

# A Survey : Study of Data Mining and Data Warehousing in Healthcare

Arvind Singh

Department of Computer Science, Awadhesh Pratap Singh University Rewa, Madhya Pradesh, India

## ABSTRACT

### Article Info

Volume 8, Issue 6

Page Number : 117-121

### Publication Issue :

November-December-2021

### Article History

Accepted : 07 Nov 2021

Published: 15 Nov 2021

Health care is one of the speedy growing areas. The Health care system contains large amount of medical data which should be mined from data warehouse. The mined data from data warehouse helps in finding the important information. Comprehensive amount of data in health care database need the growth of tools which can be used to access the data, analyze and analysis the data, discovery of knowledge, and versed use of the stored knowledge. The health care system has lot of data about the patient's details, medications etc. In this paper we have studied different data mining and warehousing techniques used in healthcare areas.

**Keywords :** Health Care System, Data Mining and Warehousing Techniques

## I. INTRODUCTION

The data mining is the extraction of the important and useful information from the data warehouses and databases. The main use of Data mining is for scientific and commercial field [2]. In this study we studied on the Data Mining applications in the scientific field. Data mining distinguishes in scientific area itself, that the character of the datasets. It is often very different from the traditional applications of data mining. We have done detailed survey on data mining applications in this work on healthcare system. Description of the type of data used and the information fetched as output.

Data mining techniques and algorithms are applied in the healthcare system that plays a vital role in the prediction of the diseases.

Finding the important knowledge from the database and data warehouse is the main goal

behind the application of data mining. Data mining is also called as knowledge finding or knowledge discovery from the large amount of data. As its name this suggests, knowledge discovery is continuous iterative process, which is a combination of developing and understanding the application.

In health care system, an institution uses data mining tools and techniques to answer the patients question quickly, that are normally very time consuming. Preparation of databases to finding the predictive information plays vital role on in these areas.

The expansion of the health care covers too many people as possible by which they can provide financial assistance to help them with the lower income

purchase coverage [1]. Reducing the health care system disparities that are the main purpose of present situation, it decreases the costs that are associated with the increased diseases load born by lot of population growth.

This paper covers the comparison of the data mining techniques with the health care system problems. The comparative study always helps to finding the accuracy. It helps to do prediction by using the data mining applications in the healthcare system. In this comparative study, it leads the aspiring researchers in the area of data mining by which data mining techniques gives a more accuracy in fetching the information from healthcare system data.

The use of Data Mining is in many areas such as engineering, marketing, customer relationship management, analysis of medicine, prediction, web mining, mobile and mobile computing.

The uses of the data mining applications are as follows:

- To detection of fraudulent phone calls
- Prediction of marketing
- Prediction of disease etc.

## II. REVIEW OF PAPERS

We have done a review on the application of data mining in health care system field. Which are the fragile marks of the present knowledge also including the methodological, constitutive contributions?

**Jayanthi Ranjan 2007 [1]**, In his paper he explains data mining how can helps in discovering and extracting knowledge by using the useful patterns of the large amount of data to find the possible observable patterns. This paper encompasses the importance and the ability of Data mining in improving the quality of the decision making process in the medical industry.

**Hian Chye Koh and Gerald Tan [1], 2005** explores data mining applications in healthcare. He also discusses data mining applications within healthcare system in important field such as the evaluation of health treatment effectiveness, management of healthcare system, customer patient relationship management, and the fraud detection. In his paper he also gives an

example of a healthcare system and data mining application involving in the identification of risk linked with the onset of diabetes. And also highlights the limitations of data mining as well as discusses future directions.

**Ruban D. Canlas Jr.2009 [3]**, in this research paper provides a survey of presently available techniques of KDD, using data mining techniques for healthcare system and public health system. Here also explains some issues and challenges linked with data mining and healthcare system in general.

Growing number of data mining techniques and applications, including analysis of health care centers for better health related policy-making, detection of disease blast and preventable hospital deaths, and detection of fraudulent claims also.

**Khitish Kumar Gadnayak et. al. 2018 [15]**, this paper author presents the different data mining and warehousing techniques as well as application used in healthcare area for the good decision making.

**M. Durairaj, K. Meena [17], 2013** discusses about a combination of the prediction system which includes Rough Set Theory and Artificial Neural Network (ANN) for the order of the medical data. The process of developing a new data mining technique and software system for assisting the competent solutions for medical data analysis has been also explained. They explain and propose a hybrid tool which incorporates RST and ANN for making a skilled data analysis and also indicative predictions. They took the experiments on data set for the prediction of excellence of different animal semen for the purpose.

**S. Dhanalakshmi et. al.[16], 2020** In this paper, they focused on various techniques to resolve the issues in healthcare system. In this paper contains various mining techniques and its procedures to identify and finding all extract errors by which error free information can be found easily from vast health related data like patient and their problems. Using various data mining techniques can able to analyze and

produce survey of appropriate technique to find out results in healthcare services.

**K. Srinivas et. al. [4], 2010** discusses on the probable use of the classification which is based on data mining techniques such as Rule Based, Decision tree, Naïve Bayes and Artificial Neural Network to the large volume of medical data. The takes the parameters that are considered here are age, sex, blood pressure and blood sugar which can predict the likelihood of patients getting a heart disease.

**Shweta Kharya [5], 2010** she has explains about the various data mining approaches and she have been utilizes for the diagnosis of breast cancer and for the prognosis of the Decision tree is found as the best predictor with the utmost accuracy of 93.62%.

**Elias Lemuye [6],** He has discussed about the disease caused by HIV that weakens the body that can no longer fight the simple infections. Their algorithm is used to find association rules. Many tools also used such as WEKA 3.6 to mine the data to implement the algorithms, J48 classifier perform the classification with an accuracy rate of 81.8%.

**Arvind Sharma et al. [7], 2012** explains about what a Data mining can contribute in the blood bank area. They used algorithm J48 and WEKA tool. They also used Classification rules to perform classification with an accuracy rate of 89.9%.

### III. WHAT IS DATA MINING?

The process of reorganizing, analyzing and thinking about data is way by which we can find what the data does. Data Analysis is a process of cleaning, inspecting transforming, and modeling various data. The main purpose of data analysis is to finding useful information, providing conclusions, and help in decision making. Data analysis consists of approaches, including various techniques under a list of names, in various business, science, and social science area. [8]

The data mining is an automatic and semiautomatic analysis of large amount of data for the extraction of interesting patterns of data records known as cluster analysis, a collection of records for anomaly detection,

and to find out dependencies i.e., association rule mining and sequential pattern mining. The database techniques are spatial indices. These patterns are used in analysis i.e., in machine learning and predictive analytics.

Data Mining is the finding and discovery of unknown information from the databases [13] [14]. Data fetching, data fishing and data snooping refer to the use of data mining method. The sample part of a larger population dataset which are too small for reliable statistical inferences to be made to validate the patterns that discovered. All these techniques can be used in the creation of new hypothesis to test data against the larger data.

The Assessment of data mining functions and products are the results of the impact from many of the disciplines, in which we can includes the databases, information retrieval, statistics, algorithms, and machine learning [8].

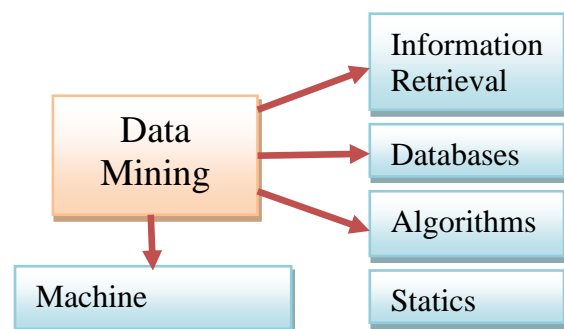


Figure: Historical Data Mining

### Development of Data Mining and Data Base

The development of Data mining system is started in early 1960s. That time data mining was just a file processing system. The next stage of it was Database management Systems which started in the year of 1970s and was still under progress till early 1980s. OLTP, Data modeling tools and Query processing worked.

There are three main categories in which a database management system worked.

1. Advanced Database Systems, which was assessed

in the 1980s to present. In this stage Data models and Application oriented work have been done.

2. In This part Data Warehousing and Data Mining worked since late 1980s to present.
3. At this part is Web based Database Systems which was started from 1990s to present. This includes Web mining and XML based database systems.

#### **Application and Task of Data Mining**

We can classify into two categories:

- Scientific Engineering & Healthcare Data
- Business & E-Commerce

In Healthcare system there are lot amounts of data about such as patient's records, health care resources, their diagnosis, medical devices and also electronic patient records etc. Application of Data mining in healthcare system can be grouped as the assessment into broad categories [1, 9].

Many technologies used here to help pharmaceutical institutions to manage their records and inventories by which the can develop new services and products. Using these records pharma institutions can take better organizational decision making. Hospital also can generate huge amount of data from data warehouses and can take better decisions to help patients [11].

Some of the other type of databases in which lot of data may be Biological database that belong to various type of fields [12].

#### **Task of Data Mining**

There are two categories of tasks of Data mining:

- Predictive and
- Descriptive

#### **IV. CONCLUSION**

In the study of data mining techniques that are used for the data analysis and Knowledge discovery in healthcare system is carried out. In this paper we studied only for the comparison of the different types of data mining applications in the healthcare system for the fetching useful information and data. It is a difficult task for the prediction of diseases using Data

Mining applications but this effort decreases the human efforts and increases the diagnostic quality.

#### **V. REFERENCES**

- [1]. Jayanthi Ranjan, Application techniques in pharmaceutical]ournal of Theoretical and Applied Technology, (2007).
- [2]. Hian Chye Koh and Gerald Tan, Data Mining Applications in Healthcare, Journal of Healthcare Information Management –Vol 19, No 2.
- [3]. Ruban D. Canlas Jr., MSIT., Healthcare: Current applications 2009.
- [4]. K. Srinivas , B. Kavitha Rani and Dr. A. Govrdhan, Applications of Data Mining Healthcare and Prediction - International Journal on Computer Science and Engineering (2010).
- [5]. Shweta Kharya singh ,Data Mining—U Techniques ForDiagnosis And Prognosis - International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.
- [6]. Elias Lemuye, status Predictive—Hiv Modelings using Data Mining Technology.
- [7]. Arvind Sharma and P.C. Guhan - Number of Blood Donors through their Age and Blood Group by using Data Mining techniques - International Journal of Communication and Computer Technologies Volume 01 –No.6, Issue: 02 September 2012.
- [8]. Boros E., P.L. Hammer, T. Ibaraki, A. Kogan.(1997). Logical Analysis of Numerical Data. Mathematical Programming, 79:163-190.
- [9]. Boros E., P.L. Hammer, T. Ibaraki, A. Kogan, E. Mayoraz, I. Muchnik.( 2000). An Implementation of Logical Analysis of Data. IEEE Transactions on knowledge and Data Engineering, 12(2):292-306.
- [10].Crama Y., P.L. Hammer, T. Ibaraki. (1988). Cause-effect Relationships and Partially Defined

- Boolean Functions. *Annals of Operations Research*, 16:299-325.
- [11].David Wai-Lok Cheung, Vincent T. Ng, Ada Wai-Chee Fu, and Yongjian Fu. (December 1996). Efficient Mining of Association Rules in Distributed Databases, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp. 911-922.
- [12].E. Boros, P. L. Hammer, T. Ibaraki, A. Kogan, E. Mayoraz and I. Muchnik(December 1996) An implementation of logical analysis of data, *RUTCOR Research Report RRR 22-96*, Rutgers University, 1996., pp. 911-922.
- [13].Hammer P.L.(1986). The Logic of Cause-effect Relationships, Lecture at the International Conference on Multi- Attribute Decision Making via Operations Research-based Expert Systems, Passau, Germany.
- [14].Proceedings of 4th international conference on statistical sciences Volume (15), University of Gujrat Pakistan, 15: 78
- [15].Khitish Kumar Gadnayak et. al., A Survey on Data Mining & Warehousing Algorithms and its Application in Medical Science, *International Journal of Research in Engineering and Science (IJRES)*, Volume 6 Issue 7 || July. 2018 || PP.63-68
- [16].S. Dhana lakshmi et. al., Survey on Information Mining Procedures Utilized in Healthcare Services, *International Journal of Engineering Research & Technology (IJERT)*,2020
- [17].M. Durairaj, Ranjani, Data Mining Applications In Healthcare Sector: A Study *International Journal Of Scientific & Technology RESEARCH* VOLUME 2, ISSUE 10, OCTOBER 2013.
- Cite this article as :**  
Arvind Singh, "A Survey : Study of Data Mining and Data Warehousing in Healthcare", *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 6, pp. 117-121, November-December 2021. Available at doi : <https://doi.org/10.32628/IJSRSET218618>  
Journal URL : <https://ijsrset.com/IJSRSET218618>