

**Fifth Semester B.C.A. Degree Examination,  
October/November 2019**

(CBCS Scheme)

**Computer Science**

**DATA MINING**

Time : 3 Hours]

[Max. Marks : 90

Instructions to Candidates : Answers should be written in English only.

**SECTION – A**

1. Answer any **TEN** questions : (10 × 1 = 10)
- (a) What is OLTP?
  - (b) Define data mining.
  - (c) Expand KDD.
  - (d) Mention any two applications of data mining.
  - (e) State Baye's theorem.
  - (f) Mention any two classification methods.
  - (g) What is clustering
  - (h) What is dendrogram?
  - (i) What is web mining?
  - (j) What is search engine?
  - (k) Define big data.
  - (l) Mention any two tools used in big data.



**SECTION – B**

Answer any **FIVE** questions : (5 × 3 = 15)

- 2. Define missing value, noisy data and duplicate data.
- 3. List any three differences between data mining and data warehouse.
- 4. Explain the associative classification with an example.
- 5. Explain the popular classification software.
- 6. Explain different smoothing techniques.

## Q.P. Code – 68506

7. Explain the web terminologies used in web mining.
8. Write any three applications of big data.

### SECTION – C

Answer any **SIX** questions :

(6 × 5 = 30)

9. Explain the OLAP operations.
10. Explain the architecture of data warehouse with a neat diagram.
11. Explain rule based classifier used to solve the classification problem.
12. How decision tree classifier works, explain with an example.
13. Given data set  $K = \{2, 3, 4, 10, 11, 12, 20, 25, 30\}$  using K-mean algorithm divide the data set into 2 clusters.
14. Explain the hierarchical clustering technique with an example.
15. List and explain the factors affecting the search engine ranking.
16. Explain the five characteristics of big data.

### SECTION – D

Answer any **FIVE** questions :

(5 × 7 = 35)

17. Explain the architecture of data mining with a neat diagram.
18. (a) Explain the steps involved in KDD process. (4)  
(b) Define machine learning. List its categories. (3)
19. Consider the given data set. Apply Navie-Baye's algorithm and predict that if a fruit has the following properties then which type of fruit it is,

Fruit = { yellow, sweet, long }

Frequency table :

Fruit	Yellow	Sweet	Long	Total
Mango	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

**Q.P. Code – 68506**

20. What is agglomerative clustering? Cluster the following data using agglomerative approach and represent through dendrogram.

	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

21. Mention any three cluster analysis methods. Explain any one in detail.
22. Explain in detail, the architecture of search engine and its working.
23. Explain map reduce technique. Explain with an example.

