International Engineering Journal For Research & Development
Vol.5
Issue 5
E-ISSN NO:-2349-0721
Impact factor : 6.549

HEALTHCARE DECISION SUPPORT SYSTEM FOR DISEASE PREDICTION

[1]Prof. Swati Powar, [2]Ms. Ashwini Patil, [3]Ms. Shrushti Desai, [4]Mr.Ashish Singh
Information Technology Finloex Academy of Management and Technology
Ratnagiri, Maharashtra, India

ABSTRACT

Data Mining is an famous and powerful technology which is of high interest in today's computer world. It uses already existing data in different databases and transform it into new technology and research. It extracts new patterns for large datasets and the knowledge associated with these patterns. There is a large amount of data available within the healthcare due to availability of computer systems. The most important and popular data processing techniques are classification, association, clustering, prediction and patterns. In healthcare concern businesses, data processing plays a crucial role in early prediction of diseases. In general, to detect a disease numerous health related tests must be conducted in a patient. The usage of knowledge mining techniques in disease prediction is to scale back the test and increase the accuracy of rate of detection of disease. This research paper intends to supply a survey of current techniques of data discovery in databases using data processing techniques that are in use in today's medical research particularly in Diabetes and liver Disease Prediction. The major objective of this paper is to evaluate data mining techniques in healthcare application to develop an accurate decisions.

Keywords— Pre-processing, prediction, classification.

INTRODUCTION

Medical data mining has great potential for exploring the hidden patterns within the data sets of the medical domain. The data mining tools are useful for predicting the various diseases in the healthcare field. Disease prediction plays a crucial role in data processing. This paper analysis diabetes and liver disease prediction using classification algorithms. Diabetes is considered as one of the deadliest and chronic diseases which causes an increase in blood sugar [3]. Many complications occurred if diabetes remain untreated and undefined. Liver is one of the most important organs in the human body but due to unhealthy lifestyle and excessive alcohol intake, liver disease has been increase at an alarming rate globally. Hence it involves an instantaneous attention to predict the disease before it's too late. Healthcare industry generally large amount of complex data's such as patient history, Hospital resources, electronic records, information about medical devices etc. These data's serves as a key resources to process and analysis for knowledge extraction that enables the decision making and to save cost [6].

LITERATURE SURVEY

Medical Diagnosis is a difficult process which requires experience and proficiency to deal with medical data [1]. Various diseases like heart condition, diabetes, carcinoma and liver disorder are diagnosed using various data processing techniques. Data mining provides better leads to disease diagnosis when appropriate tools and techniques are applied. Several target values are combined to obtain disease prediction using various clustering

and classification methods [1]. Classification is a major task in disease diagnosis. Various classification algorithms are used to predict liver diseases at early stage Classification algorithms like Decision Tree, SVM, Naïve Bayes, ANN, Hill climbing etc are applied on the dataset. Further these algorithms are compared on the basis of performance measures [4]. Pre-processing helps to improve the accuracy of classification algorithms. Pre-processing technique is applied on the info set to get rid of noise and cluster the data. The pre-processed data is applied to various classification algorithms and their performance is compared [6].

METHODOLOGY

Classification is the process of identifying a new observation category set on the basis of training set of data that contains observations whose category is known. Cluster analysis technique is employed to group the objects according to its similarity. Studies are made to compare the different techniques of classification which have been developed so far. In this study we have compared different classifiers Naïve Bayes, Decision tree (ID3).

A. NAÏVE BAYES CLASSIFIER:

Naive Bayes classifiers are a set of classification algorithms supported Bayes' Theorem. In this algorithm all of the attributes share a common principle, i.e. every pair of features being classified is independent of every other. Baye's Theorem finds the probability of an event occurring given the probability of another event that has already occurred. If B represents the dependent event and a represents the prior event, Baye's theorem can be stated as follows.

Prob (B given A) = Prob (A and B)/Prob (A)

The information in the patient record are pre-processed initially using data mining techniques and then the attributes are classified using a Naïve Bayes classification Algorithm. In the classification 13 attributes are given as input to the Naïve bayes classifier to determine the risk of disease.Since it's supported contingent probability it's considered as a strong algorithm employed for classification purpose. It works well for the data with imbalancing problems and missing values.

B.DECISION TREE CLASSIFIER:

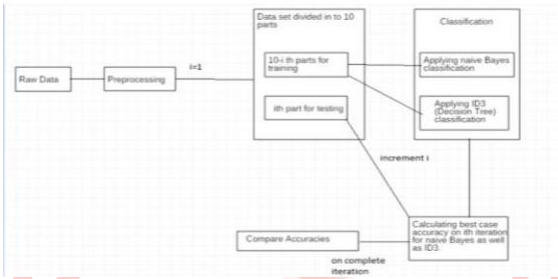
Decision Tree may be a supervised machine learning algorithm wont to solve classification problems. The main objective of using Decision Tree during this research work is that the prediction of target class using decision rule taken from prior data. It uses nodes and internodes for the prediction and classification. Root nodes classify the instances with different features. Root nodes can have two or more branches while the leaf nodes represent classification. In every stage, Decision tree chooses each node by evaluating the highest information gain among all the attributes [5].

5.2. Confusions matrix Confusion matrix is used to present the accuracy of classifiers obtained through classification. It is used to show the relationship between outcomes and predicted classes.

Confusion matrix		Targeted Values	
		Positive	Negative
Model	Positive	a	B
	Negative	c	D

Table-1. Confusion matrix.

Here ‘a’ = number of correct instance that is of negative instance; ‘b’ = the number of incorrect prediction that is of positive instance. c is the number of incorrect predictions that is of negative instance. d is the correct predictions that is of positive instance. The work is divided into three parts: Feature selection, Pre-processing and Classification. The flow diagram of entire work described in this paper is mentioned in the Fig.1



EXPERIMENTAL RESULTS

Dataset values of male female related to liver disease are converted to 0, 1 to avoid value error also all dataset values in diabetes converted from (morning, noon, night), (pre, post) and (low, medium, high) to (0,1,2), (0,1) and (0,1,2) to avoid value error.

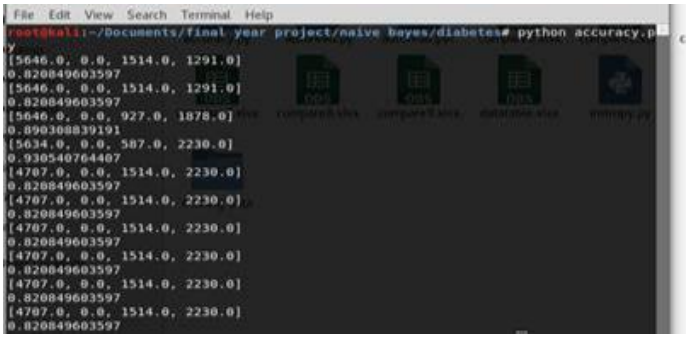
Diabetes patient records were obtained from two sources: an automatic electronic recording device and paper records. The paper records only provided logical time slots (breakfast, lunch, dinner, bedtime). Fixed times were assigned to breakfast (08:00), lunch (12:00), dinner (18:00), and bedtime (22:00). Thus paper records have uniform recording times whereas electronic records have more realistic time stamps.

Diabetes files consist of four fields per record. Each field is separated by a tab and each record is separated by a newline. File Names and format:

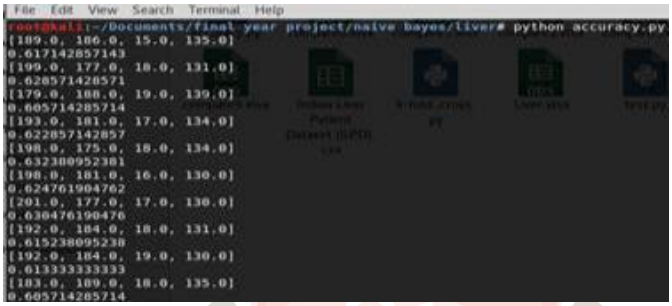
(1) Date in MM-DD-YYYY format (3) Code

(2) Time in XX: YY format (4) Value

1) Diabetes Naïve Bayes:



2) Liver Disease Naïve Bayes:



3) Diabetes Decision Tree

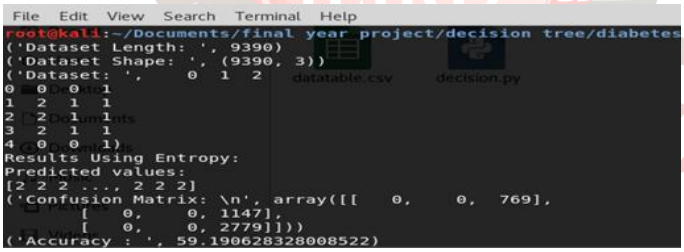
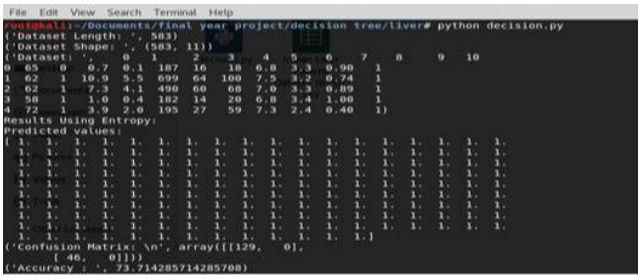


Fig.1 Architecture Diagram

4) Liver Disease Decision Tree:



The Code field is as follows:

33 = Regular insulin dose

34 = NPH insulin dose

35 = Ultra Lente insulin dose

48 = Unspecified blood glucose level

57 = Unspecified blood glucose level

58 = Pre-breakfast blood glucose level

59 = Post-breakfast blood glucose level

60 = Pre-lunch blood glucose level

61 = Post-lunch blood glucose level

62 = Pre-supper blood glucose level

63 = Post-supper blood glucose level

64 = Pre-snack blood glucose level

65 = Hypoglycaemic symptoms

• Diseases	• Naive Bayes	• Decision Tree
• Diabetes	• 83.8724%	• 59.1906%
• Liver	• 61.9573%	• 73.7143%

Table 2. Final Classifier & Accuracy table

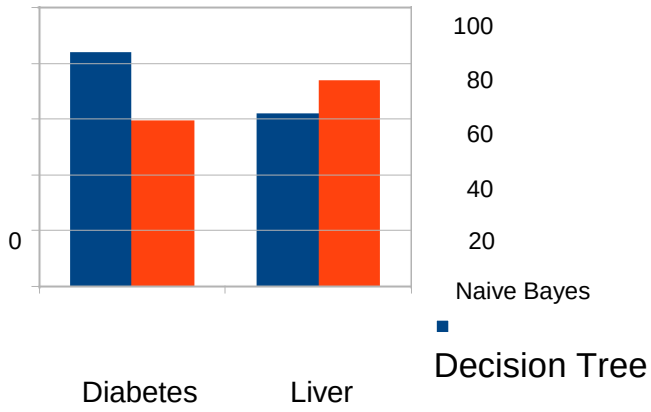


Fig.2 Classifiers comparison Graph

CONCLUSION

The main aim of this paper is to improve the accuracy of classification of diabetes and liver disorders. The aim is achieved by performing a comparative study of classification algorithms on the dataset. Pre-processing technique is used to divide the data into groups which is done using K-fold cross validation algorithm. Further the clustered dataset is applied to varied classification algorithms. The performance of each algorithm is evaluated and a comparative study has been carried out. Based on the performance comparison, it is clear that Naive Bayes is better performance for Diabetes dataset and Decision Tree is better for Liver disorder.

REFERENCES

1. Balpande, V. R., & Wajgi, R. D. (2017, February). Prediction and severity estimation of diabetes using data mining technique. In Innovative Mechanisms for Industry Applications (ICIMIA), 2017 International Conference on (pp. 576-580). IEEE..
2. Hashi, E. K., Zaman, M. S. U., & Hasan, M. R. (2017, February). An expert clinical decision support system to predict disease using classification techniques. In Electrical, Computer and Communication Engineering (ECCE), International Conference on(pp. 396-400). IEEE.
3. Bashir, S., Qamar, U., Khan, F. H., & Naseem, L. (2016). HMV: a medical decision support framework using multi-layer classifiers for disease prediction. Journal of Computational Science, 13, 10-25.
4. Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. The Kaohsiung journal of medical sciences, 29(2), 93-99.
5. Prakash Mahindrakar et al. 2013. Data Mining in Healthcare: A Survey of Techniques and Algorithms with Its Limitations and Challenges. Int. Journal of Engineering Research and Applications. 3(6): 937-941. (ISSN: 2248-9622).
6. S.Vijiyarani and S.Sudha. 2013. Disease Prediction in Data Mining Technique - A Survey. International Journal of Computer Applications and Information Technology. 2(1).
7. AbuKhousa. E. 2012. Predictive data mining to support clinical decisions: An overview of heart disease prediction system. IEEE Transaction onInnovations in Information Technology (IIT).