1. INTRODUCTION TO DATA MINING

1. Definition of Data Mining:

Data Mining is the process of **extracting useful, previously unknown, and potentially valuable patterns or knowledge** from large datasets.

It is also called **Knowledge Discovery from Data (KDD)**.

2. Data Mining Tasks:

a) Classification:

- Assigns data to predefined categories.
- Example: Email → spam or not spam.

b) Regression:

- Predicts a continuous numeric value.
- Example: Predicting house prices based on location, size, etc.

c) Clustering:

- Groups similar data points together without predefined labels.
- Example: Grouping customers based on shopping behavior.

d) Summarization:

- Provides a compact description or overview of the data.
- Example: Generating reports that summarize sales trends.

e) Dependency Modeling:

- Finds relationships between variables.
- Example: Market basket analysis (if someone buys bread, they are likely to buy butter).

2. Data Mining vs Knowledge Discovery in Databases (KDD):

Aspect	Data Mining	Knowledge Discovery in Databases (KDD)
Part of	A step in KDD process	The complete process
Focus	Applying algorithms to find patterns	From data selection to pattern interpretation
Involves	Analysis, modeling	Preprocessing, mining, postprocessing

KDD Steps:

- Selection → Preprocessing → Transformation → **Data Mining** → Evaluation

4. Data Mining Issues:

a) Mining Methodology and User Interaction Issues:

- Handling noise/incomplete data
- Incorporating user knowledge
- Interactive and iterative mining

b) Diversity of Database Types:

- Different types of data: relational, transactional, temporal, spatial, multimedia, web, etc.
- Data mining techniques must adapt accordingly.

5. Applications of Data Mining:

Field	Application Example
Marketing	Customer segmentation, targeted marketing
Banking & Finance	Fraud detection, credit scoring
Healthcare	Disease prediction, patient data analysis
Retail	Market basket analysis, inventory management
Education	Student performance prediction
Web	Recommendation systems, user behavior analysis