Paper / Subject Code: 37311 / Data Mining & Business Intelligence

May 15, 2024   02:30 pm - 05:30 pm   1T01236 - T.E.(Information Technology Engineering)
(SEM-VI)(Choice Base Credit Grading System) (R- 19) (C Scheme) / 37311 - Data Mining &
Business Intelligence   QP CODE:10055218

Duration: 3hrs   **[Max Marks: 80]**

N.B. : (1) Question No 1 is Compulsory.
        (2) Attempt any three questions out of the remaining five.
        (3) All questions carry equal marks.
        (4) Assume suitable data, if required and state it clearly.

Q.1   [20]
    A   List out stages in Data Mining with neat labelled diagram.
    B   A sales firm has reported following sales figures for FY 23-24 (i.e. March
         23 to Feb 24)
         2300, 435, 675, 543, 454, 7877, 5434, 345, 2342, 654, 567, 545.
         Show how to normalize this data series using Min-Max scaling.
    C   With an example explain Star Versus Snowflakes schema in dimensional
         modelling
    D   What is market basket analysis? Explain with a real use case.

Q.2 A   Draw and list the components of a typical Data warehouse architecture   [10]
     B   Consider we have age of 35 participants in a survey given to us in sorted [10]
          order.
          5, 10, 13, 13, 15, 16, 16, 20, 20, 20, 21, 22, 22, 22, 25, 25, 25, 25, 30, 30,
          33, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 52, 70, 85.
          Draw histograms for this data taking bin size as **5** and **8**. Explain the effect
          of bin size on the histograms you obtain.

Q.3 A   What is OLAP? Explain various OLAP operations with neat labeled diagram   [10]

     B   Explain working of decision tree based classifier? With an example explain [10]
          steps for inducting tree using ID3 algorithm.

Q.4 A   Use the Apriori algorithm to identify the frequent item-sets in the following [10]
          database.

Tid	a	b	c	d	e	f	g
Items	1,2,4,5,6	2,3,5	1,2, 4,5	1,2,4, 5	1,2,3,4,5, 6	2,3,4	1,2,4,5

          Consider Minimum Support as 75% and confidence at 85% level. Write
          down all strong association rules.
     B   What is an outlier? Explain various methods for performing outlier analysis. [10]

Q.5 A   Explain steps in hierarchical Clustering algorithm. Perform hierarchical   [10]
          clustering on following data that represents 10 points in 2 D space
          (2,3), (5,4), (9,6), (4,7), (8,1), (7,2), (6,3), (1,9), (3,6), (4,8).
          Consider you require 3 clusters.
     B   Explain mining of Multilevel association rules and Multidimensional asso-   [10]
          ciation rules.

Q.6   Write short note on following (Any 4)   [20]
    A   Navie Bayes Classifier.
    B   Boot Straping
    C   BIRCH Algoirthm.
    D   BI Architecures
    E   Types of attributes

****

55218

X525Y2FDE85X525Y2FDE85X525Y2FDE85X525Y2FDE85