K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov –Dec 2023-24
Program: B.Tech Scheme : II
Examination: TY Semester: V
Course Code: AIC502 and Course Name: Data Warehousing and Mining
Date of Exam: 30/11/2023 Duration: 2.5 Hours Max. Marks: 60

Instructions:
(1)All questions are compulsory.
(2)Draw neat diagrams wherever applicable.
(3)Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q 1	**Solve any six questions out of eight:**	**12**		
i)	Construct a box plot for the following data. 12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25	2	CO4	Ap
ii)	Justify the need for data quality.	2	CO2	An
iii)	Document database features.	2	CO1	U
iv)	Brief about fact table used in schema	2	CO3	U
v)	List different types of data used in cluster analysis.	2	CO5, CO6	U
vi)	Manhattan distance measurement.	2	CO5, CO6	U
vii)	Write formulas for the precision, F1 score.	2	CO5, CO6	U
viii)	Calculate the mean and median? Data riverside bowling score: [104,117,104,136, 189,109,113,104]	2	CO4	Ap
Q.2	**Solve any four questions out of six.**	**16**		
i)	Describe the concept of graphical representation in data analysis. How do different types of graphs, such as histograms, box plots, and line charts, aid in conveying information about the distribution and relationships within a dataset?	4	CO4	U
ii)	Justify the need of Strategic information.	4	CO1	U
iii)	Compare E-R Modeling Vs Dimensional Modeling	4	CO2	An
iv)	Justify the need of hierarchy in OLAP? Compare OLTP and OLAP.	4	CO3	An

<table>
<tr><td>v)</td><td>Implement Apriori Algorithm and write strong association rules using support 50% and confidence 70%.
<table><tr><th>TID</th><th>Items</th></tr><tr><td>100</td><td>1, 3, 4</td></tr><tr><td>200</td><td>2, 3, 5</td></tr><tr><td>300</td><td>1, 2, 3, 5</td></tr><tr><td>400</td><td>2, 5</td></tr></table></td><td>4</td><td>CO5, CO6</td><td>U</td></tr>
<tr><td>vi)</td><td>What are different ensemble methods? Explain in details any 2.</td><td>4</td><td>CO5, CO6</td><td>U</td></tr>
<tr><td>Q.3</td><td>Solve any two questions out of three.</td><td>16</td><td></td><td></td></tr>
<tr><td>i)</td><td>Construct FP tree with minimum support 3 and write frequent patterns generated.
<table><tr><th>TID</th><th>Items bought</th><th>Ordered frequent Items</th></tr><tr><td>100</td><td>{f, c, a, d, g, i, m, p}</td><td>{f, c, a, m, p}</td></tr><tr><td>200</td><td>{a, b, c, f, l, m, o}</td><td>{ f, c, a, b, m}</td></tr><tr><td>300</td><td>{b, f, h, j, o, w}</td><td>{f, b }</td></tr><tr><td>400</td><td>{b, c, k, s, p}</td><td>{c, b, p }</td></tr><tr><td>500</td><td>{a, f, c, e, l, p, m, n}</td><td>{f, c, a, m, p}</td></tr></table></td><td>8</td><td>CO5, CO6</td><td>Ap</td></tr>
<tr><td>ii)</td><td>Explain the main components of the ETL process (Extract, Transform, Load) and their individual roles in preparing data for analysis.</td><td>8</td><td>CO3</td><td>U</td></tr>
<tr><td>iii)</td><td>Write in detail about unstructured data</td><td>8</td><td>CO1</td><td>U</td></tr>
<tr><td>Q.4</td><td>Solve any two questions out of three.</td><td>16</td><td></td><td></td></tr>
<tr><td>i)</td><td>Design a data cube schema for Amul company sales analysis system. The system should be able to track sales data based on product categories, geographical regions, and time periods. Include the necessary dimensions, hierarchies, and measures to support common OLAP operations such as roll-up, drill-down, slice-and-dice, and pivot.</td><td>8</td><td>CO3</td><td>Ap</td></tr>
<tr><td>ii)</td><td>Describe the Knowledge Discovery in Databases (KDD) process. What are the key stages involved, and why is each stage important in data mining?</td><td>8</td><td>CO4</td><td>U</td></tr>
<tr><td>iii)</td><td>Use the following database. Follow the single linkage technique to find the clusters in database. Use the Euclidian distance measure.
<table><tr><td></td><td>X</td><td>Y</td></tr><tr><td>P1</td><td>0.40</td><td>0.53</td></tr><tr><td>P2</td><td>0.22</td><td>0.38</td></tr><tr><td>P3</td><td>0.35</td><td>0.32</td></tr><tr><td>P4</td><td>0.26</td><td>0.19</td></tr><tr><td>P5</td><td>0.08</td><td>0.41</td></tr><tr><td>P6</td><td>0.45</td><td>0.30</td></tr></table></td><td>8</td><td>CO5, CO6</td><td>Ap</td></tr>
</table>

****