| Q.1	**What is Data mining?**
	Data mining is the process of identifying valid, novel, potentially useful and ultimately comprehensible information from databases that is used to make crucial business decisions.
Q.2	Explain use of data mining for Market Analysis and Management.
	Listed below are the various fields of market where data mining is used: • **Customer Profiling** - Data mining helps determine what kind of people buy what kind of products. • **Identifying Customer Requirements** - Data mining helps in identifying the best products for different customers. It uses prediction to find the factors that may attract new customers. • **Cross Market Analysis** - Data mining performs Association/correlations between product sales. • **Target Marketing** - Data mining helps to find clusters of model customers who share the same characteristics such as interests, spending habits, income, etc. • **Determining Customer purchasing pattern** - Data mining helps in determining customer purchasing pattern. • **Providing Summary Information** - Data mining provides us various multidimensional summary reports.
Q.3	What is Knowledge Discovery? (KDD)
	Some people don't differentiate data mining from knowledge discovery while others view data mining as an essential step in the process of knowledge discovery. Here is the list of steps involved in the knowledge discovery process: • **Data Cleaning** - In this step, the noise and inconsistent data is removed. • **Data Integration** - In this step, multiple data sources are combined. • **Data Selection** - In this step, data relevant to the analysis task are retrieved from the database. • **Data Transformation** - In this step, data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations. • **Data Mining** - In this step, intelligent methods are applied in order to extract data patterns. • **Pattern Evaluation** - In this step, data patterns are evaluated. • **Knowledge Presentation** - In this step, knowledge is represented. The following diagram shows the process of knowledge discovery:
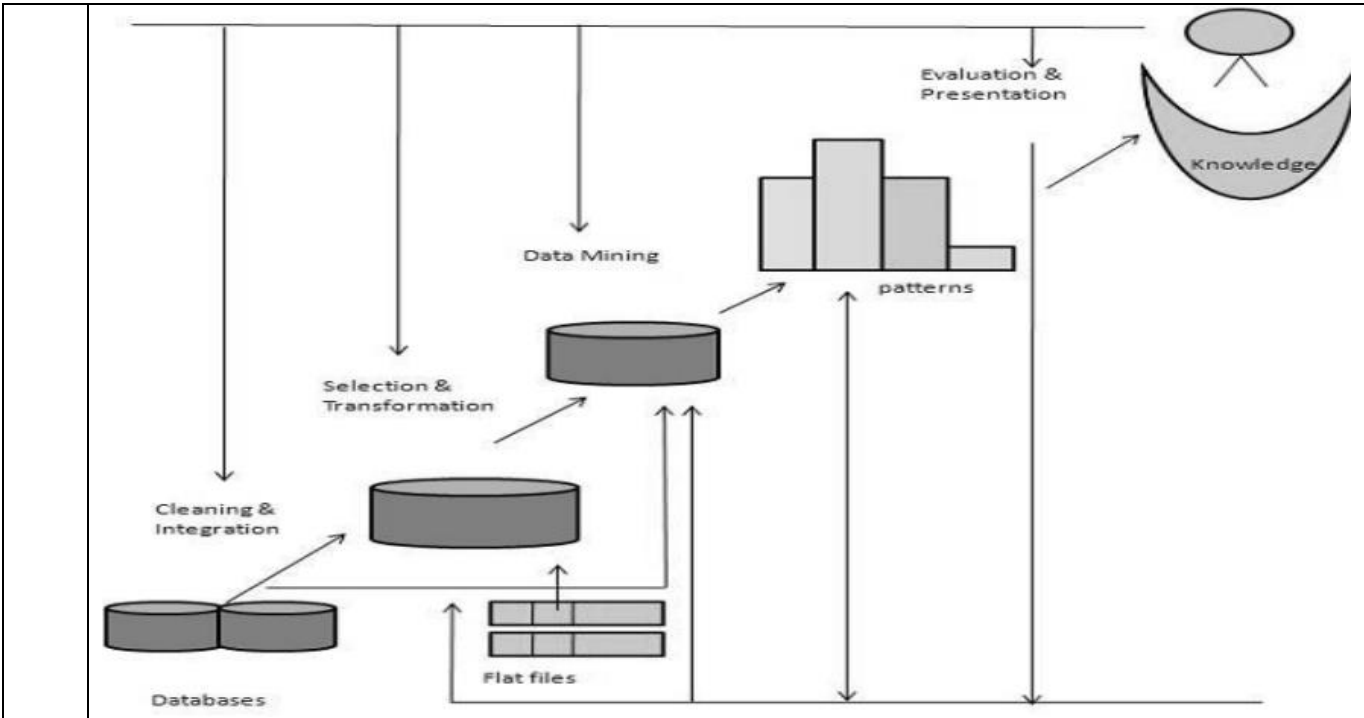
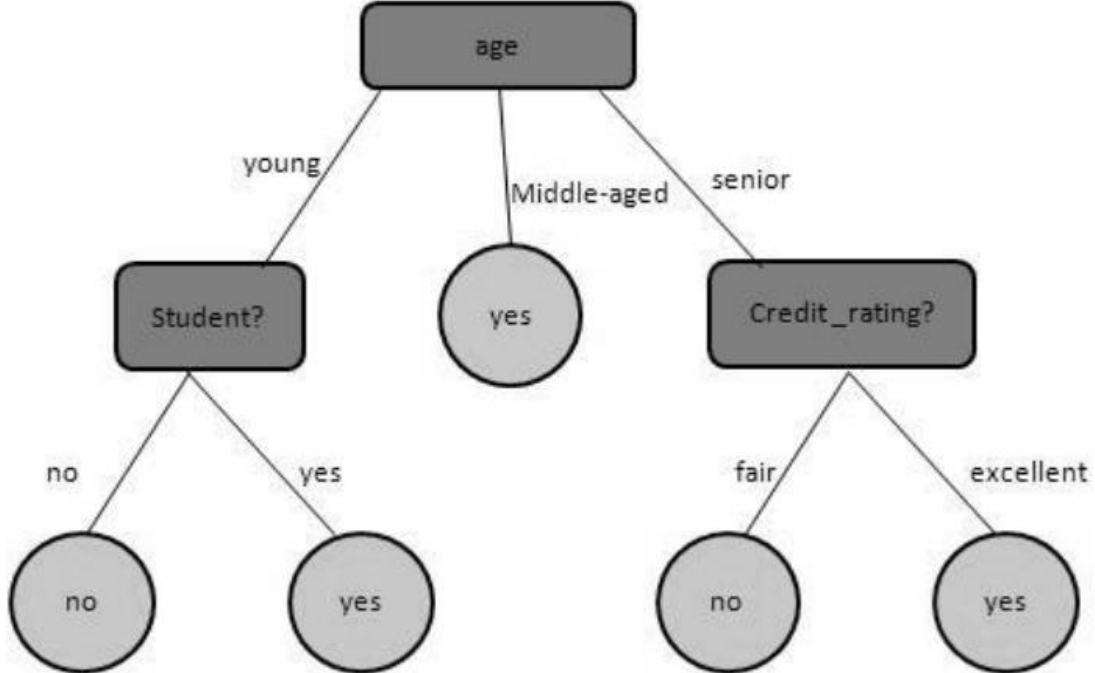| Q.4	What is Classification?
	Following are the examples of cases where the data analysis task is Classification:
	- A bank loan officer wants to analyze the data in order to know which customer (loan applicant) are risky or which are safe. - A marketing manager at a company needs to analyze a customer with a given profile, who will buy a new computer. In both of the above examples, a model or classifier is constructed to predict the categorical labels. These labels are risky or safe for loan application data and yes or no for marketing data.
Q.5	What is Prediction?
	Following are the examples of cases where the data analysis task is Prediction:
	Suppose the marketing manager needs to predict how much a given customer will spend during a sale at his company. In this example we are bothered to predict a numeric value. Therefore the data analysis task is an example of numeric prediction. In this case, a model or a predictor will be constructed that predicts a continuous-valued-function or ordered value.
	Note: Regression analysis is a statistical methodology that is most often used for numeric prediction.
Q.6	Comparison of Classification and Prediction Methods
	Accuracy - Accuracy of classifier refers to the ability of classifier. It predict the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data. **Speed** - This refers to the computational cost in generating and using the classifier or predictor. **Robustness** - It refers to the ability of classifier or predictor to make correct predictions from given noisy data. **Scalability** - Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data. **Interpretability** - It refers to what extent the classifier or predictor understands.

Q.7	What is Decision Tree?
	A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. The following decision tree is for the concept buy_computer that indicates whether a customer at a company is likely to buy a computer or not. Each internal node represents a test on an attribute. Each leaf node represents a class.
Q.8	What are benefits of Decision Tree?
	The benefits of having a decision tree are as follows: It does not require any domain knowledge.• It is easy to comprehend.• The learning and classification steps of a decision tree are simple and• fast.
Q.9	What are different Data Mining Tasks?
	Data mining involves six common classes of tasks: Anomaly detection (Outlier/change/deviation detection) – The identification of unusual data records, that might be interesting or data errors that require further investigation. Association rule learning (Dependency modelling) – Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis. Clustering – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data. Classification – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam". Regression – attempts to find a function which models the data with the least error Summarization – providing a more compact representation of the data set, including visualization and report generation.
Q.10	Architecture of Data Mining

User Interface

Pattern Evaluation

Knowledge Base

Data Mining Engine

Database or
Data Warehouse Server

data cleaning, integration and selection

Database

Data Warehouse

World Wide Web

Other Info Repositories