

# APPLICATION OF DATA MINING IN MEDICAL FIELDS

[<sup>1</sup>]Mrs.J.Arockia Jeyanthi,[<sup>2</sup>]Mrs.Sowmya Fernandez &[<sup>3</sup>]Mrs.A.Jane

[<sup>1</sup>]AssistantProfessor,[<sup>2</sup>]AssistantProfessor &[<sup>3</sup>]AssistantProfessor

Department of Computer Science,A.P.C. Mahalaxmi College For Women,Thoothukudi-2.

## ABSTRACT:

The major objective of Data Mining , relatively new field of research is to acquire knowledge from large amounts of data.A large amount of data is becoming available due to regulations and due to the availability of computers in medical and health care areas. On the one hand, the expectations with the practitioners is to use all this data in their work ,but at the same time, such a large amount of data cannot be processed by humans in a short time to make diagnosis, prognosis and treatment schedules. In this paper we shall discuss the applications of data mining in medical fields is evaluated.

**KEYWORDS:** Objective,Acquire knowledge ,Large amount of data,Evaluates tools

## Introduction

Among the most information intensive industry is the Healthcare industry. On a daily basis knowledge , data and Medical information keep growing. In a year as per the estimation carried out five terabytes of data are generated in an acute care hospital. It is very crucial to extract useful information for quality healthcare.

In the use of clinical data a very important role is played by Medical informatics. In these years after year. Useful patterns of information crucial discoveries like the study of different can be found within the data with the help of patterns found when classification of data takes place and for the diagnosis of new diseases pattern recognition is very important.

For quality decision making and to avoid human error computer assisted information retrieval may help support. Though human decision-making is often finest, when there are huge amounts of

data to be classified it is poor. Also when humans are put into stress and immense work efficiency and accuracy of decisions will decrease.

This paved the way to the usage of data mining in medical informatics that is dealing with the database that is found in the hospitals, namely, the hospital information systems (HIS) containing massive amounts of information which includes data from laboratories ,patients information which keeps on increasing rapidly in these years after year. Useful patterns of information can be found within the data with the help of patterns found when classification of data takes place and for the diagnosis of new diseases pattern recognition is very important. These patterns are utilized for further research and evaluation of reports. But how to classify or group this massive amount of data is the other question that arises . Based on similarities present in the data Automatic classification is done. The automatic classifier is proven fruitful only if the conclusion drawn is acceptable to the clinician or the end user.

This at last paves the way for an requirement of some of the data mining algorithms that make use of rules for categorization. The various data patterns within a database.

mining algorithms extracts patterns present in

the training data set, based on which Rules are obtained . The rules thus obtained can be stored in a Personal Digital Assistant (PDA) . In a medical field inputs regarding the patient can be fed to the PDA and classification of the input can take place in real time based on the rules stored in the device.

The difference between data mining and knowledge discovery is that the latter is the application of different intelligent algorithms to extract patterns from the data whereas that is involved in discovering knowledge from data.

## Goals and Objectives

The main aim of this paper is to explain the application of data mining to medical databases, to predict or classify the data with a sensible accuracy.

The extraction of rules required for classification is carried out and stored on a mobile computing device like a handheld computer after the classification of testing data with sensible accuracy. Once the data is inputted based on the rules that are stored into the hand held classification can be done. The end result is classification of data based on the rules that does not need a lot of computation and is most suitable for PDAs.

## Knowledge Discovery in databases [KDD] and data mining

To find patterns or extract knowledge from the database traditional methods use manual analysis.

For example in the case of health care, the health organizations analyze the trends in diseases and the occurrence rates. This helps to take precautions in future in decision making and planning of health care management by the health organizations.

We state the findings (patterns or rules) as data

The other steps like data preprocessing, data selection, data cleaning and data visualization are also a part of the KDD process.

## Data mining

As we have seen before that data mining is one among the most important steps in the knowledge discovery process. Data mining can be considered the heart of the KDD process.

This is the area, which deals with the application of intelligent algorithms to get useful patterns from the data.

Out of the four different types of learning methods we will be only concentrating on two, namely the classification learning and association rules.

There are a number of different types of classification and association techniques . The other name for Classification type of learning is supervised learning and clustering is also known as un-supervised learning.

## Data preprocessing

Each algorithm needs data to be submitted in a specified format. Preprocessing is called as the generation of raw data into machine understandable format.

Transformation of the attributes in the database into a single scale and the

replacement of all the missing values in the data are the other steps that are performed during preprocessing.

### **Machine understandable format**

Raw data can be stored in several formats including Excel ,text, or other database types of files. Sometimes the raw data will not be in any format.

Having data already in a format understandable by algorithms can result in better time efficiency with respect to processing of the data.

In many cases a single case is represented by rows and the attributes that are present within this case are represented by columns.

In some of the free databases that are available online most of them are in comma separated value (CSV) format. CSV format is one where all the attributes are separated by commas and two commas simultaneously stands for a missing data attribute.

Sometimes when attributes are missing, we may find a question mark in place of the missing attribute instead of finding an empty space.

### **Raw data**

The raw data usually has a great deal of noise. With the machine-learning algorithms raw data cannot be used directly for processing. They first need to be preprocessed into machine understandable format.

### **Filling up missing and incomplete values**

Sometimes there are attributes which are missing or incomplete. A most common method of representing missing data, is inputting values that cannot be found in the data e.g. represent missing data as “-1”. Usually one may think that the case is less useful than the rest of the cases in the data set if an attribute is empty. This is not correct as

each of the other attributes contributes useful information towards the set of attribute category. When there are missing values there are a number of methods that can be used for filling these missing attributes, instead of leaving them as missing.

Having efficient methods to fill up missing values extends the applicability in terms of accuracy for many data mining methods. With a larger training set , better rules and decision trees can be developed which contributes towards better classification of the data which results in the increase in the accuracy of the tool

The most common method of filling the attributes quickly and without too much computation is to replace all the missing values with the arithmetic mean or the mode with respect to that attribute.

The other methods are to run a clustering algorithm and replace the missing attributes with the attributes of cases that appear close in an n-dimensional space.

### **Rule based classification**

Another alternative in data mining to the decision tree method is Rule based classification. Thus a rule can be broken up into two parts, the condition (IF) can be considered as one of the tests that are used at the decision node of the decision tree and the conclusion (THEN) that is drawn stands for the classification of the case when this rule is considered.

Another point that needs to be made is that there exists another kind of rule-based classification called the association rule. Although the association rule is very similar to the classification rule, a difference is that association rule can predict any attribute as well as the

final classification and it can be also used to predict any combination of attributes. Thus there can be a number of association rules that are obtained from a small database.

## Conclusion

For automatic classification of data there are a number of data mining algorithms that are found useful. Most of them produce results that are variable in nature. While running one type of data as compared to the rest . Thus finding the best type of algorithm is an interesting and time consuming work.

