Data mining and Data warehousing:

1) Data Warehouse is:
=) A Subject - Oriented integrated time variant non - volatile collection of data in support of management.

2) Fact tables are describes by which of the following?
=) Completely normalized.

3) Dimensionality refers to: =) The data that describes the transcation in fact table

4) A Start Schema has what type relanship from dimension to fact table =) One to many.

5) Facts are numeric measurments or Values that represent a specific business aspect or activity

6) MDDB stand for =) multidimensional databases.

7) ETL is =) Excration / Transformation / Loading

8) Which of the following is not componenend of data warehouse?
=) Meta data.

9) The data from the operational environment enter current detail data of data warehouse.

10) Which of the following is not a component of data warehouse
=) Ligtly Summarized data.

11) The data Stored, retrived & updated in =) OLTP

12) Data mining System classification consist of =) Information science, Database Technology, Machine Learning.

13) Dimensionally reduction reduces the data set size by removing =) Irrelavant attribute

14) The Bottom Tier the tre tier architecture of a data warehouse consist of.
=) Data repository.

15) KDD Involves =) Transforming & Reducing, Selecting & Sampling Data, Cleaning and Pufrocessing

16) Pattern Evaluation is Converted to =) knowlege based.

17) Merging the Data from multiple Source into coherent data method is =) Data Integration.

18) Efficient and Scalabhty of data mining algouthm issue come under ? Performance Issue.

19) Data Selection is => The Stage of Selecting the right data for KDD process.

20) Relational data can be accessed by <u>queries</u> Written as relational query Language.

21) A Relational database is collection of table each of which is assigned a unique name which uses the entity relationship (ER) data model.

22) To remove noise and inconsistent data, Data cleaning is needed.

23) Multiple data Source may be combained is called as Data Integration

24) A Data Warehouse is described by which of the following?
=> Organized around important Subjects areas.

25) A Business metadata contains information that gives user an easy to understand prospective of the information Stored in data warehouse.

26) Operational Metadatabase defines the structure of the data held in operation database and used by operatable application.

27)

Data mining & Data Warehousing (Quiz-2)

1) Support $(A \Rightarrow B) = \Rightarrow P(A \cup B)$

2) confidence $(A \Rightarrow B) \Rightarrow P(B/A)$

3) Tree pruning is $\Rightarrow$ To identify and remove such branches.

4) Minimum confidence threshold are called $\Rightarrow$ strong Association rule

5) I one of the frequent item - Set Example it observed that if milk and bread are brought then egg are also purchased by the coustomer after generating an association rule among the given set of item it is infered.
The item set {milk, bread} is consequent and is antecedent

6) How can we best represent 'support' for the following association rule if X and y then 'Z'
{x, y, z} / (Total number of transcation)

7) A database have five transcation of our transaction consist milk & bread, further of the given four trans thru trasc include cheese bread and Support percentage for the following association rule if milk and bread are purchased then chese purchased $\Rightarrow 60\%$.

8) Knowledge type of constraints to be mined based on Association None of the option (Both A & B)

9) high level declarative Data mining query Language and user inter face based on $\Rightarrow$ Dimensional level constraints / rule contrains / Data constrains.

10) Bayesian classifier are statisticaly classifier they can predict -the $\Rightarrow$ Membership probablities. (Both A & B)

11) Mining frequent item set and association has been extended Various way include mining $\Rightarrow$ Multi dimensional association rule. None of the option you above

12) Method for mining frequent item set can be extended for the mining of $\Rightarrow$ closed frequent itemset

13 Users allows to focous the search for rule by providing mela rule consist $\Rightarrow$ constraint based rule mining

14) What pattern are pattern appear frequent data set $\Rightarrow$ frequent

15) Finding frequent pattern association correlations, or casual structure among set of item or object in transacation database relational databases and other information repositories => **Association rule mining.**

16) classification predicts => **categorical Data**

17) The training data are accompanied by the label indicating. The classes observation is called => **Supervised Learning**

18) Which of the following is disadvantage of descion tree => **Descision tree prone to overfit.**

19) Which of the following algorithm are not an Example of ensemble learning algorithm => **Decsion tree.**

20) If Linear regression model perfectly first ie train error is zero then => **Couldn't comment on Test Error.**

21) Which of the following metrics can be used for evaluating regression models? => **RMSE / MSE / MAE**

22) A regression model in which more than one independent variable is used to predict the dependent variance called => **multiple regression models.**

23) The Bayesian Network can be used to answer any query by using => **Joint Distribution.**

24) The training data are accompained by the label indicating the class observation called => **Supervised Learning**

25) What type of class label training data unknown => **Unsupervised**

26) Decision Tree are => **High dimensional data with good accuracy**

27) Bagging is => **Reduce Variance of descsion tree classifier.**

28) Descsion tree indication is learning of descsion tree each branch represent => **Out come of the tree**

29) Preprocessing data in order to reduce noise and handling missing value => **Data cleaning**

30) Which is constructed in top down recursive divide and conquer manner => **Decision Tree.**