Post Graduate Diploma in Management (Hospital & Health Management)
PGDM – 2021-23 Batch

Term – 2nd Year Vth Term End Examination

Course & Code	:	**Basics of Data Mining and Data Warehousing**	**Reg. No.**	:
Term & Batch	:	**V, 2021-23**	**Date**	: **11th January 2023**
Duration	:	**3 Hrs**	**Max. Marks**	: **70**

Instructions:
- Budget your time as per the marks given for each question and write your answer accordingly.
- Don't write anything on the Question Paper except writing your Registration No.
- Mobile Phones are not allowed even for computations.

Part A: Q. 1 to Q.10 (10 questions*1 marks = 10 marks).

One liner, MCQs, True/False

Q.1 The accuracy identifies a relation between.

(1) Sum of correct positive predictions to the sum of all predictions

(2) Sum of correct predictions to the sum of all predictions

(3) Sum of correct negative predictions to the sum of all predictions

(4) None of these

Q.2 The PPDM stands for.

(1) Privacy Prediction Data Mining

(2) Privacy Protected Data Mining

(3) Privacy Provisioning Data Mining

(4) None of these

Q.3 Identify the correct option which defines Datamart.

(1) A subgroup of Data warehouse

(2) Another type of Data warehouse

(3) Not related to Data warehouse

(4) None

Q.4 State whether True or False: Data warehouse is generally updated in real-time.

(1) True

(2) False

Q.5 Where is data warehousing used?

(1) Transaction System

(2) Logical system

(3) Decision support system

(4) None

Q.6 Small logical units where data warehouses hold large amounts of data is known as _____

(1) Access layers

(2) Data marts

(3) Data storage

(4) Data miners

Q.7 Choose the incorrect property of the data warehouse.

(1) Collection from heterogeneous sources

(2) Subject Oriented

(3) Time variant

(4) Volatile

Q.8 The data is stored, retrieved, and updated in _____.

(1) SMTP

(2) FTP

(3) OLAP

(4) OLTP

Q.9 Identify the incorrect option related to the surrogate key.

(1) It is join key between fact and dimension table

(2) The primary key of the fact table

(3) It is a list of running members

(4) None

Q.10 Which of the following is not a clustering method.

(1) K nearest neighbor method

(2) K mean method

(3) Self-organizing feature map method

(4) Agglomerative method.

Part B: Q.11 to Q.15 (4 questions *5 Marks =20 Marks) Attempt any four.

Short Notes

Q.11 Explain the advantages and disadvantages of Data Mining?

Q.12 Define ETL, and the steps involved in Data Cleaning?

Q.13 Describe the steps to build the data warehouse?

Q.14 Define data privacy? List any two data privacy techniques that can be performed in data mining?

Q.15 What is a confusion matrix and what is meant for?

Part C: Q.16 to Q.20 (4 questions *10 Marks =40 Marks) Attempt any four.

Long Notes

Q.16 List the various categories in data mining algorithms, with conceptual explanation of any two algorithms?

Q.17 Define PPDM techniques and explain any two of its techniques?

Q.18 Describe the KDD and its types during data mining?

Q.19 What is Association rule mining and explain rule evaluation metrics?

Q.20 Explain Decision trees and its pros and cons of decision trees?