

Code No: P18CSE04

HALL TICKET NUMBER

--	--	--	--	--	--	--	--	--	--



PACE INSTITUTE OF TECHNOLOGY & SCIENCES::ONGOLE  
(AUTONOMOUS)

III B.TECH I SEMESTER END SUPPLEMENTARY EXAMINATIONS, MARCH/APRIL-2023  
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) How is a data warehouse different from a database	[2M]	1	1
	b) Define data quality	[2M]	2	1
	c) What is confidence	[2M]	3	1
	d) List the limitations of hierarchical clustering	[2M]	4	1
	e) What is Ubiquitous data 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 discrimination and classification	[5M]	1	2
	b) Explain the major challenges of mining a huge amount of data in comparison with mining a small amount of data	[5M]	1	2
OR				
3.	a) Compare and Contrast Operational Database Systems and Data Warehouses	[5M]	1	2
	b) Explain the three-tier data warehousing architecture with neat sketch	[5M]	1	2
UNIT-II				
4.	a) Explain in brief about Data cleaning	[5M]	2	2
	b) Explain Major Tasks in Data Pre-processing	[5M]	2	2
OR				
5.	a) Describe about Data discretization	[5M]	2	2
	b) Explain about Dimensionality reduction methods	[5M]	2	2
UNIT-III				
6.	a) Explain Rule Generation in Apriori Algorithm	[5M]	3	2
	b) Explain the maximal frequent itemset with example	[5M]	3	2
OR				



7.		Apply FP-Growth algorithm to the following transactional data to find frequent itemsets. List all frequent itemsets with their support count	[10M]	3	3																						
		<table border="1"> <thead> <tr> <th>TID</th> <th>List of Item IDs</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>I1,i3,i5,i7</td> </tr> <tr> <td>2</td> <td>I2,i4,i6,i8</td> </tr> <tr> <td>3</td> <td>I1,i3,i5,i7</td> </tr> <tr> <td>4</td> <td>I9,i7,i5,i1</td> </tr> <tr> <td>5</td> <td>I2,i4,i6,i7</td> </tr> <tr> <td>6</td> <td>I1,i2,i3,i4</td> </tr> <tr> <td>7</td> <td>I3,i4,i5,i6</td> </tr> <tr> <td>8</td> <td>I7,i8,i6,i1</td> </tr> <tr> <td>9</td> <td>I8,i5,i3,i2</td> </tr> <tr> <td>10</td> <td>I1,i3,i4,i6</td> </tr> </tbody> </table>	TID	List of Item IDs	1	I1,i3,i5,i7	2	I2,i4,i6,i8	3	I1,i3,i5,i7	4	I9,i7,i5,i1	5	I2,i4,i6,i7	6	I1,i2,i3,i4	7	I3,i4,i5,i6	8	I7,i8,i6,i1	9	I8,i5,i3,i2	10	I1,i3,i4,i6			
TID	List of Item IDs																										
1	I1,i3,i5,i7																										
2	I2,i4,i6,i8																										
3	I1,i3,i5,i7																										
4	I9,i7,i5,i1																										
5	I2,i4,i6,i7																										
6	I1,i2,i3,i4																										
7	I3,i4,i5,i6																										
8	I7,i8,i6,i1																										
9	I8,i5,i3,i2																										
10	I1,i3,i4,i6																										
UNIT-IV																											
8.		Explain different types of Clusters	[10M]	4	2																						
OR																											
9.	a)	Explain the DBSCAN algorithm.	[5M]	4	2																						
	b)	Explain the Bisecting K-means algorithm with an example of four clusters	[5M]	4	2																						
UNIT-V																											
10.	a)	Explain the Similarity Search in Time-Series Data	[5M]	5	2																						
	b)	Explain the Hidden Markov Model for Biological Sequence Analysis	[5M]	5	2																						
OR																											
11.	a)	Explain about Statistical Modeling of Networks	[5M]	5	2																						
	b)	Discuss about web mining	[5M]	5	6																						

\*\*\*\*\*