

Term End Examination

Dec 2023

COS3127B - Data Mining and Warehousing

Question Paper ID: 025837

Faculty/School	Faculty of Engineering and Technology	Term	V
Program	TYBCA	Duration	2 Hours 30 Minutes
Specialization		Max. Marks	70

Read the instructions provided for every question properly before attempting the answer.

Section - 1 : contain(s) **10** question(s) and each question carries **10** mark(s). You can answer any **7** questions out of **10**.

Click **Finish** only after completion of the Exam.

Section - 1 (7 X 10 Marks)

Answer any 7 questions

1	What are predictive and descriptive models? Explain data mining techniques that are suitable for descriptive analysis.	10 marks	CO1, CO2, CO3	Remembering										
2	Define data warehouse. Compare OLAP and OLTP for a university.	10 marks	CO1	Remembering										
3	Discuss the possible ways to deal with the missing values. Explain any two techniques to smooth out noise from data using suitable example.	10 marks	CO1, CO4	Remembering										
4	Consider the market basket transactions shown below. Assuming the minimum support=40% and minimum confidence = 80% i) Find all frequent item sets using Apriori Algorithm ii) Find all association rules using Apriori Algorithm <table><tr><td>Transaction ID</td><td>Items</td></tr><tr><td>T1</td><td>M,A,B,D</td></tr><tr><td>T2</td><td>A,D,C,B,F</td></tr><tr><td>T3</td><td>A,C,B,F</td></tr><tr><td>T4</td><td>A,B,D</td></tr></table>	Transaction ID	Items	T1	M,A,B,D	T2	A,D,C,B,F	T3	A,C,B,F	T4	A,B,D	10 marks	CO1, CO3, CO4	Remembering
Transaction ID	Items													
T1	M,A,B,D													
T2	A,D,C,B,F													
T3	A,C,B,F													
T4	A,B,D													

5	1.Explain measure ‘confidence’ for association rule using suitable example. 2.Construct FP tree along with FP header for below transaction(min_support=40%) <table><tr><th>Transaction ID</th><th>Items</th></tr><tr><td>T1</td><td>A,B,C</td></tr><tr><td>T2</td><td>A,B,C,D,E</td></tr><tr><td>T3</td><td>A,C,D</td></tr><tr><td>T4</td><td>A,C,D,E</td></tr><tr><td>T5</td><td>A,B,C,D</td></tr></table>	Transaction ID	Items	T1	A,B,C	T2	A,B,C,D,E	T3	A,C,D	T4	A,C,D,E	T5	A,B,C,D	10 marks	CO1, CO3	Remembering																																																																														
Transaction ID	Items																																																																																													
T1	A,B,C																																																																																													
T2	A,B,C,D,E																																																																																													
T3	A,C,D																																																																																													
T4	A,C,D,E																																																																																													
T5	A,B,C,D																																																																																													
6	Define Classification. Consider the below training data from sales database. Let ‘buys_computer ’ be the class level attribute which can be ‘yes’ or ‘no’. <table><tr><th>RID</th><th>age</th><th>income</th><th>student</th><th>credit_rating</th><th>Class: buys_computer</th></tr><tr><td>1</td><td>youth</td><td>high</td><td>no</td><td>fair</td><td>no</td></tr><tr><td>2</td><td>youth</td><td>high</td><td>no</td><td>excellent</td><td>no</td></tr><tr><td>3</td><td>middle_aged</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>4</td><td>senior</td><td>medium</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>5</td><td>senior</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>6</td><td>senior</td><td>low</td><td>yes</td><td>excellent</td><td>no</td></tr><tr><td>7</td><td>middle_aged</td><td>low</td><td>yes</td><td>excellent</td><td>yes</td></tr><tr><td>8</td><td>youth</td><td>medium</td><td>no</td><td>fair</td><td>no</td></tr><tr><td>9</td><td>youth</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>10</td><td>senior</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>11</td><td>youth</td><td>medium</td><td>yes</td><td>excellent</td><td>yes</td></tr><tr><td>12</td><td>middle_aged</td><td>medium</td><td>no</td><td>excellent</td><td>yes</td></tr><tr><td>13</td><td>middle_aged</td><td>high</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>14</td><td>senior</td><td>medium</td><td>no</td><td>excellent</td><td>no</td></tr></table> Consider tuple (age=senior,income=high,student=no,credit_rating=fair)), compute a naïve Bayesian classification of tuple.	RID	age	income	student	credit_rating	Class: buys_computer	1	youth	high	no	fair	no	2	youth	high	no	excellent	no	3	middle_aged	high	no	fair	yes	4	senior	medium	no	fair	yes	5	senior	low	yes	fair	yes	6	senior	low	yes	excellent	no	7	middle_aged	low	yes	excellent	yes	8	youth	medium	no	fair	no	9	youth	low	yes	fair	yes	10	senior	medium	yes	fair	yes	11	youth	medium	yes	excellent	yes	12	middle_aged	medium	no	excellent	yes	13	middle_aged	high	yes	fair	yes	14	senior	medium	no	excellent	no	10 marks	CO1, CO3, CO4	Remembering
RID	age	income	student	credit_rating	Class: buys_computer																																																																																									
1	youth	high	no	fair	no																																																																																									
2	youth	high	no	excellent	no																																																																																									
3	middle_aged	high	no	fair	yes																																																																																									
4	senior	medium	no	fair	yes																																																																																									
5	senior	low	yes	fair	yes																																																																																									
6	senior	low	yes	excellent	no																																																																																									
7	middle_aged	low	yes	excellent	yes																																																																																									
8	youth	medium	no	fair	no																																																																																									
9	youth	low	yes	fair	yes																																																																																									
10	senior	medium	yes	fair	yes																																																																																									
11	youth	medium	yes	excellent	yes																																																																																									
12	middle_aged	medium	no	excellent	yes																																																																																									
13	middle_aged	high	yes	fair	yes																																																																																									
14	senior	medium	no	excellent	no																																																																																									

7	Consider below training data from sales database. Calculate information gain for all feature. Comment on , which attribute can be selected at the root level node for constructing decision tree. <table border="1"> <thead> <tr> <th>Instnace</th><th>Gender</th><th>Flight shift</th><th>Ticket Class</th><th>Survived</th></tr> </thead> <tbody> <tr><td>1</td><td>Male</td><td>Day</td><td>Second</td><td>Yes</td></tr> <tr><td>2</td><td>Female</td><td>Day</td><td>First</td><td>Yes</td></tr> <tr><td>3</td><td>Male</td><td>Night</td><td>Second</td><td>No</td></tr> <tr><td>4</td><td>Male</td><td>Night</td><td>Third</td><td>No</td></tr> <tr><td>5</td><td>Female</td><td>Day</td><td>Second</td><td>Yes</td></tr> <tr><td>6</td><td>Female</td><td>Day</td><td>Third</td><td>No</td></tr> <tr><td>7</td><td>Male</td><td>Night</td><td>First</td><td>No</td></tr> <tr><td>8</td><td>Female</td><td>Day</td><td>First</td><td>Yes</td></tr> <tr><td>9</td><td>Female</td><td>Day</td><td>Second</td><td>No</td></tr> <tr><td>10</td><td>Male</td><td>Night</td><td>First</td><td>Yes</td></tr> <tr><td>11</td><td>Male</td><td>Night</td><td>Second</td><td>No</td></tr> <tr><td>12</td><td>Female</td><td>Day</td><td>Third</td><td>No</td></tr> </tbody> </table>	Instnace	Gender	Flight shift	Ticket Class	Survived	1	Male	Day	Second	Yes	2	Female	Day	First	Yes	3	Male	Night	Second	No	4	Male	Night	Third	No	5	Female	Day	Second	Yes	6	Female	Day	Third	No	7	Male	Night	First	No	8	Female	Day	First	Yes	9	Female	Day	Second	No	10	Male	Night	First	Yes	11	Male	Night	Second	No	12	Female	Day	Third	No	10 marks	CO1, CO3	Remembering
Instnace	Gender	Flight shift	Ticket Class	Survived																																																																	
1	Male	Day	Second	Yes																																																																	
2	Female	Day	First	Yes																																																																	
3	Male	Night	Second	No																																																																	
4	Male	Night	Third	No																																																																	
5	Female	Day	Second	Yes																																																																	
6	Female	Day	Third	No																																																																	
7	Male	Night	First	No																																																																	
8	Female	Day	First	Yes																																																																	
9	Female	Day	Second	No																																																																	
10	Male	Night	First	Yes																																																																	
11	Male	Night	Second	No																																																																	
12	Female	Day	Third	No																																																																	
8	Discuss the significance of attribute selection measures in decision tree algorithm. Explain the process of post-pruning using suitable example	10 marks	CO1, CO3	Remembering																																																																	
9	Suppose the data for analysis includes the attribute age .The age values for the data tuples are (in increasing order) 8,10,15,15,18,20,20,22,23,25,25,27,30,32,35 1.Calculate mean , median and range of the data. 2.Find mode of the data? Comment on the data's modality 2.Explain binning and clustering can be used to discretize the data using suitable example	10 marks	CO1, CO3	Remembering																																																																	
10	Explain the key features and functionalities of the Weka tool in the context of machine learning. Outline the main steps involved in using Weka for a typical machine learning workflow.	10 marks	CO2, CO3, CO4	Remembering																																																																	

END OF QUESTION PAPER