

Study of Data Mining and Understanding its Types, Methods and Techniques

Monika Soni

President, Auricle Global Society of Education and Research,
Bikaner, Rajasthan (India).
president@agser.org

Abstract: - Data mining is the process of identification of the patterns and trends in large and complex data volumes. The normal data processing will not be helpful in handling such complex data. Hence data mining is used to handle such large volumes of the data. This paper will explain the steps of data mining process, types of data mining. It will also explain the importance and advantages of the data mining and challenges involved in the implementation of the data mining.

Keywords: Data Mining, Steps, Types of data mining, Advantages, Challenges.

Introduction: - In an organisation, there is a lot of data which is recorded in databases. Some companies will have very large volumes of data to be stored on a daily basis. In such scenarios, in order to store complex data complex and large data base is needed. To access information and data from these types of databases, we need techniques which makes the task easy and makes easy to access data on time with less efforts. To support such process of handling the data, the technique of data mining comes in the picture. Hence, the process of handling large amount of data and complex data from complex data basis is called as Data Mining process. Data mining helps to access information easily from complex databases. The time and efforts reduce to large extent by using data mining technique to handle large volume of data. [1]

Data Mining Process: - [2]

There is a certain process of performing data mining in order to access the data from large complex databases. The data analyst should perform following steps in order to fetch data from the databases. Each stage is important and should be followed. There is step by step procedure of performing data mining process. Following are the key stages of the data mining process.

1. **Understanding Goals:** - In this first step, the data miners will try to understand the type of business data and information which is exchanged daily in and out of the organisation> He will prepare the document and write

down each and every single detail that from where the data is coming and going and also the data communication of the business. He will try to understand the business goals and requirements.

2. **Data Identification:** - In this stage the miner will try to understand the information gathered in the first stage. He will go thorough all the details collected and try to sort the data and information according to many factors like complexity of the data, importance of data, its usage in the business etc.



Figure 1 Data Mining Process.

3. **Prepare data:** - In this step after understanding the data type the data miner will arrange the data. The data can be

arranged using rows, tables, columns. Basically, the data miner will arrange the data according to its type, complexity, importance etc so that it can be easily understood and accessed. For arranging the data, the data miner can also use different models available which makes his task easy.

4. Evaluation of the data: - In this stage the data miner will evaluate the data gathered and arranged. He will test whether it is according to the business needs and specifications.
5. Fix issues: - In this stage it is tested and fixed if there is any missing data and information. If there is any modification also, then that also can be done in this stage other than fixing the issue.
6. Present Data: - In this stage, finally the data miner will arrange the data and the information in such a way that is useful for all those who will access this information. It should be such that it is easily accessible by all.

Importance of Data Mining: - [3]

- Data mining is the process of analysing data, identification of data, process of handling data from large and complex databases. It is very important concept as it makes the data accessing easy and helps the business to achieve goals in less time.
- It helps to reduce the overall cost of the data analysis.
- It also helps to automate the data queries which makes the data processing efficient.
- It helps to attract more clients as the response to query of customers is faster.
- It can be used in following applications: -
 - . Risk Analysis
 - . Filtration of spam
 - . Identification of security attacks.
 - . Marketing

Data Mining Techniques: - [4]

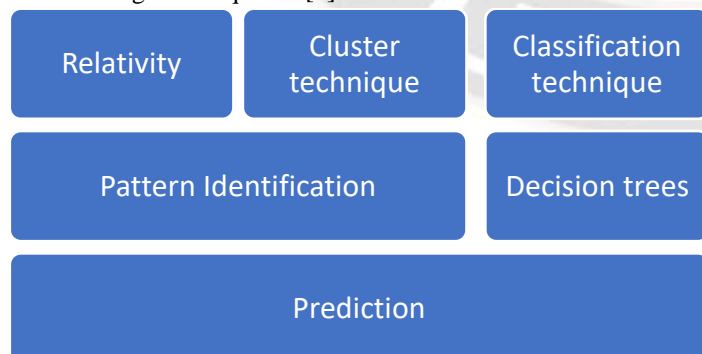


Figure 2 Methods of Data Mining.

- Relativity: - In this type of data mining it is found that how the data is interrelated to each other and the pattern of relation of data is identified. This type of

technique can be used in business to study the pattern with which the users buy or sell something.

Challenges of Relativity Technique: - It has complex algorithm to identify the relation between various data and hence it is difficult for those who are not experts in the data mining techniques.

- Cluster Technique: - In this type of technique the data is placed in different clusters where all type of data will have same characteristics. It makes it easy for the analyst to identify the data and the performance will also increase. It is like arranging similar type of grains at one place in a grocery store. In grocery store same type of oils, grains, dairy etc are arranged in specific corners and hence the user can directly go to the corner where he wants specific items. Same way the data which has same specifications will be arranged in clusters.

Challenges of Cluster Technique: - The main challenge of this technique is that if the particular cluster is corrupted then it will be very difficult, almost impossible to access data from such clusters. So, this technique is not reliable in case of corrupted cluster.

- Classification Technique: - This technique is used to classify the data based on its type. Like the different variables or objects can be classified in different categories of classes which makes it easy to access. For example, in a school the same colour toys can be kept in same cupboard and other colour toys in other and so on.

Challenges of Classification data mining technique: It is easy technique to implement but it requires more storage space. It has performance issues if the class in which the data is defined is not accessible.

- Pattern Identification: - This type of technique is used to identify the patterns which occurs over a certain period of time. For example, it is seen that the purchase of umbrella increases during rainy season so it is a pattern. Similarly, in this technique the data which is accessed again and again is identified and the pattern is noted and then it is automated for further use.

Challenges of Pattern identification technique: - The pattern identification can sometimes become difficult to identify in large and complex databases and therefore it will be time consuming to identify the sequential patterns in such scenarios. This type

pf technique is feasible in small database which is not complex.

5. Prediction: - This technique uses the prediction to predict the communication and relation between dependant components and independent components. Based on the previous study the prediction can be made.

Challenges of Prediction Data Mining: -

If the study of the previous information and pattern is not done properly and if there is any error in doing research then the prediction of future data can not be done properly. It will definitely give bugs and errors. Hence, this technique is not so feasible and reliable.

Application use of Data Mining: - [5]

1. Communication Industry: - Since these days everything is happening over internet for all types of communication. In this field Data mining can be applied to increase the efficiency of the application.
2. Finance Department: - In the field of finance like banks etc the type of data is very crucial and important and should be kept very safe and secretly. Here in this field the data mining can be used to arrange the data easily and in appropriate manner which can be accessed easily.
3. Manufacturing industries: - The data mining can be used to study the relationship between the components of the product being manufactured and how they interact with each other.
4. Health Management system: -In this domain also data mining can be used to determine the patients records, their previous health issues, prescriptions etc

List of Tools useful for Data Mining: - [6]

- Orange: - This tool has components called widgets which helps to identify, extract, analyse the data. With the help of this tool the modification and analysis of data becomes easy and faster.
- KNIME: - This tool is useful in the area of business intelligence, pharmacy etc. This helps in data analytics etc.
- Weka: - This tool helps to write SQL queries to perform analysis of the data using data mining technique.
- Rattle: - This tool uses the R language and it has the graphical user interface which makes easy for the process of data mining. It also helps in analyses of mathematical operations.
- Data Melt: -This tool is largely used by students, engineers etc for their own use. The data can be analysed and also be seen.

Conclusion: - It is clear that in large and big organizations, it is necessary to have data mining concept as it helps to access the data from complex database easily and quickly. There are many techniques which a data miner can select based on the business needs. The advantage of implementing data mining in business is that it reduces the time and effort to gather the needed data and hence helps in achieving customer satisfaction.

References: -

1. <https://www.javatpoint.com/data-mining>
2. <https://zipreporting.com/en/data-mining/data-mining-process.html>
3. <https://www.quora.com/What-is-data-mining-and-why-is-it-so-important>
4. <https://www.upgrad.com/blog/data-mining-techniques/>
5. <https://www.upgrad.com/blog/12-most-useful-data-mining-applications-of->
6. <https://www.educba.com/data-mining-tool>