

End Semester Examination

May-June 2023

CET0012B - Foundations of Data Warehousing and Data Mining

Schedule ID: 19287

Faculty/School	Faculty of Engineering and Technology	Term	Semester IV
Program	SY Int. BTech CSE/CSE-DS	Duration	1 Hours 30 Minutes
Specialization	--	Max. Marks	40

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

Section - 1 : contain(s) **10** question(s) and each question carries **5** mark(s). You can answer any **8** questions out of **10**.

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

Section - 1 (8 X 5 Marks)

Answer any 8 questions

1	Explain the Structure of Data Warehouse with Diagram.	5 marks	CO1	Understanding
2	Differentiate between OLAP and OLTP.	5 marks	CO1	Remembering
3	List and Explain Data Visualisation Techniques.	5 marks	CO2	Applying
4	Explain Decision Tree along with diagram and relevant example	5 marks	CO3	Understanding

5	1. Calculate the Information Gain for the attribute Outlook Refer to the below given decision tree dataset. <table><tr><th>Day</th><th>Outlook</th><th>Temp</th><th>Humidity</th><th>Wind</th><th>Play Tennis</th></tr><tr><td>D1</td><td>Sunny</td><td>Hot</td><td>High</td><td>Weak</td><td>No</td></tr><tr><td>D2</td><td>Sunny</td><td>Hot</td><td>High</td><td>Strong</td><td>No</td></tr><tr><td>D3</td><td>Overcast</td><td>Hot</td><td>High</td><td>Weak</td><td>Yes</td></tr><tr><td>D4</td><td>Rain</td><td>Mild</td><td>High</td><td>Weak</td><td>Yes</td></tr><tr><td>D5</td><td>Rain</td><td>Cool</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>D6</td><td>Rain</td><td>Cool</td><td>Normal</td><td>Strong</td><td>No</td></tr><tr><td>D7</td><td>Overcast</td><td>Cool</td><td>Normal</td><td>Strong</td><td>Yes</td></tr><tr><td>D8</td><td>Sunny</td><td>Mild</td><td>High</td><td>Weak</td><td>No</td></tr><tr><td>D9</td><td>Sunny</td><td>Cool</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>D10</td><td>Rain</td><td>Mild</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>D11</td><td>Sunny</td><td>Mild</td><td>Normal</td><td>Strong</td><td>Yes</td></tr><tr><td>D12</td><td>Overcast</td><td>Mild</td><td>High</td><td>Strong</td><td>Yes</td></tr><tr><td>D13</td><td>Overcast</td><td>Hot</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>D14</td><td>Rain</td><td>Mild</td><td>High</td><td>Strong</td><td>No</td></tr></table>	Day	Outlook	Temp	Humidity	Wind	Play Tennis	D1	Sunny	Hot	High	Weak	No	D2	Sunny	Hot	High	Strong	No	D3	Overcast	Hot	High	Weak	Yes	D4	Rain	Mild	High	Weak	Yes	D5	Rain	Cool	Normal	Weak	Yes	D6	Rain	Cool	Normal	Strong	No	D7	Overcast	Cool	Normal	Strong	Yes	D8	Sunny	Mild	High	Weak	No	D9	Sunny	Cool	Normal	Weak	Yes	D10	Rain	Mild	Normal	Weak	Yes	D11	Sunny	Mild	Normal	Strong	Yes	D12	Overcast	Mild	High	Strong	Yes	D13	Overcast	Hot	Normal	Weak	Yes	D14	Rain	Mild	High	Strong	No	5 marks	CO3	Evaluating
Day	Outlook	Temp	Humidity	Wind	Play Tennis																																																																																									
D1	Sunny	Hot	High	Weak	No																																																																																									
D2	Sunny	Hot	High	Strong	No																																																																																									
D3	Overcast	Hot	High	Weak	Yes																																																																																									
D4	Rain	Mild	High	Weak	Yes																																																																																									
D5	Rain	Cool	Normal	Weak	Yes																																																																																									
D6	Rain	Cool	Normal	Strong	No																																																																																									
D7	Overcast	Cool	Normal	Strong	Yes																																																																																									
D8	Sunny	Mild	High	Weak	No																																																																																									
D9	Sunny	Cool	Normal	Weak	Yes																																																																																									
D10	Rain	Mild	Normal	Weak	Yes																																																																																									
D11	Sunny	Mild	Normal	Strong	Yes																																																																																									
D12	Overcast	Mild	High	Strong	Yes																																																																																									
D13	Overcast	Hot	Normal	Weak	Yes																																																																																									
D14	Rain	Mild	High	Strong	No																																																																																									
6	Explain various Data mining Techniques	5 marks	CO3	Applying																																																																																										
7	Explain Naive Bayesian Classification with example	5 marks	CO4	Understanding																																																																																										
8	Explain Naive Bayesian equation to calculate the Posterior and Prior probability for each each class, For given dataset below <table><tr><th>Outlook</th><th>Temp</th><th>Humidity</th><th>Windy</th><th>Play?</th></tr><tr><td>Sunny</td><td>Hot</td><td>High</td><td>False</td><td>No</td></tr><tr><td>Sunny</td><td>Hot</td><td>High</td><td>True</td><td>No</td></tr><tr><td>Overcast</td><td>Hot</td><td>High</td><td>False</td><td>Yes</td></tr><tr><td>Rainy</td><td>Mild</td><td>High</td><td>False</td><td>Yes</td></tr><tr><td>Rainy</td><td>Cool</td><td>Normal</td><td>False</td><td>Yes</td></tr><tr><td>Rainy</td><td>Cool</td><td>Normal</td><td>True</td><td>No</td></tr><tr><td>Overcast</td><td>Cool</td><td>Normal</td><td>True</td><td>Yes</td></tr><tr><td>Sunny</td><td>Mild</td><td>High</td><td>False</td><td>No</td></tr><tr><td>Sunny</td><td>Cool</td><td>Normal</td><td>False</td><td>Yes</td></tr><tr><td>Rainy</td><td>Mild</td><td>Normal</td><td>False</td><td>Yes</td></tr><tr><td>Sunny</td><td>Mild</td><td>Normal</td><td>True</td><td>Yes</td></tr><tr><td>Overcast</td><td>Mild</td><td>High</td><td>True</td><td>Yes</td></tr></table> Problem: Apply Naive Bayes to the weather training data. The hypothesis space is V yes no . Classify the following new instance: <table><tr><td>Outlook</td><td>Temp</td><td>Humidity</td><td>Windy</td><td>Play</td><td></td></tr><tr><td>sunny</td><td>cool</td><td>high</td><td>true</td><td>?</td><td></td></tr></table>	Outlook	Temp	Humidity	Windy	Play?	Sunny	Hot	High	False	No	Sunny	Hot	High	True	No	Overcast	Hot	High	False	Yes	Rainy	Mild	High	False	Yes	Rainy	Cool	Normal	False	Yes	Rainy	Cool	Normal	True	No	Overcast	Cool	Normal	True	Yes	Sunny	Mild	High	False	No	Sunny	Cool	Normal	False	Yes	Rainy	Mild	Normal	False	Yes	Sunny	Mild	Normal	True	Yes	Overcast	Mild	High	True	Yes	Outlook	Temp	Humidity	Windy	Play		sunny	cool	high	true	?		5 marks	CO4	Evaluating													
Outlook	Temp	Humidity	Windy	Play?																																																																																										
Sunny	Hot	High	False	No																																																																																										
Sunny	Hot	High	True	No																																																																																										
Overcast	Hot	High	False	Yes																																																																																										
Rainy	Mild	High	False	Yes																																																																																										
Rainy	Cool	Normal	False	Yes																																																																																										
Rainy	Cool	Normal	True	No																																																																																										
Overcast	Cool	Normal	True	Yes																																																																																										
Sunny	Mild	High	False	No																																																																																										
Sunny	Cool	Normal	False	Yes																																																																																										
Rainy	Mild	Normal	False	Yes																																																																																										
Sunny	Mild	Normal	True	Yes																																																																																										
Overcast	Mild	High	True	Yes																																																																																										
Outlook	Temp	Humidity	Windy	Play																																																																																										
sunny	cool	high	true	?																																																																																										
9	Explain Apriori Algorithm with example	5 marks	CO5	Understanding																																																																																										

10	Explain the following concepts --Finding Frequent Item sets by Confined, Candidate Generation, Generating Association Rules from Frequent Item sets using Apriori Algorithm for given below example <table><tr><th>T_{ID}</th><th>Items in transaction</th></tr><tr><td>T1</td><td>I₁, I₂, I₅</td></tr><tr><td>T2</td><td>I₂, I₄</td></tr><tr><td>T3</td><td>I₂, I₃</td></tr><tr><td>T4</td><td>I₁, I₂, I₄</td></tr><tr><td>T5</td><td>I₁, I₃</td></tr><tr><td>T6</td><td>I₂, I₃</td></tr><tr><td>T7</td><td>I₁, I₃</td></tr><tr><td>T8</td><td>I₁, I₂, I₃, I₅</td></tr><tr><td>T9</td><td>I₁, I₂, I₃</td></tr></table>	T _{ID}	Items in transaction	T1	I ₁ , I ₂ , I ₅	T2	I ₂ , I ₄	T3	I ₂ , I ₃	T4	I ₁ , I ₂ , I ₄	T5	I ₁ , I ₃	T6	I ₂ , I ₃	T7	I ₁ , I ₃	T8	I ₁ , I ₂ , I ₃ , I ₅	T9	I ₁ , I ₂ , I ₃	5 marks	CO5	Applying
T _{ID}	Items in transaction																							
T1	I ₁ , I ₂ , I ₅																							
T2	I ₂ , I ₄																							
T3	I ₂ , I ₃																							
T4	I ₁ , I ₂ , I ₄																							
T5	I ₁ , I ₃																							
T6	I ₂ , I ₃																							
T7	I ₁ , I ₃																							
T8	I ₁ , I ₂ , I ₃ , I ₅																							
T9	I ₁ , I ₂ , I ₃																							

END OF QUESTION PAPER