



## Prediction of Kidney Disease by using Data Mining Techniques

Neha Sharma, Er. Rohit Kumar Verma  
Department of Computer Science & Engineering  
India

**Abstract**— Data mining is the term used to describe the process of extracting value from a database. A data-warehouse is a location where information is stored. The type of data stored depends largely on the type of industry and the company. Many companies store every piece of data they have collected, while others are more ruthless in what they deem to be “important”. The kidney disease of a patient is analyzed and the results are to compute automatically using the data set of the patient. The kidney diseases in the patient are predicted using rule based prediction method conventionally. The probabilistic theory cannot be used for getting the results because it is concerned with the life of a patient and accurate results are required. Chronic kidney disease is a large and growing problem among aging populations. Detection & prediction of kidney disease is important for providing proper & right treatment to the patient. The conventional systems that were used for detection of the kidney diseases used data sets of the patient and generated results using if-then rules along with and-or mechanism. This new technique uses both fuzzy systems and neural networks called as neuro-fuzzy system that will generate results on the basis of obtained input data set. This new system that is made from the combination of both fuzzy system and neural network generates results by mathematical computation and not on the basis of probabilistic theory. The results that are generated using mathematical computation tend to have higher accuracy

**Keywords**— Data Mining, Knowledge Discovery, Fuzzy, Neural Network, ANFIS.

### I. INTRODUCTION

Data mining is a process in which important or meaningful values are extracted from database. Data warehouse is a kind of storage of data where the data is kept for the purpose of fetching it in near future when it is required. Data Warehouse can store any kind of data particularly the type of data depends upon the kind of industry for which it is being used. Most of the industries keep record of each and every kind of data whereas some companies only store that information which is beneficial and meaningful for them. The data stored in a warehouse is helpful in decision support system. On the basis of historical data the decision regarding the future schemes can be taken easily or much effectively [9].

The example of a banking organization can effectively define the purpose of the data warehouse for an organization. For banking organization income is a very important source for describing the socio-economic situation. The strategies for the benefits of the customer can only be created if and only if the bank has any record regarding the income of its customers. [22] If the bank has a record regarding the income of the customers then they can offer the various schemes like concession on interest rates on various loans. Bank can also alter the limit of credit card issued to the customer. Bank can know the income of the customer only in two ways such as amount deposited by the customer in his account on the daily basis and second way is to access the purchasing habits of the customer by using credit card issued to him.

There is another example which shows the failure of the institution of insurance products to access the data warehouse in a proper manner. By having the access to the information of transaction the institution can have information regarding whether the payments done by the customer to any other broker or not. In this way they can take decision regarding their prospectus. All of this can be achieved only if the data is stored in data warehouses and data mining is performed to extract valid data [9].

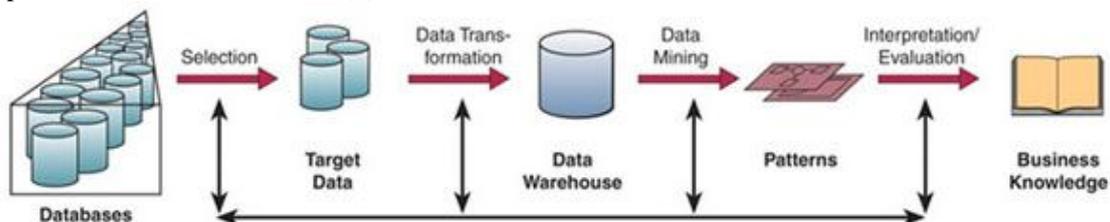


Figure 1. Data mining process

- **Tasks In Data Mining [21]**

Data mining is a process of extracting the interesting information from the large amount of data sets available. Data mining particularly perform the following tasks:

1. Classification
2. Estimation
3. Prediction
4. Affinity grouping or association rules
5. Clustering
6. Description and visualization

Classification, Estimation and Prediction is performed for directed data mining. Directed Data Mining is a term which describes the process when the given data in database is used to create patterns or model that defines the single or multiple meaningful attributes as compare to rest of the attributes.

Whereas Association, Clustering and Description defines the process of undirected data mining. Undirected data mining is the process in which the relationship between attributes is developed [6].

#### **A. Classification**

Classification is a process which is performed to evaluate the characteristics of given object and then these properties are allotted to the existing objects. Classification is done by using classes along with a training set of which includes reclassified objects or examples. The main aim of the classification is to classify the unclassified patterns or data. Examples of classification are as follows:

- Classification of credit applicants on the basis of low, high or medium risk.
- Classification of vegetables and fruits such as mushroom as edible or hazardous.
- To determine the telephone lines connected to internet.

#### **B. Estimation**

Estimation is the second process or task to perform after classification. In this the output is estimated on the basis of input pattern. The estimation is performed on the basis of given input parameters and these variables or parameters are unknown. Example of estimation process is as follows:

- From the mother's qualification as an input variable estimating the number of children in a family.
- On the basis of number of vehicles in a house estimating the total income of that house.

#### **C. Prediction**

Prediction is a process which can be considered either as classification or estimation. It is the process in which the outcome is predicted on the basis of some historical behavior or future values. The example of prediction process is as follows:

- On the basis of buying behavior of the customer to predict whether he will leave in near future or not.
- On the basis of user's behavior and activated plans predicting that which customer would like to have value added services on his connection [7].

#### **D. Association Rules**

Association rule is used to describe the relationship between set of objects. It defines that how various objects are linked or associated to each other with in a database. To understand the association rule let's consider an example, in a given set of transactions each and every transaction consist of items and X and Y are two data items. The relationship among them can be described as the database contained an X tends to contain Y.

#### **E. Clustering**

Clustering is a technique in which the data is divided into sub classes. These sub classes are known as clusters. These clusters contain the related data sets only. This leads to select the data easily when needed. Hence the data is divided into categorical clusters according to the nature of the data sets. There are many algorithms which are used to divide the data into various clusters. Some of the algorithms used for clustering are partitioning method, hierarchy method, grid based method of clustering. There is a difference between classifications and clustering is that the clustering is not based on the predefined data sets or variables whereas the classification of data is performed on the basis of predefined values [6].

## **II. PRESENT WORK**

The kidney disease of a patient is analysed and the results are computed using the data set of the patient. The kidney diseases in the patient are predicted using rule based prediction method conventionally. Probabilistic theory has been used in order to detect whether patient is suffering from kidney disease or not. Technique used for this work was rule based method. Attributes had been extracting from such method and basis on which results had been concluded. Results evaluated from the present work are based on the probability totally due to which it remain unstable. To avoid this problem, a proposed system have introduced in which combination of fuzzy and Neuro have used in which results are evaluated on the basis of mathematical model rather than probabilistic theory.

## **III. PROBLEM FORMULATION**

The probabilistic theory cannot be used for getting the results because it is concerned with the life of a patient and accurate results are required. Statistical approach, Bayesian classifier or associative prediction mechanism cannot be employed for prediction as these algorithms compute results on the basis of probabilistic theory and the obtained results

are not accurate. Accuracy of the results is the prior reason for the development of a new technique for prediction of the kidney disease. The disease prediction needs to be accurate for saving the life of a patient and also it will improve the healthcare facilities. The techniques that are used till date for computing results are based on the if-then rules along with and-or mechanism, but the accuracy of these results is not desired. So, need for a new technique was felt that could generate results with higher accuracy so that efficiency of the medical techniques could be improved.

#### IV. PROPOSED WORK

Chronic kidney disease is a large and growing problem among aging populations. Detection & prediction of kidney disease is important for providing proper & right treatment to the patient. The conventional systems that were used for detection of the kidney diseases used data sets of the patient and generated results using if-then rules along with and-or mechanism. The disadvantage of using this conventional was less accuracy and high probability of getting wrong results. Wrong prediction of the disease could lead to loss of life of the patient. A new technique is proposed for prediction & detection of kidney disease using data set of the patient. This new technique uses both fuzzy systems and neural networks called as neuro-fuzzy system that will generate results on the basis of obtained input data set. This new system that is made from the combination of both fuzzy system and neural network generates results by mathematical computation and not on the basis of probabilistic theory. The results that are generated using mathematical computation tend to have higher accuracy and hence improving the efficiency of the system. A system is developed that accurately predicts kidney diseases.

#### V. METHODOLOGY

In this proposed method the new technique is proposed for the data mining process for obtaining the normalized results. The ANFIS system is used in the proposed in this work. The methodology of the proposed work is defined below:

1. The initial step is to load the data set that is to be processed. The data set consist of the various information of the particular subject that is to be normalized.
2. Next step after loading the data set is to initialization the ANFIS system, the ANFIS system is combination of the artificial neural network and the fuzzy logics. The system is most efficient system for computing the results.
3. After the system is initialized, next step is to initialize the parameters of the system. These parameters can be number of inputs required and the outputs. After this the member ship function are defined. With the help of these membership functions the calculation is done.
4. In this step after parameters and the member functions are defined training of the ANFIS system is done.
5. After training the system the calculation of the performance parameters is done. The performance parameters will show the efficiency of the designed system.
6. Next step is to optimize the parameters of the MNFIS system designed in order to achieve more better and efficient results.
7. Finally the calculation of the optimized results is done and the comparison is performed. The comparison will show the best results.

#### Block Diagram:

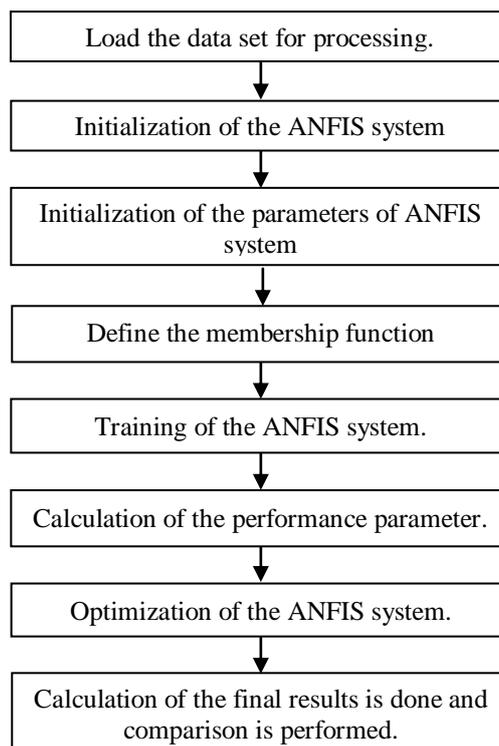


Figure 2 Block diagram of proposed technique

## VI. RESULT AND DISCUSSION

In this section of the paper, results have been discussed using proposed technique. In the proposed work dataset has been taken, parameters have been evaluated in order to check the performance of the proposed and the traditional technique.

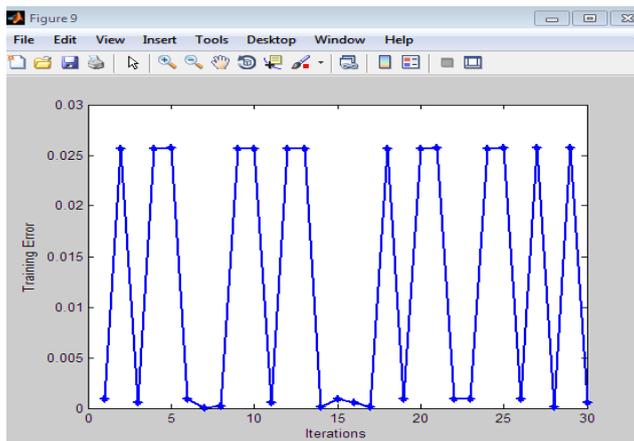


Figure 3 Training errors along its iterations.

In figure 3 the graph shows the number of iterations and corresponding training errors. The graph is observed after doing simulation work of proposed work.

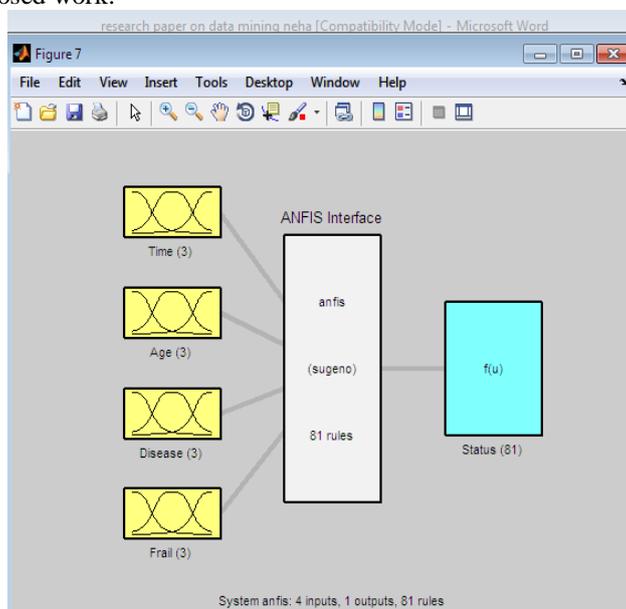


Figure 4 ANFIS interface

Above graph shows the ANFIS interface of the proposed technique where number of inputs are four such as Time, Age, Disease and frail. Sugeno model has been used to evaluate the results and output parameter 'Status' shown the results either patient is suffering from kidney disease or not.

Table 1 Error table

| S. No. | Error          | Old Approach | Proposed Approach |
|--------|----------------|--------------|-------------------|
| 1      | Testing Error  | 1.0327       | 0.04              |
| 2      | Training Error | 3.1668       | 0.02578           |

## VII. CONCLUSION AND FUTURESCOPE

Data mining is an automatic process which is used to remove the meaningful information from the data storage and further use this removed information for various purposes. The extraction of meaningful data can be performed by matching pattern. As the size of the data is increased various methods have been proposed for the data mining. In this proposed work the ANFIS system is proposed which is considered to be better for obtain useful data. Along with this the optimizations if the system is done to improve the efficiency of the results obtained. From the results obtained it is concluded that this method is better and efficient than the traditional systems. The designed system will help in extraction of the useful data from the data set.

As it is concluded from the results that this proposed system is effect approach for the data mining process. In future the method can be advanced by using various optimization algorithms that can increase the efficiency of the system. Also hybrid approaches can be used for obtaining more precise results.

## REFERENCES

- [1] K. Amarendra “Research on Data Mining Using Neural Networks” Special Issue of International Journal of Computer Science & Informatics (IJCSI), , Vol.- II, Issue-1, 2, Pp2231–5292
- [2] Dr. T. Karthikeyan(2012) “ A Study on Ant Colony Optimization with Association Rule” International Journal of Advanced Research in Computer Science and Software Engineering,
- [3] DR. YASHPAL SINGH “NEURAL NETWORKS IN DATA MINING” Journal of Theoretical and Applied Information Technology
- [4] Rafael S. Parpinelli “Data Mining with an Ant Colony Optimization Algorithm”
- [5] Bijaya Kumar Nanda “Class Based Rule Mining using Ant Colony Optimization” Special Issue of International Journal on Advanced Computer Theory and Engineering (IJACTE)
- [6] K.Rajeswari,(2011), “A Novel Risk Level Classification of Ischemic Heart Disease using Artificial Neural Network Technique - An Indian Case Study “ternational Journal of Machine Learning and Computing, Vol. 1, No. 3,Pp 231-235
- [7] Miss. Manjusha B. Wadhonkar(2013), “ Classification of Heart Disease Dataset using Multilayer Feed forward backpropogation Algorithm ” ( IJAIEM) Volume 2, Issue 4,Pp 214-220
- [8] D.Shanthi, “Designing an Artificial Neural Network Model for the Prediction of Thrombo-embolic Stroke” International Journals of Biometric and Bioinformatics (IJBB), Volume (3) : Issue (1),Pp 10-18
- [9] S. Vaishnavi (April, 2014) “Artificial Intelligence Approach for Disease Diagnosis and Treatment”,IJIRCCE, Vol. 2, Issue 4, pp 4000-4007
- [10] Beant Kaur (October, 2014) “Review on Heart Disease Prediction System using Data Mining Techniques”, IJRITCC, Volume: 2 Issue: 10, pp 3003-3008
- [11] Abdel-Motaleb (2012) “Artificial intelligence algorithm for heart disease diagnosis using Phonocardiogram signals”, IEEE, ISSN :2154-0357, pp 1-6
- [12] Ankita R. Mokashi (September, 2015) “Review on Heart Disease Prediction using Ann and Classifier”, IJERT, Vol: 4, Issue:9
- [13] M. Akhil Jabbar (2013) “Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection”, Global Journal of Computer Science and Technology Volume XIII Issue III Version I, pp 5-14
- [14] Jyoti Soni Junw,( 2011) “Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers”, IJCSE, Vol. 3 No. 6, pp 2385-2392
- [15] R.Chitra (October, 2013) “Heart Attack Prediction System Using Fuzzy C Means Classifier”, IOSR-JCE, Volume 14, Issue 2, pp 23-31
- [16] Velu C. M “Heart Disease Diagnosis Using Multiple Kohonen Self Organizing Maps”
- [17] Dimple “Comparative Study of Different Techniques for Heart Disease Prediction System”, SSIJMAR, Vol. 2, No. 2
- [18] Neelamadhab Padhy (2012), “The Survey of Data Mining Applications And Feature Scope” International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.3,Pp 43-58
- [19] Y.Elovici ,”Using Data Mining Techniques for Detecting Terror-Related Activities on the Web”
- [20] Hongjun Lu(1996), “Effective Data Mining Using Neural Networks” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 8, NO. 6,Pp 957- 961
- [21] Anand V. Saurkar,(2014) , “A Review Paper on Various Data Mining Techniques”, International Journal of Advanced Research in Computer Science and Software Engi neering, Volume 4, Issue 4,Pp 98-101
- [22] Arno J. Knobbe,” Multi-Relational Data Mining”
- [23] Abhishek (2013) , “Heart Disease Prediction System Using Data Mining Techniques “ORIENTAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGYVol. 6, No. (4): Pgs. 457-466
- [24] Boshra Bahrami,(2015), “ Prediction and Diagnosis of Heart Disease by Data Mining Techniques “Journal of Multidisciplinary Engineering Science and Technology (JMEST) ,Vol. 2 Issue 2,Pp 164-168
- [25] Deepali Chandna(2014), “Diagnosis of Heart Disease Using Data Mining Algorithm” (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1678-1680