REVIEW PAPER ON HEALTHCARE DECISION SUPPORT SYSTEM FOR DISEASE PREDICTION

[1]Prof. Swati Powar, [2]Ashwini Patil, [3]Shrushti Desai, [4]Ashish Singh

[1]Assistant Professor, [2]Student, [3]Student, [4]Student [1,2,3,4]Department of Information Technology,

[1,2,3,4]Finloex Academy of Management and Technology, Ratnagiri, Maharashtra, India

Abstract: With illness in common people, a vast amount of data is available there for the purpose of analysis in the field of healthcare and diagnostics, patients consult to doctor in order to cure there wounds, injuries and diseases. The idea in this paper is to highlight application of classifying and predicting data generated in the field of medical healthcare, which helps a doctor to provide a decision support system. The system will prevent human error via doctor while working parallel to them so this type of treat to patient will be quite precise and most efficient.

Index Terms **- Component, formatting, style, styling, insert.**

I. INTRODUCTION

One of the trending field in computer science is data mining. The data mining uses kdd (knowledge discovery from data) process to find many patterns. The data mining can used for either prediction, i.e. classification or regression or, for description purpose like clustering or association. The medical science is a field where lots of data is generated whether textual or non-textual. This vast data can be utilized (if not textual then converted to textual) and predictive or descriptive system can be made using algorithms. This type of system can be used to provide a decision support system which help to reduce the effort and time, meanwhile increasing accuracy. the decision support system is a classification system it can be trained via using algorithms like k-nn, naive bayes, bayesian belief network etc. it observed in the paper, data mining and visualization for prediction of multiple diseases in healthcare by ajika kunjur, harshal sawant and nuzhat f. shaikh that naive bayes is more efficient on larger datasets. The optimization in classification and prediction can be done by fp-growth or apriori algorithm. Nowadays people don't pay much attention on to their health due to their busy schedule. To take care of our health is very important. The main aim of our project is to design a web based application which will help the users to find the disease by providing symptoms as the input to the system. This system will be very helpful for those who don't get enough time to visit doctor and also it will save time and money of the user.

II. LITERATURE SURVEY

The data required for analysis has been collected from World Wide Web. Information is read from only those xls file which are associated with basic objective of our intended application [2]. we compare the performance of alternative learning algorithms for predicting health data collected from these intelligent devices in a smart home. With a growing aging population that desires to maintain their independent living, the need for smart homes arises [4]. with minimum description length(MDL) discretization the compare to popular variant of Naïve Bayes and some non-Naïve Bayes statistical classifiers Naïve Bayes classifier seems to be the best performer [3]. Box –Jekins forecasting models are based on statistical concepts and are able to model a wide spectrum behavior of time series [6]. The system is fed with various symptoms and the diseases associated with each system. The system is first taught with various symptoms and the disease associated with each system [4]. For analysing the sentiments for text we can use Stanford core-nlp API as sentiment analyser [5].

III. DATA

For training the classification models, we are using a labelled dataset of approx. 2500 tuple in decision support system of PIMA indian diabetes. The dataset is in the form of text file. We have extracted dataset from the UCI repository as well as we consulted from doctor for real-time data sets.Each record in our dataset contains following attributes

(1) Date in MM-DD-YYYY format

(2) Time in XX:YY format

(3) Code

(4) Value

Where each code have their own description, like blood glucose level before breakfast and after breakfast, etc.

IV. METHODOLOGY

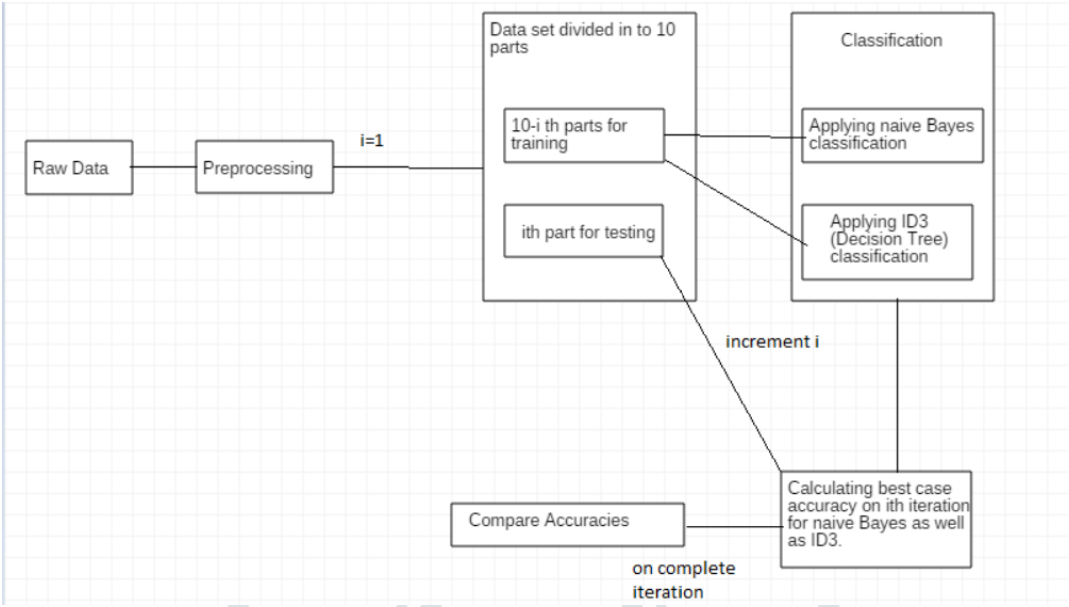
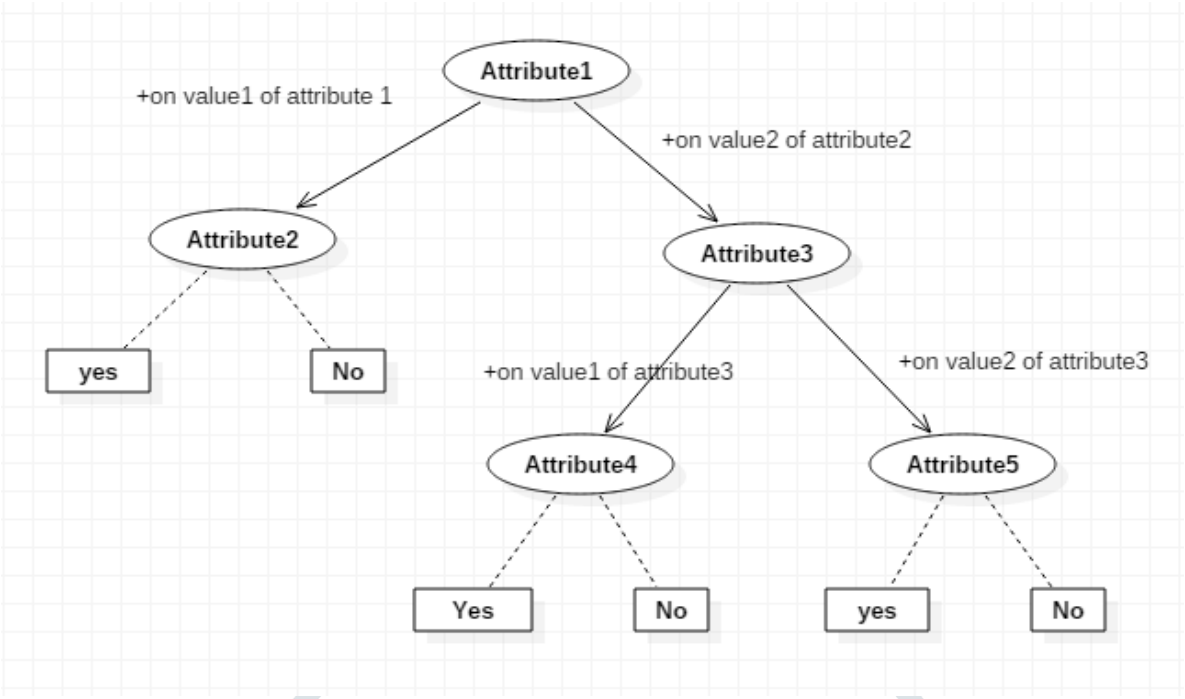


Fig.4.1 .Architecture diagram

In our methodology we first, pre-process the data as per the value of the code because some code In the dataset represents that values taken were erroneous .After that we split the large amount of data in k folds, where k=no. of tuples in dataset/10, This process is for k-fold cross validation technique.Now, after that we have to set a threshold value in the k-1 training data set on the basis of which we can do the classification, and this threshold value will be different in different cases.Now, we have table which is ready and available for data mining. As from 10 parts of preprocessed dataset we have already used 9 parts for training ,now the rest 1 part will be for training purpose.Since we are comparing the accuracy of two algorithms, which is naive-Bayes and ID3(Decision-tree) algorithms.Now we run the iteration 10 time in which 9 parts of dataset is used as training purpose and rest one part will be for testing purpose.By doing this we will get 10 accuracy values for each of two methods(algorithms). The accuracy values which represents the best real time situation (i.e. accuracy>some real time threshold value) will be selected for each algorithm and the algorithm which represents more accuracy is considered to be more efficient.

4.1 ID3 (Iterative Dichromotise) or decision tree:

This method is greedy as well as divide and conquer approach. In this method decision tree is generated and splitting of higher level node is done on the basis of entropy of attribute (i.e. attribute which has low entropy). Unfortunately, in Decision tree the main table is divided in to sub table based on distinct value of the attribute. The process stops when every leaf node is turned to class label or represents the probability of each class label.



4.2 Naïve Bayes:

It is a classification technique based on Bayes Theorem. Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. It is easy to predict class of test data set.

$$P(c \mid x) = P(x \mid c)\, P(c) / P(x) \qquad (2)$$

P(c | x) = posterior probability of the target class
P(c) = prior probability of class
P (x | c) = likelihood which is the probability of predicator class
P(x) = prior probability of predicator

4.3 k-fold cross validation:

k-fold cross validation is a method of finding accuracy while keeping some of the data as test set. In this, the whole dataset of n tuples is divided into k parts, in which k-1 parts are used for training purpose and 1 part is used for testing purpose.From this we get total of k accuracy values, for each combination of k-1 part training set. Now we compare the accuracies and the training set which gives the best accuracy is the re-presenter of the dataset. Generally value of k=10, but some times it is considered as k=sqrt(size of dataset).

V. CONCLUSION

Hence, we propose in this paper a system which uses classification algorithm and provides a decision support system to the doctor. Hence, minimizing the human error in healthcare and diagnosis field. We can compare the accuracies of implementation on two different algorithm. To predict diseases using Data mining applications is challenging task but it drastically reduces the human efforts and increases the diagnostic accuracy.In future it can be extended to n number of diseases and uses can chartthe with the doctor and intend freely in close of emergency.

REFERENCES

[1] Swati Powar, Dr. Subhash Shinde "Named Entity Recognition and Tweet Sentiment Derived from Tweet Segmentation using Hadoop" IEEE & CSI sponsored 1st International Conference on Intelligent Systems & Information Mannagement, Oct -17

[2] Vikramadity Jakkula "Predictive Data Mining to Learn Health Vitals of a Resident in a SmartHome" EECS, Washington State University, Pullman, WA-99164, 20007

[3] Puja Sarage,Trupti Sudrik, Kalyani Zodage " Health Prediction System by using DataMining" International Journal for Research in Applied Science & EngineeringT echnology(IJRASET)

[4] Sujatha R, Sumathy R, Anitha Nithya "A Survey of Health Care Prediction Using Data Mining, International Journal of Innovative Research in Science, Engineering and Technology Vol.5, Issue8, August2016

[5] Pinky Saikia Dutta , Shrabani Medhi, Sunayana Dutta, Tridisha Das, "SMART HEALTH CARE USING DATA MINING" ISSN(PRINT):2393-8374,(ONLINE):2394-0697,VOLUME-4,ISSUE-8,2017.

[6] Nikita Kamble, Manjiri Harmalkar, Manali Bhoir, Supriya Chaudhary, "Smart Health Prediction System Using Data Mining" International Journal of Scientific Research in computer science, Engineering and Information Technology ©2017.