B.C.A. Semester-V (Honours) Examination, 2022-23

BACHELOR OF COMPUTER APPLICATION

Course ID : 53317   Course Code : BCA-DSE-02

Course Title : Data Warehouse & Data Mining

Time : 3 Hours   Full Marks : 80

The figures in the right-hand margin indicate marks.

Candidates are required to give their answers in their own words as far as practicable.

GROUP–A

A.   Answer **all** the questions: 1×10=10

1.   Which of the following refers to the problem of finding abstracted patterns (or structures) in the unlabelled data?

   a)   Supervised learning

   b)   Unsupervised learning

   c)   Hybrid learning

   d)   Reinforcement learning

   e)   None of these

2.   Which one of the following refers to querying the unstructured textual data?

   a)   Information access

   b)   Information update

   c)   Information retrieval

   d)   Information manipulation

   e)   None of these

3.   Which of the following is an essential process in which the intelligent methods are applied to extract data patterns?

   a)   Warehousing

   b)   Data Mining

   c)   Text Mining

   d)   Data Selection

   e)   None of these

4.   For what purpose, the analysis tools pre-compute the summaries of the huge amount of data?

   a)   In order to maintain consistency

   b)   For authentication

   c)   For data access

   d)   To obtain the queries response

   e)   None of these

5. The self-organizing maps can also be considered as the instance of _____ type of learning.

   a) Supervised learning

   b) Unsupervised learning

   c) Missing data imputation

   d) Both (a) and (c)

   e) None of these

6. Identify the operation which can be performed in the data warehouse:

   a) Alter

   b) Modify

   c) Scan

   d) Read/Write

   e) None of these

7. On what is data warehouse based?

   a) 1D Model

   b) 2D Model

   c) 3D Model

   d) Multidimensional Model

   e) None of these

8. Identify the options below that a data warehouse can include:

   a) Database table

   b) Online data

   c) Flat files

   d) All of the above

   e) None of these

9. Identify the correct option which defines Datamart.

   a) a subgroup of data warehouse

   b) another type of data warehouse

   c) not related to data warehouse

   d) all of the above

   e) None of these

10. ETL stands for _____

   a) Effect, transfer and load

   b) Explain, transfer and load

   c) Extract, transfer and load

   d) Extract, transform and load

   e) None of these

GROUP-B

B. Answer any **ten** questions: 2×10=20

1. What is Knowledge Extraction?

2. What is Data Mart?

3. How are Meta Data useful in Data Mining?

4. What is the difference between Data and Information?

5. List the various forms of data preprocessing.

6. How binning can handle noisy data?

7. How Data Warehouse differ from DBMS?

8. Describe Data Normalization.

9. What is Categorical Variable?

10. What is Data Cleaning?

11. What is Business Intelligence?

12. Write Some Recommendation System Example.

13. What is Web Mining?

14. What is the Difference between Classification and Prediction?

15. What is Distributed and Parallel Data Mining?

GROUP-C

C. Answer any **four** questions: 5×4=20

1. Differentiate between OLAP vs OLTP.

2. Explain KDD Process.

3. What is Classification, Clustering, Regression Analysis? Explain with Example.

4. Which allows selection of the relevant information necessary for the data warehouse?

5. Write various application of Data Warehouse.

6. Differentiate between Supervised, Un-Supervised and Reinforcement Learning.

GROUP-D

D. Answer any **three** questions: 10×3=30

1. Explain with figure Data Warehouse Architecture Key Component.

2. Write a short notes on Hierarchical Clustering Algorithm.

3. Write the Concept Data Warehouse Architecture Model(2-tier,3-tier, 4-tier Model)

4. For the following Given Transaction Data-Set, generate Rules using Apriori Algorithm. Consider the values as Support=50% and Confidence=75%

Transaction ID	Items Purchased
1	Bread, Cheese, Egg, Juice
2	Bread, Cheese, Juice
3	Bread, Milk, Yogurt
4	Bread, Juice, Milk
5	Cheese, Juice, Milk

5.. Apply K-Mean Clustering Algorithm to Given Dataset (Take K=2)

K = {2,3,4,10,11,12,20,25,30)

How to Determine value of K?

6. For the following Medical Diagnosis Data, Create Decision Tree.

Sore Throat	Fever	Swollen Glands	Congestion	Headache	Diagnosis
Yes	Yes	Yes	Yes	Yes	Strep Threat
No	No	No	Yes	Yes	Allergy
Yes	Yes	No	Yes	No	Cold
Yes	No	Yes	No	No	Strep Threat
No	Yes	No	Yes	No	Cold
No	No	No	Yes	No	Allergy
No	No	Yes	No	No	Strep Threat
Yes	No	No	Yes	Yes	Allergy
No	Yes	No	Yes	Yes	Cold
Yes	Yes	No	Yes	Yes	Cold

_____