PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

BCA DEGREE EXAMINATION DECEMBER 2022
(Fourth Semester)

Branch – COMPUTER APPLICATIONS

DATA WAREHOUSING AND MINING

Time: Three Hours   Maximum: 75 Marks

SECTION-A (10 Marks)
Answer ALL questions
ALL questions carry EQUAL marks   (10 x 1 = 10)

1. What is KDD in data mining?
(i) Knowledge Discovery Database
(ii) Knowledge Discovery Data
(iii) Knowledge Data definition
(iv) Knowledge data house

2. What are the functions of Data Mining?
(i) Association and correctional analysis classification
(ii) Prediction and characterization
(iii) Cluster analysis and Evolution analysis
(iv) All of the above

3. Which one of the following statements is not correct about the data cleaning?
(i) It refers to the process of data cleaning
(ii) It refers to the transformation of wrong data into correct data
(iii) It refers to correcting inconsistent data
(iv) All of the above

4. Choose from the following that are Decision Tree nodes?
(i) Decision Nodes
(ii) End Nodes
(iii) Chance Nodes
(iv) All of the mentioned

5. A data warehouse is which of the following?
(i) Can be updated by end users.
(ii) Contains numerous naming conventions and formats.
(iii) Organized around important subject areas.
(iv) Contains only current data.

6. Significant characteristics of Distributed systems have of
(i) 5 types
(ii) 2 types
(iii) 3 types
(iv) 4 types

7. What is auto-association task in neural networks?
(i) find relation between 2 consecutive inputs
(ii) related to storage & recall task
(iii) predicting the future inputs
(iv) none of the mentioned

8. Genetic algorithms are search algorithms based on the mechanics of natural_____.
(i) systems.
(ii) genetics.
(iii) logistics.
(iv) statistics.

9. What is the use of data cleaning?
(i) to remove the noisy data
(ii) correct the inconsistencies in data
(iii) transformations to correct the wrong data.
(iv) All of the above

10. A collection of interesting and useful patterns in database is called ___
(i) knowledge.
(ii) information.
(iii) data.
(iv) algorithm

Cont...

SECTION - B (25 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** Marks   (5 x 5 = 25)

11 a) Discuss about the cleaning and transforming data.
(OR)
  b) Explain the functions of load manager with necessary theory.

12 a) Describe on tuning the data load.
(OR)
  b) Narrate on recovery testing.

13 a) Bring out the data mining issues.
(OR)
  b) Explain social implications of data mining with necessary theory.

14 a) Outline about decision tree.
(OR)
  b) Analyze about genetic algorithm.

15 a) Explain about incremental rules in detail.
(OR)
  b) Give a outline on web content mining.

SECTION -C (40 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** Marks   (5 x 8 = 40)

16 a) Analyze about query management process.
(OR)
  b) Elucidate in detail about warehouse manager.

17 a) Explain in detail about various system managers.
(OR)
  b) Elucidate on testing the data warehouse.

18 a) Give an outline of data mining metrics with necessary theory.
(OR)
  b) Analyze about knowledge discovery in database.

19 a) Expound on hierarchical algorithms.
(OR)
  b) Illustrate on neural networks.

20 a) Give a clear analysis on parallel and distributed algorithms.
(OR)
  b) Infer on spatial clustering algorithm.

Z-Z-Z   END