KUMARAGURU
college of technology
character is life

Register Number:……..

B.TECH DEGREE EXAMINATIONS: MAY 2015

(Regulation 2009)

Seventh Semester

INFORMATION TECHNOLOGY

ITY117**:** Data Warehousing and Data Minning

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. An aggregate function is -------- if there is no constant bound on the storage size needed to describe a subaggregate

   a) Distributive b) Algebraic

   c) Holistic d) Normal

2. A semantic data model such as ---------- data model, which models the database as a set of entities and their relationships, is often constructed for relational database

   a) UML b) ER

   c) CRC d) Database

3. Data mining uses _____ approach

   a) Query driven b) Update driven

   c) Module driven d) Process driven

4. ---------- involves finding the “best” line to fit two variables, so that one variable can be used to predict the other.

   a) Clustering b) Linear regression

   c) Multiple linear regression d) Binning

5. ------ computes an aggregate function over an attribute or set of attributes in order to find aggregate values above some specified threshold.

   a) Iceberg queries b) Holistic queries

   c) Aggregate queries d) Database queries

6. -------- measure is referred to as a measure of the goodness of split.

   a) Information gain b) Support

   c) Confidence d) Pruning

7. --------- is a multi resolution clustering algorithm that first summarizes the data by imposing a multidimensional grid structure onto the data space.
   a) STING b) CLIQUE
   c) WaveCluster d) GLIQUE
8. --------- approach to outlier detection assumes a distribution or probability model for the given data set and then identifies outliers with respect to the model using a discordancy test.
   a) Distance-based outlier detection b) Statistical-based outlier detection
   c) Deviation-Based outlier detection d) Sequential-Based outlier detection.
9. -------- is an effective way to discover knowledge from huge volume of data
   a) Visual data mining b) Web mining
   c) Weblog mining d) Web content mining
10. Interest increases as more and more methods for the technology are proposed. This stage of technology adoption is called
   a) Innovators b) Chasm
   c) Laggards d) Early adopters

PART B (10 x 2 = 20 Marks)

11. List down the major features of a data warehouse
12. Identify the contents of a metadata repository.
13. Calculate the information gain with 20 samples of class graduate and 30 samples of class undergraduate
14. List down the steps involved in the process of knowledge discovery.
15. Explain the terms- support and confidence
16. Explain – Association rule mining
17. Compare nominal and ordinal variables.
18. Explain- Agglomerative hierarchical clustering
19. Identify any four commercial data mining systems.
20. List out some tools for text mining.

PART C (5 x 14 = 70 Marks)

21. a) Explain the steps involved in the design and construction of data warehouse.

(OR)

b) Explain the various OLAP operations for multidimensional data with an appropriate example.

22. a) (i) List down the strategies for data reduction. Explain about dimensionality reduction with an example. (8)

(ii) Discuss about the various methods for handling missing values (6)

(OR)

b) Explain about discretization and concept hierarchy generation for numeric data with examples.

23. a) (i) Consider the following database (7)

TID	List of item_IDs
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

Find the frequent item set using Apriori algorithm.

(ii) Discuss about the process of improving the efficiency of Apriori with an example. (7)

(OR)

b) (i) Compare clustering and classification. (4)

(ii) Explain about the Naïve Bayesian classification with an example. (10)

24. a) Discuss about Partitioning and Model-based clustering methods.

(OR)

b) Discuss about Distance-based and deviation-based approaches for outlier detection.

25. a) Discuss the role of data mining in processing the Text.

(OR)

b) Discuss about mining multimedia databases.

****