NCRISET-2017 e-ISSN: 2456-3463
International Journal of Innovations in Engineering and Science, Vol. 2, No.6, 2017
www.ijies.net

A design of patient health predication system based on naïve bayes and decision tree algorithm on data mining

Amol pisode[1],Diksha Bhambore[2], Shankar Pawar[3],Prof. Rahul Shahane [4]

Dept of CSE, WCEM , RTM Nagpur University,
Dept of CSE, WCEM , RTM Nagpur University,
Dept of CSE, WCEM , RTM Nagpur University,
HOD Dept of IT, WCEM .

Abstract:
One of the rapid growing fields is health care area. The medical industries have big amount of data set collections about patient details, diagnosis and medications. To turns these information is into useful pattern and to predicting coming up trends data mining approaches are preowned in health care industries. The healthcare industry collects huge amount of healthcare data which are not "mined" to find out private information. The medical area come crossways with new treatments and medicine every day. The healthcare industries should provide best diagnosis and therapy to the patients to attain better quality of service. This paper explores different data mining techniques which are used in medicine field for good decision making. Data mining is one of the most motivating area of research that is most useful in medical field like health prediction with using some data mining algorithms. Online Health Prediction System is an end user support and online consolation project. It might have happened so many times that you need doctor help immediately, but they are not available due to some reason, that time this system is used. Here we propose a system that allows users to get instant guidance on their health issue through an intelligent health care system .Here we use some intelligent data mining technique to guess the most accurate illness could be associated with patients' symptoms .User can talk about their illness and get instant diagnosis. User can search for doctors help at any point of time . Doctor get more clients online

Key words: *Data mining, KDD, Prediction techniques, Decision making. Data Mining, Healthcare, Prediction, ,Classification, Clustering, Association, Decision tree algorithms , naïve Bayes*.

INTRODUCTION:

Data mining is the method for searching unknown values from enormous amount of data. As the patients requirement increases the medical databases also increasing every day. The process and finding of these medical data is difficult without the computer based analysis system. The computer based analysis system denoted the mechanized medical diagnosis system. This mechanized diagnosis system support the medical practitioner to make better decision in treatment and disease. Data mining is the big areas for the doctors to handling the big amount of data sets of patient's in many ways such as make sense of complex diagnostic tests, interpreting previous results, and combining the dissimilar data together. . This method diagnosis system provides to increases the quality of service provided to the patients and decreases the costs of medical. "Online Health Prediction System" is an end user support and online consolation project. It might have happened so many times that you need doctor help immediately, but they are not available due to some reason, that time this system is used. Here we propose a system that allows users to get instant guidance on their health issue through an intelligent health care system.Here we use some intelligent data mining technique to guess the most accurate illness could be associated with patient's symptom.In the early 1970's, it was very costly to store the data or information. But due to the advancement in the field of information gathering tools and WWW in the last twenty-six years, we have seen huge amount of information or data are available in electronic format. To store such a large amount of data or information the sizes of databases are increased very rapidly. Such type of databases consist most useful information. This information may be

useful for decision making process in any field. It becomes possible with the help of data Knowledge Discovery in Databases (KDD). Data mining is the process of extracting or taking the useful information from a large collection of data which was previously unknown .

Selection :- The data selection is process its selected that data it is relevant to our taskand irrelevant or unusable data are omitted

Preprocessing:- This stage removes that information which is not usable for example it is also called as data cleaning process.

Transformation:- This stage transformed only those data which are useful in a particular research for example only data related to a particular demography is useful in market research.

· Data mining:- Data mining is a stageuseful for knowledge discovery process. This stage is useful for extracting the patterns from data which is meaningful.

· Interpretation and evaluation :- The meaningful patterns identified by system are interpreted into knowledge in this stage. This knowledge may be useful for making useful decisions.

I. KNOWLEDGE DISCOVERY AND DATA MINING

This area provides an introduction to knowledge discovery and data mining.

A.Knowledge Discovery Process

The terms Knowledge Discovery in Databases (KDD) andData Mining are frequently usedinterchangeable. KDD is the process of changing the lowlevel data into high-level knowledge. Hence, KDD refers to the nontrivial removal of implicit, previously unknown and useful information from data in databases. While data mining and KDD are often use as comparable words but in real data mining is an efficient step in the KDD process.
The Knowledge Discovery in Databases process include of a few steps leading from collections of raw data to some form of new information. The iterative process is consists of the following steps:
Data cleaning: also known as data cleansing it is a phase in which noisy data and unrelated data are removed from the collection.
Data integration: at this stage, several data sources, often heterogeneous, may be shared in a common source.
Data selection: at this step, the data related to the analysis is decided on and retrieve from the data collection.
Data transformation: also known as data consolidation, it is a phase in which the select data is transformed into forms appropriate to the mining procedure.
Data mining: it is the essential step in which clever techniques are applied to extract patterns potentially useful.
Pattern evaluation: this step, very interesting patterns representing knowledge are known based on given measures.
Knowledge representation: this is the last phase in which the discovered knowledge is visually represented to the user. In this phase visualization techniques are used to help users understand data mining results.

The following figure shows data mining as a step in an iterative knowledge discovery process.

B. Data Mining Process

In the KDD process, the data mining methods are for extracting or taking patterns from data. The patterns that can be discovered depend upon the data mining tasks applied. Generally, there are two types of data mining tasks: *descriptive data mining and predictive data mining . the descriptive data mining tasks* that explain the general properties of the existing data, and *predictive data miningtasks* that try to do predictions based on available data. Data mining can be done on data which are in textual, quantitative or multimedia forms. Data mining applications can use dissimilar kind of parameters to detect the data. They include association, sequence or path analysis, classification and clustering.Data mining involves some of the following key steps:
(1) *Problem definition:* The first step is to locate goals. Based on the defined goal, the series which is correct of tools can be applied to the data to build the corresponding behavioral model.
(2) *Data exploration:* If the value of data is not suitable for an perfect model then re-commendations on future data collection and storage strategies can be built at this. For analysis, all data needs to be consolidated so that it can be treated consistently.
(3) *Data preparation:* The purpose of this step is to clean and convert the data so that missing and invalid values are treated and for more robust analysis, all known valid values are made reliable.
(4) *Modeling:*A data mining algorithm or group of algorithms is selected for analysis based on the data and the desired outcomes. These algorithms include classical techniques such as statistics, neighborhoods and clustering but also next invention techniques such as decision trees, networks

algorithms. The specific algorithm is selected based on the particular objective to be achieved and the quality of the data to be analyzed.

(5) *Evaluation and Deployment:* an analysis is included to find out key conclusions from the analysis and create a sequence of recommendations for considerationBased on the outcome of the data mining algorithms .

HEALTHCARE DATA MINING

The increasing research area in data mining technology is Healthcare data mining. Data mining holds immense promising for healthcare management to grant health system to systematically use data and analysis to progress the care and decrease the cost simultaneously could apply to as much as 30% of overall health scare spending. In the healthcare managing data mining prediction are playing various role. Some data mining techniques for prediction are as follows:

1. Neural network
2. Bayesian Classifiers
3. Decision tree
4. Support Vector Machine

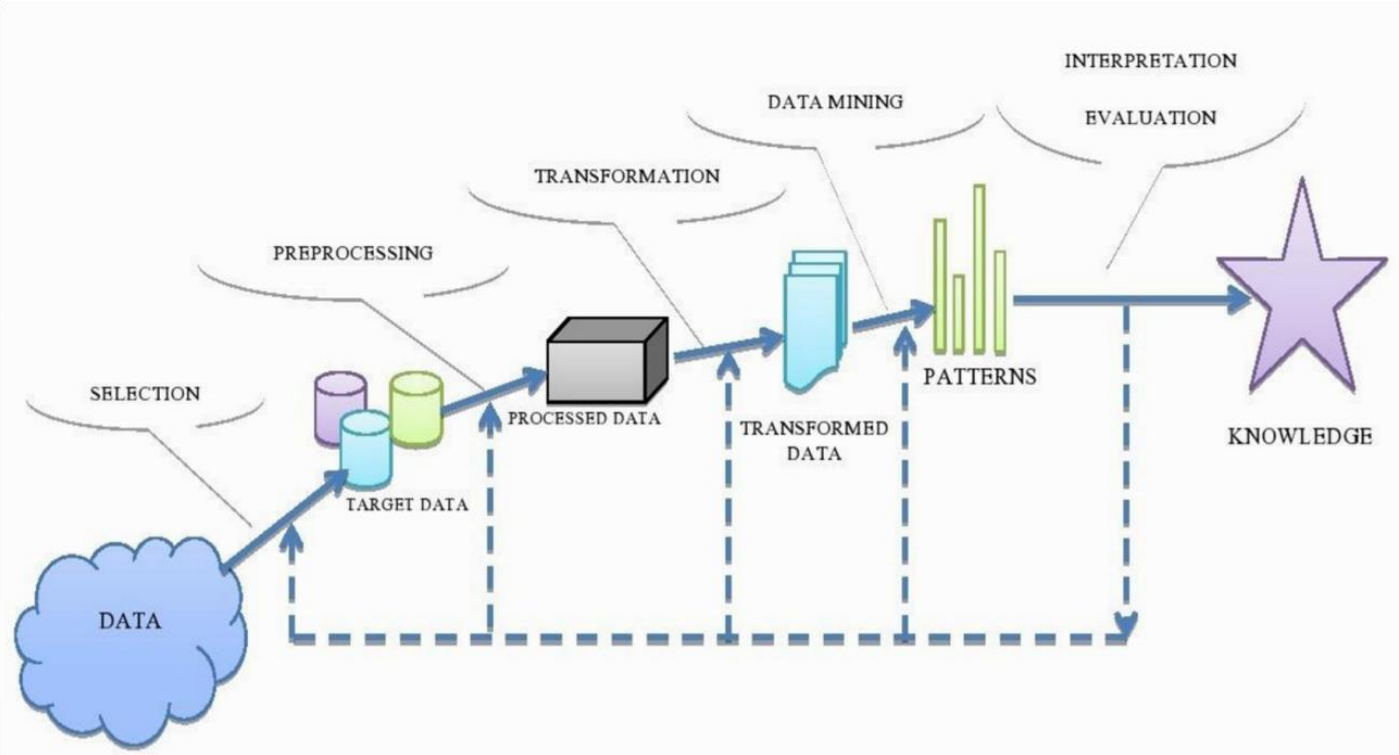
C. PREDICTION TECHNIQUES :

Releted Work :

•"Online Health Prediction System " is an end user support and online consolation project It might have happened so many times that you need doctors help immediately , but they are not available due to some reason ,that time this system is used.Health predict system allows users to get instant guidance on patient health issue .The system is not fully automated, it needs doctors for full diagnosis This project data mining technique to guess the most accurate illness could be associated with patients symptoms .User can talk about their illness and get instant diagnosis.User can search for doctors help at any point of time .Doctor get more clients online.

Application :

Online health predict system can be used by all patients or their family members who need help in emergency.Conclude form discussion that the Health Prediction System is useful in patient help, And get reference exact doctors diagnosis on patient disease.sIt can save time to search the doctor for particular diseases at any time at any place



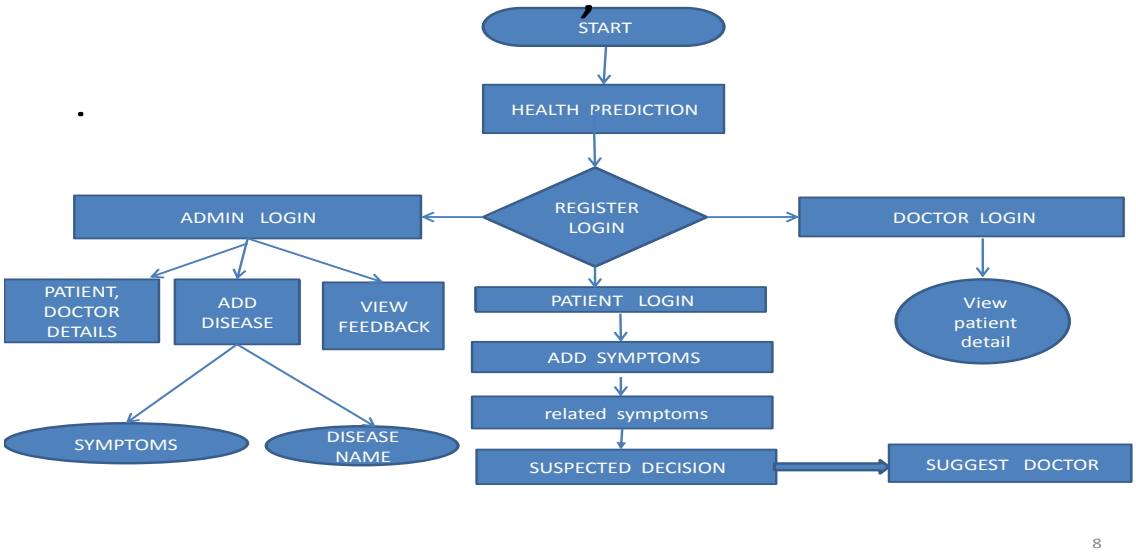


Figure2. Flow chart of Online Health Prediction System.

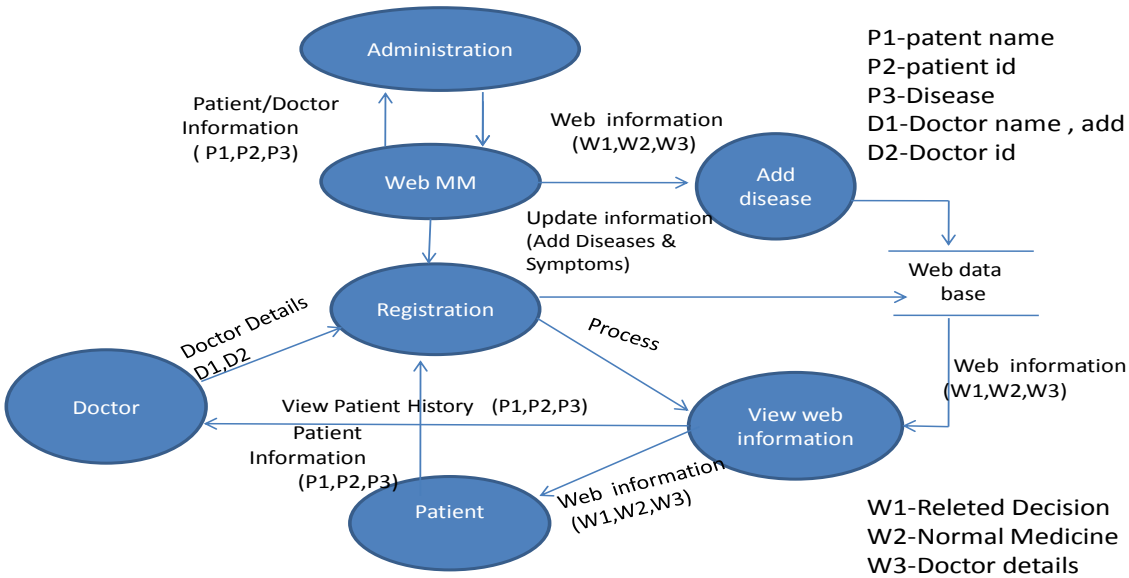


Figure3. DFD Diagram.

Theorem:

To calculate probability of A given B, P (B

2. BAYESAIN 1.CLASSIFICATION:

Bayesian classifier is a statistical classification approach based on the Bayes theorem.

For probabilistic learning method Bayesian classification is used. With the help of classification algorithm we can easily obtained it . Bayes theorem of statistics plays a very important role in it. While in medical domain attributes such as patient symptoms and their health state are related with each other but Naïve Bayes Classifier assumes that all attributes are independent with each other. This is the major disadvantage with Naïve Bayes Classifier. If attributes are independent with from each other then Naïve Bayesian classifier has shown best performance in terms of accuracy. In healthcare field they play very important roles. Hence, researchers across the world used them there are various advantages . One of them is that it helps to makes computation process very easy. Another one is that for huge datasets it has better speed and accuracy.

given A) = P (A and B)/P (A) the algorithm counts the number of cases where A and B occurs simultaneously and distribute it by the number of cases where A alone occurs. Let X be a data tuple, X is considered "Evidence", in Bayesian . Let H be some hypothesis, such that the data tuple X termsbelongs to class C. P (H|X) is posterior probability, of H conditioned on X. P (H) is the prior probability of H in contract.

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)}$$

$$Posterior = Likelihood * \frac{Prior}{Evidence}$$

2. DECISION TREE:

Decision tree uses the divide-and-conquer algorithm. In these tree structures, leaves show classes and branches signify conjunctions of features that lead to those classes. The attribute that most effectively splits case into different classes is select, at each node of the tree. A path to a leaf from the root is found depending on the determine the predicate at each node that visited, to predict the class label of an input.

Decision tree is fast and easy method since it does not require any of domain information. In the decision tree inputs are divided into two or more groups continue the steps up to complete the tree as shown on Fig.4

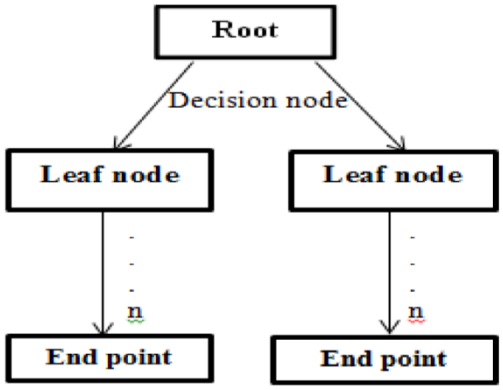


Figure 4. Decision tree Structure

Various decision tree algorithms as follows:

a) CART (Classification & Regression Tree)

b) C4.5 (Successor of ID3)

c) ID3 (Iterative Dichotomiser 3)

CHAID (CHI-squared Automatic Interaction Detector) is considered to be one of the most popular approaches for representing classifier. We can construct a decision tree by using available data which can deal with the problems related to different research areas. It is equivalent to the flowchart in which every non-leaf nodes denotes a test on a particular attribute and every branch is denotes a result of that test and every leaf node have a class label. Root node is the top most node of a decision tree. For example, with the help of medical readmission decision tree we can decide whether a particular patient requires re-admission or not. Knowledge of domain is not required for building decision regarding any problem. The most common use of Decision Tree is in operations research analysis for evaluting conditional probabilities Using Decision Tree, decision makers can choose best alternative and traversal from root to leaf indicates unique class divide based on maximum information gain . Decision Tree is widely used by many researchers in healthcare field. verious advantages of decision tree as follows: Decision trees are self–explanatory and when using very little space they are also easy to follow. Even set of rules can also be constructed with the help of decision trees. Hence, representation of decision tree plays a very important role in order to display any discrete-value classifier because it can be capable to handle both type of attributes, nominal as well as numeric input attributes. If any datasets have missing or erroneous values, such type of datasets can be easily handled by decision trees. Due to this reason decision tree can be considered to be nonparametric method. The meaning of above sentence is that there is no require to make assumptions regarding distribution of space

and structure of classifier. Decision trees have several disadvantages. These are as follows: Most of the algorithms (like I and C) require that the target attributes have only discrete values because decision trees use the divide and conquer method. If there are more difficult interactions among attributes exist then performance of decision trees is low. Their performance is better only when there exist a few highly relevant attributes. One of the reasons for this is that another classifiers can compactly describe a classifier that would be very challenging to represent using a decision tree. A simple illustration of this phenomenon is replication problem of decision trees , and the greedy characteristic of decision trees leads to another disadvantage. This is its over-sensitivity to the training set, irreverent attributes and to noise

4. Support Vector Machine (SVM)

Normally SVM is the classification technique. Initially it developed for binary type classification later extended to many classifications. This SVM creates the hyper plane on the original inputs for effective distribution of data points.

ADVANTAGES OF MINING APPLICATION IN HEALTH CARE

Information chnologies in healthcare have enabled the creation of electronic patient records display from monitoring of the patient visits. This information includes patient demographics, records on the treatment progress, details of examination, prescribed drugs, backword medical history, lab results, etc. Information system simplifies and automates the workflow of health care institution. Security of documentation and ethical use of information about patients is a important obstacle for data mining in medicine. In order for data mining to be more exact, it is necessary to make a considerable amount of documentation. Health records are secrect information, yet the use of these private documents may help in treating deadly diseases .Before data mining process can begin, healthcare organizations must

formulate a clear policy burden privacy and security of patient records. This policy must be fully implemented in order to ensure patient privacy.

Health institutions are able to use data mining applications for a various of areas, such as doctors who use patterns by measuring clinical indicators, quality indicators, customer satisfaction and economic indicators, performance of physicians from multiple perspectives to result use of resources, cost efficiency and decision developing based on evidence, identifying high-risk patients and intervene proactively, optimize health care, etc .Integration of data mining in information systems, healthcare institutions reduce subjectivity in decision-making and provide a new useful medical knowledge.

Data mining provides the link between knowledge of continuous data, such as biomedical signals collected from patients in intensive care units, and it making an intelligent monitoring system that sends reminders, warnings and alarms for the pre-selected critical conditions .Using association rules involves finding all the rules, or few part

of key subsets of rules that is characteristic of certain information as a consequences or as a antecedent. This type of problem is very interesting for health professionals who are finding for the relations between diseases and lifestyles or demographics or between survival rates and treatment. The tasks of association are used to help strengthen the arguments regarding whether to engage or reject certain rules in the knowledge model . Tasks of the managers that manage quality of the healthcare services can be display as optimization of clinical processes in terms of medical and administrative quality as well as the cost/benefit relation. Key questions of the process of healthcare quality management are quality of information, standards, plans, and treatments.

Data mining can be used by quality managers to solve the following tasks: Discovering new hypothesis for indexes of quality for data, standards, plans and treatments; finding if the given indexes of quality for data, standards, plans and treatments are still valid; Improving, strengthening and adjusting of quality indexes for data, standards, plans and treatments; These tasks can be supported by data mining if the existing knowledge in domain is seriously considered in data mining process. Data mining has most importance for area of medicine, and it represents comprehensive process that demands thorough understanding of needs of the healthcare organizations. Healthcare is one of the more important sectors which can highly benefit from the implementation and use of information system. We have provided an overview of applications of data mining in infrastructure, administrative, financial and clinical Health care system. Knowledge gained with the use of techniques of data mining can be used to make successful decisions that will improve success of healthcare organization and health of the patients.

Data mining necessary appropriate technology and analytical techniques, as well as systems for reporting and tracking which can enable measuring of results. Data mining, once started, display continuous cycle of knowledge discovery. For organizations, it presents one of the key things that help create a good business strategy. Today, there has been many efforts with the objective of successful application of data mining in the healthcare institutions. Primary potential of this technique

lies in the possibility for research of hidden patterns in data sets in healthcare domain. These patterns should be used for clinical diagnosis. However, available raw medical data are widely distributed, different and voluminous by nature. These data should be collected and stored in data warehouses in organized forms, and they can be integrated in order to form hospital information system.

Data mining technology provides customer oriented approach towards new and unfound patterns in data, from which the knowledge is being generated, the knowledge that can help in providing of medical and other services to the patients. Healthcare institutions that use data mining applications have the probality to predict future requests, needs, desires, and conditions of the patients and to make adequate and optimal decisions about their treatment. the future development of information communication technologies, data mining will achieve its full potential in the discovery of knowledge hidden in the medical data.

VI. CONCLUSION

Thispaper is presented to study about the various data mining application in the healthcare sector to discover new range of pattern information. There is verious of data mining tools and techniques are define for health care diagnosis systems that are clearly defined. This data mining based prediction system reduces the human effects and cost effective one. AT the end of this proposal we remember that this is fully unique system and we believe that it will helpful for us as well as any hospital business can add this with their existing features . Hope this system will be very demandable incoming future.

REFERENCES

[1] Muhamad Hariz Muhamad Adnan,Wahidah Husain, Nur'Aini Abdul Rashid(2012), “Data Mining for Medical Systems: A Review” ,International conferences on advances in computer and information technology. [2] V. Krishnaiah et al,(2013)” Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques”, International Journal of Computer Science and Information Technologies, Vol. 4 (1) ,39 – 45.

[3] Abdelghani Bellaachia, Erhan Guven,” Predicting Breast Cancer Survivability Using Data Mining Techniques”,Department of Computer Science, The George Washington University.

[4] Ravi Sanakal, Smt. T Jayakumari(2014),” Prognosis of Diabetes Using Data mining Approach-Fuzzy C Means Clustering and Support Vector Machine”,International Journal of Computer Trends and Technology , vol. 11 (2).

[5] L. G. Kabari and E. O. Nwachukwu(2012),” Neural Networks and Decision Trees For Eye Diseases Diagnosis”,INTECH.

[6] Qeethara Kadhim ,Al-Shayea and Itedal S. H. Bahia(2010),”Urinary System Diseases Diagnosis Using Artificial Neural Networks”, IJCSNS International Journal of Computer Science and Network security, Vol.10 No.7.

[7] Dhanashree S.Medhekar,Mayur P.Bote,Shruti D.Deshmukh(2013),”Heart Disease Prediction using Naïve Bayes”, International Journal Of Enhanced Research In Science Technology & Engineering ,Vol.2 Issue 3. [8] Ms.Rupali R.Patil,(2014) “Heart disease prediction system using Naïve Bayes and Jelinek-mercer smoothing”, International Journal Advanced Research in Computer and Communication Engineering, Vol.3, Issue 5.

[9] A.H. Hadjahmadi, and Taiebeh J. Askari(2012)” A Decision Support System for Parkinson's Disease Diagnosis using Classification and Regression Tree”, The Journal of Mathematics and Computer Science Vol.4(2),.257 – 263. [10] Hian Chye Koh and Gerald Tan.” Data Mining Applications in Healthcare”,Research Gate.

[11] M. Duraira and V. Ranjani(2013),” Data Mining Applications In The Healthcare Sector: A Study”, International Journal Of Scientific & Technology Research Vol. 2, Issue 10.

[12] Hlaudi Daniel Masethe and Mosima Anna Masethe(2014),” Prediction of Heart Disease using Classification Algorithms”, Proceedings of the World Congress on Engineering and Computer Science Vol II , 2224.

[13] Jyoti Soni, Ujma Ansari and Dipesh Sharma(2011),” Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”, International Journal of Computer Applications ,Vol. 17– No.8.

[14] R. Chitra and V. Seenivasagam(2013),” Review Of Heart Disease Prediction System Using Data Mining And Hybrid Intelligent Techniques”, ICTACT journal on soft computing, Vol. 03, Issue 04.

[15] S. Syed Shajahaan, S. Shanthi and V. ManoChitra(2013),” Application of Data Mining Techniques to Model Breast Cancer Data”, International Journal of Emerging Technology and Advanced Engineering, Vol. 3, Issue 11.

[16] Vikas Chaurasia and Saurabh Pal(2014),” A Novel Approach for Breast Cancer Detection using Data Mining Techniques”,International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1.

[17] Ahmad LG et al.,(2013),” Machine Learning Techniques for Predicting Breast Cancer Recurrence”, Health & Medical Informatics, Health Med Inform 2013.

[18] Ronak Sumbaly, N. Vishnusri and S. Jeyalatha(2014),” Diagnosis of Breast Cancer using Decision Tree Data Mining Technique”, International Journal of Computer Applications, Vol. 98– No.10.

[19] K. Rajalakshmi & Dr. S. S. Dhenakaran(2015),” Analysis of Datamining Prediction Techniques in The Healthcare Management System”, International Journal of Advanced Research in Computer Science and Software Engineering,Vol.5,Issue4.