Application, Issues and Challenges

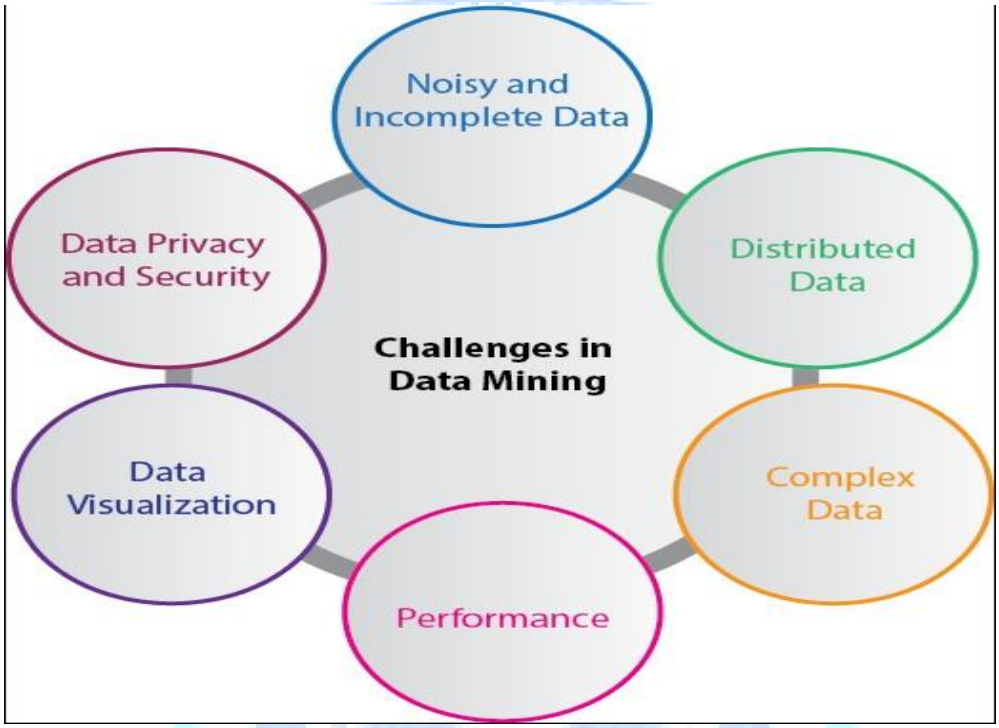
Applications of Data Mining

Data Mining is primarily used by organizations with intense consumer demands- Retail, Communication, Financial, marketing company, determine price, consumer preferences, product positioning, and impact on sales, customer satisfaction, and corporate profits. Data mining enables a retailer to use point-of-sale records of customer purchases to develop products and promotions that help the organization to attract the customer.

Applications	Usage
Communications	Data mining techniques are used in the communication sector to predict customer behavior to offer highly targeted and relevant campaigns.
Insurance	Data mining helps insurance companies to price their products profitable and promote new offers to their new or existing customers.
Education	Data mining benefits educators to access student data, predict achievement levels and find students or groups of students who need extra attention. For example, students who are weak in a science subject.
Manufacturing	By using the help of Data Mining Manufacturers can predict wear and tear of production assets. They can anticipate maintenance which helps them reduce them to minimize downtime.
Banking	**Data mining helps the finance sector to get a view of market risks and manage regulatory compliance. It helps banks to identify probable defaulters to decide whether to issue credit cards, loans, etc.**
Retail	Data Mining techniques help retail malls and grocery stores identify and arrange most sellable items in the most attentive positions. It helps store owners to come up with the offer which encourages customers to increase their spending.
Service Providers	Service providers like mobile phone and utility industries use Data Mining to predict the reasons when a customer leaves their company. They analyze billing details, customer service interactions, complaints made to the company to assign each customer a probability score and offer incentives.
E-Commerce	E-commerce websites use Data Mining to offer cross-sells and up-sells through their websites. One of the most famous names is Amazon, which uses Data mining techniques to get more customers into their eCommerce store.

Challenges of Implementation in Data mining

Although data mining is very powerful, it faces many challenges during its execution. Various challenges could be related to performance, data, methods, and techniques, etc. The process of data mining becomes effective when the challenges or problems are correctly recognized and adequately resolved.



Incomplete and noisy data:

- The process of extracting useful data from large volumes of data is data mining. The data in the real-world is heterogeneous, incomplete, and noisy. Data in huge quantities will usually be inaccurate or unreliable. These problems may occur due to data measuring instrument or because of human errors.
- Suppose a retail chain collects phone numbers of customers who spend more than $ 500, and the accounting employees put the information into their system. The person may make a digit mistake when entering the phone number, which results in incorrect

data. Even some customers may not be willing to disclose their phone numbers, which results in incomplete data. The data could get changed due to human or system error. All these consequences (noisy and incomplete data)makes data mining challenging.

Data Distribution:

- Real-worlds data is usually stored on various platforms in a distributed computing environment. It might be in a database, individual systems, or even on the internet. Practically, It is a quite tough task to make all the data to a centralized data repository mainly due to organizational and technical concerns.
- For example, various regional offices may have their servers to store their data. It is not feasible to store, all the data from all the offices on a central server. Therefore, data mining requires the development of tools and algorithms that allow the mining of distributed data.

Complex Data:

Real-world data is heterogeneous, and it could be multimedia data, including audio and video, images, complex data, spatial data, time series, and so on. Managing these various types of data and extracting useful information is a tough task. Most of the time, new technologies, new tools, and methodologies would have to be refined to obtain specific information.

Performance:

The data mining system's performance relies primarily on the efficiency of algorithms and techniques used. If the designed algorithm and techniques are not up to the mark, then the efficiency of the data mining process will be affected adversely.

Data Privacy and Security:

- Data mining usually leads to serious issues in terms of data security, governance, and privacy.
- For example, if a retailer analyzes the details of the purchased items, then it reveals data about buying habits and preferences of the customers without their permission.

Data Visualization:

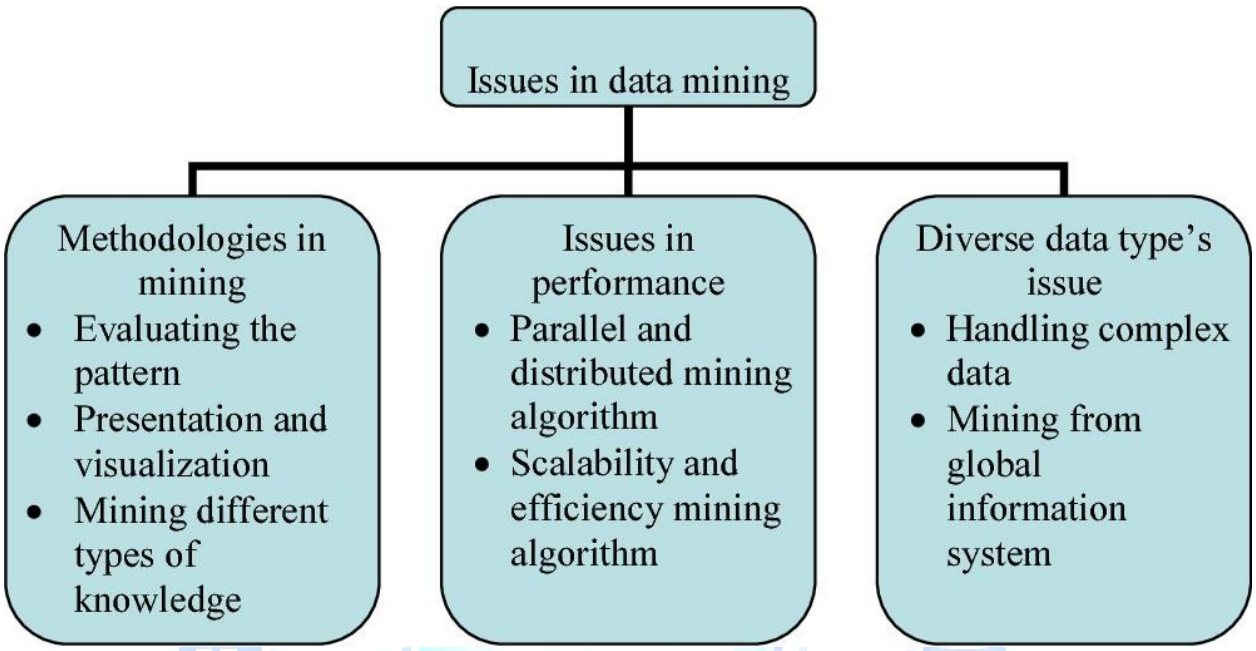
- In data mining, data visualization is a very important process because it is the primary method that shows the output to the user in a presentable way. The extracted data should convey the exact meaning of what it intends to express. But many times, representing the information to the end-user in a precise and easy way is difficult The input data and the output information being complicated, very efficient, and successful data visualization processes need to be implemented to make it successful.
- There are many more challenges in data mining in addition to the problems above-mentioned. More problems are disclosed as the actual data mining process begins, and the success of data mining relies on getting rid of all these difficulties.

Major Issues in Data Mining

- Major Issues in Data Mining: Data Mining is not very simple to understand and implement. As it is already evident that Data Mining is a process which is very crucial for various researchers and businesses.
- But in data mining, the algorithms are very complex and on top of that, the data is not readily available at one place. Every technology has flaws or issues. But one needs to always know the various flaws or issues that technology has.

Major Issues in Data Mining

- User interaction
- Data mining query languages and ad-hoc mining
- Expression and visualization of data mining results n
- Interactive mining of knowledge at multiple levels of abstraction
- Applications and social impacts
- Domain-specific data mining & invisible data mining
- Protection of data security, integrity, and privacy



Mining Methodology and User Interaction Issues:

- Mining different kinds of knowledge in databases
  - Interactive mining of knowledge at multiple levels of abstraction
  - Incorporation of background of knowledge
  - Data mining query languages and ad hoc data mining
  - Presentation and visualization of data mining results
  - Handling noisy or incomplete data
- Performance Issues
  - Efficiency and Scalability of data mining algorithm
  - Parallel, distributed and incremental mining algorithm
- Diverse Data Types Issues
  - Handling of relational and complex types of data
  - Mining information from heterogeneous databases and global information systems

Mining different kinds of knowledge in databases

This issue is responsible for addressing the problems of covering a big range of data in order to meet the needs of the client or the customer. Due to the different information or a different way,

it becomes difficult for a user to cover a big range of knowledge discovery task.

Interactive mining of knowledge at multiple levels of abstraction

Interactive mining is very crucial because it permits the user to focus the search for patterns, providing and refining data mining requests based on the results that were returned. In simpler words, it allows user to focus the search on patterns from various different angles.

Incorporation of background of knowledge

The main work of background knowledge is to continue the process of discovery and indicate the patterns or trends that were seen in the process. Background knowledge can also be used to express the patterns or trends observed in brief and precise terms. It can also be represented at different levels of abstraction.

Data mining query languages and ad hoc data mining

Data Mining Query language is responsible for giving access to the user such that it describes ad hoc mining tasks as well and it needs to be integrated with a data warehouse query language.

Presentation and visualization of data mining results

In this issue, the patterns or trends that are discovered are to be rendered in high level languages and visual representations. The representation has to be written so that it is simply understood by everyone.

Handling noisy or incomplete data

For this process, the data cleaning methods are used. It is a convenient way of handling the noise and the incomplete objects in data mining. Without data cleaning methods, there will be no accuracy in the discovered patterns. And then these patterns will be poor in quality.

Performance Issues

It has been noticed several times before also that there are performance related issues in data mining as well. These issues are listed as follows

Efficiency and Scalability of data mining algorithm

Efficiency and Scalability is very important when it comes to data mining process. It is also very necessary because with the help of using this, the user can withdraw the information from the data in a more effective and productive manner.

Parallel, distributed and incremental mining algorithm

There are a lot factors which can be responsible for the development of parallel and

distributed algorithms in data mining. These factors are large in size of database, huge distribution of data, and data mining method that are complex. In this process, the first and foremost step, the algorithm divides the data from database into various partition. In the next step, that data is processed such that it is situated in parallel manner. Then the last step, the result from the partition is merged.

Diverse Data Types Issues

The issues in this type of issue are given below:

Handling of relational and complex types of data

The database may contain the various data objects for example, complex, multimedia, temporal data, or spatial data objects. It is very difficult to mine all these data with the help of a single system.

Mining information from heterogeneous databases and global information systems:

The problem in this kind of issue is to mine the knowledge from various data sources. These data are not available as a single source instead these data are available at the different data sources on LAN or WAN. The structures of these data are different as well.