PVP 20

Code: 20IT2601A

III B.Tech - II Semester – Regular / Supplementary Examinations
APRIL 2024

INTRODUCTION TO DATA MINING
(Common to All Branches)

Duration: 3 hours Max. Marks: 70

Note: 1. This paper contains questions from 5 units of Syllabus. Each unit carries 14 marks and have an internal choice of Questions.
2. All parts of Question must be answered in one place.

BL – Blooms Level CO – Course Outcome

			BL	CO	Max. Marks
		UNIT-I			
1	a)	What is data warehouse? How is it differs from DBMS.	L2	CO1	4
	b)	Describe the various phases in knowledge discovery process with a neat diagram.	L2	CO1	10
		OR			
2	a)	Define Data Mining? Explain applications of data mining.	L2	CO1	5
	b)	Explain data mining functionalities.	L2	CO1	9
		UNIT-II			
3	a)	What is data cleaning? What are the different techniques for handling missing values?	L3	CO2	7

	b)	Discuss about any two measures of similarity.	L3	CO2	7
		OR			
4		List and explain various Data reduction strategies.	L2	CO2	14
		UNIT-III			
5	a)	Can we overcome the draw backs of Apriori algorithm? Discuss.	L2	CO3	6
	b)	Explain the methods of Frequent Itemset Generation and Rule Generation.	L3	CO3	8
		OR			
6		Illustrate the frequent itemset generation using the Apriori algorithm.	L3	CO3	14
		UNIT-IV			
7		Explain decision tree induction algorithm for classifying data tuples with suitable example.	L3	CO3	14
		OR			
8	a)	Discuss about metrics for evaluating classifier performance.	L3	CO3	7
	b)	Illustrate about Rule Induction using a Sequential covering algorithm with an example.	L3	CO4	7

UNIT-V					
9		What are the advantages and disadvantages of k-means clustering against model based clustering? You are given a set of numbers {2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377}. Use the following techniques to find two clusters from this data set. (i) K-Means with initial centroids {1} and {378} (ii) K-Means with initial centroids {21} and {34}.	L3	CO3	14
OR					
10	a)	Compare K-Means with Agglomerative Hierarchical Clustering.	L3	CO4	7
	b)	Define Clustering. Explain about types of Data in Cluster Analysis.	L2	CO4	7