Paper / Subject Code: 48894 / Data Warehouseing & Mining

1T01875 - T.E. Computer Science & Enginering (Artificial Intelligence & Machine Learning) (Choice Based) (R-2019-20'C' Scheme) SEMESTER - V / 48894 - Data Warehouseing & Mining

QP CODE: 10014142   DATE: 30/11/2022

Time: 3 hours   **Max. Marks: 80**

N.B. (1) Question one is Compulsory.
      (2) Attempt any 3 questions out of the remaining.
      (3) Assume suitable data if required.

Q. 1 (a) Explain features of data warehouse.	05
(b) Demonstrate with diagram the process of KDD.	05
(c) What is Market basket analysis?	05
(d) Explain with example confusion matrix, accuracy and precision.	05

Q. 2 a) Suppose that a data warehouse for Big_University consists of the four dimensions Student, Course, Semester and Instructor, and two measures count and avg_grade, where count is the number of students and average grade is the course grade of the student.

    Perform following tasks:
  i.   Design the star schema for the Big_University.
  ii.   Create a base cuboid for the Big_university database and apply different OLAP operations.   10

  b) What is clustering? Explain K-mean clustering algorithm.Suppose that the data mining task is to cluster the following items into two clusters.
{2, 4, 10, 12, 3, 20, 30, 11, 25}. Apply k-means algorithm.   10

Q. 3 a) Explain ETL process in detail.   10
  b) Consider the transaction database given below:
    Use Apriori Algorithm with min-support count= 2 and min-confidence = 60%, to find frequent itemset andstrong association rules.   10

TID	Items
10	1, 3, 4
20	2, 3, 5
30	1, 2, 3, 5
40	2 , 5
50	1, 3, 5

Q. 4 a) Illustrate any one classification technique for the following dataset.
    Show how we can classify new tuple **(Homeowner=YES, Status=Employed, Income= Average)**.   10

Sr. No	Homeowner	Status	Income	Defaulted
1	Yes	Employed	High	No
2	No	Business	Average	No
3	No	Employed	Low	No
4	Yes	Business	High	No
5	No	Unemployed	Average	Yes
6	No	Business	Low	No
7	Yes	Unemployed	High	No
8	No	Employed	Average	Yes
9	No	Business	Low	No
10	No	Employed	Average	Yes

  b) What is web mining? Explain web content mining in detail   10

AB22FDF6FF764326E703F5E35DBFE2C7

Q. 5 a) Explain different data cleaning techniques.   10

    b) Clearly explain the working of DBSCAN algorithm using appropriate diagram   10

Q.6 a) Explain Multidimensional and multilevel rule mining with example.   10

    b) Explain with example different data sampling techniques.   10

———————————————