Code : 051822

B.Tech 8th Semester Exam., 2015

DATA MINING

Time : 3 hours   AKUbihar.com   Full Marks : 70

Instructions :

(i) The marks are indicated in the right-hand margin.

(ii) There are **NINE** questions in this paper.

(iii) Attempt **FIVE** questions in all.

(iv) Question No. **1** is compulsory.

1. Choose the correct option of any *seven* of the following : 2×7=14

   (a) Data mining is also referred to as

      (i) knowledge discovery in databases

      (ii) data cleaning

      (iii) data extraction

      (iv) data management

   (b) Data about data is called

      (i) table   AKUbihar.com

      (ii) database

      (iii) metadata

      (iv) integration

   (c) To represent any n-dimension data we need a series of _____ dimension cubes.

      (i) $(n-1)$

      (ii) n AKUbihar.com

      (iii) $n+1$

      (iv) $n+2$

   (d) The _____ operation performs a selection on one dimension of the given cube, resulting in a subcube.

      (i) pivot

      (ii) slice

      (iii) roll-up

      (iv) drill-down

   (e) _____ servers support multidimensional views of data through array-based multidimensional storage engines.

      (i) ROLAP

      (ii) MOLAP

      (iii) Data warehouse AKUbihar.com

      (iv) Database

   (f) The _____ software gives the user the opportunity to look at the data from a variety of different dimensions.

      (i) query tools

      (ii) multidimensional analysis

      (iii) data mining tools

      (iv) None of the above

AKUbihar.com

(g) _____ techniques can be used to reduce the number of values for a given continuous attribute by dividing the range of the attribute into intervals.

   (i) Discretization

   (ii) Transformation

   (iii) Smoothing

   (iv) Generalization

(h) FP tree growth algorithm can be implemented in

   (i) one phase

   (ii) two phases

   (iii) three phases

   (iv) four phases   **AKUbihar.com**

(i) Consider a scenario where a bin contains values 4, 8 and 15. If smoothing by bin-means method is applied to clean the data, then each of the original value in the bin will be replaced by

   (i) 8

   (ii) 9

   (iii) 15

   (iv) 4

(j) _____ is simple text files that are automatically generated every time someone accesses one Web site.

   (i) Server session

   (ii) Log file   **AKUbihar.com**

   (iii) User session

   (iv) None of the above

2. (a) State what data cleaning routine attempts to fill the missing data.   7

  (b) Explain the market-basket analysis problem. State the technique to analyse this type of problem.   7

3. (a) Why is tree pruning useful in tree induction method?   5

  (b) What is the classification of association rules based on various criteria?   5

  (c) What is data mining? Explain the steps in knowledge discovery.   4

4. (a) What are descriptive and predictive data mining? Explain the data mining functionalities. **AKUbihar.com**   5

  (b) Explain the various data mining repositories on which mining can be performed.   5

(c) Define data warehouse. What is the need of data warehouses? Define multidimensional data model. 4

5. (a) List out the OLAP operations and explain the same with an example. 5

(b) How are association rules mined from large databases? Explain. 5

(c) Explain about generalization and summarization. 4

6. (a) What are Bayesian classifiers? 3

(b) Write a note on k-means algorithm. 4

(c) Discuss classifier accuracy with examples. 4

(d) Explain outlier analysis. 3

7. (a) Explain the process of mining the World Wide Web. **AKUbihar.com** 5

(b) Explain spatial mining and time series mining. 4

(c) Discuss about the grid-based methods. 5

8. Write short notes on the following : 5+5+4=14

(a) Data preprocessing

(b) Data mining query language

(c) Prediction

9. (a) What is hierarchical method? 3

(b) Differentiate between agglomerative and divisive hierarchical clustering. 4

(c) Define density-based method. 3

(d) Discuss the major differences between the star schema and the snowflake schema. 4

★ ★ ★