

QUESTIONS AND ANSWERS ON THE CONCEPT OF DATA MINING

Q1- What is Data Mining?

Ans- Data mining can be termed or viewed as a result of natural evolution of information technology. So data mining refers to extracting or mining knowledge from large amount of data. It is a computational procedure of finding patterns in the bulk of data and extracting the required useful data which is further transformed in understandable information. In data mining, the sets of useful data are extracted involving methods like database systems, machine learning, statistics and artificial intelligence. Various data mining tools are used for the process of data mining. Please have a look at this PPT on Data mining for better understanding,

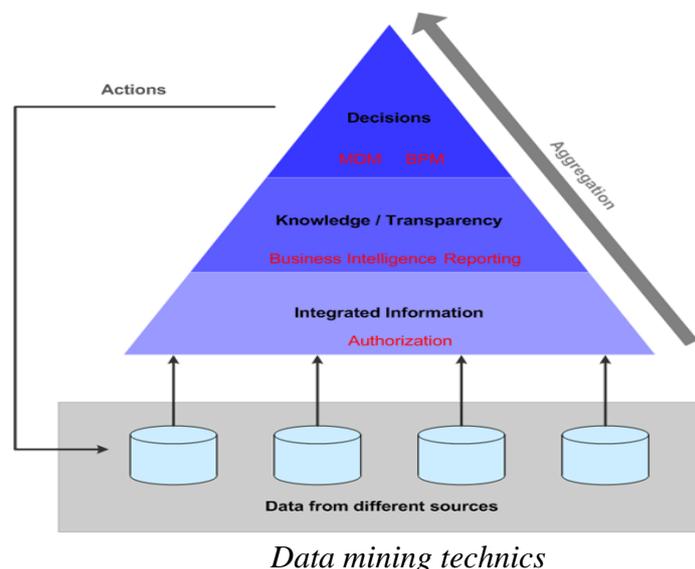
Q2- What are the different functions of data mining?

Ans- Following are the different functions of data mining-

- Characterization
- Association and correlation analysis
- Classification
- Prediction
- Clustering analysis
- Evolution analysis

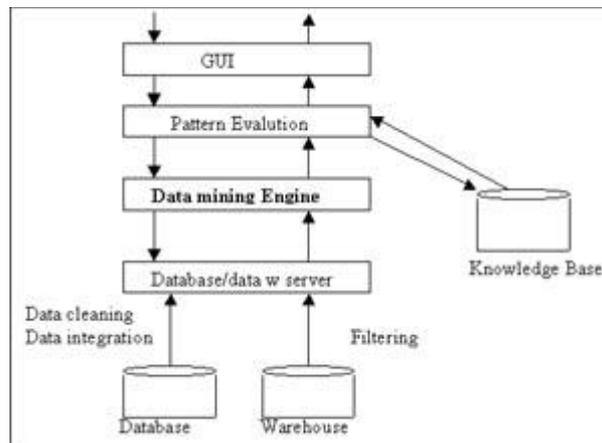
Q3- What is Data Aggregation and Generalization?

Ans- Aggregation is the technique wherein , as the name suggest , summary or aggregation operation are applied to the data for example , we can construct a data cube for analysis of the data at multiple granularities .Generalizing the data basically deals with the concept wherein low level or primitive data is replaced by high level concept.



Q4- What do you mean by Analytical characterization?

Ans- Data mining has proved to be an effective means of integrating the analysing the diverse collection of variables. If the mined concept descriptions involve many attribute, an analytical characterization should be performed.



Association Data Mining Rule

Q5- Explain the association mining rule in brief.

Ans- Mining association rule provides association or correlation among a large set of item. The association rule provides a great help in decision making process.

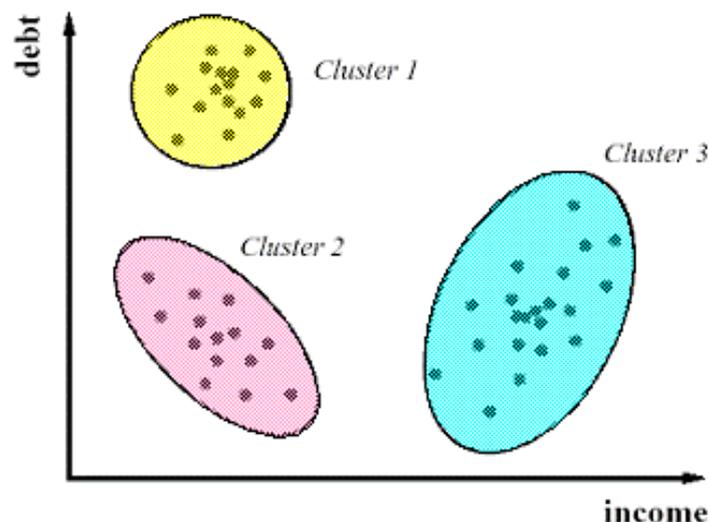
Q6- Explain the following terms in brief-

Learning

Classification

Ans- In Learning, a model is used to analyze a training data set is composed of training data samples that are randomly selected from a sample population.

In Classification, the model is used for estimation using class-labelled, randomly selected test samples. Classification is usually associated with finding a known data class for the given unknown data, which is analogous to labelling the unlabelled data. *"Classification is the process that finds the common properties along a set of objects in a data set and classifies them into different classes according to classification model."*



Cluster Analysis in Data Mining

Q7- What do you mean by Cluster Analysis?

Ans- Cluster analysis is an important human activity. Cluster analysis has been widely used in numerous applications, including market research, pattern recognition, data analysis and image processing.

Q8- What is OLAP?

Ans- OLAP (Online Analytical Processing): In the multidimensional model, data are organized into multiple dimensions and each dimension contains multiple levels of abstraction defined by concept hierarchies. OLAP provides a user friendly environment for interactive data analysis.

Q9- What do you mean by Tuning Data Warehouse explain?

Ans- Tuning a data warehouse is more difficult than tuning an OLTP environment because of the ad hoc and unpredictable nature of the load. A data warehouse, of its very nature, evolves over time. As a result, many aspects of the warehouse environment will change as the profile and the usage of the data change.

Q10- What is "Data Mining Interface"?

Ans- For better feedback to user during the construction of a query data mining interface is used in GUI form. The GUI of data mining query improves the quality of the query.

