



RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF ENGINEERING

DATA MINING & WAREHOUSEING
LECTURE-32

MR. DHIRENDRA

ASSISTANT PROFESSOR

RAMA UNIVERSITY

OUTLINE

- ❖ **CLASS CHARACTERIZATION: AN EXAMPLE**
- ❖ **PRESENTATION OF GENERALIZED RESULTS**
- ❖ **PRESENTATION—GENERALIZED RELATION**
- ❖ **PRESENTATION—CROSSTAB**
- ❖ **IMPLEMENTATION BY CUBE TECHNOLOGY**
- ❖ **MCQ**
- ❖ **REFERENCES**



Class Characterization: An Example Example

Initial Relation

Name	Gender	Major	Birth-Place	Birth_date	Residence	Phone #	GPA
Jim Woodman	M	CS	Vancouver,BC, Canada	8-12-76	3511 Main St., Richmond	687-4598	3.67
Scott Lachance	M	CS	Montreal, Que, Canada	28-7-75	345 1st Ave., Richmond	253-9106	3.70
Laura Lee	F	Physics	Seattle, WA, USA	25-8-70	125 Austin Ave., Burnaby	420-5232	3.83
...
Removed	Retained	Sci,Eng, Bus	Country	Age range	City	Removed	Excl, VG,...

Prime Generalized Relation

Gender	Major	Birth_region	Age_range	Residence	GPA	Count
M	Science	Canada	20-25	Richmond	Very-good	16
F	Science	Foreign	25-30	Burnaby	Excellent	22
...

Gender \ Birth_Region	Canada	Foreign	Total
	M	16	14
F	10	22	32
Total	26	36	62

Presentation of Generalized Results

Generalized relation

– Relations where some or all attributes are generalized, with counts or

other aggregation values accumulated.

- Cross tabulation

– Mapping Mapping results into cross tabulation form (similar to contingency tables).

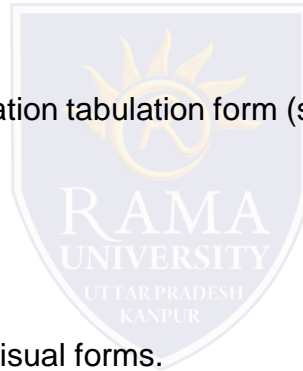
– Visualization techniques:

– Pie charts, bar charts, curves, cubes, and other visual forms.

- Quantitative characteristic rules

– Mapping generalized result into characteristic rules with quantitative

information associated with it



Presentation—Generalized Relation

location	item	sales (in million dollars)	count (in thousands)
Asia	TV	15	300
Europe	TV	12	250
North_America	TV	28	450
Asia	computer	120	1000
Europe	computer	150	1200
North_America	computer	200	1800

Presentation—Crosstab

location \ item	TV		computer		<i>both_items</i>	
	sales	count	sales	count	sales	count
Asia	15	300	120	1000	135	1300
Europe	12	250	150	1200	162	1450
North_America	28	450	200	1800	228	2250
<i>all_regions</i>	45	1000	470	4000	525	5000

Implementation by Cube Technology

Construct a data cube on-the-fly for the given data

mining query

– Facilitate efficient drill-down analysis

– May increase the response time

– A balanced solution: precomputation of “subprime

relation

• Use a predefined & precomputed data cube

– Construct a data cube beforehand

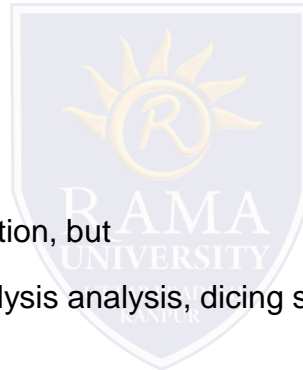
– Facilitate not only the attribute-oriented induction, but

also attribute relevance analysis, dicing, slicing, roll-up

and drill-down

– Cost of cube computation and the nontrivial storage

overhead



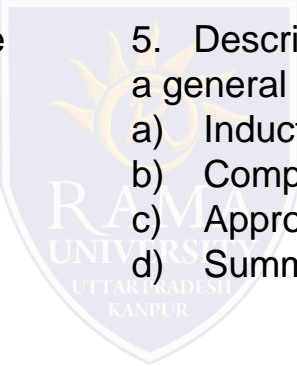
Multiple Choice Question

1. Various visualization techniques are used in _____ step of KDD.
 - a) selection
 - b) transformaion
 - c) data mining.
 - d) interpretation.

 2. Extreme values that occur infrequently are called as _____.
 - a) outliers
 - b) rare values.
 - c) dimensionality reduction.
 - d) All of the above.

 3. Box plot and scatter diagram techniques are _____.
 - a) Graphical
 - b) Geometric
 - c) Icon-based.
 - d) Pixel-based.

 4. _____ is used to proceed from very specific knowledge to more general information.
 - a) Induction
 - b) Compression.
 - c) Approximation.
 - d) Substitution.

 5. Describing some characteristics of a set of data by a general model is viewed as _____.
 - a) Induction
 - b) Compression
 - c) Approximation
 - d) Summarization
- 
- The watermark is a shield-shaped logo for Rama University. It features a stylized sun or flame symbol at the top, with the text 'RAMA UNIVERSITY' in the center and 'UTTAR PRADESH KANPUR' at the bottom.

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- <https://www.geeksforgeeks.org/>
- <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf> DATA MINING BOOK WRITTEN BY Micheline Kamber
- <https://www.javatpoint.com/three-tier-data-warehouse-architecture>
- M.H. Dunham, “ Data Mining: Introductory & Advanced Topics” Pearson Education
- Jiawei Han, Micheline Kamber, “ Data Mining Concepts & Techniques” Elsevier
- Sam Anahory, Dennis Murray,” data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, “ Pearson Education
- Mallach,” Data Warehousing System”, TMH
- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE’97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997
- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB’96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD’97
- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.
- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.