

(6 pages)

Reg. No. :

Code No. : 21046

Sub. Code : GMCA 6 B

B.C.A. (CBCS) DEGREE EXAMINATION, APRIL 2017.

Sixth Semester

Computer Application — Main

Major Elective — DATA MINING

(For those who joined in July 2012-2015)

Time : Three hours

Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. Which one of the following questions can be easily answered without using any data mining tools?
 - (a) How many different types of customers are there?
 - (b) What customers are the most profitable?
 - (c) Which state in India did most students come from?
 - (d) Which patient has a high risk of breast cancer?

2. _____ is not a step in the data mining process.

- (a) Visualization
- (b) Report production
- (c) Requirements analysis
- (d) Testing and validation

3. Discovery of cross-sales opportunities are called _____.

- (a) segmentation (b) visualization
- (c) correction (d) association

4. Three frequent pairs (p, q) , (r, s) and (t, u) have been found in association rule mining. What is the maximum number of rules that may be derived from these three pairs?

- (a) 3 (b) 4
- (c) 6 (d) 8

5. _____ software provides new splitting criteria, non-greedy search, new partitions, extraction of several and different solutions.

- (a) Quadstone (b) OC1
- (c) SMILES (d) NBC



6. _____ classification in which the classes are determined based on the given data.
 (a) unsupervised (b) supervised
 (c) apriori (d) decision tree
7. The _____ algorithm can be applied in cleaning data.
 (a) search (b) pattern recognition
 (c) learning (d) clustering
8. _____ makes the clustering problem difficult to solve.
 (a) Too much data
 (b) Too many attributes
 (c) Not all attributes are numeric
 (d) The number of clusters is not given
9. _____ pages assist the user to navigate through the web site.
 (a) Home (b) Index
 (c) Reference (d) Content
10. _____ is not part of web mining.
 (a) Web content mining
 (b) Web structure mining
 (c) Web quality mining
 (d) Web usage mining

Page 3 Code No. : 21046

PART B — (5 × 5 = 25 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 250 words.

11. (a) What kind of tasks is data mining suitable for?
 Or
 (b) What is an OLTP system?
12. (a) Give an approach for improving the Apriori algorithm.
 Or
 (b) If (a, b) , (c, b) and (e, c) are the only 2-itemsets that are frequent, how many 3-itemsets are possible? What are they?
13. (a) What is the aim of the split algorithm?
 Or
 (b) Comment on information gain.
14. (a) Mention the desired features of cluster analysis.
 Or
 (b) What is the role of distance between cluster in Hierarchical methods?

Page 4 Code No. : 21046
 [P.T.O.]



15. (a) Comment on Web communities.

Or

- (b) Give an algorithm for finding hubs and authorities.

PART C — (5 × 8 = 40 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 600 words.

16. (a) Narrate the important guidelines for successful data mining.

Or

- (b) Explain how data mining software should be evaluated and selected for purchase.

17. (a) Show how the Apriori algorithm works on an example data set.

Or

- (b) What is Dynamic Itemset Counting (DIC)? Explain the algorithm.

18. (a) Describe the building algorithm for the decision tree using information theory.

Or

- (b) Explain the software used for classification.

Page 5 Code No. : 21046

19. (a) Describe the features of K-Means clustering methods.

Or

- (b) How to deal with large databases? Explain.

20. (a) Narrate the step by step procedure of HITS algorithm.

Or

- (b) Write note on Web content mining.

Page 6 Code No. : 21046

