COMPUTER ENGINEERING DEPARTMENT

SUBJECT: DATA WAREHOUSING & MINING

COURSE: T.E. **YEAR: 2020-2021** **SEMESTER: VI**
DEPT: COMPUTER ENGINEERING
SUBJECT CODE: CSC603 **EXAMINATION DATE: 07/06/2021**

==========

DATA WAREHOUSING & MINING
ANSWER SHEET

NAME : AMEY MAHENDRA THAKUR

SEAT NO. : 61021145

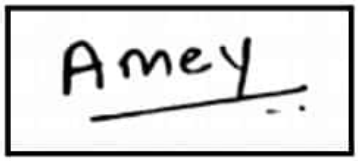
EXAM : SEMESTER VI

SUBJECT : DATA WAREHOUSING & MINING

DATE : 07-06-2021

DAY : MONDAY

STUDENT SIGNATURE:



Q2

1]

Metadata:

① Metadata is simply defined as data about data.

② The data that is used to represent other data is known as metadata

③ For example, the index of a book serves as a metadata for the contents in the book.

④ In other words, we can say that metadata is the summarized data that leads us to detailed data.

⑤ In terms of data warehouse, we can define metadata as follows:

(A) Metadata is the road map to a data warehouse.

(B) Metadata in a data warehouse defines the warehouse objects.

(C) Metadata acts as a directory. This directory helps the decision support system to locate the contents of a data warehouse.

Roles of Metadata:

① It is used for query tools.

② It is used in extraction and cleansing tools.

③ It is used in reporting tools.

④ It is used in transforming tools.

⑤ It plays an important role in loading functions

07-06-2024 STUDENT SIGNATURE: Amey

Q2.

Metadata of book store like Amazon will contain.

- Name of a book
- Summary of a book
- Assessments about book
- Date of publication
- High level description of what is contained
- Publisher details
- How can you find the book
- Book Availability
- This information helps to
  a. search a book
  b. Access the book
  c. Understand about book before you access or buy it.

07-06-2021

STUDENT SIGNATURE: Amey

Q 2

B]

Soln:

Dataset = { 6, 9, 12, 13, 15, 25, 50, 70, 72, 92, 204, 232 }

No. of bins = b = 3

No. of elements = 12

$$\text{Frequency} = \frac{n}{b} = \frac{12}{3} = 4$$

Dataset is already sorted.

Bin 1:

bin 1 = { 6, 9, 12, 13 }
bin 2 = { 15, 25, 50, 70 }
bin 3 = { 72, 92, 204, 232 }

Lets smooth the values by bin method

For Bin 1,

$$\text{Mean of bin1} = \frac{6 + 9 + 12 + 13}{4} = 10$$

Lets replace all values of bin 1 by 10.

bin 1 = { 10, 10, 10, 10 }

07 - 06 - 2021 STUDENT SIGNATURE: Amey

For bin 2

Mean of bin 2 = $\frac{15+25+50+70}{4}$

= 40

Replace by 40

bin 2 = {40, 40, 40, 40}

For bin 3

Mean of bin 3 = $\frac{72+92+204+232}{4}$

= 150

Replace all elements in bin 3 by 150

bin 3 = {150, 150, 150, 150}

07-06-2021

STUDENT SIGNATURE: Amey

Q2.

D]

Soln:

2, 3, 6, 8, 9, 12, 15, 18, 22

Assign

K_1 = 2, 8, 15 - mean = 8.3

K_2 = 3, 9, 18 - mean = 10

K_3 = 6, 12, 22 - mean = 13.3

Re-assign

K_1 = 2, 3, 6, 8, 9 - mean = 5.6

K_2 = mean = 0

K_3 = 12, 15, 18, 22 - mean = 16.75

Reassign

K_1 = 3, 6, 8, 9 - mean = 6.5

K_2 = 2 - mean = 2

K_3 = 12, 15, 18, 22 - mean = 16.75

Reassign

K_1 = 6, 8, 9 - mean = 6.5

K_2 = 2, 3 - mean = 2.5

K_3 = 12, 15, 18, 22 - mean = 16.75

Reassign

K_1 = 6, 8, 9 - mean = 7.6

K_2 = 2, 3 - mean = 2.5

K_3 = 12, 15, 18, 22 - mean = 16.75

Last two groups are same

∴ Finally we got clusters

Cluster 1 = {6, 8, 9}

Cluster 2 = {2, 3}

Cluster 3 = {12, 15, 18, 22}

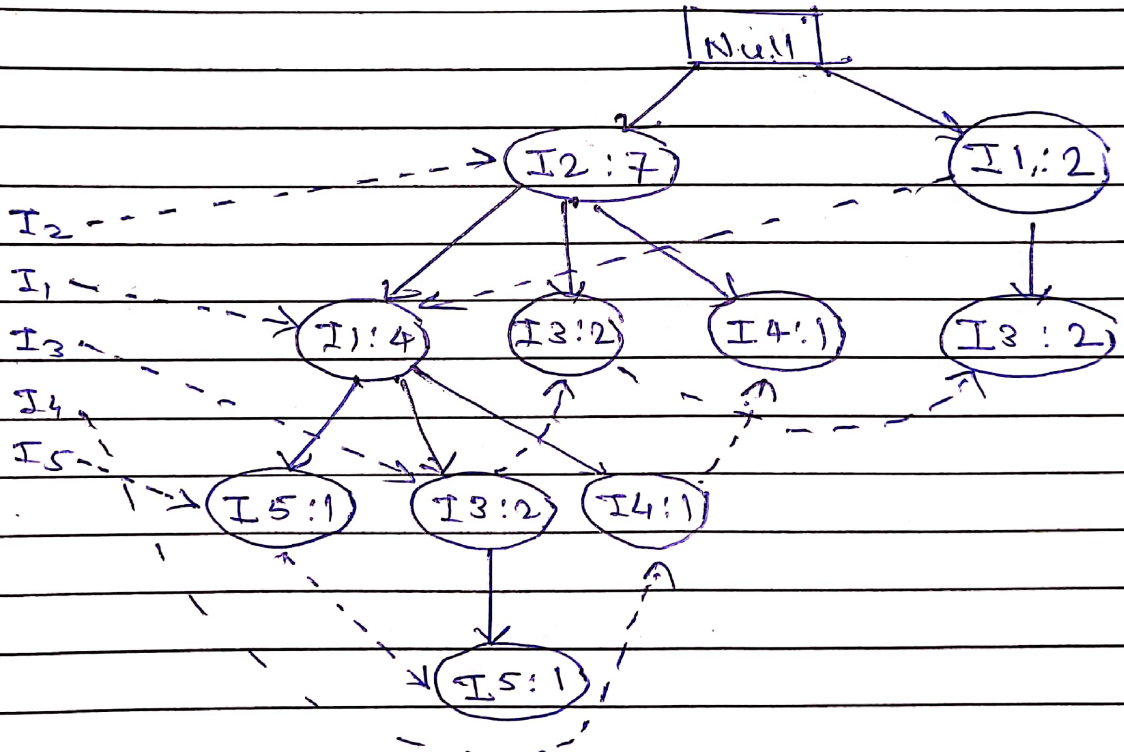
07-05-2021 STUDENT SIGNATURE: Amey

Q2.

E] Minimum Support = 2

Soln:

Item ID	Support count
I_2	7
I_1	6
I_3	6
I_4	2
I_5	2



07-06-2021 STUDENT SIGNATURE: Amey

Q2.

F]

Spatial Clustering Technique : CLARNS

CLARNS (A clustering algorithm based on randomized search)

- Clustering is a descriptive task that seeks to identify homogeneous groups of objects based on the values of their attribute.
- In spatial data sets, clustering permits a generalization of the spatial component like explicit location and extension of spatial objects which defines implicit relations of spatial neighbourhood.

- CLARNS improves on CLARA by using multiple different samples
- For every step of search CLARNS draws a sample neighbour
- So it is not confining a search to localized area
- It use two additional parameters numlocal and maxneighbour.
- Numlocal indicates number of samples to be taken.
- Maxneighbour is the number of neighbours of a node to which any specific node can be compared.

07-06-2021 STUDENT SIGNATURE: Amey