No. of Printed Pages—4

871 (DWD)

[20-HRGPNW–D5S(N)21A]

BCA 5th Semester (New) Exam., 2020 (W)

DATA WAREHOUSING AND DATA MINING

[GCA S5–04]

Full Marks : 80

Time : 3 hours

The figures in the margin indicate full marks for the questions

1. *(a)* Answer the following questions : 1×8=8

(i) What is data mining?

(ii) Define OLAP.

(iii) Slice operation is used normally in a database. Is it correct? Justify.

(iv) What is cross validation?

(v) Write a difference between star schema and snow flake schema.

(vi) What is data mart?

(vii) What is frequent item set?

(viii) Give some examples of binary attribute.

(b) Choose the correct option : 1×4=4

(i) Confidence can be defined using

1. one variable
2. two variables
3. three variables
4. any number of variables

(ii) A set of data divided into some groups based on some predefined classes is known as

1. grouping
2. prediction
3. classification
4. All of the above

(iii) Which of the following initially is developed for binary classification?

1. Decision tree
2. Neural network
3. K-Nearest neighbours
4. Support vector machine

(iv) To improve the classification, accuracy, which of the following techniques is suitable?

1. Neural network
2. Decision tree
3. Ensemble method
4. Support vector machine

(c) Fill in the blanks : 1×4=4

(i) The algorithm that attempts to automatically partition the data space into several regions is called _____.

(ii) The process of extracting information from the content of web document is called _____.

(iii) The forecasting for future on the basis of the part is called _____.

(iv) The set of objects which does not belong to any cluster is called as _____.

2. Answer any *five* from the following questions : 2×5=10

(a) Explain the concept of visualization of data.

(b) What are the differences between data warehouse and database?

(c) Explain roll-up operation using a suitable example.

(d) What is data reduction?

(e) Give one example of star schema.

(f) Give some examples of metadata.

3. Answer any *six* from the following questions : 4×6=24

(a) Explain frequent patterns with example.

(b) What is attribute? Explain different types of attribute.

(c) Apply apriori algorithm, considering the following market basket dataset. Assume minimum threshold is 3 (60%) :

Transaction ID	*Items bought*
T1	{Mango, Onion, Eggs, Meat, Maggi}
T2	{Car, Onion, Eggs, Fish, Maggi}
T3	{Mango, Apple, Eggs, Meat}
T4	{Mango, Onion, Eggs, Fish}
T5	{Corn, Onion, Eggs, Ice-cream}

(d) Explain at least two measures for attribute selection.

(e) Using the following table, find the probability that a randomly selected person is male, given that they own a pet?

	Have pets	*Do not have pets*	*Total*
Male	0·41	0·08	0·49
Female	0·45	0·06	0·51
Total	0·86	0·14	1

(f) Explain 10-fold validation with a suitable example.

(g) There are four types of fruit and each fruit have two attributes or features as shown in the table below. Considering *K*-means clustering make two clusters from these four fruits :

Fruits	*Attribute 1 (X) weight index*	*Attribute 2 (Y) pH value*
Apple	1	1
Mango	2	1
Orange	4	3
Papaya	5	4

4. Answer any *five* from the following questions : 6×5=30

(a) Explain categories of web data mining.

(b) Explain sequence data mining with a suitable example.

(c) *(i)* Write the benefits of IOT.

(ii) Who and why people uses data warehouses?

(d) Discuss the differences between OLTP and OLAP.

(e) How do data transformation operations contribute toward the success of the mining process?

(f) *(i)* Give one example of multidimensional data model.

(ii) List the steps of data warehouse design.

★ ★ ★