Q.P. Code: 16CS531 R16

Reg. No: | | | | | | | | | | |

SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR

(AUTONOMOUS)

B.Tech IV Year I Semester Supplementary Examinations August-2021

DATA WAREHOUSING AND DATA MINING

(Computer Science & Engineering)

Time: 3 hours Max. Marks: 60

(Answer all Five Units **5 x 12 = 60** Marks)

UNIT-I

1		Define Data mining? Explain about data mining on what kind of data?	**12M**
		OR	
2	**a**	Describe about Data discretization?	**6M**
	b	Write about Dimensionality reduction methods?	**6M**

UNIT-II

3	**a**	What are steps in designing the data warehouse? Explain	**7M**
	b	Compare OLTP and OLAP	**5M**
		OR	
4		Discuss in detail about the case study of Data Warehouse for the Government of Andhra Pradesh	**12M**

UNIT-III

5		You are given the transaction data shown in the Table below from a fast food restaurant. There are 9 distinct transactions (order:1 – order:9). There are a total of 5 meal items that are involved in the transactions. Forsimplicity we assign the meal items short names (M1 – M5) rather than thefull descriptive names (e.g., Big Mac). Meal Item List of Item IDs Meal Item List of Item IDs Order: 1 M1, M2, M5 Order: 6 M2, M3 Order: 2 M2, M4 Order: 7 M1, M3 Order: 3 M2, M3 Order: 8 M1, M2, M3, M5 Order: 4 M1, M2, M4 Order: 9 M1, M2, M3 Order: 5 M1, M3 Compute the following : a) Apply the Apriori algorithm to the dataset of transactions and identify all frequent itemsets with minimum support count as 2. Clearly explain the step by step procedure of the algorithm. b) Find all strong association rules of the form: X ΛY →Z and note their confidence values.	**12M**
		OR	
6	**a**	Write short notes of Multidimensional Association rule.	**5M**
	b	Discuss about the FP Growth Algorithm.	**7M**

UNIT-IV

7		Explain decision tree induction algorithm for classifying data tuples and discuss suitable example.	**12M**
		OR	
8	**a**	What is prediction? Explain about Linear regression method.	**6M**
	b	Discuss about Accuracy and Error measures.	**6M**

UNIT-V

9	**a**	Write K-means clustering algorithm.	**6M**
	b	Write the key issue in hierarchical clustering algorithm.	**6M**
		OR	
10	**a**	Differentiate between AGNES and DIANA algorithms.	**6M**
	b	How to access the cluster quality?	**6M**

*** END ***