WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analysis of Data Mining Tools and Techniques in Medical Field

Mr. K. M. Pradeepan[1], Mr. A. Ayyanar[2], Mr. K. Ajith[3], Ms. G. Anitha[4], Ms. R. Akalya[5], Ms. V. X. Bhavyasri[6]

[1] Assistant Professor, Department of MCA, K.S.R. College of Engineering, Tiruchengode, pradeepan85@gmail.com
[2,3,4,5,6] Student, Department of MCA, K.S.R. College of Engineering, Tiruchengode, ayyanarlovely4344@gmail.com, ktajithkumar2001@gmail.com, anigopal2901@gmail.com, akalyaramasamy2001@gmail.com vxbhavyasri@gmail.com

ABSTRACT

In medical field the organization collects large amount of data that are not mined to find out unknown information. In medical industries everyday come new treatments and medicine to produce well conclusion and remedy to patients to achieve the quality of service. Data mining techniques and tools are can be used in many fields in different ways. Various organizations are used data mining is a valid for data analysis. The main purpose of this paper is explain information about what type of tools and techniques are used to clarify accuracy level of various problems. This data mining techniques reduces the human effects and cost effective.

Keywords: Data mining, Knowledge discovery in databases, Techniques.

1. Introduction

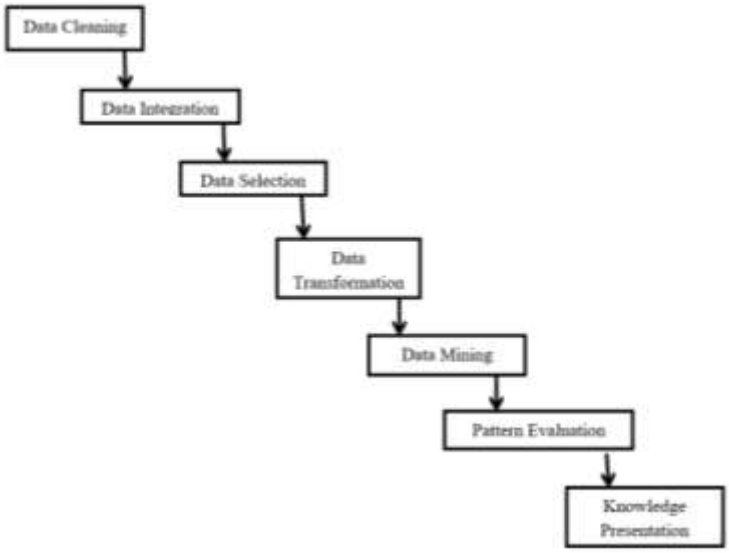
Data mining is defined as observe new information from a data warehouses or databases. Data mining is also called as knowledge discovery in databases (KDD). In medical industry data mining offers more benefits. In medical industry data mining techniques are used to estimate and finding of the diseases. Data mining applications are used in medical industry such as pharmaceutical industry and hospital management. Data mining application is used to catch the valuable and unknown information from the databases.

The health organization is made data for analyse the data in order to mark vital conclusion regarding patient health. This data contain details about hospitals, patients, cost of treatment, medical claims etc., Data mining is used to create a powerful tool to analysing and extracting information from this complex data. The analysis of health data is used to improve the presentation of management jobs. In medical industry the data mining tools are used to find out how much patients are having a similar type of diseases, which helps to that industry is for provides an active treatments to patients. So the tools of data mining in medical field are to improve the medical services in cost effective manner.

2. Data mining process

i. Data Cleaning
ii. Data Integration
iii. Data Selection
iv. Data transformation
v. Data Mining
vi. Pattern Evaluation
vii. Knowledge Presentation

Figure: 1 Process of Data Mining



(i) Data cleaning

Data cleaning is one of the processes which is used to establish inaccurate and incomplete data and then develop the quality of data by correction of data. Data cleaning is mainly used to improve the quality and reduce error.

(ii) Data Integration

Data integration is called as data pre-processing technique that combines the data from multiple unrelated data sources into a reasonable data store. It is used to combine multiple data sources.

(iii) Data selection

Data selection is the process, which is used to take relevant data from the database for study task. Before the data selection, the data consolidation and data transformation was performed.

(iv) Data Transformation

Data transformation process is used to transform consolidated data into particular format for mining by carry out form for mining by performing brief statement.

(v) Data Mining

Data mining is the process of reduce large data sets to identify patterns and initiate relationships to reduce problems through data analysis.

(vi) Pattern Evaluation

It is critical step in which intelligent techniques are applied to extract patterns potentially useful. Surely interesting patterns representing knowledge are identified based on given measures.

(vii) Knowledge Presentation

It is the final step in which the uncover knowledge is visually represented to the user. This step uses visualization techniques to help users understand the data mining results.

3. Data Mining Tools

Six best open source data mining tools are

(i) R-Programming

(ii) Weka

(iii) Knime

(iv) Orange

(v) Rapid Miner

(vi) Natural Language Toolkit

4. Data Mining Techniques

4.1 Association

Association is one of the most important and well known techniques of data mining. A pattern is learned based on association between data in the similar transaction. Association technique is used to classify a set of products that customer regularly purchase together by basket analysis in market. Investigation buyer`s buying lifestyles by using association techniques.

4.2 Clustering

Clustering is one of the most important data mining techniques that make a valuable cluster of objects which have same characteristics using the automatic technique. This technique explains the classes and put objects in all classes while in the classification techniques, objects are allotted in to predefined classes. The clustering technique is particularly used in library to take or identify particular author of book from more than one book.

4.3 Classification

This is the common data mining technique depends on machine learning. Classification is used to divide each data in a set of data into one of a predefined set of groups. This techniques are used some mathematical logics such as linear programs, statistics, neural network etc., New software is introduced by classification technique for how to classify the data items into the group.

4.4 Sequential patterns

It is used to recognize associated patterns or fashions in transaction data over a business period.

5. Data Mining Techniques in Medical

Neural Network, Decision tree, K-nearest neighbour, K- means clustering these data mining techniques are widely used in medical field.

5.1 Neural Network

In neural network the classification is one of the most consequential and scientific research areas. The neural network utilizes the prognostic decision making in medical field by the detailed set of rules. Powerful contrivance provided by the neural network to help the medical practitioner to review and model and make critical clinical data in medical applications.

5.2 Decision Tree

Some standard tree algorithms are implemented in data mining for medical data; c4.5 and ID3. C4.5 is denoted as called version j48. Mostly j48 is providing a better result with compare than c4.5. Weka data mining tool, j48 is a strengthen version of c4.5. The c4.5 generated the decision tree used for classification process.

5.3 K-Nearest Neighbour

K-means clustering is a simplest, measurable and repetitive method used for entirely enormous sets of data. An algorithm is used to maintain the objects based on attribute into number of objects (clusters). These centres' should be position by a distorted means as different place need different result.

5.4 Naïve Bayes

One of mature and historical classification algorithm is the naïve Bayes algorithm. This algorithm is especially attractive because of its clarity and elegance, activeness. Naïve Bayes is a supervised classification method it can rectify difficulty require both unqualified and continuous evaluate attributes.

Techniques	Advantages	Dis advantages
Artificial Neural Network	1. Error susceptible 2. Efficient 3. Robust in noise domain	1. Critical model with high duration of timing
KNN	1. Need fast training	1. Minimum area to storage

	2. Easy to understand	2. Testing time is slow
K-means clustering	1. Efficient 2. Fast 3. Easy to use	1. For non-linear data set the algorithm was failed 2. Unable to handle large data
Naïve Bayes	1. Perfect Outcomes 2. Easy to Handle	1. Loss of Validity

6. Conclusion

In this paper present an overview of the present research carried out by using techniques of data mining for the prognosis and diagnosis of various diseases. The main purpose of this study is to discover the most

7. References

data mining algorithms. By using the data mining techniques the result will be most accurate. More number of data mining algorithms is applied in the experiments and produce good results in medical field. This study explains about the data mining techniques and advantages and dis advantages of those algorithms.

[1] Mittal P. et al. 2013. “Study and Analysis of Predictive Data Mining Approaches for Clinical Dataset”, International Journal of Computer Applications, Volume 63, No. 3.

[2] Naib M. et al. 2014. “Predicting Primary Tumors using Multiclass Classifier Approach of Data Mining”, International Journal of Computer Applications (0975 – 8887), Volume 96, No. 8.

[3] Qarzaie S. et al. “Using the Data Mining Techniques for Breast Cancer Early Prediction”.

[4] Rani K. 2011. “Analysis of Heart Dksiseases Dataset Using Neural Network Approach”, International Journal of Data Mining and Knowledge Management Process (IJDKP), Vol. 1, No. 5.

[5] Saurkar A. et al. 2014. “A Review Paper on various Data Mining Techniques”, International Journal of Advance Research in Computer Science and Software Engineering, Volume 4, Issue 4, ISSN: 2277 128X, pp. 98-101.

[6] Brosette SE, Spragre AP, Jones WT, Moser SA. A data mining system for infection control surveillance. Methods Inf Med,39: 303-310, 2000.

[7] M. Ridinger, “American Healthways uses SAS to improve patient care”, DM Review, vol. 12, no.139, 2002.

[8] M. Durairaj, V.Ranjani, Data mining applications in healthcare sector: A Study, International Journal Of Scientific & Technology ResearchnVolume 2, Issue 10, ISSN 2277-8616, October 2013.

[9] Anonymous. Texas Medicaid Fraud and Abuse Detection System recovers $2.2 million, wins national award. Health Management Technology, vol. 20, no. 10, 1999.

[10] H. C. Koh and G. Tan, “Data Mining Application in Healthcare”, Journal of Healthcare Information Management, vol. 19, no. 2, 2005