

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

Supplementary

End Semester Examination: ~~November-December~~ 2025 *Jan/Feb*
 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~~ *30/01/26* 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.

Ques. No.	Question	Max. Marks	CO	BT Level																	
Q1	Solve <u>any two</u> questions out of three: (05 marks each)	10																			
a)	Explain Data Mining and Data Warehousing with examples.		1	U																	
b)	Differentiate classification and prediction by analyzing their use in solving a business or scientific problem.		3	A																	
c)	Calculate the support of the itemset {Smart TV, Streaming Device} given it appears in 4 out of 10 transactions. Also calculate the confidence of the rule {Smart TV} ⇒ {Streaming Device} given Smart TV appears in 5 transactions and both Smart TV and Streaming Device appear together in 3 transactions.	5		A																	
Q2	Solve <u>any two</u> questions out of three: (05 marks each)	10																			
a)	Consider the given data of daily maximum temperature (°C) recorded over 11 days in a city: [34, 36, 32, 35, 33, 37, 39, 36, 38, 35, 50]. Sketch a boxplot for the data. Identify the minimum, Q1, median, Q3, maximum, and any outliers.		2	A																	
b)	Differentiate Agglomerative and Divisive clustering.		4	A																	
c)	A multi-specialty hospital wants to implement Big Data Analytics to improve patient care and predict disease outbreaks. Explain the 5Vs of big data the hospital must consider while designing this solution.	6		A																	
Q3	Solve <u>any two</u> questions out of three. (10 marks each)	20																			
a)	Sketch and explain steps involved in Knowledge Discovery from Data (KDD).		1	U																	
b)	Apply Hierarchical Clustering using single-link distance for the following GPS coordinates of wildlife sensors in a forest monitoring application and sketch the Dendrogram tree: <table border="1" style="margin-left: 40px;"> <thead> <tr> <th>Sensor ID</th> <th>Longitude (X)</th> <th>Latitude (Y)</th> </tr> </thead> <tbody> <tr><td>S1</td><td>10</td><td>12</td></tr> <tr><td>S2</td><td>11.5</td><td>13</td></tr> <tr><td>S3</td><td>20</td><td>25</td></tr> <tr><td>S4</td><td>17</td><td>21</td></tr> <tr><td>S5</td><td>14</td><td>20</td></tr> </tbody> </table>	Sensor ID	Longitude (X)	Latitude (Y)	S1	10	12	S2	11.5	13	S3	20	25	S4	17	21	S5	14	20	4	A
Sensor ID	Longitude (X)	Latitude (Y)																			
S1	10	12																			
S2	11.5	13																			
S3	20	25																			
S4	17	21																			
S5	14	20																			

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

Supplementary End Semester Examination: ~~November–December 2025~~ *Jan/Feb.*
 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: ~~25/11/2025~~ *30/01/26* Duration: 2.5 Hours Max. Marks: 60

c)	<p>A survey was conducted to determine whether the presence of live music at a restaurant influences the ratings given by customers. Apply the Chi-square test to analyze if attributes <i>Live_Music</i> and <i>Customer_Ratings</i> are correlated:</p> <table border="1" data-bbox="408 595 1066 792"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="2"><i>Customer_Ratings</i></th> <th rowspan="2">Total</th> </tr> <tr> <th>Good</th> <th>Average</th> </tr> </thead> <tbody> <tr> <th rowspan="2"><i>Live_Music</i></th> <th>Yes</th> <td>180</td> <td>420</td> <td>600</td> </tr> <tr> <th>No</th> <td>120</td> <td>980</td> <td>1100</td> </tr> <tr> <th colspan="2">Total</th> <td>300</td> <td>1400</td> <td>1700</td> </tr> </tbody> </table> <p>Consider that for 1 degrees of freedom, the chi-square value to reject the hypothesis at 0.001 significance level is 10.828.</p>			<i>Customer_Ratings</i>		Total	Good	Average	<i>Live_Music</i>	Yes	180	420	600	No	120	980	1100	Total		300	1400	1700	2	A											
				<i>Customer_Ratings</i>			Total																												
		Good	Average																																
<i>Live_Music</i>	Yes	180	420	600																															
	No	120	980	1100																															
Total		300	1400	1700																															
Q4	Solve <u>any two</u> questions out of three. (10 marks each)																																		
a)	<p>Apply Naïve Bayes classifier on the following hypothetical data to predict the class of a new tuple (Female, Short, Intermediate). Consider Selection for Tournament as the class label.</p> <table border="1" data-bbox="368 1055 1150 1379"> <thead> <tr> <th>Gender</th> <th>Height</th> <th>Training Level</th> <th>Selection for Tournament</th> </tr> </thead> <tbody> <tr><td>Male</td><td>Short</td><td>Elementary</td><td>No</td></tr> <tr><td>Male</td><td>Medium</td><td>Intermediate</td><td>No</td></tr> <tr><td>Female</td><td>Tall</td><td>Intermediate</td><td>Yes</td></tr> <tr><td>Male</td><td>Short</td><td>Complete</td><td>Yes</td></tr> <tr><td>Female</td><td>Medium</td><td>Elementary</td><td>Yes</td></tr> <tr><td>Male</td><td>Tall</td><td>Complete</td><td>Yes</td></tr> <tr><td>Female</td><td>Short</td><td>Intermediate</td><td>No</td></tr> </tbody> </table>	Gender	Height	Training Level	Selection for Tournament	Male	Short	Elementary	No	Male	Medium	Intermediate	No	Female	Tall	Intermediate	Yes	Male	Short	Complete	Yes	Female	Medium	Elementary	Yes	Male	Tall	Complete	Yes	Female	Short	Intermediate	No	20	3 A
Gender	Height	Training Level	Selection for Tournament																																
Male	Short	Elementary	No																																
Male	Medium	Intermediate	No																																
Female	Tall	Intermediate	Yes																																
Male	Short	Complete	Yes																																
Female	Medium	Elementary	Yes																																
Male	Tall	Complete	Yes																																
Female	Short	Intermediate	No																																
b)	Apply the first two steps of the Knowledge Discovery from Data (KDD) process on a financial dataset, such as transaction logs for fraud detection.	6	A																																
c)	<p>Consider a transactional dataset:</p> <table border="1" data-bbox="576 1536 919 1760"> <thead> <tr> <th>TID</th> <th>Items</th> </tr> </thead> <tbody> <tr><td>T1</td><td>I1, I2, I4</td></tr> <tr><td>T2</td><td>I2, I3, I4</td></tr> <tr><td>T3</td><td>I1, I2</td></tr> <tr><td>T4</td><td>I2, I4</td></tr> <tr><td>T5</td><td>I1, I2, I3</td></tr> </tbody> </table> <p>If minimum support = 30% and minimum confidence = 75%, apply Apriori algorithm to find all possible strong association rules.</p>	TID	Items	T1	I1, I2, I4	T2	I2, I3, I4	T3	I1, I2	T4	I2, I4	T5	I1, I2, I3	5	A																				
TID	Items																																		
T1	I1, I2, I4																																		
T2	I2, I3, I4																																		
T3	I1, I2																																		
T4	I2, I4																																		
T5	I1, I2, I3																																		
