24-04-2023

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL/MAY 2023
Seventh Semester
Computer Science and Engineering
CS8075 – DATA WAREHOUSING AND DATA MINING
(Regulations 2017)

Duration: 3 Hours   Max. Marks: 100

PART - A (10 × 2 = 20 Marks)
Answer ALL Questions

		Marks, K-Level,CO
1.	How is data warehouse different from database? How they are similar?	2,K1,CO1
2.	List the distinct features of OLTP with OLAP.	2,K2,CO1
3.	What is Euclidean distance?	2,K2,CO2
4.	What are the major tasks done in data preprocessing?	2,K2,CO2
5.	List out the various kinds of attributes involved in data mining.	2,K2,CO3
6.	Define outlier analysis.	2,K1,CO3
7.	What is CMAR?	2,K1,CO4
8.	What is pattern evaluation?	2,K2,CO4
9.	List the phases of outlier detection method.	2,K1,CO5
10.	How data will be read from the database using Weka tool.	2,K2,CO6

PART - B (5 × 13 = 65 Marks)
Answer ALL Questions

11.	a)	What is data warehouse? Give the Steps for design and construction of Data Warehouses and explain with three tier architecture diagram.	13,K2,CO1
		OR	
	b)	(i) Explain the typical OLAP Operations with necessary diagram.	6,K2,CO1
		(ii) Analyze the information needed to support DBMS schemas for Decision support.	7,K2,CO1
12.	a)	Explain the techniques used in development of data mining methods.	13,K2,CO2
		OR	
	b)	Demonstrate in detail about data mining steps in the process of knowledge discovery.	13,K2,CO2

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create   **11804**

13. a) Design a star-schema, snow-flake schema and fact- constellation schema for the data warehouse that consists of the following four dimensions (Time, Item, Branch andLocation). Include the appropriate measures required for the schema.   *13,K2,CO3*

OR

b) (i) Explain the major issues in data mining.   *6,K2,CO3*

(ii) Explain mapping data warehouse with multiprocessor architecture with the concept of parallelism and data partitioning.   *7,K2,CO3*

14. a) Discuss the various Pattern evaluation methods and compare their measures.   *13,K2,CO4*

OR

b) (i) Discuss about mining association rules using the apriori algorithm in detail.   *6,K2,CO4*

(ii) Define Market Basket Analysis. Describe about Frequent Itemsets, Closed Item set and Association Rules.   *7,K2,CO4*

15. a) (i) What is outlier mining method? Explain the different methods of outlier detection.   *6,K2,CO5*

(ii) Explain hierarchical based method and density based method.   *7,K2,CO5*

OR

b) Illustrate k-means algorithm on the above data set. Consider 5 points as a two dimensional sample for clustering:
D1=(0,2) D2=(1,0) D3=(2,1) D4=(4,1) D5=(5,3). The required number of clusters is 2.   *13,K2,CO5*

PART - C (1 × 15 = 15 Marks)

16. a) (i) Explain the learning algorithm in Weka tool.   *8,K2,CO6*
   (ii) Illustrate the exploration of Weka explorer.   *7,K2,CO6*

OR

b) Explain the process of clustering algorithm using Weka tool.   *15,K2,CO6*