Code No: P18CSE04

HALL TICKET NUMBER

PACE INSTITUTE OF TECHNOLOGY & SCIENCES::ONGOLE
(AUTONOMOUS)
III B.TECH I SEMESTER END REGULAR EXAMINATIONS, DEC/JAN – 2022/23
DATA WARE HOUSING AND DATA MINING
(CSE Branch)

Time: 3 hours   Max. Marks: 60

Note: Question Paper consists of Two parts (Part-A and Part-B)

PART-A
Answer all the questions in Part-A (5X2=10M)

Q.No.		Questions	Marks	CO	KL
1.	a)	What is data mining?	[2M]	1	1
	b)	Define Data discretization.	[2M]	2	1
	c)	Define Support.	[2M]	3	1
	d)	What is clustering?	[2M]	4	1
	e)	What is web mining?	[2M]	5	1

PART-B
Answer One Question from each UNIT (5X10=50M)

Q.No.		Questions	Marks	CO	KL
		UNIT-I			
2.	a)	Explain the difference and similarity between characterization and clustering, and between classification and regression.	[5M]	1	2
	b)	Explain the major issues in data mining.	[5M]	1	2
		OR			
3.	a)	Explain the key features of data ware house.	[5M]	1	2
	b)	Explain the three-tier data warehousing architecture with neat sketch	[5M]	1	2
		UNIT-II			
4.	a)	Why do we preprocess the data? Explain.	[5M]	2	2
	b)	Explain the data transformation methods.	[5M]	2	2
		OR			
5.		Explain Data reduction in detail.	[10M]	2	2
		UNIT-III			
6.	a)	How is association rules generated from frequent itemsets? Explain.	[5M]	3	2
	b)	Explain the limitations of apriori algorithm.	[5M]	3	2
		OR			

7.		Find the frequent itemsets and strong association rules for the following transactional database table using Apriori algorithm. Consider the thresholds as support = 30% and confidence = 40%.	[10M]	3	3
		<table><tr><td>TID</td><td>ITEM IDs</td></tr><tr><td>1</td><td>I1,i2,i3,i5</td></tr><tr><td>2</td><td>I2,i5,i7,i9</td></tr><tr><td>3</td><td>I1,i3,i5,i7</td></tr><tr><td>4</td><td>I2,i4,i6,i8</td></tr><tr><td>5</td><td>I1,i2,i3,i4</td></tr><tr><td>6</td><td>I2,i3,i4,i5</td></tr><tr><td>7</td><td>I3,i4,i5,i6</td></tr><tr><td>8</td><td>I4,i5,i6,i7</td></tr><tr><td>9</td><td>I5,i6,i7.i8.i9</td></tr><tr><td>10</td><td>I9.i1.i2.i5</td></tr><tr><td>11</td><td>I8,i2,i9,i7</td></tr><tr><td>12</td><td>I5,i6,i3,i2</td></tr></table>			
		UNIT-IV			
8.		Explain the basic K means algorithm with iterations.	[10M]	4	2
		OR			
9.	a)	Explain the DBSCAN algorithm.	[5M]	4	2
	b)	Discuss about key issues in Hierarchical clustering.	[5M]	4	2
		UNIT-V			
10.	a)	Discuss about Regression and Trend Analysis in Time-Series Data.	[5M]	5	2
	b)	Explain about Sequential Pattern Mining in Symbolic Sequences.	[5M]	5	2
		OR			
11.	a)	Discuss about Graph Pattern Mining.	[5M]	5	2
	b)	Compare and Contrast of Clustering, Ranking, and Classification of homogeneous and Heterogeneous Networks.	[5M]	5	3

******