

(6 pages)

Reg. No. : .....

Code No. : 9063

Sub. Code : HCAM 52

M.C.A. (CBCS) DEGREE EXAMINATION,  
NOVEMBER 2016.

Fifth Semester

Computer Applications

DATA MINING AND WAREHOUSING

(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. \_\_\_\_\_ is a process for assembling and managing data from various sources for the purpose of gaining single detailed view of an enterprise
- (a) Data warehouse
  - (b) Operational data stores
  - (c) Data mining
  - (d) None

2. \_\_\_\_\_ is data about data or documentation about the data that is needed by the users
- (a) Meta data
  - (b) Cluster
  - (c) Both (a) and (b)
  - (d) None
3. \_\_\_\_\_ is the algorithm used for mining frequent patterns without candidate generation
- (a) Apriori algorithm
  - (b) FP growth
  - (c) DIC algorithm
  - (d) Clustering algorithm
4. A \_\_\_\_\_ is a popular classification method
- (a) Decision tree
  - (b) Web content mining
  - (c) Both (a) and (b)
  - (d) None
5. \_\_\_\_\_ deals with discovering and modeling the link structure of the web
- (a) Web usage mining
  - (b) Web structure mining
  - (c) Web content mining
  - (d) None
6. \_\_\_\_\_ deals with understanding user behaviour in interacting with the web or with a website
- (a) Web usage mining
  - (b) Web structure mining
  - (c) Web content mining
  - (d) None



7. In Baye's theorem, the expression \_\_\_\_\_ refers to the probability that event A will occur
- (a)  $P(B)$  (b)  $P(A)$   
(c)  $P(B/A)$  (d)  $P(A/B)$
8. The process of grouping given data into meaningful clusters is called as \_\_\_\_\_
- (a) Classification (b) Grouping  
(c) Prediction (d) Cluster analysis
9. The \_\_\_\_\_ summarizes data by replacing relatively low level values with higher concepts
- (a) Data generalization  
(b) Specialization  
(c) Both (a) and (b)  
(d) None
10. \_\_\_\_\_ mining refers to the extraction of knowledge spatial relationships, or other interesting patterns not explicitly stored in spatial databases
- (a) Web (b) Spatial  
(c) Special (d) None

Page 3 Code No. : 9063

PART B — ( $5 \times 5 = 25$  marks)

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

Each answer should not exceed 250 words.

11. (a) Explain various data mining techniques in detail.

Or

- (b) What is data warehousing?

12. (a) Define support and confidence for an association rule.

Or

- (b) How to build a decision tree using tree induction algorithm?

13. (a) How do you improve accuracy of classification methods?

Or

- (b) What is cluster analysis?

14. (a) What are the different terminology available in web mining?

Or

- (b) Draw a neat diagram of search engine.

Page 4 Code No. : 9063  
[P.T.O.]



15. (a) With short note on generalization and specialization.

Or

- (b) Explain spatial rules in detail.

PART C — (5 × 8 = 40 marks)

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

Each answer should not exceed 600 words.

16. (a) Explain the importance of meta data in data warehousing. Give a brief note on data cube.

Or

- (b) What is data mining? Explain data mining process in detail.

17. (a) Explain split algorithm based on Gini index.

Or

- (b) Explain direct hashing and pruning in detail.

18. (a) How do you estimate predictive accuracy of classification methods? Explain the various methods for estimating the accuracy of a method.

Or

- (b) What is the use of partitioning methods in cluster analysis? Explain with an example.

Page 5

Code No. : 9063

19. (a) Explain web structure mining in detail.

Or

- (b) With an example, explain how to rank the web page.

20. (a) Write short note on spatial clustering algorithm.

Or

- (b) Write short notes on the following

- (i) Spatial data overview
- (ii) Spatial data mining primitives.

Page 6

Code No. : 9063

