

(6 pages)

Reg. No. :

Code No. : 40386 E

Sub. Code : JMCS 64/
JMSE 64

B.Sc. (CBCS) DEGREE EXAMINATION,
NOVEMBER 2019.

Sixth Semester

Computer Science/Software Engineering — Main

DATA MINING

(For those who joined in July 2016 onwards)

Time : Three hours

Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. _____ is useful when the classes in the data are not already known and the training data is not available
 - (a) Association rule mining
 - (b) Classification
 - (c) Cluster Analysis
 - (d) Web data mining

2. _____ has data mining software called knowledge STUDIO
- (a) Angoss (b) Mineset
(c) Mentas (d) Data Miner
3. Another name for association rule mining is
- (a) Relationship marketing
(b) Customer profiling
(c) Customer segmentation
(d) Market basket analysis
4. Confidence of $X \rightarrow Y$ is
- (a) $P(X \cup Y) / P(X)$ (b) $P(Y) / P(X)$
(c) $P(X \cap Y) / P(X)$ (d) $P(X) / P(X \cap Y)$
5. The _____ method requires a training set and a test set
- (a) Holdout
(b) Random Sub-Sampling
(c) Leave - one - out
(d) Bootstrap

6. _____ is a simple Naive Bayes classifier software.
- (a) SMILES (b) C4.5
(c) CART5.0 (d) NBC
7. In _____ the classes are not pre defined
- (a) Classification (b) Regression
(c) Clustering (d) Association Rules
8. _____ methods obtain a nested partition of the objects resulting in a tree of clusters
- (a) Partitional (b) Hierarchical
(c) Density-based (d) Grid-based
9. Copying an entire website is called _____
- (a) Replication (b) Mirroring
(c) Sample (d) Pattern
10. A Cyber-Community may be a collection of web pages that form a
- (a) Complete graph (b) Connected graph
(c) Bipartite graph (d) Cyclic graph

PART B.— (5 × 5 = 25 marks)

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

Each answer should not exceed 250 words.

11. (a) Explain the future of Data mining.

Or

- (b) Explain the following Data mining software

(i) Mineset

(ii) Miningmart

(iii) Oracle

12. (a) Write a note on Apriori-TID.

Or

- (b) Explain the performance Evaluation of Algorithms.

13. (a) Write about building a decision tree using the tree induction algorithm.

Or

- (b) Write about improving accuracy of classification methods.

14. (a) Write about computing distance in Cluster Analysis.

Or

- (b) Explain the following cluster analysis software.

- (i) Autoclass
- (ii) Cluster 3.0
- (iii) CViZ Cluster Visualization

15. (a) Define some of the web terminology based on the work of world wide wEB construction.

Or

- (b) Write about Finger printing.

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 Data mining process.

Or

- (b) List and explain Data mining Applications.

17. (a) Write in detail about improving the efficiency of the Apriori Algorithm.

Or

- (b) Write a note on mining frequent patterns without candidate generation (FP-Growth).

18. (a) Write a note on estimating predictive accuracy of classification methods.

Or

- (b) Discuss in detail on other Evaluation Criteria for classification methods.

19. (a) List and explain the desired features of cluster Analysis.

Or

- (b) Write about dealing with large databases.

20. (a) Discuss about Locality and Hierarchy in the Web.

Or

- (b) Write an elaborate note on web usage mining.
-