*International Journal of Engineering Research and Modern Education (IJERME)*
*Impact Factor: 7.018, ISSN (Online): 2455 - 4200*
*(www.rdmodernresearch.com) Volume 3, Issue 2, 2018*

# A LITERATURE SURVEY ON DATA MINING TECHNIQUES AND CONCEPTS

**N. Thinaharan*, B. Chitradevi*, P. Malathi** & K. Kalpana****
* Assistant Professor in Computer Science, Thanthai Hans Roever College (Autonomous), Elambalur, Perambalur, Tamilnadu
** Research Scholar, Department of Computer Science, Thanthai Hans Roever College (Autonomous), Elambalur, Perambalur, Tamilnadu

**Cite This Article:** N. Thinaharan, B. Chitradevi, P. Malathi & K. Kalpana, "A Literature Survey on Data Mining Techniques and Concepts", International Journal of Engineering Research and Modern Education, Volume 3, Issue 2, Page Number 1-3, 2018.



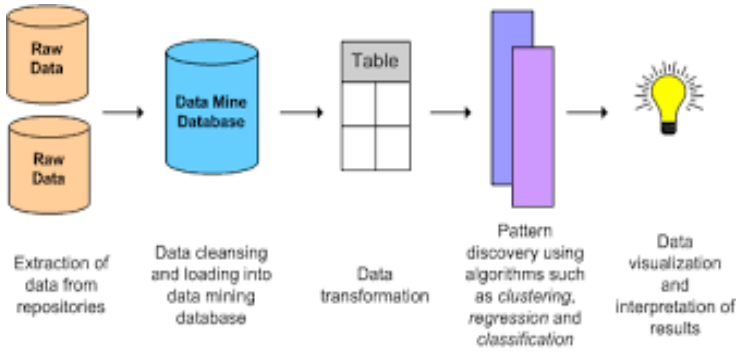
**Abstract:**

Data mining is a multidisciplinary field, drawing work from areas including database technology, machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge-based systems, artificial intelligence, high-performance computing, and data visualization. Data mining is the process of analyzing data from different views and summarizing it into useful data. "Data mining, also popularly referred to as knowledge discovery from data (KDD), is the automated or convenient extraction of patterns representing knowledge implicitly stored or captured in large databases, data warehouses, the Web, other massive information repositories or data streams.".

**Key Words:** Data mining, KDD, Classification, Sequencing & Clustering

**Introduction:**

Today, Data mining helps different organization focus on the information in the data they have collected about the behavior of their customer's. From last few years, research in data mining continues growing in various fields of organization such as Statistics, Machine Learning, Artificial Intelligence, Pattern Recognition, business, education, medical, scientific etc. In this paper, discusses the concept of data mining, important issues and applications. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information the patterns, associations, or relationships among all this data can provide information. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. As a advancement of information technology in various fields of human life has increased to the large amount of data storage in various ways like records, documents, images, sound recordings, videos, scientific data, and many new data formats.




**Process of Data Mining:**

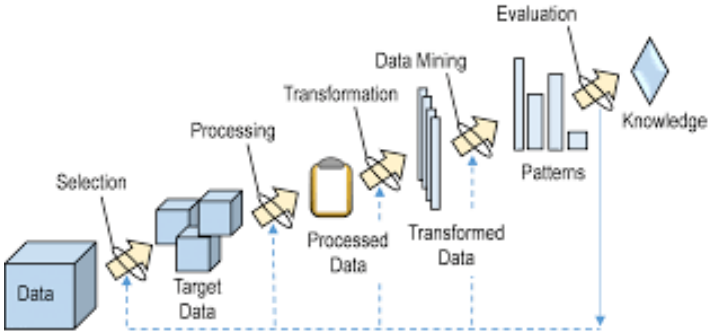
- ✓ Data preparation Data preparation generally consists of two processes: data collection and data collation. Data collections the first step of data mining, and the data can come from the existing transaction processing systems, also can be obtained from the data warehouse; data collation is to eliminate noise or inconsistent data, it is the necessary link of data mining. The data obtained from the phase of the data collection may have a certain degree of "pollution", which refers to that in the data may be its own inconsistency, or some missing data, so the collation of the data is essential. At the same time, through data collation the data can be done on a simple generalization processing, thus on the basis of the original data more rich data information will be obtained, which will facilitate the next data mining step.
- ✓ Data mining Data mining is the core stage of the entire process, it mainly uses the collected mining tools and techniques to deal with the data, thus the rules, patterns and trends will be found.
- ✓ Information expression Information expression is to use visualization and knowledge information expression technology to provide the mined knowledge information for users, is an important means to

show the data mining results. Clear and effective mining result information expression will greatly facilitate the accuracy and efficiency of the decision-making.

- ✓ Analysis and decision-making The ultimate goal of data mining is to assist the decision making. Decision-makers can analyze the results of data mining and adjust the decision-making strategies combining with the actual situation.


**Tasks of Data Mining:**

- ✓ Classification: Classification is finding models that analyze and classify a data item into several predefined classes.
- ✓ Sequencing: Sequencing is similar to the association rule. The relationship exists over a period of time such as repeat visit to supermarket.
- ✓ Regression: Regression is mapping a data item to a real-valued prediction variable.
- ✓ Clustering: Clustering is identifying a finite set of categories or clusters to describe the data.
- ✓ Dependency Modeling: Dependency Modeling (Association Rule Learning) is finding a model which describes significant dependencies between variables.
- ✓ Deviation Detection: Deviation Detection (Anomaly Detection) is discovering the most significant changes in the data.
- ✓ Summarization: Summarization is finding a compact description for a subset of data.

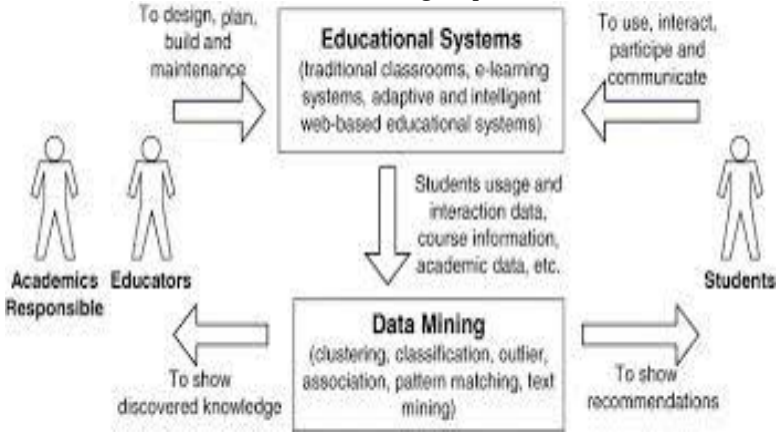



**Data Mining Applications:**

The Data mining applications are widely used in diverse areas such as retail stores, hospitals, banks, and insurance companies [2]. Many domains like health care, finance insurance, retail stores combines the data mining applications with statistics, pattern recognition, and other important tools to perform data analytics. Data mining is used primarily for decision making.

- ✓ **Medicare and Health Care:** Data mining in medicine enables to characterize patient activities to see incoming office visits. Data mining helps identify the patterns of successful medical therapies for different illnesses.
- ✓ **Education:** Educational Data Mining is a blooming field which provides knowledge from educational Environment data. The goals of EDM are identified as predicting students' learning behavior, emotions and skills [3]. This study improves the educating methods by understanding the ward and to take accurate decisions respectively.
- ✓ **Market Basket Analysis:** Market basket analysis is a technique that uses association rule mining to understand the purchasing behavior of the customer. It also allows the seller to understand his business, customer's needs and to make profitable change accordingly.
- ✓ **Financial Banking:** Data mining can contribute to solving business problems in banking and finance by finding patterns, causalities, and correlations in business information and market prices. The managers may find this information for better segmenting, targeting, acquiring, retaining and maintaining a profitable customer.

- ✓ **Research Analysis:** Data mining is very useful in data pre-processing and integration of databases. Data mining allows the researchers to identify co-occurring sequences and the correlation between any activities .Data visualization and visual data mining help the researcher with a clear view of the data.




- ✓ **Fraud Detection:** The traditional fraud detection methods are expensive, time consuming and complex. Data mining aids in providing meaningful patterns and turning data into information. Valid and useful information is called as knowledge. The results are categorized into fraudulent or non-fraudulent.
- ✓ **Transportation:** Data mining helps determine the distribution schedules among warehouses and outlets and analyze loading patterns.
- ✓ **Agriculture:** Data mining is emerging technology in agriculture field for crop yield analysis with respect to four parameters namely year, rainfall, production and area of sowing. Yield prediction is a very important agricultural problem that remains to be solved based on the available data. The yield prediction problem can be solved by employing Data Mining techniques such as K Means, K nearest neighbor (KNN), Artificial Neural Network and support vector machine
- ✓ **Cloud Computing:** Data Mining techniques are used in cloud computing. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage. Cloud computing uses the Internet services that rely on clouds of servers to handle tasks. The data mining technique in Cloud Computing helps to perform efficient, reliable and secure services for their users.

**Conclusion:**

According to the techniques of data mining listed above, it is learned that this a powerful and essential technique for performing manipulation of data that is data mining gives proper and targeted outcome from large and vastly growing data worldwide. This paper discusses the idea of data mining, the process of KDD, different techniques such as clustering, association, classification, prediction and so on. Data mining brings a lot of benefits to businesses, society, governments as well as individual. However privacy, security and misuse of information are the big problem if it is not address correctly.

**References:**

1. Berry, M.and Linoff, G., Master Data Mining: The Art and Science of Customer Relationship Management, Wiley Publisher, 2000.
2. Smita, Priti and Sharma, "Use of Data Mining in Various Field: A Survey Paper" IOSR Journal of Computer Engineering, 8727Volume 16, Issue 3, Ver. V (May-Jun. 2014)
3. Ming-Syan Chen, Jiawei Han, Philip S yu. Data Mining: An Overview from a Database Perspective [J]. IEEE Transactions on Knowledge and Data Engineering, l996, 8(6):866-883.
4. Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management System`s) 3rd Edition
5. Lefait, G. and Kechadi, T, (2010) "Customer Segmentation Architecture Based on Clustering Techniques" Digital Society, ICDS'10, Fourth International Conference, 10-02-2010.
6. Marcel Brun, Chao Sima, Jianping Hua, James Lowey, Brent Carroll, Edward Suh, Edward R. Dougherty, (2007) Model-based evaluation of clustering validation measures, Pattern Recognition, v.40 n.3, March p.807-824.