K18U 0183

Reg. No. : .....

Name : .....

VI Semester B.C.A. Degree (CBCSS – Reg./Supple./Imp.) Examination, May 2018
Core Course
6B18BCA : DATA MINING AND DATA WAREHOUSING
(2014 Admn. Onwards)

Time : 3 Hours Max. Marks : 40

SECTION – A

1. **One** word answer : (8×0.5=4)
   a) Data are stored in a data warehouse to provide a historical perspective. This property is termed as _____ .
   b) A popular data model that influences data warehouse architecture is _____.
   c) The term _____ refers to the process of extracting relevant and useful information from large datasets.
   d) An itemset that was a border set before update and becomes a frequent set after update, is called a _____ itemset.
   e) For two clusters C_1 and C_2 with centroids $O_{centroid,\ 1}$ and $O_{centroid,\ 2}$ respectively, the average intercluster distance between these clusters is defined as _____ .
   f) An agglomerative hierarchical clustering algorithm normally uses _____ based representatives to determine the similarity between clusters.
   g) The dataset used to measure the accuracy of the classifier is called _____
   h) In decision tree induction algorithm ID3 _____ is used to as an attribute selection measure to select the split attribute.

SECTION – B

Write short notes on **any seven** of the following questions : (7×2=14)

2. Define a data warehouse.
3. Explain the term "Meta Data", associated with data warehouse.
4. List out the different stages of a KDD process.

P.T.O.

5. What is text mining ?
6. Define the terms Support and Confidence.
7. Explain the apriori property.
8. Define Manhattan intercluster distance between two clusters.
9. What is a decision tree ?
10. Define the terms entropy and information gain.
11. What is supervised learning ?

SECTION – C

Answer **any four** of the following questions : (4×3=12)

12. Differentiate between MOLAP and ROLAP.
13. Write a short note on Spatial Data Mining.
14. Describe the working of FP-Tree Growth Algorithm.
15. Describe the underlying principle behind the CLARA clustering algorithm.
16. Explain about various splitting criteria usually adopted in decision tree constructions algorithms.
17. What do you mean by overfitting ? How it will affect the classification using a decision tree ? How overfitting can be handled in a decision tree ?

SECTION – D

Write an essay on **any two** of the following questions : (2×5=10)

18. Explain about various data warehouse schemas.
19. What are the various issues and challenges in Data Mining ?
20. Explain apriori association rule mining with an example dataset.
21. Write short notes on :
    a) DBSCAN clustering algorithm.
    b) C4.5 decision tree induction algorithm.

———————