



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA

Approved by AICTE, Ministry of HRD, Government of India

Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

Data mining - Concept and Techniques

DMCT

5AID3-01

MM-100

(IA-30, ETE-70)

End Term Exam - 3 hr

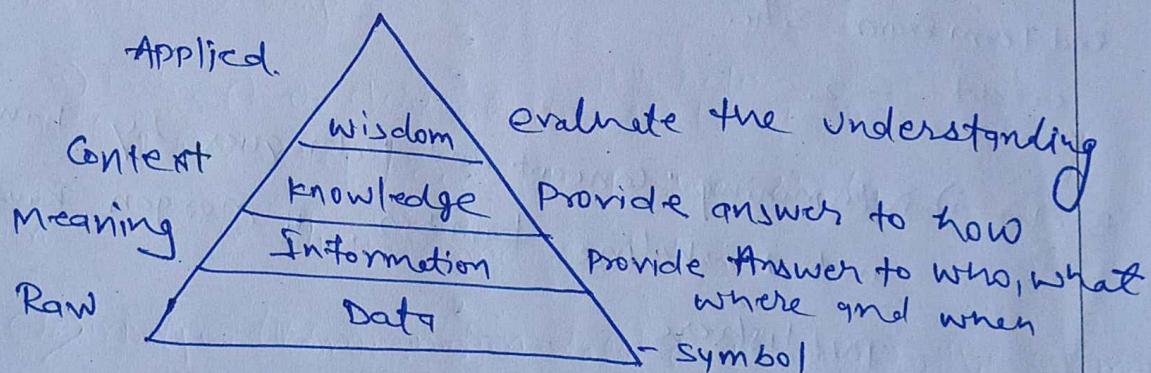
Credit - 2

Books - i) Data mining : Concept and Techniques by Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers.

ii) Data mining and Data warehousing by Udit Agarwal, KAISON & SONS publication.

iii) Introduction to Data mining , Second edition, Pearson, 2019, Poole-Ning Tan Michael Steinbach, Anuj Karpatne

Data mining :— DM means extracting or mining the knowledge from the large ~~is~~ amount of Data. DM is basically an essential step in the process of knowledge discovery.



Ex:

- i) marketing — users what interest them based on their searches.
- ii) Retail — purchasing pattern.
- DM also detect which offer are most valued by customers.
- iii) Banking — for understand market risk, anti fraud system.
- iv) medicine — DM enable more accurate diagnostics.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

- v) Television & Radio — valuable knowledge for their advertisement
- vi) E-commerce — "Frequently bought together" in amazon & flipkart.
- vii) Crime prevention
- viii) Research.

Step's of Data ^{mining} Processing :-

- i) Data collection — In Row form & valid source
- ii) Data cleaning — Remove noise data & inconsistency
- iii) Data Integration — multiple data combined.
- iv) Data selection — relevant data selection.
- v) Data Transformation — change in appropriate format.
- vi) pattern evaluation —
- vii) knowledge representation —



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

⇒ Data mining functionality :—

① Concept / class Description :—

Data is always associated with class/concept such descriptions of a class or a concept are called class/concept description. These description can be done in 2 ways—

a) Data characterization — It is refer to the summary of the class/concept of the study.
o/p — General overview.

b) Data discrimination — It is compare the common feature of the class.
o/p — bar, chart, curves, graph etc.

② Mining Frequent pattern :—

Things which are found most commonly in data. The data may be item set, sub sequences and substructure.



Swami Keshvanand Institute of Technology, Management & Gramothan,

Ramnagar, Jagatpura, Jaipur-302017, INDIA

Approved by AICTE, Ministry of HRD, Government of India

Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

③ Association rule ^{Analyisis} — It is a way of identify the relation b/w various items.

Ex: used to determine rules of item that are frequently purchase together.

④ Correlation Analysis —

— mathematical techniques

— show how strongly pair of attributes are related together.

Ex: Tall people tend to have more weight

①

②

⑤ Classification and Regression for predictive Analysis : —

prediction of missing data.
wrong

↳ process of finding a model that different data items.

— decision Tree is used for classification.

Regression — statistical methodology that is used for numeric prediction for missing data. and it is done based on previous data.

(5)



Swami Keshvanand Institute of Technology, Management & Gramothan,

Ramnagar, Jagatpura, Jaipur-302017, INDIA

Approved by AICTE, Ministry of HRD, Government of India

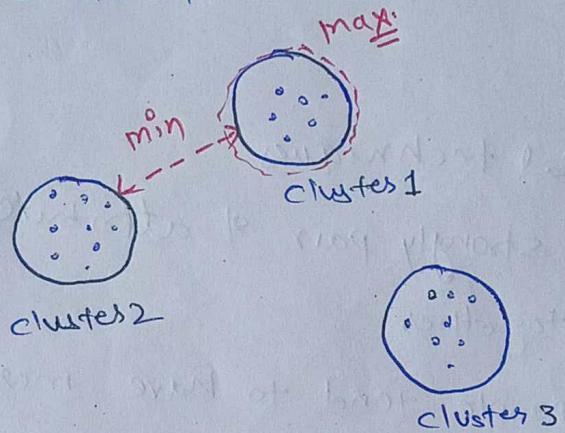
Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

⑥ Cluster Analysis:-

The data items are clustered based on the principle of maximizing the intra-class (within same) similarity & minimizing the inter-class (bw two different class) similarity.



⑦ Outlier Analysis:-

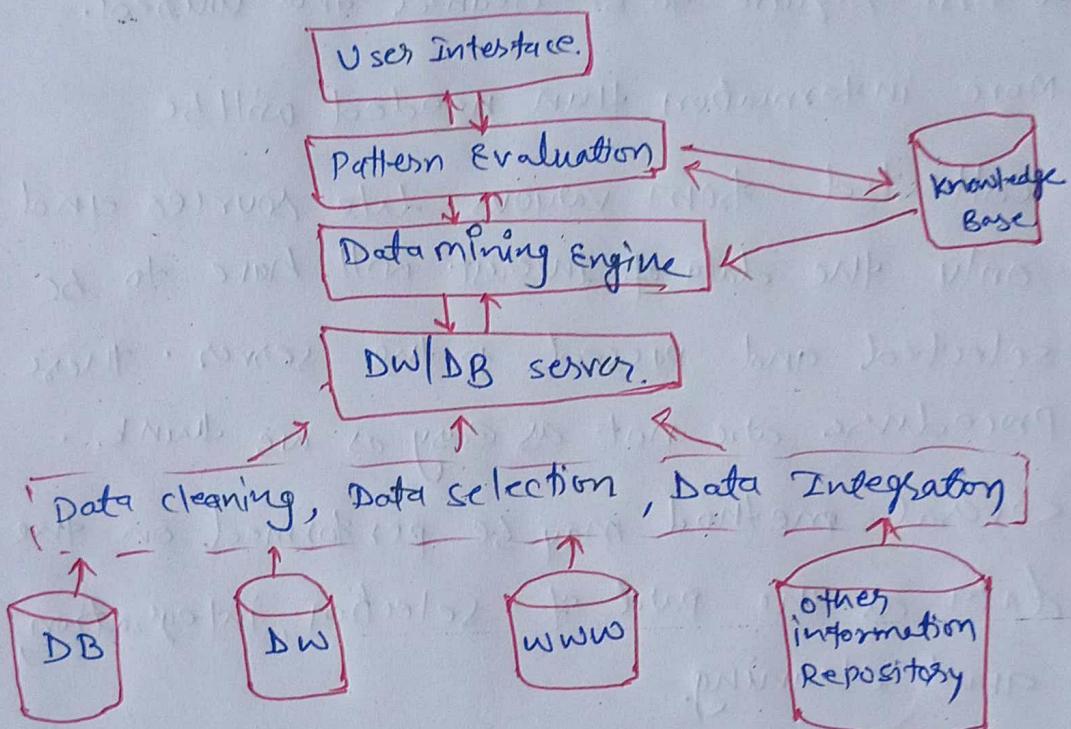
Among the data items in a Database, there may be some items, which do not follow the general behaviour of data.

$$\underline{\text{ex}} \quad \{2, 4, 6, 7, 8, 10, 12, \dots\}$$

{7 is outlier}

Data mining Architecture :-

The significant component of data mining system are a data source, data mining engine, DW server, the pattern evaluation engine module, GUI and knowledge base.



⇒ Data Source — The actual source of data is the DB, DW, WWW text files, and other ~~files~~ documents.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA

Approved by AICTE, Ministry of HRD, Government of India

Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

⇒ different process — Before passing the data to the DB or DW server, the data must be cleaned, integrated and selected.

As the information comes from various sources and different format, it can't be used directly for the DM procedure.

because the data may not be complete and accurate.

so first ^{the data} require to be cleaned and unified.

More information than needed will be

collected from various data sources and only the data of interest will have to be selected and passed to the server. These procedures are not as easy as we think.

several methods may be performed on the data as the part of selection, integration, and cleaning.

⇒ DB or DW servers — It consists the original data that is ready to be processed. Hence the server is cause for retrieving the relevant data that is based on DM as per user's request.



Swami Keshvanand Institute of Technology, Management & Gramothan,

Ramnagar, Jagatpura, Jaipur-302017, INDIA

Approved by AICTE, Ministry of HRD, Government of India

Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

→ DM Engine — It is the major component of any DM system. It contains several modules for operating DM tasks including association, characterization, classification, clustering, prediction, time series analysis etc.

→ Pattern Evaluation — This module is primarily responsible for the measure of investigation of the pattern by using a threshold value. It collaborates with the DM engine to focus the search on exciting patterns.

User Interface — It is communication b/w the DM system and user. This module helps the user to easily and efficiently use the system without knowing the complexity of the process. This module cooperates with the DM system when the user specifies a query or a task and display the result.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

- knowledge base → the knowledge base is helpful in the entire process of DM.

It might be helpful to guide the search or evaluate the stack of the result pattern.

It may contain user view and data from user experiences that might be helpful in the DM process. DM engine may receive the help from the knowledge base to make the result more accurate and reliable.

(1)
 (2)
 (3)
 (4)
 (5)
 (6)
 (7)
 (8)
 (9)
 (10)

Classification of Data mining system —

- ① classification based on the mined Database
- ② " " " " " knowledge
- ③ " " " " statistics
- ④ " " " " machine learning
- ⑤ classification based on visualization
- ⑥ " " " " on information science.
- ⑦ " " " " utilized techniques
- ⑧ " " " " adapted application.

1) classification Based on the mined Database

- classify as the ~~DB~~. Type of Database that have been mined
 - Relational
 - Transactional
 - object relational
 - Data warehouse

2) classification based on type of knowledge mined -

- characterization
- discrimination
- Association and correlation analysis.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

- classification
- prediction
- Outlier Analysis
- evaluate Analysis

iii) classification based on kind of techniques used -

- ML, statistics, Neural Network, pattern recognition,
- data ware house, oriented techniques. etc.

4) classification based on application adapted.

- Finance
- Telecommunication
- DNA
- stock market
- email etc.



**Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA**

Approved by AICTE, Ministry of HRD, Government of India

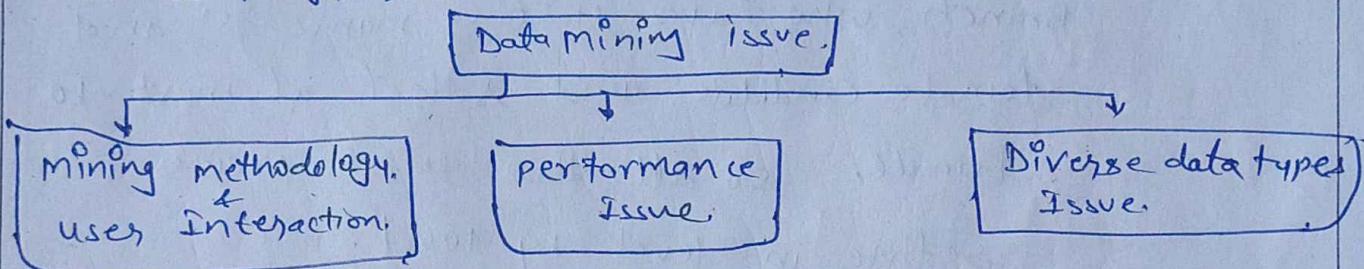
Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

MAJOR ISSUE IN DATA MINING:

DM is not an easy task, as the algorithms used can get very complex. and data is not always available at one place. So we need data integration from various sources. These factors also create some issue - there are some major issue.



① mining methodology & user interaction Issues →

It refers to different kind of issue.

- a) Mining different different kind of knowledge in Database:— Different users may be interested in different kind of knowledge. Therefore it is necessary for DM to cover a broad range of knowledge discovery range.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

b) Interactive mining of knowledge at multiple levels of abstraction -

The DM process needs to be interactive because users to focus the search for pattern, providing and refining the DM request based on the returned result.

Ex If we want to search roll No. from AI Branch, who have starting name 'S' and female candidate, and select at most 10 female. So all ~~this~~ these process define in level by level.

c) Incorporation of background knowledge of domain -

To guide discovery process and to express the discovered pattern, the background knowledge can be used. {Background knowledge required,

Ex passing any exam required notes, PPT, Previous Years' Questions etc.}

d) presentation and visualization of data mining result -

Once the patterns are discovered it needs to be expressed in high level language and visual representation. These representation should be easily understandable.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

e) Handling noisy or incomplete data -

The data cleaning method are required to handle the noise and incomplete objects, while mining the data regularities. If the data cleaning method are not there then the accuracy of the discovered patterns will be poor.

f) Pattern evaluation - The pattern discovered should be interesting because either they represent common knowledge or lack novelty.

② Performance Issues -

a) efficiency and scalability of data mining algorithms. → In order to effectively extract the information from huge amount of data in Database, DM Algorithms must be efficient and scalable.

Ex: If any ~~any~~ algo implement on 100 data, or it may be 1000 data, this algo. work as well as on 10000 data, ~~some operation~~.



⇒ parallel, distributed and incremental mining algorithms -

The factors such as huge size of DB, wide distribution of data and complexity of DM methods motivate the development of parallel, distributed DM algo. These algo divide the data, into partition which is further processed in a parallel fashion.

Then the result from the partitions is merged.

The incremental algorithms, update the DB without mining the data again from scratch.

③ Diverse Data Type Issues -

a) Handling of relational and complex type of data -

The DB may contain complex data object, multimedia object, spatial data, temporal data objects etc., it is not possible for one system to mine all these kind of data.



Swami Keshvanand Institute of Technology, Management & Gramothan,

Ramnagar, Jagatpura, Jaipur-302017, INDIA

Approved by AICTE, Ministry of HRD, Government of India

Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

⑥ Mining Information from heterogeneous DB and global information system — The data is available at different data source on LAN or WAN. These data source may be structured, semi structured or unstructured. Therefore mining the knowledge from them adds challenges to data mining.

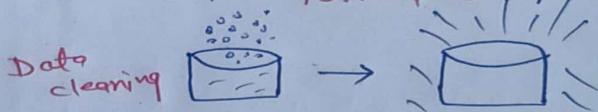
Data Processing ^{pre} : — The process of transforming raw data into an understanding format.

Data preprocessing phases focus on the preprocessing steps that produce the data to be mined. It is one of the most important step in data mining.

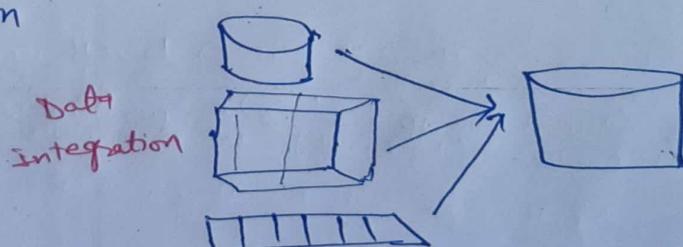
The mined result are much more accurate. This means that this step is also very critical for success of D.M. method.

Data mining preprocessing involves many methods of data preprocessing ↓

i) Data cleaning



ii) Data Integration

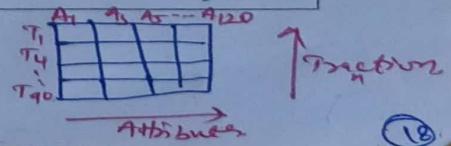


iii) Data Transformation

Data Transformation
 $-2, 63, 100, 59, 48 \rightarrow -0.02, 0.63, 1.00, 0.59, 0.48$

	A ₁	A ₂	A ₃	... A ₁₂₀
T ₁				
T ₂				
T ₃				
T ₄				
T ₅				
T ₆				
T ₇				
T ₈				
T ₉				
T ₁₀				

Data reduction



(18)



Swami Keshvanand Institute of Technology, Management & Gramothan,
Rammagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

i) Data cleaning — It is define to clean the data by filling in the missing value, smoothing noisy data, analysing and removing inconsistency in the data.

There are various type of data cleaning

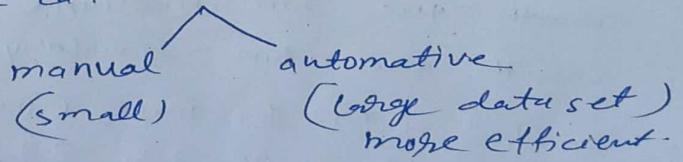
- a) missing value
- b) Noisy Data
- c) Inconsistency data.

a) Handling & Missing value — In place of missing value, we can replace with ~~NA~~.

" " mean value (Normal Distribution)
" " median value (Non normal distribution)

- Sometimes replaced with most probable values.

- missing value can be filled in ② ways-





Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

b) Noisy data — Noise is a random error or variance in a measured variable.

There are some common methods for removing noisy data.

- Binning
- Regression
- Clustering
- Combined computer and human inspection.

Binning — In this method, smooth the sorted data value by consulting the 'neighborhood' of value around it. The sorted value are distributed into a number of 'bucket' or bins.

There are three method for smooth data values —

- smooth the bin means. (Each bin value is replaced by mean of value)
- smooth the bin median. (" " " median of value)
- smooth the bin boundary. (" " " boundary value) by closest

Ex Let's sorted data for price (in dollars).

3, 5, 10, 20, 20, 23, 27, 29, 34

Step 1 Partition value in equal bins —

Bin 1 3, 5, 10

Bin 2 20, 20, 23

Bin 3 27, 29, 34



Swami Keshvanand Institute of Technology, Management & Gramothan,
Rammagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Step-2:- Smoothing data by bin means -

Bin 1 : 6, 6, 6

Bin 2 : 21, 21, 21

Bin 3 : 30, 30, 30

Step 3 :- Smoothing by bin boundary

Bin 1: 3, 3, 10

Bin 2: 20, 20, 23

Bin 3: 27, 27, 34

ii) Regression — Data can be smoothed by fitting

~~predict value~~ data in a function, such as with regression.

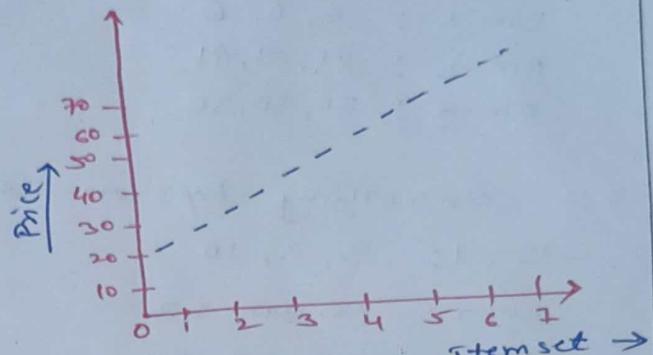
$$\text{Ex- } y = mx + c$$

- linear regression — find the best line to fit two attributes. So one attribute can be used to predict the other.
- multiple regression — It is the extension of linear regression, where more than two variables are involved and the data are fit to a multidimensional surface.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

a different itemset	P-Price.
1	16
6	56
5	40
7	60
4	40
2	35
3	48



a)

b)

These are some regression method -

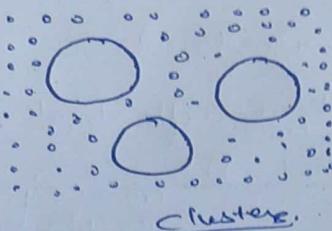
- i) multiple linear regression.
- ii) logistic regression.
- iii) neural net modelling.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

iii) clusters:-

- Outliers may be detected by clustering - where similar data or objects are organized into a group or 'clustered'.
- The values which fall outside of the set of clusters may be considered outliers. So clustering is a method of grouping data into different groups. So that data in each group share similar trends and patterns.



iv) Combined Computer and Human inspection - Outliers may be identified through a combination of computer and human inspection.

These methods find the suspicious value using the computer's program and then they are verify by human experts. By this process all outliers are checked.



Swami Keshvanand Institute of Technology, Management & Gramothan,

Ramnagar, Jagatpura, Jaipur-302017, INDIA

Approved by AICTE, Ministry of HRD, Government of India

Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

c) Inconsistent Data —

There may be inconsistencies in the data recorded for some transactions.

It is a situation where there are multiple tables within a DB that deal with the same data but may receive it from different inputs.

⇒ How to minimize the data inconsistency



→ Central semantic store

→ master reference store



② Data Integration — Data integration combine the data from multiple sources into a coherent data store.

$$\xrightarrow{\text{dm} + \text{dw/db}}$$

Integration → association / combining.

If there are no integration — no communication with db.

There are four schemes for integration —

① No coupling → - No dm & No db/dw
- dm system will not use any function.

→ Loose Coupling — will use some of the functionalities (only upto some extent)
- better than No coupling
- suitable for small dataset

upto 50%

Some Coupling → linked to be db/dw — 25%

- some of the dm primitives are also implemented in db.

Tight Coupling → - dm system is completely linked to db/dw
- most efficient method among all.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Rammagarin, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

③ Data Transformation: — It is a data preprocessing techniques that transformed or consolidated the data into alternate form for mining.

There are some process for transformation-

a) Smoothing — remove the noisy data
techniques - Binning, clustering, regression.

b) Aggregation — where summary or aggregation operations are applied to the data.

Ex: The daily sales data may be aggregated so as to compute monthly and annual total amount.
— It is construct a data cube for analysis the data at multi~~s~~ dimensional.

c) Generalization — In this, the low level or primitive (raw) data are replaced by higher level concept through the use of concept hierarchies.
Ex: In address, the street is replaced by city names.



Swami Keshvanand Institute of Technology, Management & Gramothan,

Ramnagar, Jagatpura, Jaipur-302017, INDIA

Approved by AICTE, Ministry of HRD, Government of India

Recognized by UGC under Section 2(f) of the UGC Act, 1956

Tel : +91-0141-3500300 Fax: +91-0141-2759555

E-mail: info@skit.ac.in Web: www.skit.ac.in

d) Normalization — where the attributes values are normalized by scaling their value so that they fall in specified range.

Ex: 2,40,500, 1,50, 900 -----

we can normalize the data b/w 0 and 1

- all data range b/w 0 and 1.

it is scale by any formula

There are two methods —

i) Min-Max Normalization —

$$v' = \frac{v - \min_x}{\max_x - \min_x}$$

Ex: 2 is replaced by

$$= \frac{2-1}{900-1} = \frac{1}{899}$$

v' new value

v = originally attribute value

\min_x = minimum value of attribute

\max_x = max. value of "

ii) Z-score Normalization —

$$v' = \frac{v - \bar{x}}{\sigma_x}$$

v' = new value

v = original attribute value

\bar{x} = mean of attribute

σ_x = standard derivation of attribute

c) Attribute Construction — where new attributes are constructed and added from the given set of attributes to help the mining process.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

- 4) Data Reduction— It is a preprocessing techniques that helps in obtaining reduced representation of data set from the available data set.
- Integrity of the original data should even after reduction in data volume.
 - It should produce same analytics result as on original data.

There are some method for data reduction—

- 5) Data cube aggregation— Used to aggregate data in a simpler form.

Ex: Year 2015	
HY	Sales
HPI	500
HPL	300

Year 2016.

Year 2016	
HY	Sales
HPI	300
HPL	400

→

Years	Sales
2015	800
2016	700



**Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA**
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

- b) Dimension reduction —
- remove redundant attributes.
- remove weakly important attributes
- c) Data Compression — Encoding of data
reduce the size. (lossless) → compression
 (lossy) → method.
- d) Numerosity Reduction — data is replaced by estimated / alternative
attributes
Ex histogram, sampling
- e) Generalization and Concept hierarchy —
↓
replaced by range.
↑
higher level
conceptual.



**Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagar, Jagatpura, Jaipur-302017, INDIA**
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Question Bank

- Q.1. Describe Data Integration and Transformation in details [RTU 2016] 2015,
- Q.2. Write short notes on Dimensionality reduction [RTU 2016]
- Q.3. What is data mining, explain the form of Data preprocessing along with Data cleaning [RTU 2016] 2014
- Q.4. Explain the concept of Association and data generalization in details [RTU 2016] 2015, 2014
- Q.5. Explain the discretization and concept hierarchy in details [RTU 2015]
- Q.6. Describe the various data reduction strategies. [RTU 2015]
- Q.7. What are the different classification techniques. [RTU 2014]
- Q.8. Describe the data mining functionalities [RTU - 2013]
- Q.9. Explain the term data cleaning & missing value. [RTU 2013]
- Q.10. Discuss the Binning method with suitable example [RTU 2013]
- Q.11. What is data mining, explain its example. [RTU 2012, 2016]
- Q.12. Write down the steps of Data mining processing.
- Q.13. Draw & explain the Architecture of Data mining.
- Q.14. How to classify the Data mining system.



**Swami Keshvanand Institute of Technology, Management & Gramothan,
Raminagar, Jagatpura, Jaipur-302017, INDIA**
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel : +91-0141-3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Q.15 what are the major issues of Data mining

Q.16 How to smooth the following data set for Ben 3
by binning method.

3, 5, 10, 20, 20, 23, 27, 29, 34

Q.17 explain the data transformation.

Q.18 How to data reduction works, explain all the
method's which are used for reduction.

Q.19 - write difference b/w

- discrimination & classification
- characterisation & clustering
- classification & prediction