

Data Mining.

Q. How is data warehouse diff. from a database? How are they similar?

<u>Ans.</u>	Database.	Data warehouse
	→ Is designed to record.	→ Is designed to analyse
	→ The database uses the online transactional Processing (OLTP).	→ Data warehouse uses online Analytical Processing (OLAP).
	→ The database helps to perform fundamental operations for your business.	→ Data warehouse allows you to analyse your business.
	→ Is an application-oriented collection of data.	→ It is subject oriented collection of data.

The similarity b/w data warehouse & database is that both the system maintain data in form of table, indexes, columns, values & keys. Also data is retrieved in both by using SQL queries.

Q. Define each of the following data mining functionalities with examples?

Ans. (i) Classification: It is a process of finding a model that describes & distinguishes data classes & concepts. The model is derived based on the analysis of set of training data. The model is used to predict the class label of ~~analysis~~ objects for which the class label is unknown.

(ii) Regression: It is a ⁺ statistical methodology that is most often used for numeric prediction, although other methods exists as well. Regression also encompasses the identification of ~~databases~~ distribution trends based the available data.

(iii) Clustering: It analyses data objects without consulting class labels. In many cases, class label data may simply not exist at the beginning. clustering can be used

to generate labels for a group of data. The objects are clustered or grouped based on the principle of maximizing the intracluster similarity & minimizing the intercluster similarity.

(iv) Outlier Analysis: A data set may contain objects that do not comply with the general behaviour or, model of the data. These data objects are outliers. Many data mining methods discard outliers as noise or exceptions. However in some applications the rare events can be more interesting than the more regularly occurring ones. The analysis of outlier data is referred to as outlier analysis.