Date _____

Expt. No. _____ Page No. _____

UNIT-I

Notes Prepared By
P.MANSA DEVI (PMD)
Assistant Professor
CSE Department
GITAM University

Define Data Mining :-

- Data Mining is a process of extracting knowledge from massive volume of data. It refers to a way of finding significant and useful information from an organization's Database.
- The knowledge which is extracted can include pattern types, association rules & different trends.
- Data Mining is not confined to particular organization, instead it has techniques to explore the knowledge hidden in any data.

Data Mining Vs KDD :-

Data Mining and KDD are treated as synonyms by most people and are often used interchangeably but there are some dissimilarities between the two terms.

→ KDD refers to a process of extracting useful knowledge which may include correlation patterns between different data objects.

→ KDD extracts data by first sampling it from a huge database, performing data cleaning, transformation and applying data mining algorithms to generate specific data.

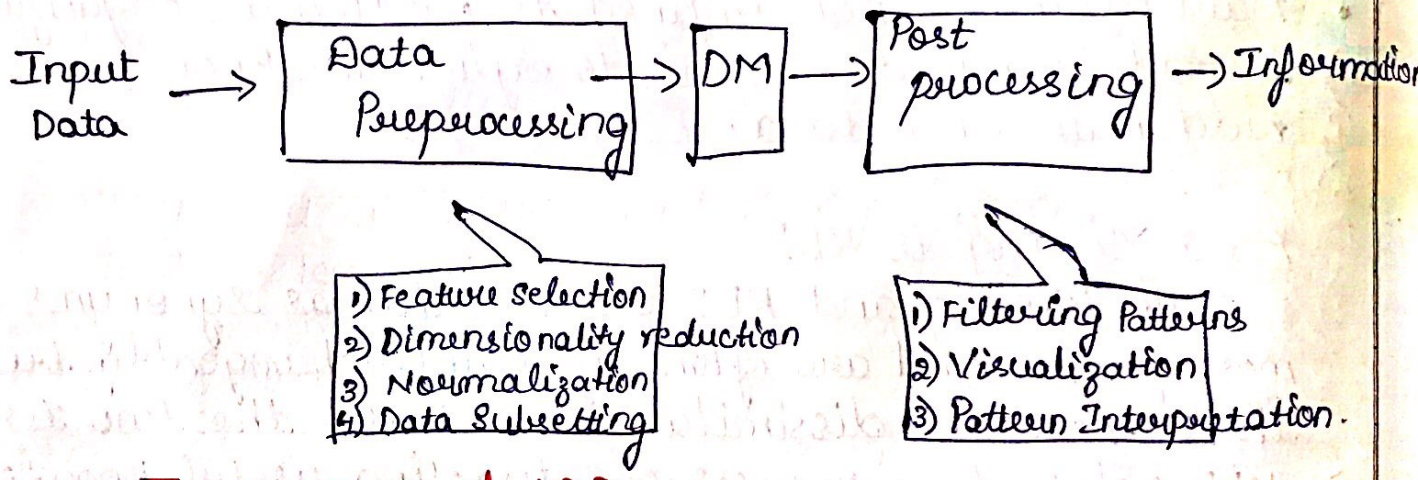
→ Data Mining simply refers to one of the stages in KDD process which is responsible for selecting algorithms.

→ These algorithms are used to discover the patterns, trends one-by-one derived by KDD process under valid calculation efficiency restrictions.

Teacher's Signature : _____

KDD - Knowledge Discovery in Databases:-

Data Mining is an integral part of Knowledge discovery in database (KDD), which is the overall process of converting raw data into useful information. This process consists of a series of transformation steps, from data preprocessing to postprocessing of data mining results.



The process of KDD.

→ There are seven different stages in KDD process. This process takes raw data as input and provides useful information desired by users as the output.

→ The objective of KDD process is to attain a good understanding about the dynamic organization's surroundings.

a) Data cleaning & preprocessing stage
b) Data Integration stage
c) Data selection stage

Date .....

Expt. No. ..... Page No. .....

(d) Data Transformation & Reduction stage
(e) Data Mining Discovery stage
(f) Pattern Interpretation & Analysis stage.
(g) Knowledge Visualization Stage.

(a) Data cleaning and Preprocessing stage :-
→ Real world data can be inconsistent, incomplete & noisy. Data cleansing is a process of removing unnecessary and inconsistent data from the databases. When data is extracted from different sources, there are chances that same information may be presented using different metrics and types.
→ The main purpose of preprocessing is to improve the quality of the data by filing out missing values, configuring data to make sure that it is inconsistent format.

(b) Data Integration Stage :-
In this stage, numerous data sources are combined (integrated) to form a larger database.

(c) Data Selection Stage :-
→ Data which is required for data mining process can be extracted from multiple & heterogeneous data sources such as databases, file & non-electronic sources.
→ Data selection is a process where the appropriate data required for analysis are fetched from the databases.

(d) Data Transformation & Reduction Stage :-
→ In the transformation stage, data extracted from multiple data sources are converted into an appropriate

Teacher's Signature : _____

format for data mining process by performing aggregation function.

→ Data reduction is used to decrease the number of possible values of data, which are being considered without affecting the integrity of data.

(e) Data Mining Discovery stage :-

Data Mining is an important process where expert techniques are applied so as to extract the hidden data pattern for evaluation.

(f) Pattern Interpretation and Analysis stage :-

→ In this stage, patterns generated as output from data mining stage are transformed into knowledge which is used in decision support system.

→ Both data mining and analysis stage can be performed repeatedly, as analysis stage is dependent on particular kind of mining algorithm.

(g) Knowledge Visualization stage :-

→ In this stage, different visualization and presentation methods are used to represent the knowledge obtained from analysis stage.

→ Presentation of knowledge can be in the form of charts, graphs, free form texts.

→ Presentation also deals with storing useful information in the knowledge base for iterative use.

Date .....

Expt. No. ..... Page No. .....

* Data Mining - Motivating Challenges :-

The challenges that motivated the development of data mining includes.

* Scalability
* High Dimensionality
* Complex Heterogeneous data
* Data ownership and Distributed data
* Non-Traditional Analysis

(i) Scalability :-

Data mining algorithms must be capable to handle and incorporate huge volumes of data as advanced data-generation & collection techniques produce data set in excess of giga bytes. These algorithms can be made more scalable by.

(i) Sampling Data
(ii) Implementing Data Structure
(iii) Developing distributed & parallel algorithms

(ii) High Dimensionality :-

Data sets of present era has several hundreds of attributes for instance.

(i) Spatial data involve very high volume of dimensions
(ii) BioInformatics micro array technology generated gene data coupled with thousands of dimensions

The number of dimensions grow with technology advancem
These data sets cannot be handled by classical data analysis techniques.

Teacher's Signature : _____

(iii) Complex Heterogeneous Data :-

→ The growth in various fields such as science, medic and finance produced large complex heterogeneous and non-traditional data.

→ Some of such data include semi-structured text, unstructured text, hyperlinks, multimedia.

→ This type of data cannot be handled by classical data analysis techniques which are capable of handling by only homogeneous data sets.

→ Therefore, new techniques were needed that were capable of handling graph connectivity, spatial auto-correlation ~~of handling graph connect~~ and relation among XML Documents.

(iv) Data ownership and Distributed Data :-

Data needed for analysis in certain circumstances does not belong to single owner (or) stored in single geographic location. This distributed data analysis needs new techniques which is faced by several challenges such as,

(i) Techniques to minimize resources needed for distributed computing.

(ii) Integration of data mining results from heterogeneo sources.

(iii) Handling data security.

Date .....

Expt. No. ..... Page No. .....

(v) Non-Traditional Analysis:-

→ Traditional statistical and analysis methods are based on hypothesis and testing, wherein an experiment is performed based on hypothesis to collect data and analyzes it. This method need high volume of resources which becomes difficult in present data mining scenario, where data sets involve distributed data & non-traditional data types.

→ Data Analysis for such data need evaluation of large volume of hypothesis. Therefore, hypothesis generation & evaluation process need to be automated.

* Origins of Data Mining:-

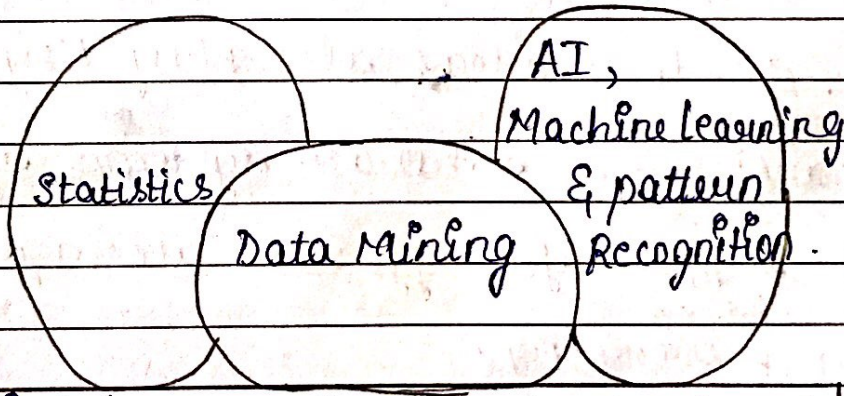


Fig:- Origin of Data Mining

→ Brought, together by the goal of meeting the challenges, researchers from different disciplines began to focus on developing more efficient and scalable tools that can handle types of data.

Teacher's Signature : _____

→ This work, which culminated in the field of data mining, built upon the methodology and algorithms that researchers had previously used.

→ In particular, data mining drawn upon ideas, such as

(a) Sampling, estimation, and Hypothesis testing from statistics.

(b) Search Algorithms, modelling techniques, & learning theories from AI, PR, & Machine learning.

→ DM has also been quick to adopt ideas from other areas, including optimization, evolutionary computing, information theory, signal processing, visualization and information retrieval.

→ A number of other areas also play key supporting roles. In particular, database systems are needed to provide support for efficient storage, indexing and query processing.

→ Techniques from high performance (parallel) computing are often important in addressing the massive size of some data sets.

→ Distributed techniques can also help address the issue of size & are essential, when the data cannot be gathered in one location.

Date .....

Expt. No. .....

Page No. .....

* Data Mining Tasks :-

DM tasks are generally divided into two major categories :-

(a) Predictive Tasks
(b) Descriptive Tasks

(a) Predictive Task :-

→ The objective of these tasks is to predict the value of a particular attribute based on the values of other attributes.

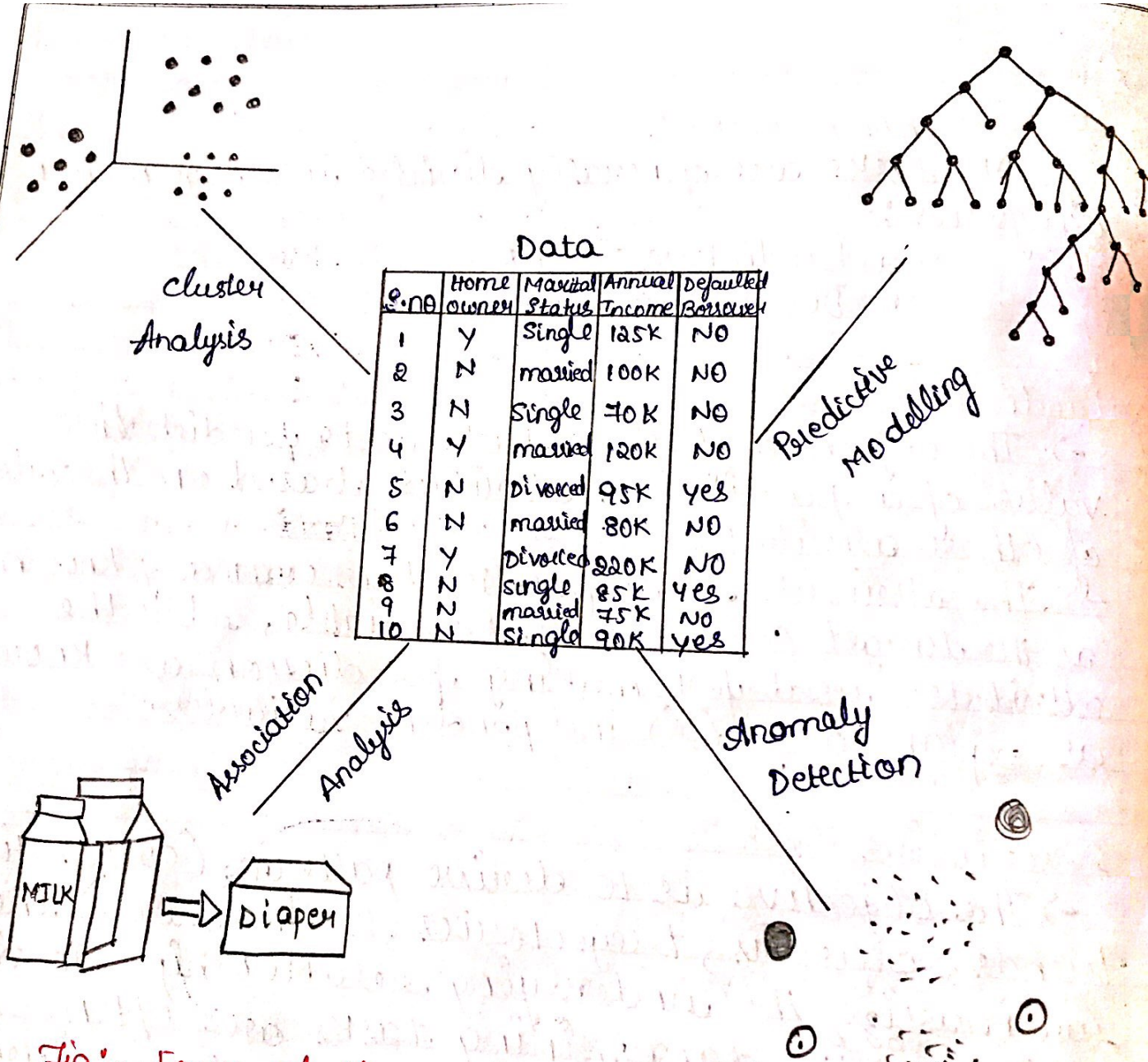
→ The attribute to be predicted is commonly known as the target (or) dependent variable, while the attributes used for making prediction are known as explanatory (or) independent variables.

(b) Descriptive Task :-

→ The objective is to derive patterns (correlations, trends, clusters, trajectories, & Anomalies) that summarize the underlying relationships in data.

→ Descriptive data mining tasks are often exploratory in nature & frequently require postprocessing techniques to validate & explain the results.

Teacher's Signature : _____



S.no	Home owner	Marital Status	Annual Income	Defaulted Borrower
1	Y	Single	125K	NO
2	N	married	100K	NO
3	N	Single	70K	NO
4	Y	married	120K	NO
5	N	Divorced	95K	yes
6	N	married	80K	NO
7	Y	Divorced	220K	NO
8	N	single	85K	yes
9	N	married	75K	NO
10	N	Single	90K	yes

Fig:- Four of the core data mining tasks

Predictive Modeling :-

It refers to the task of building a model for the target variable as a function of the explanatory variables. There are two types of predictive modeling tasks :-

(a) classification :- which is used for discrete target variables.

(b) Regression :- which is used for continuous target

Date .....

Expt. No. ..... Page No. .....

→ For example, predicting whether a web user will make a purchase at an online bookstore is a classification task because the target variable is binary-valued.

→ On the other hand, forecasting the future price of a stock is a regression task because price is a continuous valued attribute.

Association Analysis :-

→ It is used to discover patterns that describe strongly associated features in the data.
→ The discovered patterns are typically represented in the form of implication rules (or) feature subsets.

Cluster Analysis :-

Cluster Analysis seeks to find groups of closely related observations, so that observations that belong to the same cluster are more similar to each other than observations that belong to other clusters.

Anomaly Detection :-

It is a task of identifying observations whose characteristics are significantly different from the rest of the data. Such observations are known as anomalies (or) outliers.

Descriptive Function :-

The descriptive function deals with the general properties of data in the database.

- → class/ Concept Description
- → Mining of Frequent Patterns
- → Mining of Associations
- → Mining of Correlations.
- → Mining of clusters.

* class/ Concept Description :-

→ class/ concept refers to the data to be associated with the classes (or) concepts.

→ For example, In a company, the classes of items for sales include computer & printers, & concepts of customers include big spenders & budget spenders.

→ Such descriptions of a class (or) a concept are called class/concept descriptions. These descriptions of a class (or) a concept are called class/concept descriptions. These descriptions can be derived by the following two ways.

a) Data characterization

b) Data Discrimination.

* Mining of Frequent patterns :-

Frequent patterns are those patterns that occur frequently in transactional data. Here is the list of kind of frequent patterns.

Date .....

Expt. No. ..... Page No. .....

(i) Frequent Item Set :-

It refers to a set of items that frequently appear together, for Eg milk & bread (or) bread & jam

(ii) Frequent subsequence :-

A sequence of patterns that occur frequently such as purchasing a camera is followed by memory card.

(iii) Frequent sub structure :-

Substructure refers to different structural forms, such as graphs, trees (or) lattices.

* Mining of Association :-

Associations are used in retail sales to identify patterns that are frequently purchased together. This process refers to the process of uncovering the relationship among data & determining association rules.

For Eg:, a retailer generates a association rule that shows that 70% of time.

* Mining of Correlations :-

It is a kind of additional analysis performed to uncover interesting statistical correlations between associated - attribute-value pairs (or) between two item sets to analyze that if they have positive, -ve (or) no effect on each other.

Teacher's Signature : _____

* Mining of clusters :-

Clusters refers to a group of similar kind of objects. cluster analysis refers to forming group of objects that are very similar to each other but are highly different from the objects in other clusters.

* classification & Prediction :-

classification is the process of finding a model that describes the data classes (or) concepts. The purpose is to be able to use this model to predict the class of objects.

→ The list of functions involved in these processes are as follows :-

a) classification.
b) Prediction
c) outlier analysis
(d) Evolution Analysis.

a) classification :-

It predicts the class of objects whose class label is unknown. Its objective is to find a derived model that describes and distinguishes data classes (or) concepts.

b) Prediction :- It is used to predict missing (or) unavailable numerical data values rather than class labels

Date .....

Expt. No. ..... Page No. .....

c) Outlier Analysis :-

Outliers may be defined as the data objects that do not comply with general behavior (or) model of the data available.

(d) Evolution Analysis :-

Evolution Analysis refers to the description & model regularities (or) trends for objects whose behavior changes over time.

* Data Mining Applications :-

(e) Data Mining is highly useful in the following domains :-

(a) Market Analysis & Management
(b) Corporate Analysis & Risk Management
(c) Fraud Detection.

→ Apart from these, data mining can also be used in the areas of production control, customer retention, science exploration, sports, astrology & Internet.

Teacher's Signature : _____

Data.

Types of Data :-

a) Attributes & Measurement

b) Types of Data sets.

a) Attributes & Measurement :-

→ Define Attribute.

→ Consider what we mean by the Type of an attribute.

→ Describe the Types of Attributes

→ Describing Attributes by the number of values.

→ Assymetric Attributes.

b) Types of Data sets :-

→ General characteristics of Data sets

→ Different types of record data sd

→ Different graph based data

→ ordered Data.

Date .....

Expt. No. ..... Page No. .....

Data :-

Collecting the information from different sources and analysing the data for result.

Data set :-

A data set is a set of data objects, records, events, patterns, entities (or) vectors.

→ Data objects are described by a number of attributes ie., that capture the basic characteristics of an object, such as the mass of a physical object (or) the time at which an event occured.

→ Other names for an attribute are variable, characteristic, field, feature (or) Dimension.

Eg:- Data set (Student Information).

→ Data set is a file, in which the objects are records (or) rows in the file & each field (or) column corresponds to an attribute.

Student id	year	Grade
22	super Senior	8.20
23	Senior	8.90
24	Junior	7.80

→ Each row correspond to a student & each column is an attribute that describes some aspect of a student such as GPA.

Teacher's Signature : _____

→ This example is contain ie., called as record based datasets.

→ These record based datasets are common, either in files (or) relational database systems.

* Attributes & Measurement :-

In this concept, we describe data by considering what types of attributes are used to describe data objects.

a) Define Attribute

b) Consider what we mean by the type of an attribute.

c) Describe the Types of attributes.

→ Define Attribute :-

Attribute is a property (or) the characteristic of an object that may vary, either from one object to another (or) from one time to another.

Eg:- ① Hair colour / eye colour may varies from person to person ie., objed to objed.

② while the temperature of an object varies over time.

Date .....

Expt. No. .....

Page No. .....

Note:- (i) Eye color / Hair color is a symbolic attribute with a small number of possible values {Brown, black, grey}.

(ii) while temperature is a numerical attribute with a potentially unlimited number of values.

→ But, Attributes are not about numbers (or) Symbols. However to discuss & analyze the characteristics of objects; we assign numbers (or) symbols to them.

→ To analyze the process, we required a measurement scale.

* Measurement Scale :-

It is a rule (or) function that it defines a relation between an objects, attribute by associating either a symbol (or) numerical value. It is called measurement scale.

Eg:- Measurement of Height & Weight.

Consider an Example, if you conducting a meeting in a room, now you count the how many chairs in a room to see, if there will be enough to seat all the people coming to a meeting.

→ In this case, the physical value of an attribute of an object is mapped to a numerical (or) symbolic value.

Teacher's Signature : _____

* Type of an Attribute :-

→ It is an important concept, that it determines a particular data analysis technique is consistent with a specific type of attribute.

→ The properties of an attribute need not be the same as the properties of the values used to measure it.

Consider Eg :- (Employee ID & Employee Age).

→ Two Attributes that might be associated with an employee ID & age (in years)

→ Both of these attributes can be represented as Integers

→ Now we can discuss about the average age of an employee, but no need to calculate the average emp ID This emp id is to test whether the data belongs to particular person (or) not.

* Different Types of an Attribute :-

This concept will describe (or) specify the type of an attribute is to identify the properties of numbers that correspond to underlying properties of the attribute.

Expt. No. ..... Page No. .....

→ There are some properties (operations) of numbers are used to describe attributes.

a) Distinctness = and ≠
b) order <, ≤, > and ≥
c) Addition + & -
d) Multiplication * & /.

→ By given these properties, define 4 types of attributes

a) Nominal } Qualitative
b) ordinal }
c) Interval } Quantitative
d) Ratio }

→ Each attribute type possesses all of the properties & operations of the attribute types.

* Explain the different Types of Attributes?

a) Qualitative Data
b) Quantitative Data.

A) Qualitative Data (or) Categorical Data :-
It is a information about qualities, ie., The information that can't actually measured. Eg of Qualitative data are the softness your skin, color of your eyes.

→ Nominal & ordinal are consider as categorical (or) qualitative attributes.

Teacher's Signature : _____

(b) Quantitative Data (or) Numeric Data :-

→ It is an information about quantities, ie; information that can be measured & written down with numbers.

Eg:- Height; shoe size; length of your finger nails.

→ Interval and ratio are referred as Quantitative (or) Numeric attributes.

Eg:- ① Age of your car → Quantitative
② The number of chairs in a room → Quantitative
③ Color of the sky → Qualitative
④ Softness of a cat → Qualitative.

Attribute Type		Description	Examples	Operations
Categorical (Qualitative)	Nominal	The values of a nominal attribute are different names; ie, nominal values provide only enough information.	Zip codes, employee ID numbers, eye color, gender	mode, entropy, contingency correlation, χ^2 test
	Ordinal	The values of an ordinal attribute provide enough information to order objects (<, >)	Hardness of minerals, {good, better, best}, grades, street numbers.	median, percentiles, rank correlation, run tests, sign tests
Numeric (Quantitative)	Interval	For interval attributes, the differences b/w values are meaningful (+, −)	Calendar dates, temperature in celsius (or) Fahrenheit	mean, std deviation, t & F tests.
	Ratio	For ratio variables, both differences & ratios are meaningful. (*, /)	Temp in Kelvin, counts, age, mass, length, electrical current.	Harmonic mean, geometric mean, Percent variation

Date .....

Expt. No. ..... Page No. .....

* **Describing Attributes by the Number of Values:-**

It is an independent way of distinguishing between attributes is by the number of values they can take.

a) Discrete
b) Continuous.

a) **Discrete:-**

→ A discrete attribute has a finite (or) countably infinite set of values. Such attributes can be categorical, such as zip codes (or) ID numbers (or) Numeric.

→ Discrete Attributes are often represented using Integer variables.

→ In this, Binary attributes are the special case of discrete attributes & assume only two values.

Eg:- True/False ; yes/NO ; Male/Female ; (or) 0/1.

→ Binary Attributes are often represented as Boolean variables (or) as Integer variables that only take the values 0 (or) 1.

b) **Continuous:-**

A continuous attribute is one whose values are real numbers.

Eg:- Include attributes such as temperature, height (or) weight.

→ Continuous attributes are represented as floating point variables.

Teacher's Signature : _____

* Types of Data sets :-

* General characteristics of Data sets :-

① Dimensionality

② Sparsity

③ Resolution.

① Dimensionality :-

→ The dimensionality of a dataset is the number of attributes that the objects in the data set possess.

→ Data with a small number of dimensions tends to be Qualitatively different than moderate (or) high dimensional data.

→ Because of this, an important motivation in preprocessing the data is dimensionality reduction.

② Sparsity :-

→ For some data sets, such as those with assymmetric features, most attributes of an object have values of 0; In many cases, fewer than 1% of the entries are non-zero.

> Sparsity is an advantage because usually only the non-zero values need to be stored & manipulated.

Date .....

Expt. No. ..... Page No. .....

③ Resolution :-

⇒ It is frequently possible to obtain data at different levels of resolution, & often the properties of the data are different resolutions.

→ As if the resolution is too coarse then the pattern becomes invisible, while if the resolution is too fine the pattern may be buried due to noise.

* Record Data :-

→ Data Mining work assumes that dataset is a collection of records (data objects), each of which consists of a fixed set of data fields (attributes).

→ Record data is usually stored in flat files (or) in relational databases.

a) Transaction (or) Market Basket Data
b) Data Matrix
c) Sparse Data Matrix

a) Market Basket Data (or) Transaction :-

It is a one of the form of record data that involves collection of items.

Eg:- Consider a grocery store. The set of products purchased by a customer during one shopping trip constitutes a transaction, while the individual products that were purchased are the items. This

Teacher's Signature :

type of data is called market basket data, because the items in each record are the products in a person's "market Basket".

→ Transaction data is a collection of set of items, but it can be viewed as a set of records whose fields are asymetric attributes.

→ Most often, the attributes are binary, indicating whether (or) not an item was purchased, but more generally, the attributes can be discrete (or) continuous such as the number of items purchased (or) the amount spent on those items.

Tid	Items
1	Bread, soda, milk
2	Bread
3.	Soda, milk.
4.	Soda, Bread, biscuit

Transaction Data.

* Data Matrix :-

If the data objects in the collection of data all have the same fixed set of numeric attributes, then the data objects can be consider as points in

Date .....

Expt. No..... Page No.....

a multidimensional space, where each dimension represents a distinct attribute describing the object.

→ A set of such data objects can be interpreted as an m×n matrix, where are m rows, for one each object & n columns, one for each attribute. This matrix is called data matrix (or) a pattern matrix.

→ A data matrix is a variation of record data, but it consists of numeric attributes, std matrix operation can be applied to transform & manipulate the data. Therefore, the data matrix is the std data format for most statistical data.

Projection of x load	Projection of y load	Distance	load	Thickness
10.23	5.27	15.22	27	1.2
12.65	6.25	16.22	22	1.1
13.54	7.23	17.34	23	1.2
14.27	8.43	18.45	25	0.9

* Sparse Data Matrix :- (or) Document-Term Matrix :-

① It is a special case of a data matrix in which the attributes are of the same type & are assymetric i.e., only non-zero values are important.

② Transaction data is an example of a sparse data

Teacher's Signature : _____

matrix that has only 0-1 entries. Another common example is document data.

(3) We can represent the data in the collection of documents, we called as document term matrix.

	team	coach	play	ball	score	win	lost	timeout	season
Document 1	3	0	5	0	2	6	0	2	2
Document 2	0	7	0	2	1	0	0	3	0
Document 3	0	1	0	0	1	2	2	0	0

→ The documents are the rows of this matrix, while the terms are the columns.

Graph - Based Data :-

A graph representation be convienient & importa we consider two cases :-

(a) The graph captures relationships among data objects.

(b) The data objects themselves are represent as graphs.

Data Quality :-

Data Mining focusson (a) The Detection & correction of data Quality problems

(b) The use of algorithms that can tolerate poor data Quality.

→ The first step, detection and correction, is often data cleaning.

* Measurement and Data Collection Issues
* Measurement and Data collection errors

* Measurement and Data collection Issues :-

→ It is unrealistic to expect that data will be perfec There may be problems due to human error, limitations of measuring devices in the data collection process.

→ Values (or) even entire data objects may be missing

→ In other cases, there may be duplicate objects.

→ Multiple objects that are correspond to a single real object.

→ For example, there might be two different record for a person who has recently lived at two different addresses.

Date .....

Expt. No.....

Page No.....

Even if all data is present and looks fine, there may be inconsistencies - a person has a height of 2 meters, but weighs only 2kgs.

* Measurement and Data collection Errors :-

→ The term measurement error refers to any problem resulting from the measurement process.

→ The difference between true value and recorded value is called measurement error.

→ The term data collection error refers to errors such as omitting data objects (or) attribute values.

For example, a study of animals of a certain species might include animals of a related species that are similar in appearance to the species of interest.

→ Both measurement errors & data collection errors can be either systematic (or) random.

* In Measurement & Data Collection errors & then consider a variety of problems that involve measurement error :-

a) Noise
b) Artifacts
c) Bias
d) Precision
e) Accuracy.

Teacher's Signature : _____

(i) Noise :-

Noise is a critical factor that influence measurement by either adding new objects (or) Exchanging existing objects.

(ii) Artifacts :- An artifact is a distortion (changes due to external forces).

For eg; streaks on photograph (lines that have different background than the photo).

(iii) Precision :-

The nearness of some attribute repeatedly being measured for one object to another object is called precision. It is measured using std deviation of values.

(iv) Bias :-

→ The difference in measurement of one attribute being measured is called bias.

→ It is measured by considering the value of attribute and its difference with mean value.

(v) Accuracy :-

The nearness of the produced results of measurement to the actual value of any attribute is called accuracy.

Expt. No. ..... Date .....

Page No. .....

* Data Quality Issues may involve both measurement and data collection problems :-

1. Eliminate Data objects (or) Attributes
2. Estimate missing values
3. Ignore missing values in Analysis
4. Inconsistent values
5. Duplicate Data

① Eliminate Data objects (or) Attributes :-

→ A data set that possess only few objects with missing values (or) its attributes can be discarded (deleted).

→ However, care should be taken while deleting such objects because sometime these attributes can carry critical information for analysis.

→ Deleting such attributes may effect the quality of Analysis!

② Estimate Missing Values :-

→ In this method, the user himself should try to find out the tuples with missing values & fill in those tuples manually.

→ This method generally is not advantageous as it consumes more time & is not suitable to use when massive volume of data contain missing values.

Teacher's Signature : _____

③ Ignore Missing Values in Analysis :-

The missing values can be ignored during the analysis without affecting the analysis.

Eg:- Consider two objects which are used to determine similarity between them. If one object have some missing values then these values can be filled by using the other object values. However, this can be done only when the number of attributes are small.

④ Inconsistent values :-

Data objects can have values that varies from a data set to another.

⑤ Duplicate Data :-

A data set can be a duplicate copy of another dataset which need to be detected & eliminated. This can be done by considering two issues known as de-duplication.

(a) If a single object is represented by two different objects then the corresponding attribute may vary. This type of inconsistency need to be resolved.

Expt. No. ..... Page No. .....

(6) Two distinct objects that have similar characteristics should not be combined together while performing de-duplication. However, in certain circumstances two objects can have similar features, these should not be considered as duplicates.

Teacher's Signature : _____