**Unit II Data Warehousing and Data Mining**

Data Mining:- Definition, Characteristic, Benefits, Date Mining Functions, Data Mining Applications, Data Mining techniques and tools. Text Mining, Web Mining.

**PART-B**

1. **Data Mining: Definition:**

Data Mining is a process of extracting knowledge from massive volumes of data. It refers to a way of finding significant and useful information from an organization’s database. The knowledge which is extracted can include: -Pattern types, Association rules and different trends.

Data Mining is not confined to a particular organization, instead it has techniques to explore the knowledge hidden in any data. The different techniques used for digging out data are Artificial Intelligence, Statistical and Mathematical Techniques and Pattern Recognition Techniques.

1. **Characteristic of Data Mining:**

**The following are the Characteristics of Data Mining:**

1. Data Mining cleanse the data to be stored in the warehouse.
2. It works on a simple client/server or web-based information system architecture.
3. It includes modern tools such as visualization n synchronization tools which eliminate the unnecessary information included in the corporate files or records.
4. Data mining user is usually the end user provided with certain additional powerful tools that help in retrieving the data quickly.
5. Data from data mining tools can be easily integrated with spreadsheets and other softwares, so that they can be accessed easily.
6. It also offers parallel processing.
7. **Benefits of Data Mining:**

Data mining helps in analyzing and summarizing different elements of information. Mining process is a form where in which all the data and information can be extracted for the purpose of future benefit.

**Benefits or Advantages of Data Mining Techniques:**

There are several types of benefits and advantages of data mining systems. One of the essential matters of these mining creates a complete structure of analysis of mining techniques.

**1. It is helpful to predict future trends:**

Most of the working nature of the data mining systems carries on all the informational factors of the elements and their structure.

One of the common benefits that can be derived with these data mining systems is that they can be helpful while predicting future trends, and that is quite possible with the help of technology and behavioral changes adopted by the people.

**2. It signifies customer habits:**

For example, while working in the marketing industry one can understand all the matters of customer behaviour and their habits, and this is possible with the help of data mining systems.

As these data mining systems handle all the information acquiring techniques. It is helpful in keeping the track of customer habits and their behavior.

**3. Helps in decision making:**

There are some people who make use of these data mining techniques to help them with some kind of decision making.

Nowadays, all the information about anything can be determined easily with the help of technology and similarly, with the help of such technology one can make a precise decision about something unknown and unexpected.

**4. Increase company revenue:**

As it has been explained earlier that data mining is a process where in which it involves some sort of technology to acquire some information about anything possible, and this type of technology makes things easier for their profit earning ratio.

As people can collect information about the marketed products online, which eventually reduces the cost of the product and their services.

**5. It depends upon market-based analysis:**

Data mining process is a system where in which all the information has been gathered on the basis of market information.

Nowadays, technology plays a crucial role in everything and that casualties can be seen in these data mining systems. Therefore, all the information collected through these data mining is basically from marketing analysis.

**6. Quick fraud detection:**

Most parts of the data mining process is basically from information gathered with the help of marketing analysis. With the help of such marketing analysis one can also find out those fraudulent acts and products available in the market.

Moreover, with the help of it one can understand the importance of accurate information.

Therefore, to sum up the benefits of Data Mining: -

1. It helps to identify the shopping patterns
2. It helps in making effective marketing campaigns
3. It helps in Determining customer groups
4. It helps to measure profitability factors
5. It is helpful to predict future trends: ...
6. It signifies customer habits & pattern ...
7. It helps in decision making…
8. It helps in increasing the company’s revenue...
9. It depends upon market-based analysis..
10. It helps in Quick fraud detection
11. **Data Mining Functions:**

The functions of Data Mining are classified into the following: -

1. **Prediction:**

The term prediction refers to the action of predicting future. It involves various factors such as suggestions, expert advices, related information etc.

1. **Association:** Associations are used to identify the variables present in the databases which are related based on certain condition. The technique of using association is also called as association rule learning. In this field, bar code machines and point of scale systems are most commonly used for gathering data that helps in generating knowledge.

However in retail field, it is called as market-based analysis.

* + - Ex: Barcode machines.. 
1. **Clustering:** clustering is a technique of combining a group of physical objects into classes of homogeneous objects. Clustering of data objects into single class is equivalent to data compression.

Different algorithms generate different clusters because these algorithms might consider different factors while generating the clusters.

1. **Data Mining Applications:**

The Major Application areas for Data Mining are as follows:-

1. Customer Relationship Management: Data mining can be used to manage data such as sales, services, enquiries etc., which are related to CRM.

Data mining helps in performing the following operations in CRM: -

* It determines the customers who interestingly participates in buying the goods/products.
* It helps in gaining customer retention by determining the causes of customer attrition.
1. **Banking:**

Data mining can be used in the field of banking in the following ways: -

* It helps in making the loan application process automatic
* It can capture the unauthorized transactions done over credit card or some other means of payment
* It helps to forecast the flow of cash in various banking systems such as ATMs, banks etc.
1. **Retail Industry:**

Data mining can be used in the field of retailing in the following ways:-

* It can predict almost accurate inventory for a specific location by predicting the sales.
* It can determine the degree of consumption for various types of products.
* It can use market-based analysis to determine the products relationship thereby optimizing the sales.
1. **Manufacturing:**

Data mining can be used in the field of manufacturing in the following ways:-

* It can improve quality of goods by identifying novel patterns.
* It can make use of sensory data for predicting the errors in the machines.
* It can optimize the capabilities of manufacturing by determining the commonalities and anomalies.
1. **Brokerage and Trading:**

Data mining can be used in the field of brokerage and trading in the following ways:

* It can used for predicting the time and reason behind the change in share prices.
* It can predict the impact of various issues on market
* It can determine fraudulent activities
1. **Computers:**

Data mining can be used in the field of computers in the following ways:

* It can determine the failure of disks and other devices
* It can capture and remove unnecessary e-mails
* It can determine security needs and unsecure softwares
1. **Military:**

Data mining can be used in the field of Military in the following ways:

* It can predict the cost required to move military equipments from one place to another
* It can predict the consumption of resources
* It can generate knowledge from data related to strategies and past experiences.
1. **Travel Industry:**

Data mining can be used in the field of travel industry in the following ways: -

* It can predict the preferences of user with respect to different services such as types of seats, types of cars etc. to increase the revenue
* It can also predict the demand with respect to various locations.
* It can determine potentially valuable customers.
1. **Data Mining techniques and tools:**

**Data Mining Techniques:**

The following are the Data Mining Techniques:-

* **Classification**
* **Clustering**
* **Regression**
* **Association Rules**
1. **Classification:**

Classification is the process of predicting the class of a new item. Therefore to classify the new item and identify to which class it belongs.

1. **Clustering:**

**Group Data into Clusters**

* Similar data is grouped in the same cluster
* Dissimilar data is grouped in the same cluster
1. **Regression:**
* “Regression deals with the prediction of a value, rather than a class.”
* Regression is a data mining function that predicts a number
* For example, a regression model could be used to predict children's height, given their age, weight, and other factors.
1. **Association Rules:**

“An association algorithm creates rules that describe how often events have occurred together.”

**Example:** When a customer buys a Computer, then 90% of the time they will buy softwares.

**DATA MINING TOOLS:**

Some of the Software tools of data mining are: -

1. **IBM Intelligent Miner:**

IBM intelligent miner offers many data mining functions such as classification, association mining, predictive modeling, clustering, deviation detection, sequential pattern analysis and regression.

1. **SPSS Clementine:**

SPSS Clementine offers an integrated development environment for developers and end users to perform data mining.

It performs various functions including prediction, clustering, association mining and classification.

1. **Oracle Data Mining (ODM):**

Oracle Data Mining tool is an add on feature of Oracle 10 Enterprise Edition Database. This tool performs various data mining functions such as classification, prediction, sequence similarity search and analysis, association mining and clustering.

1. **Classification and Regression Tree (CART):**

This tool generates regression trees for prediction and decision trees for classification. It improves the accuracy of data mining by employing boosting mechanism.

1. **Weka:**

Weka is an open-source data mining tool developed on Java platform in the university of Waikato, Newzealand.

It is capable to perform various data mining functions such as regression, classification, association mining and data processing.

1. **TEXT MINING:**

Text mining is the analysis of data contained in natural language text. The application of text mining techniques to solve business problems is called text analytics.

The main purpose of this mining is to process unstructured information and to extract meaningful numeric indices from the database, so as to make the information accessible to different data mining algorithms.

Text mining is an important part of data mining process because such mining enables the user to make comparison among several documents, provide priority to essential documents or identify the procedure of several documents.

The functioning is identical to data mining that considers the input in the text format only. The files can be MS-Word documents, PDFs, XML files etc. it is considered as effective in the fields where text data is generated in a larger amount.

**Examples** of such fields are: -

* Court orders,
* Academics,
* Medicine,
* Marketing etc.
1. **WEB MINING:**

Web mining is the application of data mining techniques to discover patterns from the World Wide Web (www).

**Types of Web Mining:**

Web mining can be divided into three different types: –

1. **Web usage mining:**

The Web usage mining refers to the process of searching the weblog records in order to determine the procedures in which users can access the Webpages.

1. **Web content mining:**

Web content mining is a process of extracting relevant information from web contents. Basically web content comprises not only textual information but also graphical information, real-time information and hyperlinks.

1. **Web structure mining:**

Web structure mining is the process of generating summary of webpages of webpages of websites.

Web structuring checks link structure. i.e. hyperlinks among various websites and classifies the webpage on the basis of hyperlinks found.

**END OF PART-B**