SPRING MIDSEM Examination-2023
School of Computer Engineering, KIIT DEEMED TO BE UNIVERSITY
DATA MINING DATAWAREHOUSING (IT-3031)

Time: 1.5 Hrs   Full Mark:20

Answer any four questions including question No.1 which is compulsory.
The figures in the margin indicate full marks.

1.   [1X5=5]
a) Point out the major differences between the star schema and the snowflake schema.
b) Compute Symmetric Matrix Coefficient and Jaccard coefficient for the two following
   binary vectors [2]
      a. $p = 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0$
      b. $q = 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1$

c) _____ is a summarization of the general characteristics or features of a target
   class of data.

d) You are given an example of seismic activity in Japan, and you want to predict the
   magnitude of next earthquake, this is an example of which data mining task?

e) A table which contains attributes of measurements stored in fact tables. This table
   consists of hierarchies, categories and logic that can be used to traverse in nodes. What is
   the name of the table?

2. a) Describe the steps of knowledge discovery process in data mining   [2.5]

   b) Suppose that the data for analysis includes the attribute age. The age values for the data
   tuples are (in increasing order): 13, 15, 16, 18, 19, 20. Use the following two methods to
   normalize the given data   [2.5]
      I.   min-max normalization by setting min=0 and max=1
      II.   Z-Score normalization

3.a) A database has five transactions. Let min_support=50% and min_confidence=75%.   [5]

TID	Items_bought
T100	Bread,Cheese,Eggs,Juice
T200	Bread,Cheese,Juice
T300	Bread,Milk,Yoghurt
T400	Bread, Juice, Milk
T500	Cheese,Juice,Milk

Find all frequent item sets using Aprioirity Algorithm.

4.
   a. Differentiate OLTP and OLAP with features .   [2.5]
   b. Determine the five-number summary, Tukey fence, and outlier (if any) for the first ten
      prime numbers. Draw the box plot to describe the distribution of data in the data set. [2.5]

5. a) Define the Slice and Dice operation in a Data cube   [2.5]
b) Describe the Architecture of Data warehouse with a neat sketch.   [2.5]