1. Regression testing

* Regression testing is defined as a type of software testing to confirm that a recent program or code change has not adversely affected existing features.
* regression testing is nothing but a full or partial selection of already executed test cases which are re-executed to ensure existing functionalities work fine.

1. Classification

* Classification constructs the classification model by using training data set.
* Classification predicts the value of classifying attribute or class label.
* For example: Classification of credit approval on the basis of customer data. University gives class to the students based on marks..

1. Black box  
   Black-box testing is a method of software testing that examines the functionality of an application without into its internal structures or workings.
2. Type 1 2 error  
   Type 1:

Type I error, also known as a “false positive”: the error of rejecting a null hypothesis when it is actually true.

Tyep 2:

Type II error, also known as a "false negative": the error of not rejecting a null hypothesis when the alternative hypothesis is the true state of nature.

1. Cleaning techniques
2. What is selenium junit  
   JUnit is an open-source unit testing tool that helps to test units of code. It is mainly used for unit testing Java projects; however, it can be used with Selenium Webdriver to automate the testing of web applications
3. White box  
   White-box testing is a method of software testing that tests internal structures or workings of an application,
4. What is information and knowledge?
5. What is data mining ?  
   Data mining is the process by which organizations detect patterns in data for insights relevant to their business needs. It's essential for both business intelligence and data science.
6. And which miniproject have you performed?
7. k means  
   k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster.
8. confusion matrix  
   A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known
9. apriori

* Apriori is an algorithm for frequent item set mining and association rule learning over relational databases.
* It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database.

1. Schema

* A schema is defined as a logical description of database where fact and dimension tables are joined in a logical manner.
* Data Warehouse is maintained in the form of Star, Snow flakes, and Fact Constellation schema

1. data preprocessing - cleaning, transformation

Cleaning: fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies.

Transformation:

1. Normalization:
   * Scaling attribute values to fall within a specified range.
   * Scaling by using mean and standard deviation
2. Aggregation: moving up in the concept hierarchy on numeric attributes.
3. market basket

Market basket analysis is a data mining technique used by retailers to increase sales by better understanding customer purchasing patterns.

It involves analyzing large data sets, such as purchase history, to reveal product groupings, as well as products that are likely to be purchased together.

1. Association rule mining

Association rule mining is a procedure which aims to observe frequently occurring patterns, correlations, or associations from datasets found in various kinds of databases such as relational databases, transactional databases, and other forms of repositories.