

Challenges of Data Mining

Nowadays Data Mining and knowledge discovery are evolving a crucial technology for business and researchers in many domains. Data Mining is developing into established and trusted discipline, many still pending challenges have to be solved.

Some of these challenges are given below.

Security and Social Challenges:

Decision-Making strategies are done through data collection-sharing, so it requires considerable security. Private information about individuals and sensitive information are collected for customers profiles, user behaviour pattern understanding. Illegal access to information and the confidential nature of information becoming an important issue.

User Interface:

The knowledge discovered is discovered using data mining tools is useful only if it is interesting and above all understandable by the user. From good visualization interpretation of data, mining results can be eased and helps better understand their requirements. To obtain good visualization many research is carried out for big data sets that display and manipulate mined knowledge.

(i) Mining based on Level of Abstraction: Data Mining process needs to be collaborative because it allows users to concentrate on pattern finding, presenting and optimizing requests for data mining based on returned results.

(ii) Integration of Background Knowledge: Previous information may be used to express discovered patterns to direct the exploration processes and to express discovered patterns.

Mining Methodology Challenges:

These challenges are related to data mining approaches and their limitations.

Mining approaches that cause the problem are:

- (i)** Versatility of the mining approaches,
- (ii)** Diversity of data available,
- (iii)** Dimensionality of the domain,
- (iv)** Control and handling of noise in data, etc.

Different approaches may implement differently based upon data consideration.

Some algorithms require noise-free data. Most data sets contain exceptions, invalid or incomplete information lead to complication in the analysis process and some cases compromise the precision of the results.

Complex Data:

Real-world data is heterogeneous and it could be multimedia data containing images, audio and video, complex data, temporal data, spatial data, time series, natural language text etc. It is difficult to handle these various kinds of data and extract the required information. New tools and methodologies are developing to extract relevant information.

(i) Complex data types: The database can include complex data elements, objects with graphical data, spatial data, and temporal data. Mining all these kinds of data is not practical to be done one device.

(ii) Mining from Varied Sources: The data is gathered from different sources on Network. The data source may be of different kinds depending on how they are stored such as structured, semi-structured or unstructured.

Performance:

The performance of the data mining system depends on the efficiency of algorithms and techniques are using. The algorithms and techniques designed are not up to the mark lead to affect the performance of the data mining process.

(i) Efficiency and Scalability of the Algorithms: The data mining algorithm must be efficient and scalable to extract information from huge amounts of data in the database.

(ii) Improvement of Mining Algorithms: Factors such as the enormous size of the database, the entire data flow and the difficulty of data mining approaches inspire the creation of parallel & distributed data mining algorithms.