Tribhuvan University

Institute of Science and Technology

MODEL QUESTION

✡

Bachelor Level/ Fourth Year/ Seventh Semester/ Science — Full Marks: 60
Computer Science and Information Technology (CSc. 451) — Pass Marks: 24
(Data Warehousing and Data Mining) — Time: 3 hours.

Candidates are required to give their answers in their own words as for as practicable.

Group - A

Attempt any two Questions **(10 x 2 = 20)**

1. Suppose that a data warehouse for Big University consists of four dimensions: student, course, semester, and instructor, and two measures count and avg-grade. When at the lowest conceptual level (eg. for a given student, course, semester, and instructor combination), the avg-grade measure stores the actual course grade of the student. At higher conceptual levels, avg-grade stores the average grade for the given combination.
   a) Draw a snowflake schema diagram for the data warehouse.
   b) Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations (eg. roll-up from semester to year) should one performing in order to list the average grade of CS courses for each Big University Student.
   c) If each dimension has five levels (including all), such as "student < major < status < university < all", how many cuboids will this cube contain (including the base and apex cuboids)?

2. A= {A1, A2, A3, A4, A5, A6}, Assume rho = 35%. Use A priori algorithm to get the desired solution.

A1	A2	A3	A4	A5	A6
0	0	0	1	1	1
0	1	1	1	0	0
1	0	0	1	1	1
1	1	0	1	0	0
1	0	1	0	1	1
0	1	1	1	0	1
0	0	0	1	1	0
0	1	0	1	0	1
1	0	0	1	0	0
1	1	1	1	1	1

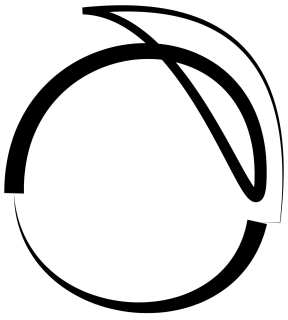
3. What kind of data processing do we need before applying datamining algorithm to any data set? Explain binning method to handle noisy data with example.

Downloaded form: https://CSITauthority.github.io

Group - B

Attempt any eight Questions. (Question no. 13 is compulsory) **(8 x 5 = 40)**

4. Explain the use of frequent item set generation process.
5. Differentiate between data marts and data cubes.
6. Explain OLAP operations with example.
7. List the drawbacks of ID3 algorithm with over-fitting and its remedy techniques.
8. Write the algorithm for k-means clustering. Compare it with k-nearest neighbor algorithm.
9. What is text mining? Explain the text indexing techniques.
10. Describe genetic algorithm using as a problem solving technique in data mining.
11. What do you mean by WWW mining? Explain WWW mining techniques.
12. What is DMQL? How do you define Star Schema using DMQL?
13. Write short notes (Any Two)
    a) Test Database Mining
    b) Back propagation Algorithm
    c) Regression
    d) HOLAP



Downloaded form: https://CSITauthority.github.io