K. J. Somaiya Institute of Technology, Sion, Mumbai
(An Autonomous Institute Affiliated to the University of Mumbai)

End Semester Examination: November - December 2025
Program: B.Tech. (Information Technology) Scheme: III
Regular Examination: TY Semester: V
Course Code: Data Mining for Artificial Intelligence Course Name: ITC502
Date of Exam: 26/11/2025 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.

<table>
<tr><th>Ques. No.</th><th>Question</th><th>Max. Marks</th><th>CO</th><th>BT Level</th></tr>
<tr><td>Q1</td><td>Solve <u>any two</u> questions out of three: (05 marks each)</td><td rowspan="4">10</td><td></td><td></td></tr>
<tr><td>a)</td><td>Using the Advertising–Sales prediction dataset given below, perform prediction analysis to determine the relationship between Advertising and Sales. Also estimate Sales value when Advertising expenditure is 22 units.
<table>
<tr><th>Advertising</th><th>Sales</th></tr>
<tr><td>5</td><td>12</td></tr>
<tr><td>10</td><td>18</td></tr>
<tr><td>15</td><td>25</td></tr>
<tr><td>20</td><td>30</td></tr>
<tr><td>25</td><td>35</td></tr>
<tr><td>30</td><td>45</td></tr>
</table></td><td>3</td><td>A</td></tr>
<tr><td>b)</td><td>Differentiate supervised learning and unsupervised learning by explaining how each technique can be applied to solve problems in the healthcare domain, such as disease prediction and patient grouping.</td><td>4</td><td>A</td></tr>
<tr><td>c)</td><td>Explain how OLAP operations such as roll-up, drill-down, slice, and dice can be applied to analyze healthcare patient data for effective clinical decision-making.</td><td>1</td><td>U</td></tr>
<tr><td>Q2</td><td>Solve <u>any two</u> questions out of three: (05 marks each)</td><td rowspan="4">10</td><td></td><td></td></tr>
<tr><td>a)</td><td>A telecom company is analyzing the call-drop counts (number of call failures recorded each day) from one of its busy network towers. To remove noise and detect service quality trends, the company uses binning-based smoothing methods. The following dataset contains call-drop counts for 24 days: 5, 7, 6, 25, 28, 26, 9, 12, 11, 50, 52, 48, 8, 6, 7, 30, 29, 27, 13, 10, 12, 55, 57, 54. Using the above dataset, divide the data into 4 equal-frequency bins, perform Smoothing by Bin Means, by Bin Median, and by Bin Boundaries.</td><td>2</td><td>A</td></tr>
<tr><td>b)</td><td>A national retail chain wants to implement Big Data Analytics to understand customer purchasing behavior and optimize inventory management. Explain the 5Vs of big data the company must consider while designing this solution.</td><td>6</td><td>A</td></tr>
<tr><td>c)</td><td>Apply multilevel association rule mining on suitable dataset.</td><td>5</td><td>A</td></tr>
<tr><td>Q3</td><td>Solve <u>any two</u> questions out of three. (10 marks each)</td><td rowspan="2">20</td><td></td><td></td></tr>
<tr><td>a)</td><td>Sketch and explain steps involved in Knowledge Discovery from Data (KDD).</td><td>1</td><td>U</td></tr>
</table>

K. J. Somaiya Institute of Technology, Sion, Mumbai
(An Autonomous Institute Affiliated to the University of Mumbai)

End Semester Examination: November - December 2025
Program: B.Tech. (Information Technology) Scheme: III
Regular Examination: TY Semester: V
Course Code: Data Mining for Artificial Intelligence Course Name: ITC502
Date of Exam: 26/11/2025 Duration: 2.5 Hours Max. Marks: 60

b)	Suppose the data mining task is to cluster the coordinates of emergency medical units for efficient disaster-response planning. The unit locations are: A1(4, 14), A2(6, 8), A3(11, 9), B1(7, 13), B2(10, 10), B3(8, 6), C1(3, 4), C2(6, 12). Each point represents a unit's location (x, y). Initially, take A1, B1, and C1 as the cluster centers. Using Euclidean distance, apply k-means algorithm to assign clusters and compute the final cluster centers.		4	A
c)	Consider a banking service usage dataset in the form <CustomerID: ServicesUsed> as: <C1: MobileBanking, ATM, FundTransfer>, <C2: ATM, CreditCard>, <C3: MobileBanking, FundTransfer>, <C4: ATM, FundTransfer>, <C5: MobileBanking, CreditCard, FundTransfer>. If the minimum support is 20% and the minimum confidence is 70%, apply the Apriori algorithm to find frequent service combinations and generate strong association rules.		5	A
Q4	**Solve <u>any two</u> questions out of three. (10 marks each)**	20		
a)	A hospital wants to study whether the type of chest disease is associated with a patient's smoking habit. Using following dataset collected from 60 patients. apply Chi-Square Test of Independence to determine whether "disease type" is associated with "smoking habit".		2	A
b)	Apply the Decision Tree algorithm to compute Information Gain and select the best root node.		3	A
c)	Apply Knowledge Discovery from Data (KDD) process on a smart city transportation dataset for traffic pattern analysis.		6	A

Disease Type	Smoker	Non-Smoker	Total
Pneumonia	12	8	20
Tuberculosis	15	5	20
Normal	3	17	20
Total	**30**	**30**	**60**

AgeGroup	ActivityLevel	DietType	ProgramFollow
Young	Low	Basic	No
Adult	Medium	Standard	No
Senior	Medium	Standard	Yes
Young	High	Advanced	Yes
Adult	Low	Basic	Yes
Senior	High	Advanced	Yes
Young	Medium	Standard	No
Adult	Medium	Basic	No
Senior	Low	Advanced	Yes

****