PVP 20

Code: 20AM4501C, 20DS4501C

III B.Tech - I Semester - Regular Examinations - NOVEMBER 2024

DATA WAREHOURSING AND DATA MINING
(Common for AIML, DS)

Duration: 3 hours Max. Marks: 70

Note: 1. This paper contains questions from 5 units of Syllabus. Each unit carries 14 marks and have an internal choice of Questions.
2. All parts of Question must be answered in one place.

BL – Blooms Level CO – Course Outcome

			BL	CO	Max. Marks
UNIT-I					
1	a)	Discuss the major challenges of mining a huge amount of data in comparison with mining a small amount of data.	L2	CO1	7 M
	b)	Discuss the significance of preprocessing in data mining. Also explain various steps involved in pre-processing.	L2	CO1	7 M
OR					
2	a)	Discuss various tasks involved in data mining with suitable example.	L2	CO1	7 M
	b)	Describe data transformation and data discretization in data mining.	L2	CO1	7 M
UNIT-II					
3	a)	What is Operational Data Base System? Explain how it differs from data warehouse?	L2	CO1	7 M
	b)	Compare and contrast the characteristics of OLTP and OLAP.	L2	CO1	7 M
OR					

<table>
<tr><td>4</td><td>a)</td><td>Explain the basic elements of data warehouse with a neat sketch.</td><td>L2</td><td>CO1</td><td>7 M</td></tr>
<tr><td></td><td>b)</td><td>What is the use of data cube? With a neat diagram, explain data-cube implementations.</td><td>L2</td><td>CO1</td><td>7 M</td></tr>
<tr><td colspan="6">UNIT-III</td></tr>
<tr><td>5</td><td>a)</td><td>The following table describes training examples to decide going to a hotel under different weather conditions. Discuss the algorithm to construct Decision tree.
<table>
<tr><th>OUTLOOK</th><th>TEMPERATURE</th><th>HUMIDITY</th><th>WINDY</th><th>Target</th></tr>
<tr><td>Rainy</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>Sunny</td><td>Mild</td><td>High</td><td>FALSE</td><td>Yes</td></tr>
<tr><td>Sunny</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>Rainy</td><td>Mild</td><td>High</td><td>FALSE</td><td>No</td></tr>
<tr><td>Rainy</td><td>Cool</td><td>Normal</td><td>FALSE</td><td>Yes</td></tr>
<tr><td>Sunny</td><td>Mild</td><td>Normal</td><td>FALSE</td><td>Yes</td></tr>
<tr><td>Rainy</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>
<tr><td>Overcast</td><td>Hot</td><td>Normal</td><td>FALSE</td><td>Yes</td></tr>
</table></td><td>L3</td><td>CO2</td><td>7 M</td></tr>
<tr><td></td><td>b)</td><td>Given a decision tree, you have the option of (i) converting the decision tree to rules and then pruning the resulting rules or (ii) pruning the decision tree and then converting the pruned tree to rules. What advantage does (i) have over (ii) Explain?</td><td>L3</td><td>CO4</td><td>7 M</td></tr>
<tr><td colspan="6">OR</td></tr>
</table>

<table>
<tr><td>6</td><td>a)</td><td>Consider the following data describing different types of Mushrooms available in a desert. Each type of Mushroom is described with attributes such as “Not Heavy”, “Smelly”, “Spotted” and “Smooth” to take a decision on the corresponding Mushroom variety is “Edible or Not”. Discuss the algorithm to construct Decision tree and derive whether the following types of Mushrooms are “Edible or Not”.
<table>
<tr><th>Type of Mushroom</th><th>Not Heavy</th><th>Smeely</th><th>Spotted</th><th>Smooth</th><th>Edible</th></tr>
<tr><td>A</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>B</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>C</td><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>D</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>E</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>F</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>G</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>H</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
</table></td><td>L3</td><td>CO2</td><td>7 M</td></tr>
<tr><td></td><td>b)</td><td>Why is tree pruning useful in decision tree induction? What is the drawback of using a separate set of tuples to evaluate pruning?</td><td>L3</td><td>CO2</td><td>7 M</td></tr>
<tr><td colspan="6">UNIT-IV</td></tr>
<tr><td>7</td><td>a)</td><td>Apply the frequent pattern growth with suitable example.</td><td>L3</td><td>CO3</td><td>7 M</td></tr>
<tr><td></td><td>b)</td><td>Explain Apriori algorithm for finding frequent item set with suitable example.</td><td>L4</td><td>CO4</td><td>7 M</td></tr>
</table>

<table>
<tr><td colspan="6">OR</td></tr>
<tr><td rowspan="2">8</td><td>a)</td><td>Explain association rules (support, confidence) for the following table.
<table>
<tr><th>TID</th><th>List of Items</th></tr>
<tr><td>T100</td><td>I1,I2,I5</td></tr>
<tr><td>T200</td><td>I2,I4</td></tr>
<tr><td>T300</td><td>I2,I3</td></tr>
<tr><td>T400</td><td>I1,I2,I4</td></tr>
<tr><td>T500</td><td>I1,I3</td></tr>
<tr><td>T600</td><td>I2,I3</td></tr>
</table></td><td>L3</td><td>CO3</td><td>7 M</td></tr>
<tr><td>b)</td><td>Use the pattern growth method for mining frequently occurring items, its approaches and application in data mining.</td><td>L4</td><td>CO4</td><td>7 M</td></tr>
<tr><td colspan="6">UNIT-V</td></tr>
<tr><td rowspan="2">9</td><td>a)</td><td>Define clustering? Explain various types of data in cluster analysis?</td><td>L1</td><td>CO1</td><td>7 M</td></tr>
<tr><td>b)</td><td>Cluster the following eight points (with (x, y) representing locations) into three clusters using k-means clustering:
A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5),A6(6, 4), A7(1, 2), A8(4, 9)</td><td>L3</td><td>CO3</td><td>7 M</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td rowspan="2">10</td><td>a)</td><td>Write a note on partitioning based clustering methods.</td><td>L2</td><td>CO1</td><td>7 M</td></tr>
<tr><td>b)</td><td>Cluster the following eight points (with (x, y) representing locations) using hierarchical clustering(Agglomerative approach):
A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4),A7(1, 2), A8(4, 9)</td><td>L3</td><td>CO3</td><td>7 M</td></tr>
</table>