Academic Year : 2023-24 **Semester : V**

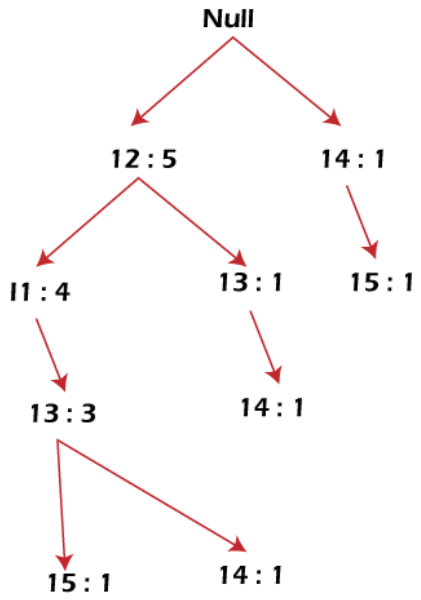
Class/Branch : TEDS **Subject : DWM**

Question Bank

<table>
<tr><td>1</td><td>What is web structure mining? List the approaches used to structure the web pages to improve effectiveness of search engine and crawlers.</td></tr>
<tr><td>2</td><td>Explain web usage mining in detail.</td></tr>
<tr><td>3</td><td>What is web content mining, and how does it differ from traditional data mining techniques?</td></tr>
<tr><td>4</td><td>How is PageRank used by search engines like Google to rank web pages?</td></tr>
<tr><td>5</td><td>Find the frequent item set in the following database of given transactions with min support 50% and confidence 70%

<table><tr><td>TID</td><td>ITEMS</td></tr><tr><td>100</td><td>1 3 4</td></tr><tr><td>200</td><td>2 3 5</td></tr><tr><td>300</td><td>1 2 3 5</td></tr><tr><td>400</td><td>2 5</td></tr></table></td></tr>
<tr><td>6</td><td>For the following transactional database construct FP tree.

<table><tr><td>TID</td><td>Items</td></tr><tr><td>1</td><td>B D C A</td></tr><tr><td>2</td><td>E D C</td></tr><tr><td>3</td><td>A B</td></tr><tr><td>4</td><td>A C D</td></tr><tr><td>5</td><td>F G D B</td></tr></table></td></tr>
<tr><td>7</td><td>For the following FP tree mine frequent patterns using the FP-growth algorithm.</td></tr>
</table>

<table>
<tr><td></td><td> 
</td></tr>
<tr><td>8</td><td>Find clusters using k-means clustering algorithm if we have several objects (4 types of medicines) and each object have two attributes or features as shown in the table below. The goal is to group these objects into k=2 group of medicine based on the two features (pH and weight index).
<table>
<tr><th>Object</th><th>Attribute 1 (X) Weight Index</th><th>Attribute 2 (Y) pH</th></tr>
<tr><td>Medicine A</td><td>1</td><td>1</td></tr>
<tr><td>Medicine B</td><td>2</td><td>1</td></tr>
<tr><td>Medicine C</td><td>4</td><td>3</td></tr>
<tr><td>Medicine D</td><td>5</td><td>4</td></tr>
</table></td></tr>
<tr><td>9</td><td>For given distance matrix draw single and complete linkage dendogram
<table>
<tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>1</td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>2</td><td>9</td><td>0</td><td></td><td></td><td></td></tr>
<tr><td>3</td><td>3</td><td>7</td><td>0</td><td></td><td></td></tr>
<tr><td>4</td><td>6</td><td>5</td><td>9</td><td>0</td><td></td></tr>
<tr><td>5</td><td>11</td><td>10</td><td>2</td><td>8</td><td>0</td></tr>
</table></td></tr>
<tr><td>10</td><td>Data for salary analysis include 30, 36,47,50,52,52,56,60,63,70,70,110</td></tr>
</table>

	Compute the first, second and third quartile for this data. Visualize using box plot
11	Data for salary analysis include 30, 36,47,50,52,52,56,60,63,70,70,110 Normalize the data by applying : Min-max , Z score , Decimal Scaling
12	Find the value of correlation coefficient from the following table
13	Explain major steps involved in data pre-processing?
14	Discuss concept hierarchy generation.
15	Design the data warehouse for Wholesale furniture Company. The data warehouse has to allow analyzing the company situation at least with respect to the Furniture , Customer and Time. More ever the company needs to anlayse the furniture with respect to its type, category and material. The customers with respect to their spatial location by considering at least cities regions and states. The company is interested in learning the quantity, income and discount of its sales.
16	Consider following dimensions for a supermarket chain: Product, Store, Time and Promotion. With respect to this business scenario, answer the following questions. Clearly state any reasonable assumptions you make. a. Design the star schema b. Can you convert the star schema to snowflake schema? If yes, justify and draw the snowflake schema. Clearly depict the fact table(s). Dimension tables their attributes and measures.
17	Draw and explain Data warehouse Architecture.
18	Discuss various OLAP operations with appropriate examples.
19	A simple example from the stock market involving only discrete ranges has profit as categorical attribute with values (UP, DOWN) and the training data is as follows:

	Age Competition Type Profit Old Yes Software Down Old No Software Down Old No Hardware Down Mid Yes Software Down Mid Yes Hardware Down Mid No Hardware Up Mid No Software Up New Yes Software Up New No Hardware Up New No Software Up Generate classification rules for the given data by applying decision tree classifier.
20	Age Income Student Credit_Rating Buys_Computer Youth High No Fair No Youth High No Excellent no Mid High No Fair yes Senior Medium No Fair Yes Senior Low Yes Fair Yes Senior Low Yes Excellent No Mid low Yes Excellent Yes Youth Medium no Fair No Youth Low Yes Fair yes Senior Medium Yes Fair Yes Youth Medium Yes Excellent Yes Mid Medium No excellent Yes Mid High Yes Fair yes Senior Medium No Excellent no

	Predict a class label using Naïve Bayes Classifier for the following tuple X = (Age =youth, income= medium, student-yes, credit rating=fair)