शिक्षा मंत्रालय
MINISTRY OF
EDUCATION

INDIAN INSTITUTE OF TECHNOLOGY
JODHPUR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NPTEL

PMRF
Prime Minister's Research Fellowship

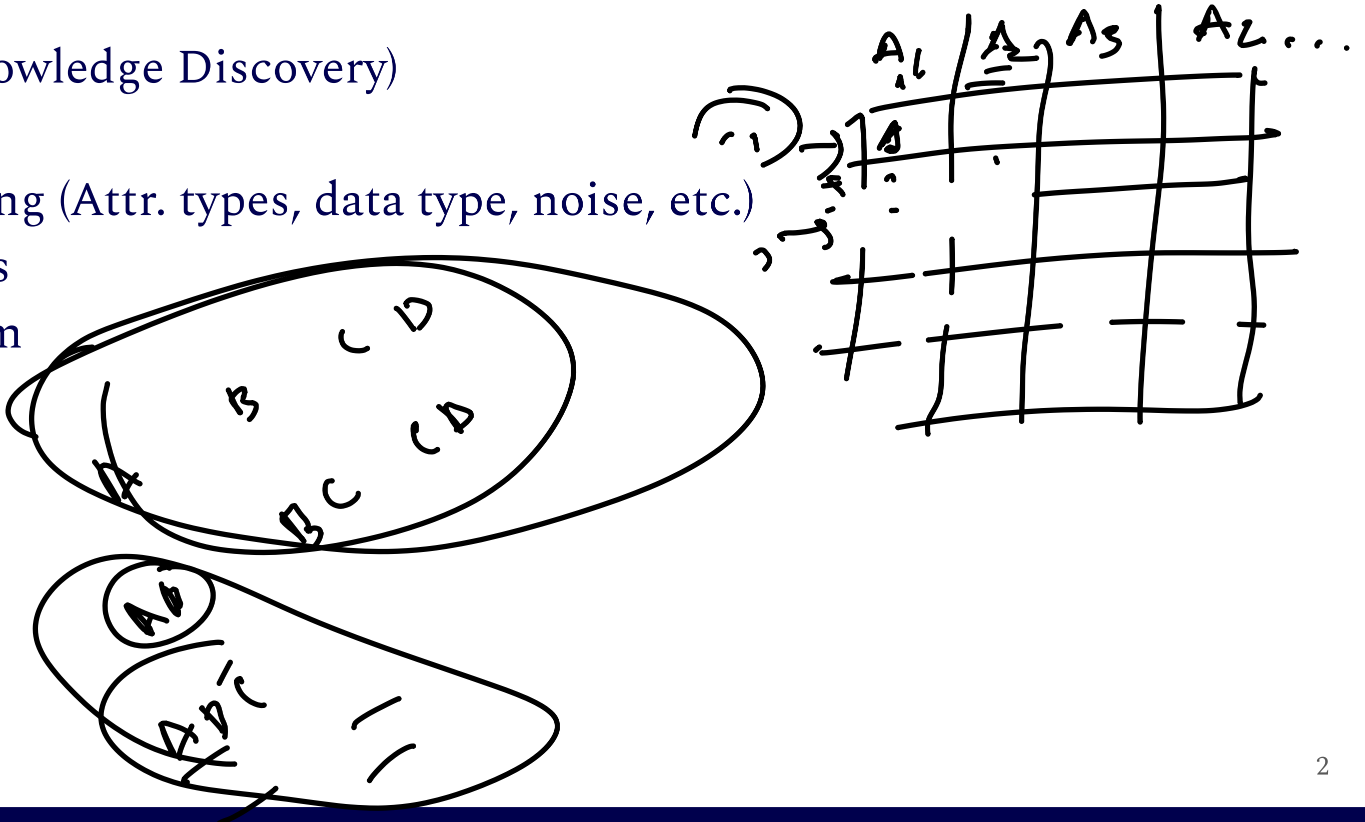
Week 1 - Live Session

Data Mining

Swapnil S. Mane
mane.1@iitj.ac.in
PMRF Research Scholar

Summary Week 1

- Data Mining (Knowledge Discovery)
- KDD Process
- Data Preprocessing (Attr. types, data type, noise, etc.)
- Association Rules
- Apriori Algorithm



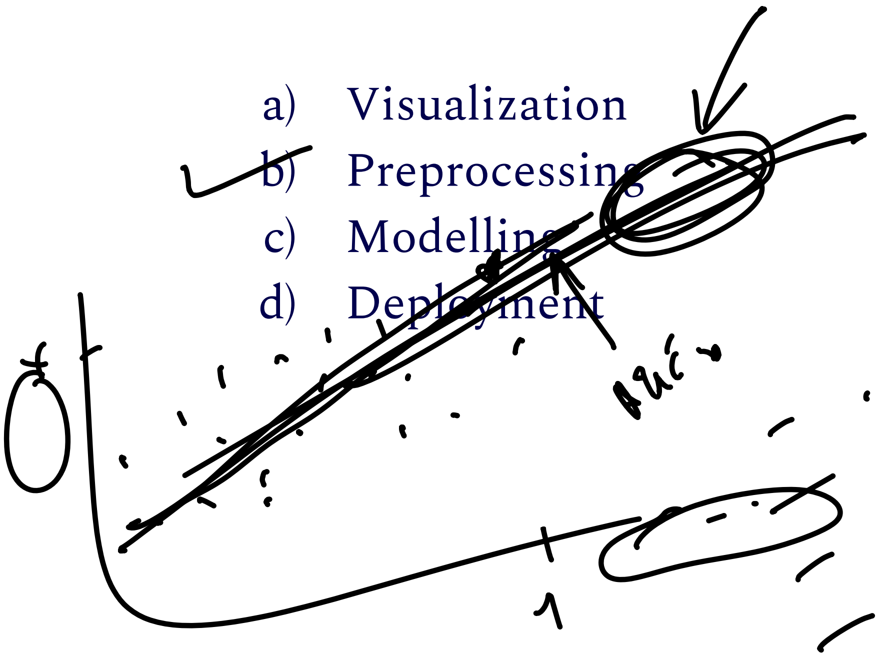
Q1. The earliest step in the data mining process is usually?

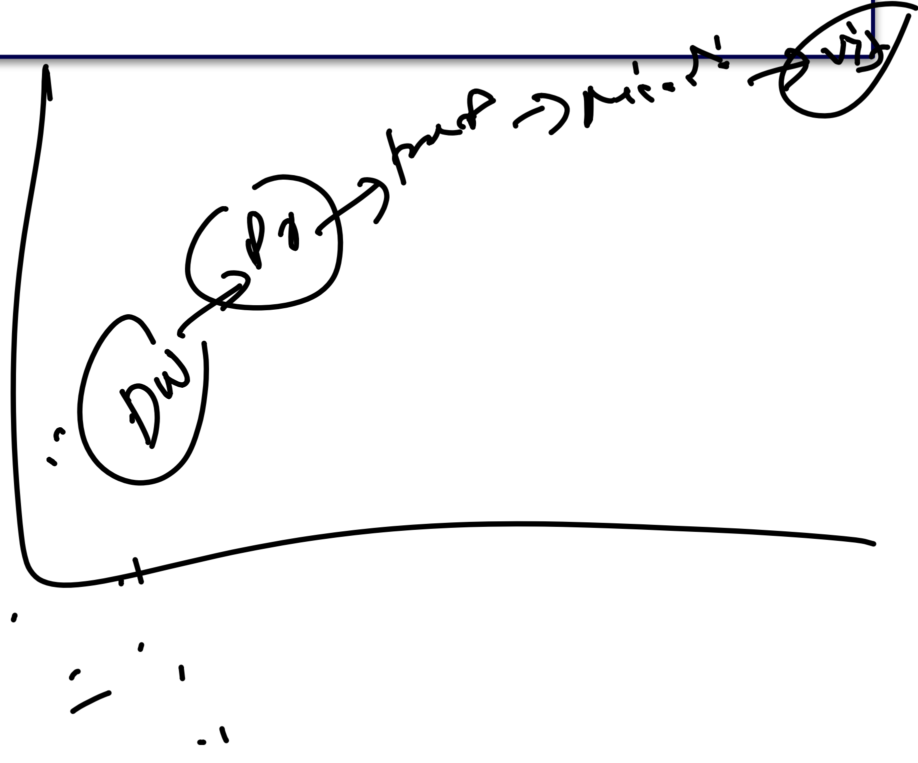
a) Visualization

[✓ b) Preprocessing]

c) Modelling

d) Deployment





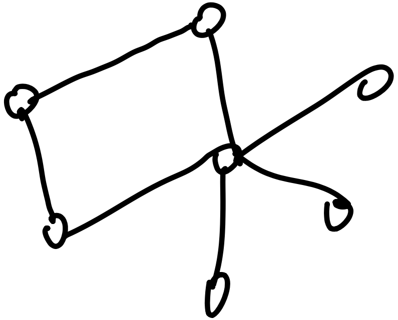
Q2. Which of the following is an example of a continuous attribute?

[✓ a)] Height of a person
b) Name of a person
c) Gender of a person
d) None of the above

Q3. Friendship structure of users in a social networking site can be considered as an example of:

a) Record data
b) Ordered data
c) Graph data
d) None of the above

G: {V, e}
V = Users
e = Friendship



Q4. Name of a person, can be considered as an attribute of type?

a) Nominal
b) Ordinal
c) Interval
d) Ratio

P1	DoE	Tx ~~DO~~	~~BO~~ 25 Jan
P2			26 Jan
P3			26 Jan

5-
5?x

Q5. A store sells 15 items. The maximum possible number of candidate 2-itemsets is:

a) 120

b) 105

c) 150

d) 2

$$^{15}C_2 = \frac{15 \times 14}{2 \times 1} = \frac{210}{2} = 105$$

A Store = 30 items = 15

t1	m, b
t2	ca, mobi
t3	

15 = {A, B, C, D, ... O}

$\frac{AB}{2}$, $\frac{BC}{2}$, $\frac{CD}{2}$, ...

$^{15}C_2$

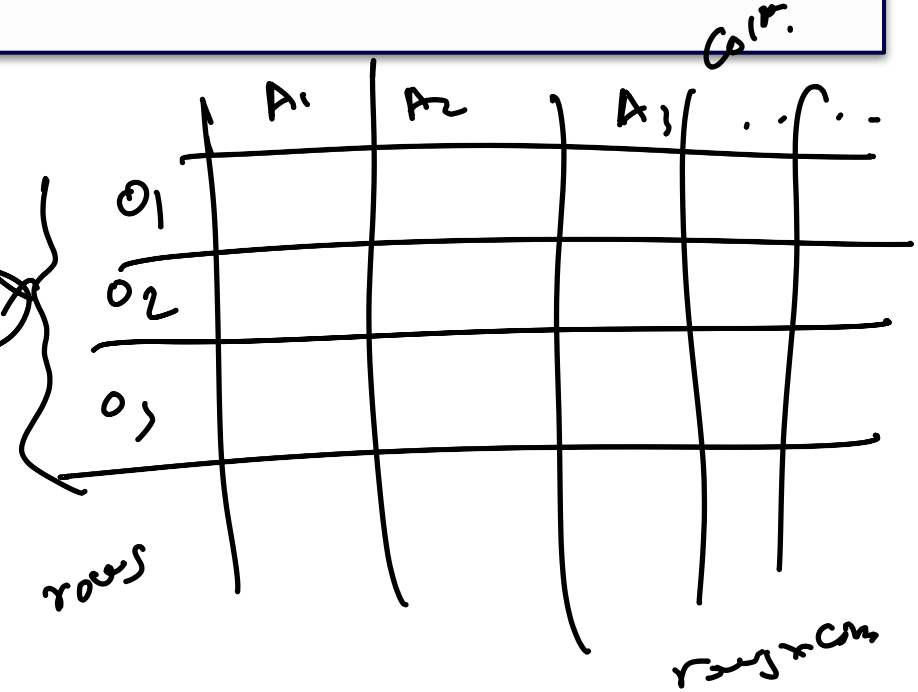
Q6. If a record data matrix has a reduced number of rows after a transformation, the transformation has performed:

a) Data Sampling

b) Dimensionality Reduction

c) Noise Cleaning

d) Discretization



Answer Q7-Q10 based on the following table:

CID	MB
1	{adebc}
2	{abcde}
3	{bcde}
4	{abcd}
5	{abde}

Customer ID	Transaction ID	Items Bought
1	1	{a,d,e}
1	2	{a,b,c,e}
2	3	{a,b,d,e}
2	4	{a,c,d,e}
3	5	{b,c,e}
3	6	{b,d,e}
4	7	{c,d}
4	8	{a,b,c}
5	9	{a,d,e}
5	10	{a,b,e}

5

e	bd	bde
1	1	1
1	1	1
1	1	1
0	1	0
1	1	1

Q7. Taking transaction ID as a market basket, support for each itemset {e}, {b,d}, and {b,d,e} is:

[✓ a) 0.8, 0.2, 0.2]

b) 0.3, 0.3, 0.4

c) 0.25, 0.25, 0.5

d) 1,0,0

$$\{e\} = \frac{\sigma(\{e\})}{T} = \frac{8}{10} = 08$$

$$\{b,d\} = \frac{\sigma(b,d)}{T} = \frac{2}{10} = 0.2$$

$$\{b,d,e\} = \frac{\sigma(b,d,e)}{T} = 0.2$$

Q8. Based on the results in (7), the confidence of association rules {b,d}->{e} and {e}->{b,d} are:

$$T_1 = \frac{\sigma(b,d,e)}{\sigma(b,d)} = \frac{2}{2} = 1$$

$$T_2 = \frac{\sigma(b,d,e]}{\sigma(e\}} = \frac{2}{8} = 0.25$$

a) 0.5, 0.5

[✓ b) 1, 0.25]

c) 0.25, 1

d) 0.75, 0.25

Q9. Repeat (7) by taking customer ID as market basket. An item is treated as 1 if it appears in at least one transaction done by the customer, 0 otherwise. Support of itemsets {e}, {b,d}, {b,d,e} are:

a) 0.3, 0.5, 0.2

b) 0.8, 1, 0.2 —

c) 1, 0.2, 0.8

[✓ d)] 0.8, 1, 0.8 —

$$\{e\} = \frac{\sigma(\{e\})}{CT} = \frac{\boxed{4}}{5} = 0.8$$

$$\{b,d\} = \frac{\sigma(\{b,d\})}{CT} = 5/5 = 1$$

$$\{b,d,e\} = \frac{\sigma(\{b,d,e\})}{CT} = \frac{4}{5} = 0.8$$

Q10. Based on the results in (9), the confidence of association rules {b,d}->{e} and {e}->{b,d} are:

[✓ a) 0.8, 1]

b) 1, 0.8

c) 0.25, 1

d) 1, 0.25

$$C_{R_1} = \frac{\sigma(\{bde\})}{\sigma(\{b,d\})} = \frac{4}{5} = 0.8$$

$$C_{R_2} = \frac{\sigma(\{bde\})}{\sigma(\{e\})} = \frac{4}{4} = 1$$