CSE- 453

Name: AYON ROY
ID : 201714018
Dept: CSE-17
Sec : B

MID

Ayon Roy
2017140018

Ans. to the ques. no.-01

Knowledge discovery procers consists of serenal steps including : Data cleaning, and Data Integration of serenal Databases and then storing on a Data warehouse and then selecting the Task-Relevant Datas and after having the Task-Relevant Data applying Data Mining Techniques we can visualyze or analyze the Patters in the data and after that we can discover the knowledge of ~~#~~ from the data.

Where, Data Mining is the procers of extracting knowledge ~~for~~ from data.

So, Data Mining is a part of the

Ayon Roy
2017140018

Knowledge Discovery process (KDD).

Data Mining plays an essential role in the Knowledge Discovery process.

Basically, Data Mining is the process where we extract the knowledge or information from some relevant data.

Where, KDD is the overall process from generating or collecting those relevant data to the extracting knowledge.

Analyzing the architecture of Data ~~Mining~~ Mining system is shown on the next page!

P.T.O.

Ayon Roy
20171401 8

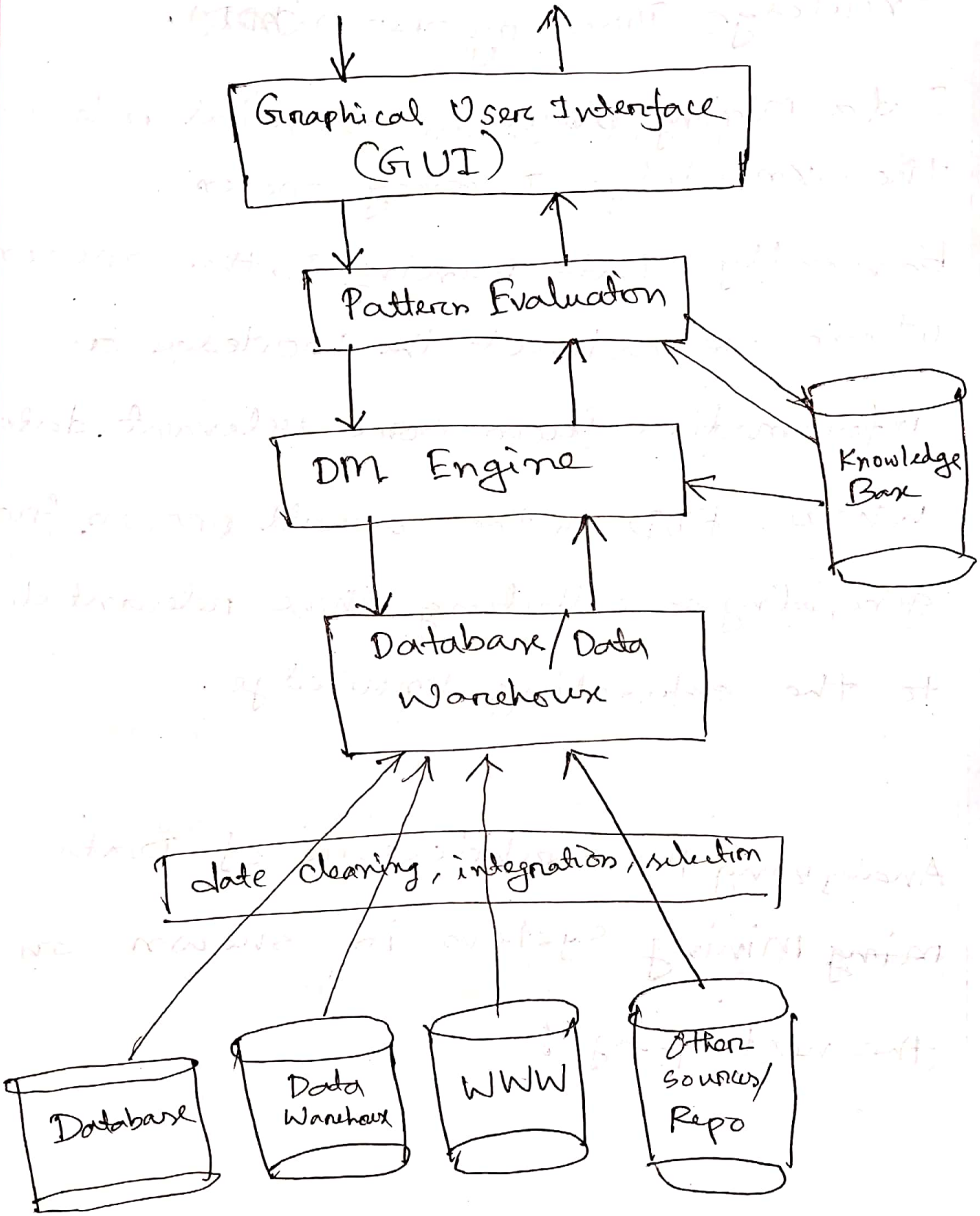


Fig: Architecture of a typical Data Mining System.

Ayon Roy
20171401B

The typical architecture of a Data Mining System first collects, cleans and integrates from Databases, internet(WWW) or other data sources and then stores to a Task relevant Database or Data Warehouse. With those data a Data Mining Engine then passes the correct formatted(Typed) Data to the Pattern Evaluation. Patten Evaluation with the help of knowledge bases shows (GUI) the underlying or extracted knowledge/information of the Data.

P.T.O

Ayon Roy
2017140018

Ans. to the ques. no.- 02(c)

Given,

Minimum support count = 2

and, the Transactional data base ;
(using Aprioni Algorithm)

T.id	items
T1	I5, I2, I3, I1
T2	I5, I3, I1
T3	I5, I3, I4
T4	I5, I2, I3
T5	I5, I2
T6	I2, I3
T7	I5, I2
T8	I3, I4
T9	I2, I3

1st scan

C_1

item	sup
I1	2
I2	6
I3	7
I4	2
I5	6

L_1

item	sup
I1	2
I2	6
I3	7
I4	2
I5	6

2nd scan

C_2

item	sup
I1 I2	1
I1 I3	2
I1 I4	0
I1 I5	2
I2 I3	4
I2 I4	0
I2 I5	4
I3 I4	2
I3 I5	4
I4 I5	1

L_2

Item	sup
I1 I3	2
I1 I5	2
I2 I3	4
I2 I5	4
I3 I4	2
I3 I5	4

(Join & Prune)

C_3

itemset
I1 I3 I5
I2 I3 I5

Ayon Roy
2017140018

① $I_1 I_3 \bowtie I_1 I_5 = \{I_1 I_3 I_5\}$
$= \{I_1 I_3\}\{I_1 I_5\}\{I_3 I_5\}$ ✓ ✓ ✓

② $I_2 I_3 \bowtie I_2 I_5 = \{I_2 I_3 I_5\}$
$= \{I_2 I_3\}\{I_2 I_5\}\{I_3 I_5\}$ ✓ ✓ ✓

③ $I_3 I_4 \bowtie I_3 I_5 = \{I_3 I_4 I_5\}$
$= \{I_3 I_4\}\{I_3 I_5\}\{I_4 I_5\}$ ✓ ✓ ✗

So,
all-3-frequent itemsets are:

		sup
①	$\{I_1 I_3 I_5\}$ →	2
②	$\{I_2 I_3 I_5\}$ →	2

Ayon Roy
20171401B

Ans. to the ques. no.- 02(b)

Various types of numeric attributes are :

(1) Interval
(2) Ratio

(1) Interval :

(a) It is measured on a scale of equal-sized-units.

(b) It has no zero-point.

(c) Example : temp in °C/°F and Calender dates etc.

(2) Ratio :

(a) It has an order of magnitude which is larger than unit of measurement.

(b) It has inherent zero-point.

(c) Example : length, age etc.

Ayon Roy
20171401B

Ans. to the ques. no.- 02(a)

Box-plot represents Five number summary:

min, Q_1, median, Q_3, max.

Data is represented in with a box.

Given data are:

1, 1, 2, 2, 4, 6, 6.8, 7.2, 8, 8.3, 9, 10, 10, 10, 11.5

Min = 1

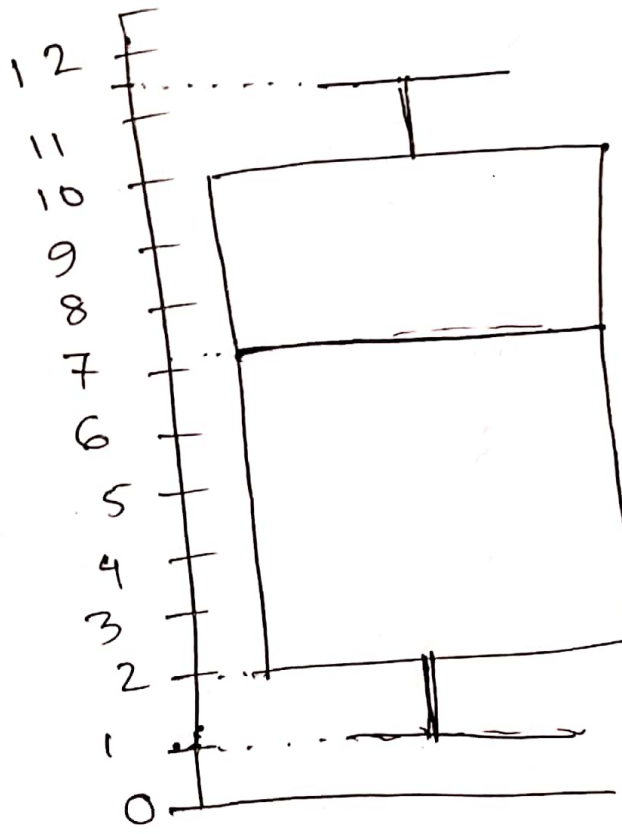
Max = 11.5

Q_1 = 2

Q_3 = 10

IQR = 8 = (10 - 2)

Median = 7.2



Box-plot