pre-processing[18]. With the application of naïve Bayes algorithm, we can know whether a drug can be prescribed to that individual or not.

A. **Advantages of proposed system**

1. it becomes easy for the inexperienced doctors to prescribe medicines

2.large amount of hospital data can be analysed easily through these techniques

3.treatment experience becomes effective with implementation of these techniques

4.making this system available to public, can improve awareness on

the side effects of drugs, and may help to take medicines in case of non-availability of doctors.

I. DESCRIPTION

The three steps involved in the process are

A. Data Pre-processing:

Before doing data pre-processing, we will perform correlation analysis by which we can know the dependency between the attributes and can eliminate some of the attributes which are irrelevant. Then we will import our data to MySQL database, then we will connect to database using python. We are having many missing values, it can be resolved through by calculating mean of particular attribute[19]. If almost all the columns having missing values then I will remove particular attribute from the dataset. We will reduce the data by deleting some unwanted data. In the outlier analysis, we will remove the outliers by using box-plot analysis or by k-means clustering. Some of data having large values, then we will make into small by applying the techniques of data normalization like decimal scaling and min-max normalization. After the data pre-processing is over we will export again to excel file to save for future use and to have the intermediate output[20].

B. Implementation

The intermediate file that has generated that again will be imported to the MySQL database. we are implementing in the language of the python[21]. Firstly, we will connect to databases. It can be implemented through the naïve Bayes classifier where we will calculate the prior probability first and then we calculate the likelihood and evidence. It has to be done with all possible ways (means cartesian product) choices. we will also implement it in decision tree induction in which we will calculate the entropy and information gain for each and every attribute having highest we will place it in the root attribute[22]. Then we will try to compare the results of both. After, which we will compare the results of both to reveal the which is the most efficient algorithm[23].

C. Visualization

In this, we should first import the packages of the python related to which it produces graphical output (like matplotlib, seaborn)[24]. we will try to get the data from MySQL database. And put in the X-axis as label and Y-axis as rest of all attributes. Based on which we obtained the results it produces the graph. Based on this graph, doctor can prescribe which medicines will be given to a patient of age grouped.
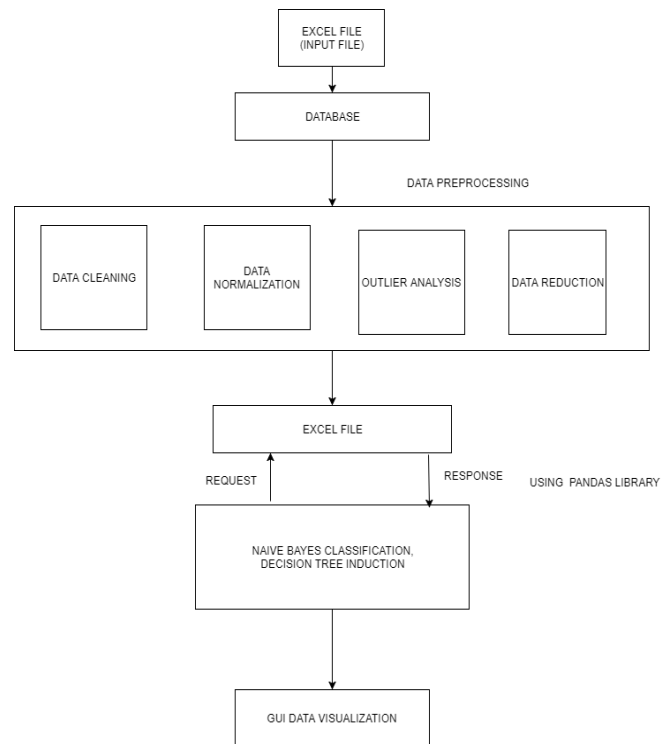
II. SYSTEM ARCHITECTURE

Fig. 1 **Architecture of the proposed system**

With the use of naïve bayes classification we have obtainted the results such that a doctor can have some suggestions from this system[25]. So,that it can be useful for treating the patient in more appropriate way and reduce the use of medicines by avoiding the ones that are not suitable for him/her.this reduces the cost of medication in the treatment.

III.CONCLUSION

We can now suggest the medicines for diabetes based on the age,gender,race,weight etc.This algorithm helps you to solve the problems of doctors who are inexperienced,they can now see the results produced by this algorithm which will be helpful to them to know how much it is affecting the patient ,and whether the particular drug is affecting a particular patient or not.We can even calculate the efficiency of a particular drug of how it is affecting the body.it is even possible to deploy it as an online system where a patient can enter his personal details and know wether a particular drug is appropriate for him.This kind of system may help many patients to reduce their medication bills ans save a lot of time ,it can help patients and many pharmacys to get appropriate drugs in special cases like unavailability of doctor.We can have admin login who can manage the databases and the queries reported by the doctors.When the Doctor send the queries it will send both his and mail and query and admin will make a priority list of that and it will be sent to the developer to make further changes in the system. We can use BigData hadoop architecture by which we can get the results faster than using the regular databases since it works parallely and since it is a health side so we want to include this and we get results much faster and here we have lesser number

of records we can even use the databases like MYSQL,mangoDB etc .We can actually combine both the naïve Bayes Classifier and Decision Induction Tree algorithm to create one new alogrithm which will have a greater efficiency than the decision Tree algorithm.By using the same algorithm we can apply to the other disease drugs by which we can also know the impact of a perticular medicine tby which it help the doctor to treat the patient appropriately.

REFERENCES

1. Wild S, Roglic G, Green A, Sicree R, King H, Global prevalence of Diabetes: Estimates for the year 2000 and paperions for 2030. Diabetes Care. 2004, 27: 1027-1053.
2. World Health Organization: Diabetes Mellitus: Report of A WHO study Group. Geneva, World Health Org. 1985. (Tech Rep. Ser. No. 727).
3. Arun Chaudhury et al., 2017, Clinical review of Antidiabetic Drugs: Implications for Type 2 Diabetes Mellitus Management, Front Endocrinol, 8: 6. Doi:10.3389/fendo.2017.00006.
4. An application of classification analysis for skewed class distribution in therapeutic drug monitoring-the case of vancomycin, Jian-xun chen, Tsang-Hsiang Cheng, A.L.F.Chan, Hue-yu Wang
5. Data Mining as a tool for detecting adverse effects of drugs, Dipali and Yogita UIET, PU, Chandigarh
6. Correlation Ratio Based Decision Tree Model For Healthcare Data Mining, Smita Roy, Samrat Mondal Department of Computer Science and Engineering, Indian Institute of Technology Patna Bihar, India
7. An Amalgam KNN to predict Diabetes Mellitus Nirmala Devi.M, Appavu alias Balamurugan.s, Swathi U.
8. Python graphs using matplotlib[Online]. Available: https://matplotlib.org/users/pyplot_tutorial.html
9. Santhi.K, Priyadarshini.C (2014). Efficiently Allocating the Virtual Machines in Cloud, International Journal of Applied Engineering Research, 9(3). P 387-392.
10. Santhi, K., Saravanan, R. (2016), Facilitate refined keywords search over encrypted data on cloud. International Journal Of Pharmacy & Technology, 8(3), p 15552-15557.
11. Santhi, K., Patel, R.(2016), Sheds: A simple and secure cost efficient data storage in heterogeneous multiple cloud. International Journal Of Pharmacy & Technology, vol.8 2016,pp.26058-26065
12. Santhi, K., Saravanan, R. (2016), A survey on queueing models for cloud computing, International Journal Of Pharmacy & Technology, 8(2), p. 3964-3977.
13. Santhi, K., Saravanan, R, (2017) Performance Analysis of Cloud Computing in Healthcare System Using Tandem Queues. International Journal of Intelligent Engineering and System. International Journal Of Pharmacy & Technology, 10(4). P.256-264.
14. Santhi, K., Saravanan, R, (2017) Performance Analysis of Cloud Computing Bulk Service Using Queueing Models. International Journal of Applied Engineering Research, 12(7), p.6487- 6492.
15. Santhi, K., & Saravanan, R. (2017). Performance Analysis of Cloud Computing Using Batch Queueing Models in Healthcare Systems. Research Journal of Pharmacy and Technology, 10(10),p.3331-3336.
16. Santhi, K. and Saravanan, R. (2019) 'Performance analysis of cloud computing using series of queues with Erlang service', International Journal. Internet Technology and Secured Transactions, Vol. 9, Nos. 1/2, pp.147–162.
17. Deepa Mani, Anand Mahendran, (2017). "Availability Modelling of Fault Tolerant Cloud Computing System". International Journal of Intelligent Engineering and Systems, Vol.10,No.1,pp.154-165.
18. Shiva Priya, K.P., Monisha, S.,Keerthiga, R., Lawanya Shri, M. , "A comparative analysis of classifier algorithm in defect prediction using cgbr framework",INTERNATIONAL JOURNAL OF APPLIED ENGINEERING RESEARCH , 2015
19. Manoj Prabhakar, K.R., Lawanya Shri, M.," Implementation of an issue tracking system in private cloud", INTERNATIONAL JOURNAL OF APPLIED ENGINEERING RESEARCH, 2014.
20. Tarun Kumar, K.S., Vignesh Kumar, P., Lawanya Shri, M.," An implementation of storage provisioning in private cloud",INTERNATIONAL JOURNAL OF APPLIED ENGINEERING RESEARCH, 2014.
21. Jothipriya, G., Lawanya Shri, M.," Database synchronization of mobile-build by using synchronization framework",INTERNATIONAL JOURNAL OF ENGINEERING AND TECHNOLOGY , 2013
22. Lawanya Shri, M., Subha, S.," An implementation of E-learning system in private cloud",INTERNATIONAL JOURNAL OF ENGINEERING AND TECHNOLOGY,2013.
23. Lawanyashri, M., Balusamy, B., & Subha, S. (2017). Energy-aware hybrid fruitfly optimization for load balancing in cloud environments for EHR applications. Informatics in Medicine Unlocked, 8, 42-50.
24. LawanyaShri, M., Subha, S., & Balusamy, B. (2017). Energy-Aware Fruitfly Optimisation Algorithm for Load Balancing in Cloud Computing Environments. International Journal of Intelligent Engineering and Systems, 10(1), 75-85.
25. LawanyaShri, M., Balusamy, B., & Subha, S. (2016). Threshold-based workload control for an under-utilized virtual machine in cloud computing. International Journal of Intelligent Engineering and Systems, 9(4), 234-241.