

(6 pages)

Reg. No. :

Code No. : 20997

Sub. Code : GMCS 63/
GMSE 63

B.Sc. (CBCS) DEGREE EXAMINATION, APRIL 2015.

Sixth Semester

Computer Science — Main

DATA MINING

(For those who joined in July 2012 and afterwards)

Time : Three hours

Maximum : 75 marks

SECTION A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. Metadata is data about data that describe the
(a) Data mining (b) Data warehouse
(c) Data mart (d) Meta data
2. KDD means
(a) Known Discovery in Database
(b) Knowledge Discovery in Database
(c) Knowledge Data in Database
(d) Known Data in Database

3. OLAP often is used in

- (a) image mining (b) audio mining
(c) text mining (d) data mining

4. _____ is the act or process of retrieving data out of data sources for further data processing or data storage

- (a) Data mining
(b) Data transformation
(c) Data cleanup
(d) Data extraction

5. _____ trees are methodologies to classify data into discrete ones using the tree-structured algorithms

- (a) AVL (b) Binary
(c) B-tree (d) Classification

6. Classification tree is also known as

- (a) Decision tree (b) Regression tree
(c) Binary trees (d) Complete trees

7. Which of the most used clustering algorithms?

- (a) k - means (b) c - means
(c) J-k means (d) L - means

Page 2 Code No. : 20997



8. Which of the clustering algorithms may not be classified as listed below?

- (a) Exclusive clustering
- (b) Overflow clustering
- (c) Hierarchical clustering
- (d) Probabilistic clustering

9. _____ analysis is a tool for exploring the structure of data

- (a) Regression (b) Cluster
- (c) Simple (d) Multiple

10. STC stands for

- (a) Suffix Tree Clustering
- (b) Side Track Code
- (c) Suffix Thread Clustering
- (d) Side Track Clustering

Page 3 Code No. : 20997

SECTION B — (5 × 5 = 25 marks)

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

Each answer should not exceed 250 words.

11. (a) Explain the importance of data mining now.

Or

(b) Discuss the future of data mining.

12. (a) Describe how to improve the efficiency of the Apriori algorithm.

Or

(b) Discuss performance evaluation of algorithms.

13. (a) Give an introduction on Naive Bayes method in classification.

Or

(b) Explain split Algorithm based on the Information theory.

14. (a) Describe in detail on hierarchical clustering.

Or

(b) Elucidate the desired features of cluster analysis.

Page 4 Code No. : 20997

[P.T.O.]



15. (a) Write short notes on the locality and hierarchy in the web.

Or

- (b) Write in detail about web content mining.

SECTION C — (5 × 8 = 40 marks)

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

Each answer should not exceed 600 words.

16. (a) Discuss in detail the data mining process.

Or

- (b) Discuss the data mining techniques.

17. (a) Explain Apriori Algorithm with example data.

Or

- (b) Illustrate mining frequent patterns without candidate.

18. (a) Explain in detail how to build decision trees.

Or

- (b) Write in detail about classification software.

Page 5 Code No. : 20997

19. (a) Briefly explain K – means algorithm.

Or

- (b) Describe in detail Agglomerative method.

20. (a) Discuss about web usage mining.

Or

- (b) What is web structure mining? Explain.

Page 6 Code No. : 20997

