

Code No. : 20997

Sub. Code : GMCS 03
GMSE 03

B.Sc. (CBCS) DEGREE EXAMINATION, APRIL 2018

Sixth Semester

Computer Science/Software Engineering — Main
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. _____ is a collection of exploratory techniques based on advanced analytical methods and tools for handling a large amount of information

- (a) Data Mining
- (b) KDD
- (c) OLTP
- (d) (a) or (b)

_____ is a typical datamining process

- (a) requirements analysis
- (b) data selection and collection
- (c) cleaning and preparing data
- (d) all the above

Confidence of $(x \rightarrow y) = ?$

- (a) $P(Y/X)$
- (b) $P(X \cap Y)$
- (c) $P(X)$
- (d) $P(XY)$

Given the association rules $x \rightarrow yz$ and $A \rightarrow BC$ which one of the following is true?

- (a) $y \rightarrow x$
- (b) $z \rightarrow y$
- (c) $A \rightarrow B$
- (d) $C \rightarrow A$

Classification is the _____ in to classes

- (a) separation of objects
- (b) separation of things
- (c) ordering of objects
- (d) all the above

6. Which one of the following is tree?
- (a) Decision tree is not predictive
 - (b) Decision tree is predictive
 - (c) Decision tree is not descriptive
 - (d) Quality of training data is not important for decision trees

7. Which of the following is a desired feature of cluster analysis?

- (a) maximal input parameters
- (b) scalability no more than three scans of large datasets
- (c) no more than three scans of large data sets
- (d) many scan of the data set

8. A distance metric used in cluster analysis must have the following property

- (a) always negative
- (b) distance from x to x is greater than Zero
- (c) distance from x to y is same as y to x
- (d) all the above

Web logs include information about _____

- (a) the referring pages
- (b) the ser information
- (c) the time a user spends and page visited
- (d) all the above

Web usage mining is _____

- (a) about web page contents
- (b) to discover the link structure of the web
- (c) about the user behaviour
- (d) a search engine

PART B — (5 × 5 = 25 marks)

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

Each answer should not exceed 250 words.

- (a) What do you understand about Data Mining? Explain.

Or

- (b) Explain Applications of Data Mining. Briefly.

- (a) Describe the basics of Association rule mining. Give examples.

Or

- (b) Explain Naive Algorithm with an example.

13. (a) What do you mean by classification? Explain decision tree with an example.

Or

- (b) Describe the NAIVE BAYES method.

14. (a) Describe the desired features of cluster analysis.

Or

- (b) Describe the types of cluster analysis methods.

15. (a) What do you mean by web mining? Describe any six web terminologies.

Or

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

PART C — (5 × 8 = 40 marks)

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

Each answer should not exceed 600 words.

16. (a) Explain Data mining Techniques in detail

Or

- (b) Describe any four data mining case studies

(a) Discuss the algorithmic aspects of the apriori algorithm.

Or

(b) Explain how to generate frequent pattern tree. Give an example.

(a) Explain split algorithm based on information theory for classification. Give an example.

Or

(b) Explain:

(i) Decision Tree rules

(ii) Evaluation criteria for classification methods.

(a) Explain partitional methods for clustering.

Or

(b) Explain Hierarchical methods for clustering.

(a) Explain about web content mining.

Or

(b) What is web structure mining? Explain HITS algorithm with an example.