

Nov – Dec 2022 (B.Tech) Program: Computer Engg. Examination: TY Semester: V Course Code: CEC504 and Course Name: DWM				
Duration: 2.5 Hours		Max. Marks: 60		
<b>Instructions:</b> (1) All questions are compulsory. (2) Draw neat diagrams wherever applicable. (3) Assume suitable data, if necessary.				
		Max. Marks	CO	BT level
<b>Q 1</b>	<b>Solve any six questions out of eight:</b>	12		
i)	Differentiate between ER Modeling and Dimensional Modeling	02	CO1	U
ii)	Explain Data Cleaning process in Data Warehouse	02	CO2	U
iii)	What is Data Discretization?	02	CO2	U
iv)	How FP growth tree overcome the drawback of Apriori algorithm. Justify.	02	CO3, CO4	U
v)	What is Web Content Mining?	02	CO5	U
vi)	Compare Agglomerative and Divisive Hierarchical Clustering.	02	CO3, CO4	U
vii)	What is Data Mart? Give one example	02	CO1	U
viii)	Differentiate between Supervised and Unsupervised Learning	02	CO3, CO4	U
<b>Q.2</b>	<b>Solve any four questions out of six.</b>	16		
i)	For the given attribute AGE values: 16,16,180,4,12,24,26,28, apply following binning techniques for smoothing the noise. (i) Bin Boundaries (ii) Bin Means	04	CO2	Ap
ii)	What is OLAP? Suppose college wants to record the grades for the courses completed by students. There are four dimensions: Course, Professor, Student and Period. On the above scenario perform: (i) Roll up (ii) Dice operation	04	CO1	Ap
iii)	How to evaluate accuracy of the classifier using Holdout and Cross validation methods.	04	CO3, CO4	U
iv)	Using K-means clustering algorithm form two clusters for the given set of values and show each step. Data: {2,4,12, 20, 30, 11, 6, 3, 25, 10}	04	CO3, CO4	Ap

v)	Explain in brief Multilevel association rule with suitable example.	04	CO3, CO4	U																																				
vi)	What do you mean by Hubs and Authorities. Explain HITS algorithm with example	04	CO5	U																																				
<b>Q.3</b>	<b>Solve any two questions out of three.</b>	16																																						
i)	Draw and explain Data Mining architecture in detail.	08	CO2	U																																				
ii)	For the following given Transaction Data-set, Generate Rules using Apriori Algorithm. Consider the values as Support=50% and Confidence=75%	08	CO3, CO4	Ap																																				
<table border="1"> <thead> <tr> <th>Transaction Id</th> <th>Items purchased</th> </tr> </thead> <tbody> <tr> <td>T001</td> <td>Bread, Cheese, Egg, Juice</td> </tr> <tr> <td>T002</td> <td>Bread, Cheese, Juice</td> </tr> <tr> <td>T003</td> <td>Bread, Milk, Yogurt</td> </tr> <tr> <td>T004</td> <td>Bread, Juice, Milk</td> </tr> <tr> <td>T005</td> <td>Cheese, Juice, Milk</td> </tr> </tbody> </table>		Transaction Id	Items purchased	T001	Bread, Cheese, Egg, Juice	T002	Bread, Cheese, Juice	T003	Bread, Milk, Yogurt	T004	Bread, Juice, Milk	T005	Cheese, Juice, Milk																											
Transaction Id	Items purchased																																							
T001	Bread, Cheese, Egg, Juice																																							
T002	Bread, Cheese, Juice																																							
T003	Bread, Milk, Yogurt																																							
T004	Bread, Juice, Milk																																							
T005	Cheese, Juice, Milk																																							
iii)	Plot a dendrogram using Complete linkage agglomerative clustering for the given dataset	08	CO3, CO4	Ap																																				
<table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>0</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td>9</td> <td>0</td> <td></td> <td></td> <td></td> </tr> <tr> <th>C</th> <td>3</td> <td>7</td> <td>0</td> <td></td> <td></td> </tr> <tr> <th>D</th> <td>6</td> <td>5</td> <td>9</td> <td>0</td> <td></td> </tr> <tr> <th>E</th> <td>11</td> <td>10</td> <td>2</td> <td>8</td> <td>0</td> </tr> </tbody> </table>			A	B	C	D	E	A	0					B	9	0				C	3	7	0			D	6	5	9	0		E	11	10	2	8	0			
	A	B	C	D	E																																			
A	0																																							
B	9	0																																						
C	3	7	0																																					
D	6	5	9	0																																				
E	11	10	2	8	0																																			
<b>Q.4</b>	<b>Solve any two questions out of three.</b>	16																																						
i)	Elaborate following concepts with suitable example: a. Crawlers b. PageRank algorithm	08	CO5	U																																				

ii)	Apply ID3 Decision tree algorithm to generate tree for following training dataset upto 1st iteration	08	CO3, CO4	Ap																																																																											
<table border="1"> <thead> <tr> <th>Sr No</th> <th>Colour</th> <th>Outline</th> <th>Dot</th> <th>Shape</th> </tr> </thead> <tbody> <tr><td>1</td><td>Green</td><td>Dashed</td><td>No</td><td>Triangle</td></tr> <tr><td>2</td><td>Green</td><td>Dashed</td><td>Yes</td><td>Triangle</td></tr> <tr><td>3</td><td>Yellow</td><td>Dashed</td><td>No</td><td>Square</td></tr> <tr><td>4</td><td>Red</td><td>Dashed</td><td>No</td><td>Square</td></tr> <tr><td>5</td><td>Red</td><td>Solid</td><td>No</td><td>Square</td></tr> <tr><td>6</td><td>Red</td><td>Solid</td><td>Yes</td><td>Triangle</td></tr> <tr><td>7</td><td>Green</td><td>Solid</td><td>No</td><td>Square</td></tr> <tr><td>8</td><td>Green</td><td>Dashed</td><td>No</td><td>Triangle</td></tr> <tr><td>9</td><td>Yellow</td><td>Solid</td><td>Yes</td><td>Square</td></tr> <tr><td>10</td><td>Red</td><td>Solid</td><td>No</td><td>Square</td></tr> <tr><td>11</td><td>Green</td><td>Solid</td><td>Yes</td><td>Square</td></tr> <tr><td>12</td><td>Yellow</td><td>Dashed</td><td>Yes</td><td>Square</td></tr> <tr><td>13</td><td>Yellow</td><td>Solid</td><td>No</td><td>Square</td></tr> <tr><td>14</td><td>Red</td><td>Dashed</td><td>Yes</td><td>Triangle</td></tr> </tbody> </table>					Sr No	Colour	Outline	Dot	Shape	1	Green	Dashed	No	Triangle	2	Green	Dashed	Yes	Triangle	3	Yellow	Dashed	No	Square	4	Red	Dashed	No	Square	5	Red	Solid	No	Square	6	Red	Solid	Yes	Triangle	7	Green	Solid	No	Square	8	Green	Dashed	No	Triangle	9	Yellow	Solid	Yes	Square	10	Red	Solid	No	Square	11	Green	Solid	Yes	Square	12	Yellow	Dashed	Yes	Square	13	Yellow	Solid	No	Square	14	Red	Dashed	Yes	Triangle
Sr No	Colour	Outline	Dot	Shape																																																																											
1	Green	Dashed	No	Triangle																																																																											
2	Green	Dashed	Yes	Triangle																																																																											
3	Yellow	Dashed	No	Square																																																																											
4	Red	Dashed	No	Square																																																																											
5	Red	Solid	No	Square																																																																											
6	Red	Solid	Yes	Triangle																																																																											
7	Green	Solid	No	Square																																																																											
8	Green	Dashed	No	Triangle																																																																											
9	Yellow	Solid	Yes	Square																																																																											
10	Red	Solid	No	Square																																																																											
11	Green	Solid	Yes	Square																																																																											
12	Yellow	Dashed	Yes	Square																																																																											
13	Yellow	Solid	No	Square																																																																											
14	Red	Dashed	Yes	Triangle																																																																											
iii)	<p>For wholesale furniture company. Analyse with respect to Furniture (type, category, material), Customer (location like cities, regions and states) and Time. The company is interested in learning the quantity, income and discount of its sales.</p> <p>i) Draw a star schema for the data warehouse</p> <p>ii) Draw a snowflake schema for the data warehouse</p>	08	CO1	Ap																																																																											

\*\*\*\*\*