



Name of the Subject: DATA ANALYSIS & INFO. EXTR. Subject Code: IT 704

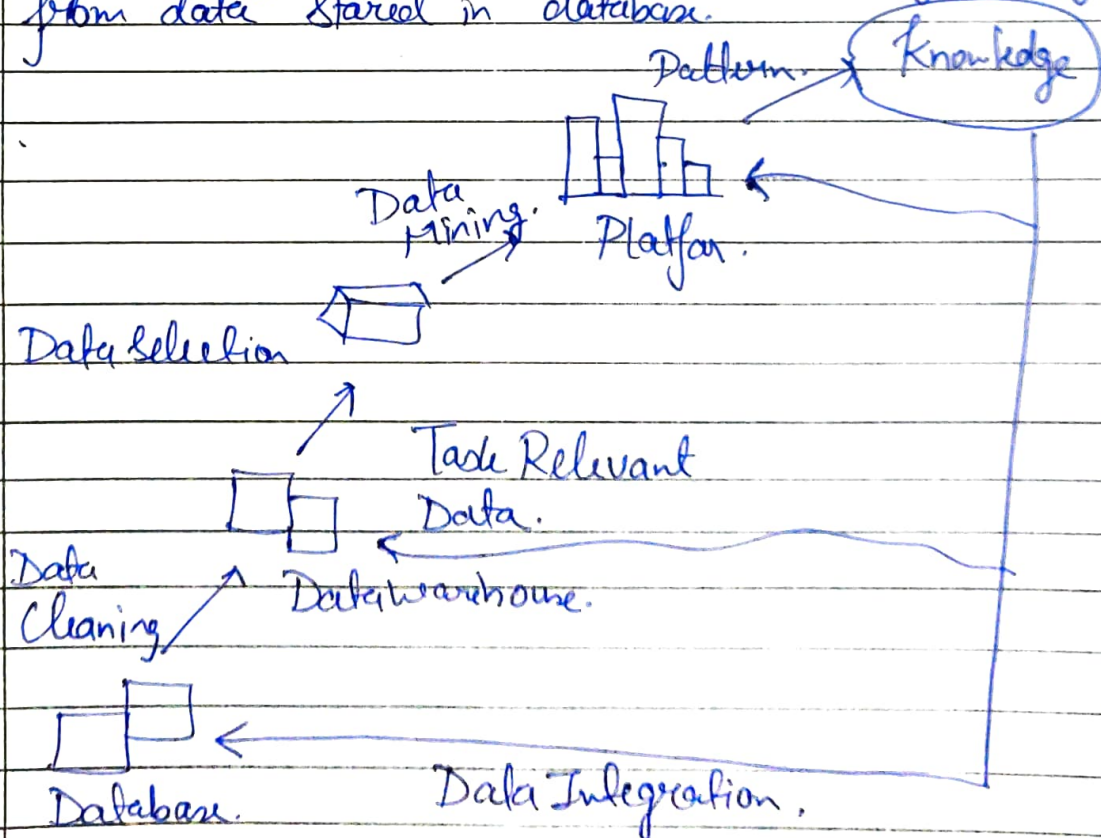
Seat No: IT076 Student ID: 18ITVB2116 Branch/Sem: IT-7

Q2 Attempt Any Two question.

a

KDD - Knowledge Discovery in Database.

- Volume of information is increasing everyday that we can handle from business transaction, & different data.
- So we need a system that will be capable of extracting essence of information available & that can generate report & making decision.
- It refers to the nontrivial extraction of implicit, previously unknown & potentially useful information from data stored in database.





Name of the Subject: DATA ANALYSIS & INFO. EXT. Subject Code: IT 704

Seat No: IT076 Student ID: 18ITURN116 Branch/Sem: IT-7

① Data Cleaning:- Removal of noisy & irrelevant data.

- Cleaning with Data discrepancy detection & Data Transformation tools

② Data Integration:- Heterogeneous data from multiple source combined in common source.

- Data Integration using ETL

③ Data Selection:- The process where data relevant to analysis is decided & retrieved from data collection.

- Data Selection using Clustering.

④ Data Transformation:- The process of transforming data into appropriate form required by mining.

- Data ~~Mapping~~ Mapping & Code generation.

⑤ Data Mining:- Techniques that are applied to extract patterns

⑥ Pattern Evaluation:- Identifying strictly increasing pattern representing knowledge base.

⑦ Knowledge Representation:- Technique which utilizes visualization tool to represent data mining result.



Name of the Subject: _____

DAIE

Subject Code: IT704

Seat No: IT076

Student ID: 18ITUBN116

Branch/Sem: IT-717

Q2

(b) Wavelet Transform

- Decomposes a signal into different frequency subbands, applicable to n-d signal.
- Data are transformed to preserve relative distance between objects at different levels of resolution.
- Also, used in Image Compression.
- It helps to shape filter which emphasizes regions where points cluster.
- Effective removal of outliers.
- And applicable to low dimensional data.

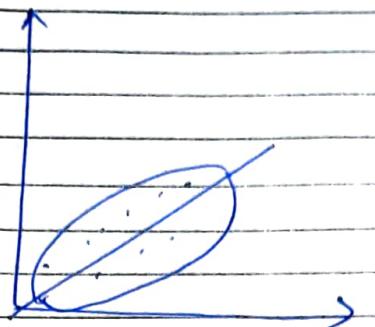
⇒ Principal Component Analysis

- It finds a projection that captures the target amount of variation in data.
- Steps to solve the PCA
 - Normalize input data.
 - Compute k orthonormal vectors.
 - Each input data is a linear combination of the principal.
- It works for numeric data only.



Name of the Subject: DAIE Subject Code: IT 704

Seat No: IT076 Student ID: 18ITUBN116 Branch/Sem: IT-7



— Projected onto a smaller space, resulting in dimensionality reduction.



Name of the Subject: DAIE Subject Code: IT704

Seat No: IT076 Student ID: 18ITUBN116 Branch/Sem: IT-7

Q3

a

Government Income Tax System.

(d) Dimension & Fact Table.

- Dimension

Branch - Branch-key
Branch-name.
Branch-type.

Tax - Tax-key
Tax-type.
Tax-year.
Tax-state.

Department - Department-key
Department-name.

Time - Time-key
day
day of the week
month
year.

Location - Location-key
Street
City
State
Country.

Due Tax - Due-key Tax-key.
Due-Tax-type.



Name of the Subject: DAIE Subject Code: IT-704
Seat No: IT-076 Student ID: 18ITURN116 Branch/Sem: IT-VII

