

## IMPROVISATION OF DATA MINING TECHNIQUES IN CANCER SITE AMONG VARIOUS PATIENTS USING MARKET BASKET ANALYSIS ALGORITHM

SUNIL YADAV<sup>1</sup> & GAJANAND SHARMA<sup>2</sup>

<sup>1</sup>Department of Computer Science & Engineering, Gyan Vihar School of  
Engineering and Technology, Jaipur, India

<sup>2</sup>Associate Professor, Department of Computer Science & Engineering,  
Gyan Vihar School of Engineering and Technology, Jaipur, India

### ABSTRACT

Hospital inspection supervision or healthcare industries are the region with reference to leadership, management and administration of hospitals, hospital networks and healthcare system. In these days healthcare business generate very large amount of complicated data they are basically found unstructured, incomplete and lot of human mistake for different type of diseases. Healthcare data collected from different type of data source like patients, electronic patient's records, medical devices, laboratory, blood bank, drug store and another hospital resources. The large amount of information, knowledge may be a key source of processed analyzed for knowledge extraction that's permits support maintenance for cost saving and higher perceptual process. Data mining introduce a collection of different type of tool and technique in order to may be applied to the current processed information to getting hidden patterns so that offer healthcare business an additional or further extra source of information for better conclusion for the administration. This paper provide various data mining algorithm used by health care analytics and also we focus on represent analytical data on systematically way to provide result using different data mining algorithm.

**KEYWORDS:** Healthcare, Hospital, Data Mining, Hidden Patterns

### INTRODUCTION

The healthcare market in India is one of the largest and fastest growing industry in the world, it consume nearly 10 per cent of the GDP (Gross Domestic Product) most of the developed or developing nation, healthcare industry contribute a major part for a country's economy. The Indian healthcare sector, one of the fastest growing industries, is expected to outstrip at a compound annual growth rate (CAGR) of 17 per cent during 2011-2020 to reach US\$ 280 billion it is expected to rank amongst the top three healthcare markets in terms of incremental growth by 2020. Indian Healthcare sector provides new and existing players with an only one and special opportunity to achieve and perform innovative research and profits. Healthcare in India also awarded as 'polio Free' country by World Health Organization (WHO). According to a research of McKinsey & Company in the next decennary, consumer awareness and demand for better service and facilities will increase and in India healthcare industry will become third largest service sector employer. Healthcare data is generally different in field and very large in content. So it's almost impossible to manage traditional /routine analytics method. Recent data mining tools and techniques can be used to discover hidden unknown facets of knowledge which may be useful to therapeutic and preventive aspects of healthcare. The new challenge in healthcare data mining is 'Big Data Analytics' supervision. Generally healthcare big data means electronic healthcare data sets which is so

complicated and very large that they are almost impossible to manage with routine / traditional machine and software technique. At present healthcare data there is large volume of data is available to detect and understand the pattern and movement inside the data hence huge amount of big data analytics has probable to renovate healthcare service like: patients care, life and cost scarcity. This paper retrace / explores the Data Mining applications, challenges, show result on different patient with their routing life and some future work / directions in health care. In particular, it discusses Data Mining algorithm and its applications within healthcare sector in major areas. This hospital based survey also providing utilities of various Data Mining techniques they show result in different way such as Clustering, Association Rule and Classification in the healthcare domain. This paper also defines the cancer side and the morphology pattern among various patients of Haryana and surrounding state with the help of above defined different data mining techniques.

## DATA MINING

Data mining is an important step to process the knowledge discovery in database by using different method in order to extract hidden pattern. At present cancer has become general reason of death in men and women in developing and developed countries. According to World Health Organization (WHO) around 40-50% of all cancer type are treatable but at primary stage. We can reduce cancer death to detect it earlier or primary stage. Predicting the sequel of a diseases or cancer is one of the major interesting and challenging jobs where to developing and used data mining algorithm. The use of data mining tool or automated tools, large volume of medical / patients data are being collected from different hospital and made available to the medical research group found result using data mining technique has become a popular research tool for medical researcher to identify and exploit patterns and relationship among large number of variable. Data mining tool holds great probable for the healthcare business to enable health system to systematically use data / information and analytics to identify inefficiencies and excellent practice that improve care and reduce costs. In general, knowledge discovery in database (KDD) and Data Mining are related term Data Mining Search knowledge out of big data there is a vast process includes in data selection, data cleaning, data integration, data transformation, data mining, knowledge representation and pattern evaluation. The Figure 1 show how Data Mining extracts unknown pattern / knowledge from a given Dataset.

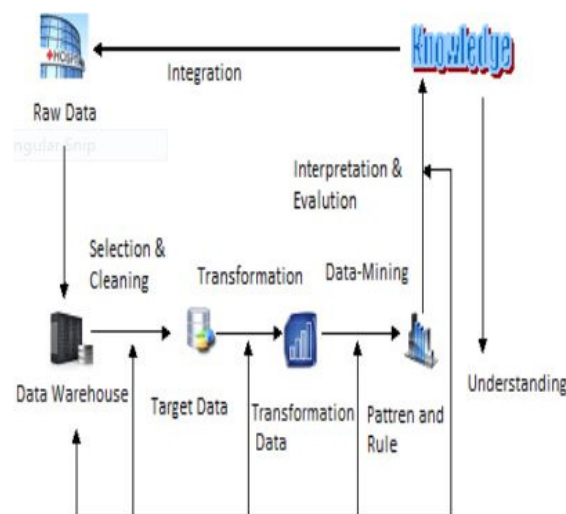


Figure 1: Knowledge Discovery Process

## HEALTHCARE ANALYTICS USING DATA MINING

Many industries successfully use data mining to develop retail industry model customer response. However, data mining in healthcare today remains, for the most part, an academic exercise with only a few pragmatic success stories. Academicians are using data-mining approaches like decision trees, clusters, neural networks, and time series to publish research. Healthcare, however, has always been slow to incorporate the latest research into everyday practice. Basically health care data set is available in flat file, RDBMS Relational database or some advance database system. Healthcare data is collected from different data source like as: OPD (outpatient department), operation theatre module, laboratory information system, chemotherapy module, radiotherapy module, blood bank and drug store. Generally healthcare data is present an unstructured way in the database for different type of cancer patients.

### Healthcare Dataset

Nature of healthcare data is very special because it's found very complicated or unstructured way. To mine healthcare data all dataset / information requirement to be changed into numeric value. With the help of numeric value it's easy to apply data mining algorithm to find hidden pattern and generate report among various patients. Initially data is collected from different source as given above then transform and load from the transaction database into Data Warehouse as show in figure 2.

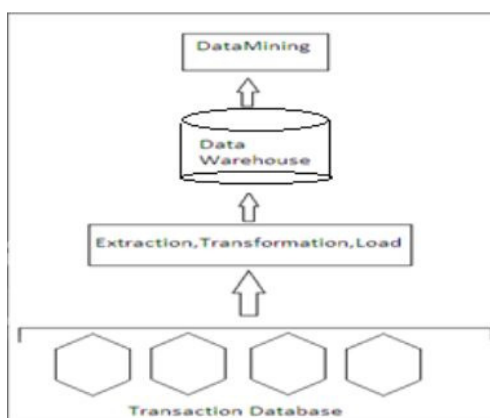


Figure 2: Healthcare Data Mining Concept

### Patient Dataset

Data mining discover knowledge from patient's dataset and these dataset is a collection of different data source as mentioned above. In this research work we have collected data source from RCCR (Regional Cancer Center Rohtak) and prepared and analysis dataset for different regions of Haryana and Surrounding states and applied Data-mining algorithm to extract hidden pattern. This helps us to determine cancer site and morphological pattern among various patients and to develop suggestive management information system to improve cancer treatment.

### Preliminary Analysis of Dataset

Preliminary analysis of dataset is an important step that is used to transform unstructured data or row data into a proper format that makes it possible to apply data mining techniques and improve the quality of data. Preliminary analysis is to identify the hospital and government needs, perform economic, technical analysis and green analytics, perform cost benefit analysis and create suggestive management information system. There should be enough expertise available for

database query and data mining algorithms and software's for doing analysis.

Before performing analysis or draw pattern the following questions emerge:

- How much time should be spent on collective and creating patient's dataset or management information system? Generally data collection is very difficult task and no such rule to find out fixed time. This is depends on dataset size, complexity end-use, contractual obligation is few parameters on which it should be decided.
- How to differentiate very large datasets? Collectively dataset in my thesis is around 64 GB of unstructured data in the database for different kind of cancer patients and diseases. So we have analysis and structured the data and created dataset for cancer patients to determine cancer site and morphology pattern among patients from various regions of Haryana and surrounding states.

Other major question that arises is which type of data mining algorithms we should apply?

Firstly we have studied various research paper related to healthcare in data mining and different type of algorithms we can use on healthcare data and then we applied as market basket algorithm and cluster on our dataset which we will discuss in further chapters.

## **MEDICAL DATA SELECTION & PREPARATION**

Medical data selection and preparation is very important task and it's also more time for processing data mining. It took around 50-60 percent of time to prepare dataset for the whole data mining process algorithm. Health care dataset made by different source of data and then analysis and processing aims to developing a common data warehouse for using different data mining algorithm. Medical data selection is basically collection of varies data sources such OPD, laboratories, operation theater, drug store, blood bank, therapy modules which holds important data for patient diseases information. After Collection data from different data source all data update centralized database according to patient's details. Then, Apply algorithm and process selected data is useful or not and then passed it Research department. Medical data preparation verifying selected data accuracy and exactness. After selected accurate data, Data cleaning is next step to reduce missing, noise and inconsistency data that affected result.

## **CANCER DIAGNOSIS USING DECISION SUPPORT SYSTEM**

Today cancer is one of most common cause of death both man and women and cancer become social health problem in world. According to World Health Organization approximate 40% of all cancer type are treatable but it is possible when found cancer sequel at primary stage. The most common cancer cause are found some pernicious habits, alcohol, smoking, adipose (weight gain). Cancer Diagnosis and Prognosis using decision support system is really an interesting and challenging part. Different type cancer prognosis using DSS to analysis patient's history and draw pattern to how many possibilities different kind of people like as: smoking and non-smoking, take alcohol and some other pernicious habit. Healthcare data is generally available in unstructured way and we convert all these data into numerical way to apply data mining algorithm. A data mining algorithm is a set of numerical calculations that creates a data mining model from data. To create data mining model, the data is primarily provided to algorithm for analyzing and then it look for specific type of patterns and trends.

## DATA MINING ALGORITHM IN CONTEXT OF HEALTH CARE

A set of large and unstructured data present at data warehouse to analysis and convert into numerical information. Data mining algorithm basically apply numerical information to create data mining model or pattern that help to patient discover predict of cancer. Primary provide numerical data to data mining algorithm for analysis and then it provide a special type of pattern. Data mining model create data help of these model:-

- Clustering show how different data set related to each other in data base.
- A decision tree provides some sequel of diseases and correlates these which will predict outcome and how different criteria affect.
- Different type of mathematical model.
- Different set of cluster produce how cluster come together in transaction.

Choosing one of correct algorithm for different analytical task is one of challenging part because when we apply different algorithm on same task every algorithm will produce different results. Selection of algorithm is purely decision on input, predict and output column.

Selection of algorithm:-

- **Classification Algorithm**

Classification is generalizing known to predict more than one variable based on different type input

- **Association Algorithm**

Association algorithm defines relationship between variable. Market basket on healthcare based on association rule on data mining. Market basket generally used on e-commerce to monitoring customer transaction and customize according to user need as well as market basket we can apply on healthcare data to find predict of cancer patients and diagnosis and prognosis cancer and related diseases.

- **Clustering Algorithm**

Clustering will discover group they have some similarities on different bases. Different cluster have some common task.

- **Regression Algorithm**

It will predict many variables based on other variable

Data mining have multiple option based on single mining structure so before going to apply different we can choice single mining solution such as association algorithm, decision model etc.

**Table 1: Selection Algorithm by Different Task**

Example on Given Task	Algorithm Used
<b>Predict a Discrete Value</b> Categorized of patients according to different outcome and explore related factor. Calculating the categories of server failure within the next four year.	Decision Trees Algorithm Naïve Bayes Algorithm Clustering Algorithm Neural Network Algorithm
<b>Predicting a Continues Attribute</b> Predicting next year sales Predicting cancer diseases according to patients pernicious habits Calculating and generating risk factor according to given demo graphs.	Decision Trees Algorithm Times Series Algorithm Linear Regression Algorithm
<b>Search Similar Item in Transaction</b> Market basket analysis Based on more patients history their pernicious habits suggested patients to save life of diseases outcome on primary stage Suggested additional product to buyer or customize page according to user.	Association Algorithm Decision Trees Algorithm

## CONCLUSIONS AND FUTURE SCOPE

Data mining in healthcare is new emerging technology. Scientists are running behind the different data mining technology to find best solution for cancer. Data mining discover one of the best solution for cancer diseases to treat patient's help of past patients history and knowledge extract from the large data collected from previous years. Data mining algorithm play important role in healthcare industries to diagnosis and prognosis cancer diseases. Healthcare industries can improve their result using data mining techniques and save money and patients life.

## REFERENCES

1. Balasubramanian T. , "An analysis on the impact of fluoride in human health (dental) using clustering data mining technique", IEEE International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME)", Salem, Tamilnadu, 370 – 375, (2012).
2. Patil, D.D., "Dynamic data mining approach to WMRHM", 7th IEEE Conference on Industrial Electronics and Applications (ICIEA), Singapore, 978, (2012).
3. Carel, Rafael and Barak, Dotan, "Utilization of Data-Mining Techniques for Evaluation of Patterns of Asthma Drugs Use by Ambulatory Patients in a Large Health Maintenance Organization", Seventh IEEE International Conference on Data Mining Workshops, 2007. ICDM Workshops, Omaha, NE, USA, 169 – 174, (2007).
4. Reyes,A.J.O.;Garcia,A.O.;Mue,Y.L. "System for Processing and Analysis of Information Using Clustering Technique", Latin America Transactions, IEEE (Revista IEEE America Latina) 12( 2 ), 364 – 371, (2014).
5. Goil, S. And Choudhary, A., "A parallel scalable infrastructure for OLAP and data mining", International Symposium Proceedings on Database Engineering and Applications, IDEAS '99, Montreal, Que, 178 – 186, (1999).

6. Shweta Kharya discussed, “Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease”, International Journal of Computer Science, Engineering and Information Technology (IJCSSEIT), 2(2), (2012).
7. Ada and Rajneet Kaur, “A Study of Detection of Lung Cancer Using Data Mining Classification Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, 3(3), (2013)
8. Abdelghani Bellaachia, Erhan Guven, “Predicting Breast Cancer Survivability Using Data Mining Techniques”, Department of Computer Science The George Washington University Washington DC available at [eguven@gwu.edu](mailto:eguven@gwu.edu).
9. Kawsar Ahmed, Abdullah-Al-Emran, TasnubaJesmin, Roushney Fatima Mukti, MdZamilurRahman, Farzana Ahmed, “ Early Detection of Lung Cancer Risk Using Data Mining”, Asian Pacific Journal of Cancer Prevention, Vol.14, (2013).
10. V.Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra “Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques”, International Journal of Computer Science and Information Technologies, Vol. 4 (1), 39 – 45 (2013)

