

2

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

Nov-Dec 2024-25		
Program: B.Tech	Scheme : II B	
Examination: TY Semester: V		
Course Code: AIC502 and	Course Name: Data Warehousing and Mining	
Date of Exam: 29/11/2024	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.

Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any two questions out of three: (05 marks each)	10		
a)	Calculate the mean and mode? Data riverside bowling score: [104,117,104,136, 189,109,113,104] Construct a box plot for the following data. 9,3,10,2,6,8,3,11,14,3,4,8,9,12,5		CO4	Ap
b)	Write short note on K-Medoids clustering		CO5	U
c)	Brief about the importance of unstructured data.		CO1	U
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	Explain the architecture of a typical data mining system. What are the main components, and how do they interact to facilitate data analysis?		CO3	U
b)	Draw the sketch for the general approach for building the classification model.		CO4	U
c)	Write short note on : E-R Modeling		CO2	U
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	Design a data cube schema for 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.		CO2	Ap
b)	Use Apriori algorithm, S=33.34% and C= 60%. Find and list all strong		CO6	Ap

	association rules.																																																																						
	<table><tr><td>Tid</td><td>Items</td></tr><tr><td>T1</td><td>Hotdogs, Buns, Ketchup</td></tr><tr><td>T2</td><td>Hotdogs, Buns,</td></tr><tr><td>T3</td><td>Hotdogs, Coke, Chips</td></tr><tr><td>T4</td><td>Chips, Coke</td></tr><tr><td>T5</td><td>Chips, Ketchup,</td></tr><tr><td>T6</td><td>Hotdogs, Coke, Chips</td></tr></table>					Tid	Items	T1	Hotdogs, Buns, Ketchup	T2	Hotdogs, Buns,	T3	Hotdogs, Coke, Chips	T4	Chips, Coke	T5	Chips, Ketchup,	T6	Hotdogs, Coke, Chips																																																				
Tid	Items																																																																						
T1	Hotdogs, Buns, Ketchup																																																																						
T2	Hotdogs, Buns,																																																																						
T3	Hotdogs, Coke, Chips																																																																						
T4	Chips, Coke																																																																						
T5	Chips, Ketchup,																																																																						
T6	Hotdogs, Coke, Chips																																																																						
c)	<table><tr><td>Example No.</td><td>Color</td><td>Type</td><td>Origin</td><td>Stolen?</td></tr><tr><td>1</td><td>Red</td><td>Sports</td><td>Domestic</td><td>Yes</td></tr><tr><td>2</td><td>Red</td><td>Sports</td><td>Domestic</td><td>No</td></tr><tr><td>3</td><td>Red</td><td>Sports</td><td>Domestic</td><td>Yes</td></tr><tr><td>4</td><td>Yellow</td><td>Sports</td><td>Domestic</td><td>No</td></tr><tr><td>5</td><td>Yellow</td><td>Sports</td><td>Imported</td><td>Yes</td></tr><tr><td>6</td><td>Yellow</td><td>SUV</td><td>Imported</td><td>No</td></tr><tr><td>7</td><td>Yellow</td><td>SUV</td><td>Imported</td><td>Yes</td></tr><tr><td>8</td><td>Yellow</td><td>SUV</td><td>Domestic</td><td>No</td></tr><tr><td>9</td><td>Red</td><td>SUV</td><td>Imported</td><td>No</td></tr><tr><td>10</td><td>Red</td><td>Sports</td><td>Imported</td><td>Yes</td></tr></table> <p>Classify using Naive Bayesian whether car will be stolen for following set of featuers.</p> <table><tr><td>Colour</td><td>Type</td><td>Origin</td><td>Stolen</td></tr><tr><td>Red</td><td>SUV</td><td>Domestic</td><td>?</td></tr></table>					Example No.	Color	Type	Origin	Stolen?	1	Red	Sports	Domestic	Yes	2	Red	Sports	Domestic	No	3	Red	Sports	Domestic	Yes	4	Yellow	Sports	Domestic	No	5	Yellow	Sports	Imported	Yes	6	Yellow	SUV	Imported	No	7	Yellow	SUV	Imported	Yes	8	Yellow	SUV	Domestic	No	9	Red	SUV	Imported	No	10	Red	Sports	Imported	Yes	Colour	Type	Origin	Stolen	Red	SUV	Domestic	?	CO4	Ap	
Example No.	Color	Type	Origin	Stolen?																																																																			
1	Red	Sports	Domestic	Yes																																																																			
2	Red	Sports	Domestic	No																																																																			
3	Red	Sports	Domestic	Yes																																																																			
4	Yellow	Sports	Domestic	No																																																																			
5	Yellow	Sports	Imported	Yes																																																																			
6	Yellow	SUV	Imported	No																																																																			
7	Yellow	SUV	Imported	Yes																																																																			
8	Yellow	SUV	Domestic	No																																																																			
9	Red	SUV	Imported	No																																																																			
10	Red	Sports	Imported	Yes																																																																			
Colour	Type	Origin	Stolen																																																																				
Red	SUV	Domestic	?																																																																				
Q.4	Solve any two questions out of three. (10 marks each)					20																																																																	
a)	Describe the Knowledge Discovery in Databases (KDD) process. What are the key stages involved, and why is each stage important in data mining?						CO3	U																																																															
b)	Justify the need of Strategic information. Compare: Operational vs Decision Support Systems						CO1	An																																																															
c)	Use the given database. Follow the single linkage technique to find the clusters in database. Use the Euclidian distance measure.						CO5	Ap																																																															
	<table><tr><td></td><td>P1</td><td>P2</td><td>P3</td><td>P4</td><td>P5</td><td>P6</td></tr><tr><td>X</td><td>0.4</td><td>0.22</td><td>0.35</td><td>0.26</td><td>0.08</td><td>0.45</td></tr><tr><td>Y</td><td>0.53</td><td>0.38</td><td>0.32</td><td>0.19</td><td>0.41</td><td>0.3</td></tr></table>						P1	P2	P3	P4	P5	P6	X	0.4	0.22	0.35	0.26	0.08	0.45	Y	0.53	0.38	0.32	0.19	0.41	0.3																																													
	P1	P2	P3	P4	P5	P6																																																																	
X	0.4	0.22	0.35	0.26	0.08	0.45																																																																	
Y	0.53	0.38	0.32	0.19	0.41	0.3																																																																	
