Data & Algorithms

AI for Engineers

दृष्टि... आने वाले समय का आभास

Vision... the shape of things to come

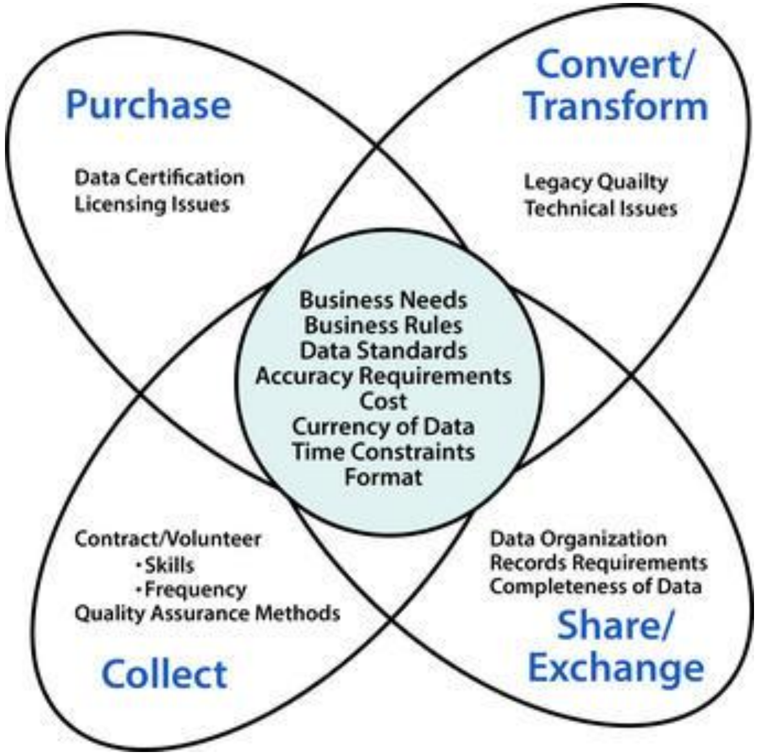
What is Data in context of AI

- Data, DataSet, Table, Spreadsheet, Relational, Big Data

Area of the House in sq feet	No. of bedrooms	Price in Million
520	1	1.5
623	1	1.6
750	2	2
1075	3	2.5
2290	4	3.75
2500	4	3.9

Acquiring Data

- Manual Labelling
- Observing Behavior- humans, machines, responses
- Download from website/get from partners

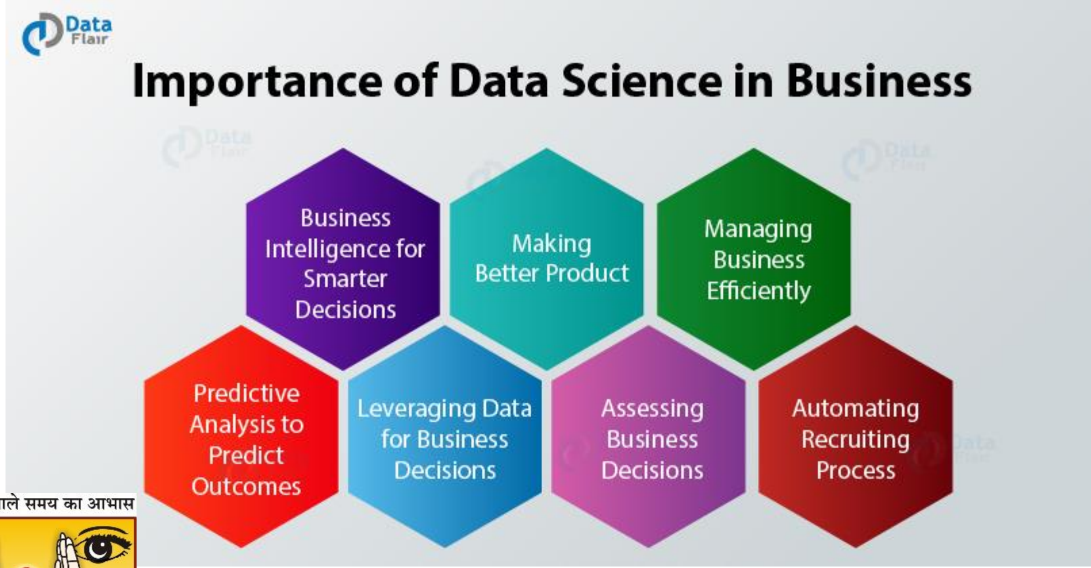


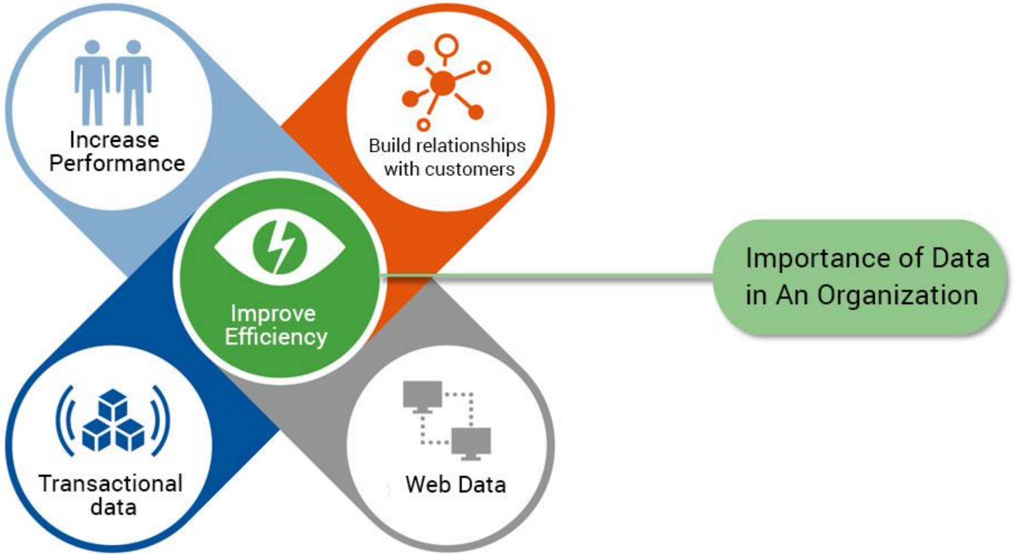
दृष्टि... आने वाले समय का आभास

Vision... the shape of things to come

Importance of Data

Data is certainly important!


Data Flair
Importance of Data Science in Business
Business Intelligence for Smarter Decisions
Making Better Product
Managing Business Efficiently
Predictive Analysis to Predict Outcomes
Leveraging Data for Business Decisions
Assessing Business Decisions
Automating Recruiting Process

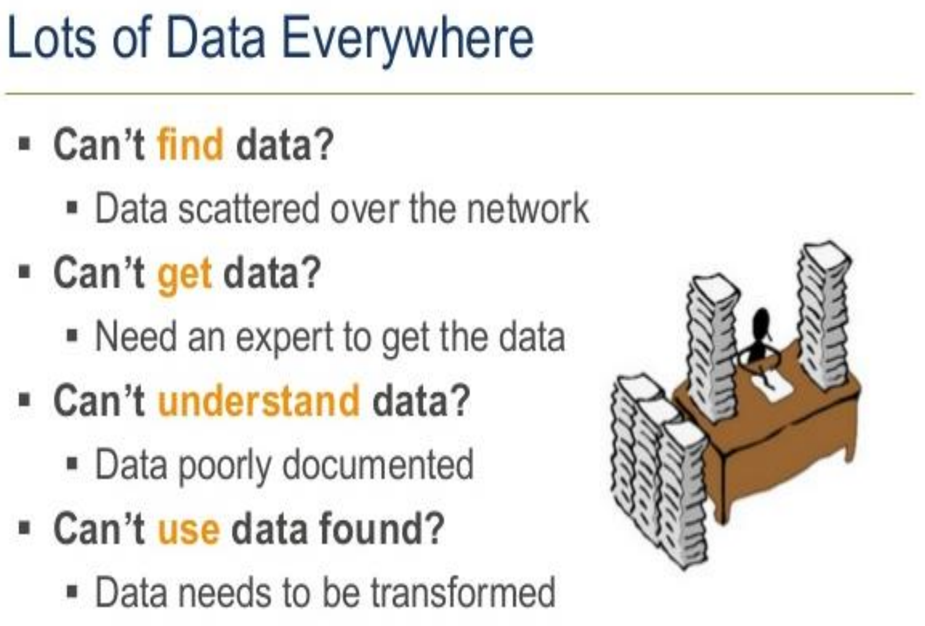

Increase Performance
Build relationships with customers
Improve Efficiency
Transactional data
Web Data
Importance of Data in An Organization

दृष्टि... आने वाले समय का आभास

Vision... the shape of things to come

Problems with Data

- Data is at times bit over hyped -there is hardly anything as perfect dataset
- Data can be misused- AI can't always work just with huge data
- Data is messy
  - GIGO
  - Data Problems
    - Incorrect Label
    - Missing values
- Threat of Data theft



Lots of Data Everywhere

- Can't find data?
  - Data scattered over the network
- Can't get data?
  - Need an expert to get the data
- Can't understand data?
  - Data poorly documented
- Can't use data found?
  - Data needs to be transformed

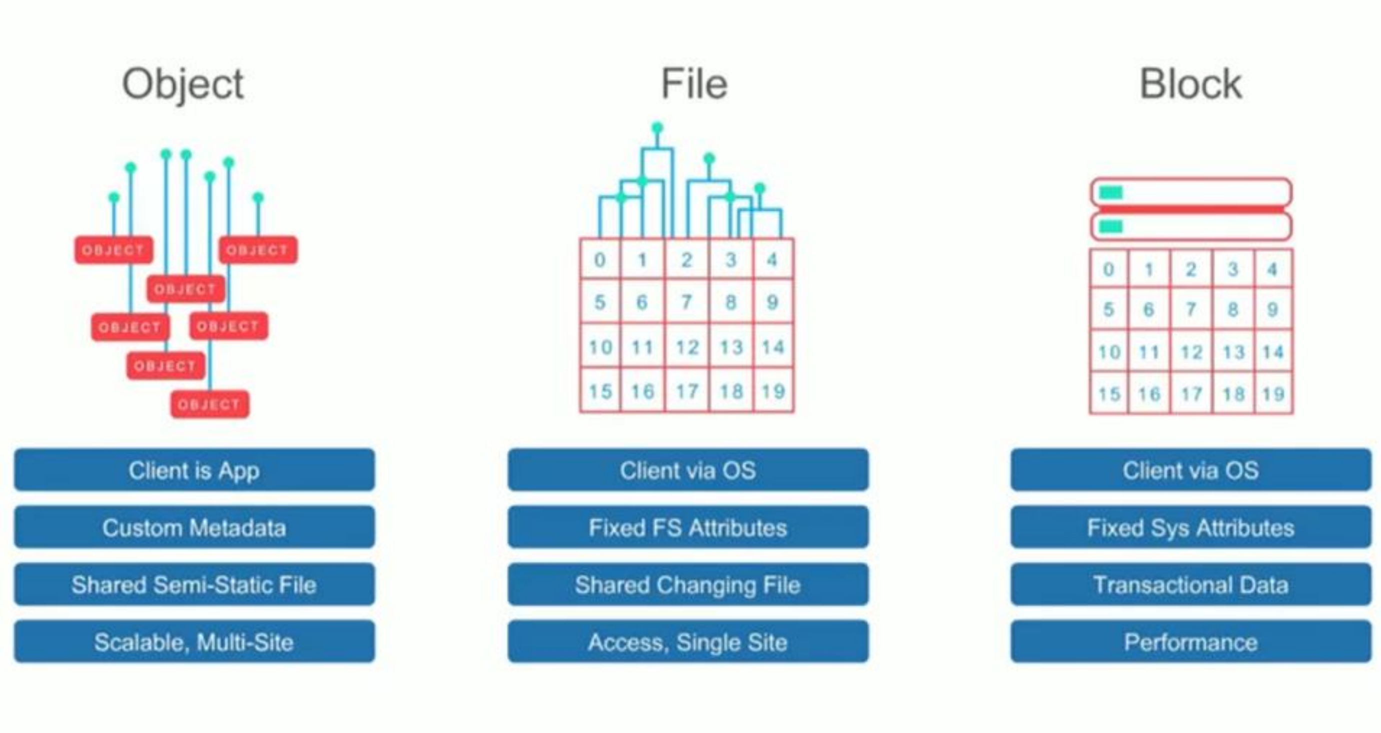
दृष्टि... आने वाले समय का आभास

Vision... the shape of things to come

Multiple type of Data

- Structured Data
  - Flat files
  - Relational database
- Unstructured Data
  - Image
  - Audio
  - text
- The techniques for dealing with unstructured data are little bit different than the techniques for dealing with structured data. But AI techniques can work very well for both of these types of data

Forms of Data Storage

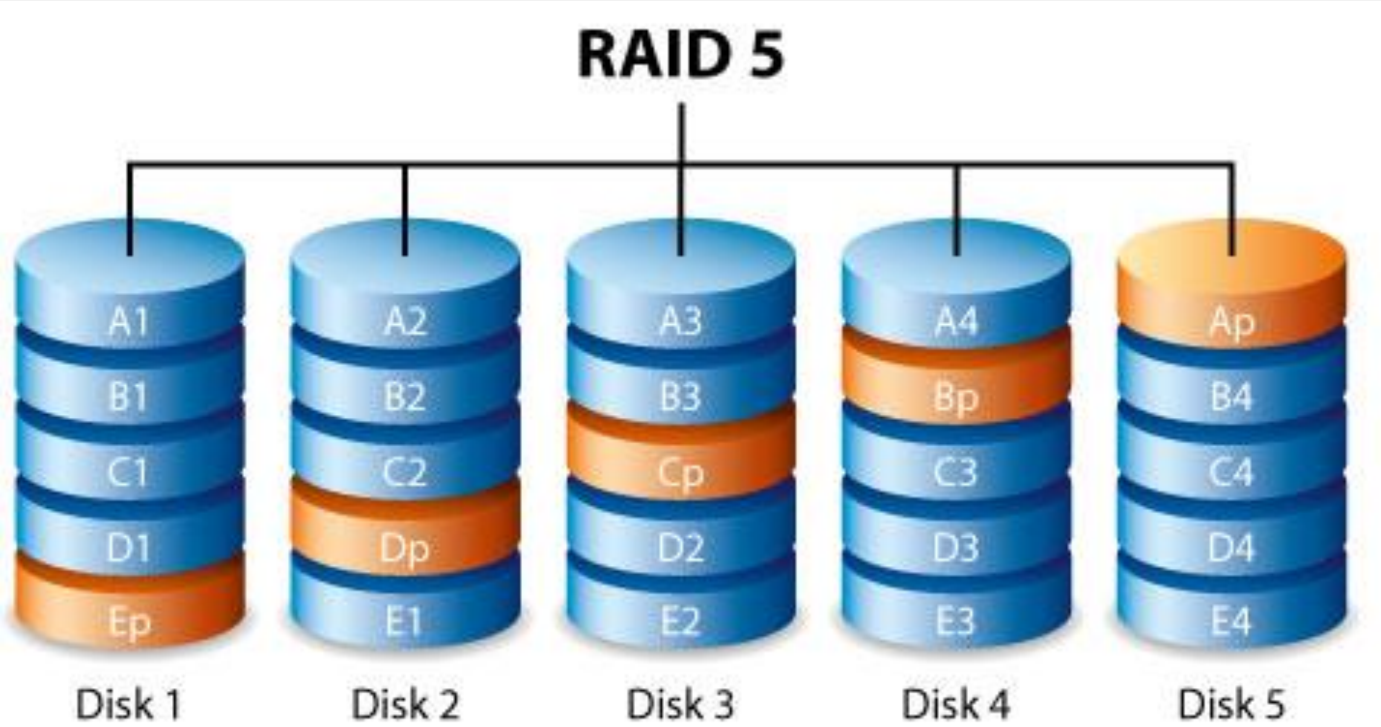


Data storage refers to the use of recording media to retain data using computers or other devices.

- File Storage
- Block Storage
- Object Storage

Types of Data Storage

- **Direct Attached Storage (DAS)**
- Physically connected to your computer like
  - Hard Drives
  - Solid-State Drives (SSD)
  - CD/DVD Drives
  - Flash Drives
- More affordable & accessible
- Data sharing cumbersome.



Data Storage – Types (contd…)

Network Attached Storage (NAS)

- Multiple machines to share storage over a network.
- Using RAID configuration.
- Centralize data and improve collaboration.
- NAS solutions more costly than DAS solutions, still very affordable as storage technology has advanced significantly.

Data Storage – Types (contd…)

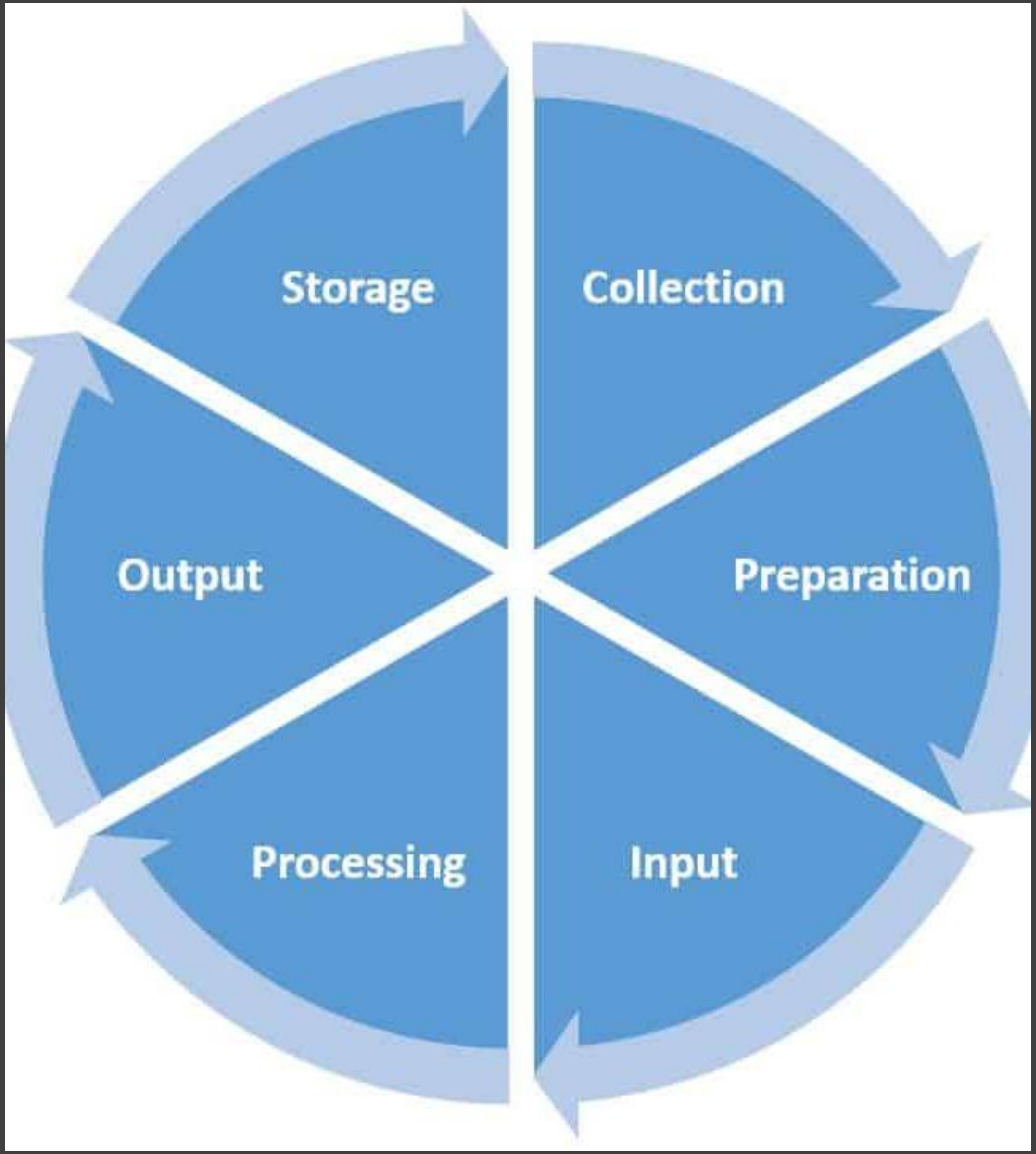
Data on Internet

Besides LAN or WAN, the data can also be stored over internet in form of

- Cloud Storage, &
- Hybrid Data Storage.

Benefits of Efficient Data Storage

- Reliable data preservation
- Data continuity and accessibility
- Quicker and easier data recovery
- Flexible price points and capacity options
- Effective security for protected files

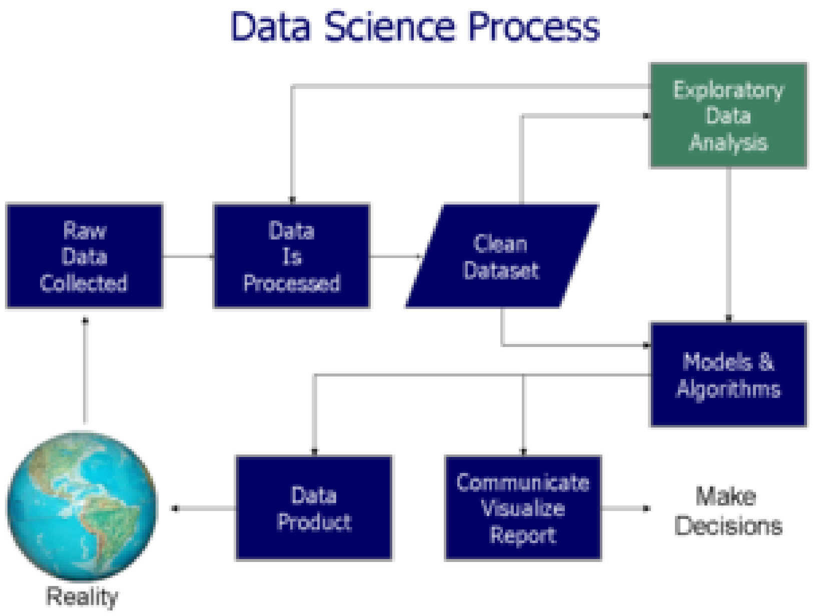


The Stages of data processing

- 1. Data collection
- 2. Data preparation
- 3. Data input
- 4. Processing
- 5. Data output/interpretation
- 6. Data storage

Data Visualization

- **Data visualization** is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
- Data visualization is one of the steps in analyzing data and presenting it to users.



दृष्टि... आने वाले समय का आभास

Vision... the shape of things to come

Data Visualization should …

Show	show the data
Induce	induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production or something else
Avoid	avoid distorting what the data has to say
Present	present many numbers in a small space
Make	make large data sets coherent
Encourage	encourage the eye to compare different pieces of data
Reveal	reveal the data at several levels of detail, from a broad overview to the fine structure
Serve	serve a reasonably clear purpose: description, exploration, tabulation or decoration
Be	be closely integrated with the statistical and verbal descriptions of a data set.

Regression Analysis

- In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables
- Regression analysis is a reliable method of identifying which variables have impact on a topic of interest.

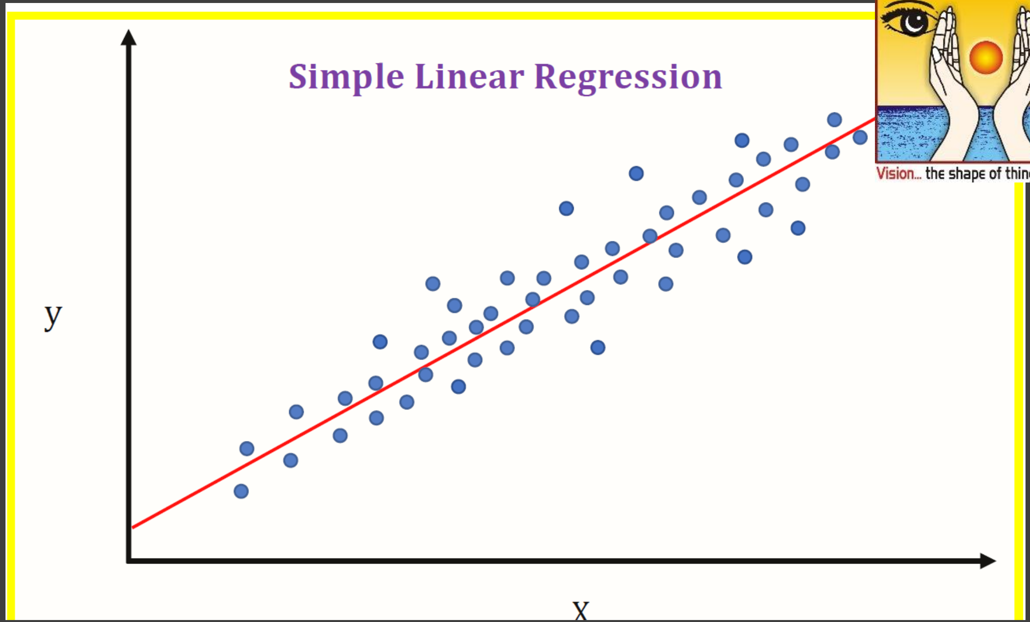
दृष्टि... आने वाले समय का आभास

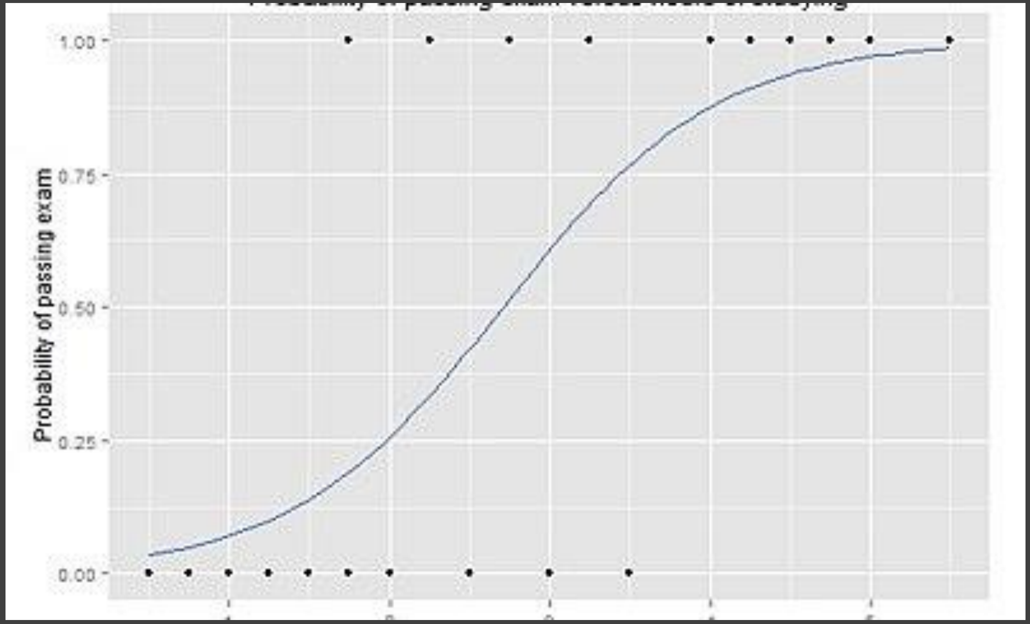
Vision... the shape of things to come

Regression Analysis- Most used types

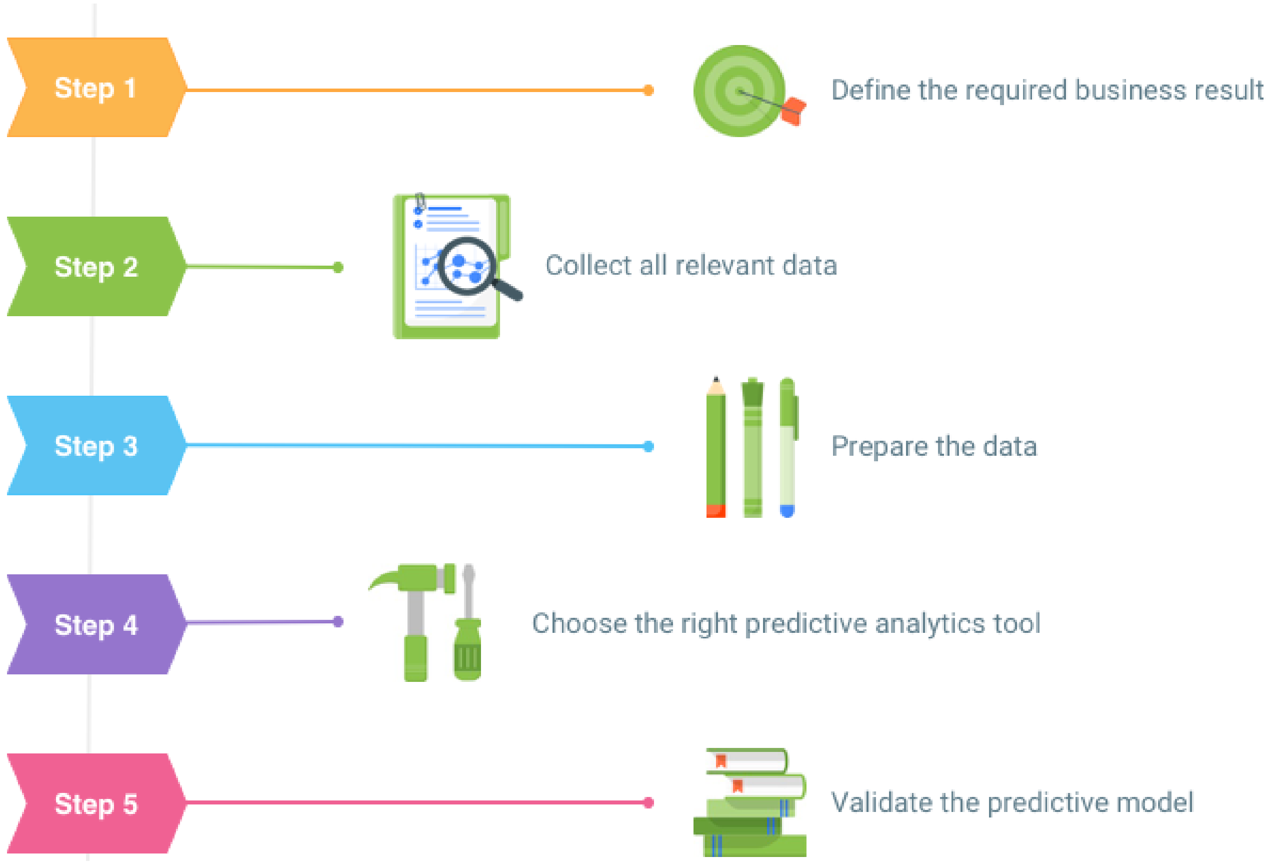
- Linear and Logistic regressions

- In statistics, linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables.
- In statistics, the logistic model is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick.





Data Prediction



Step 1 Define the required business result

Step 2 Collect all relevant data

Step 3 Prepare the data

Step 4 Choose the right predictive analytics tool

Step 5 Validate the predictive model

Predictive analytics encompasses a variety of statistical techniques from data mining, predictive modelling, and machine learning, that analyze current and historical facts to make predictions about future or otherwise unknown events.

We may invoke 5 steps for data prediction.

Data Classification

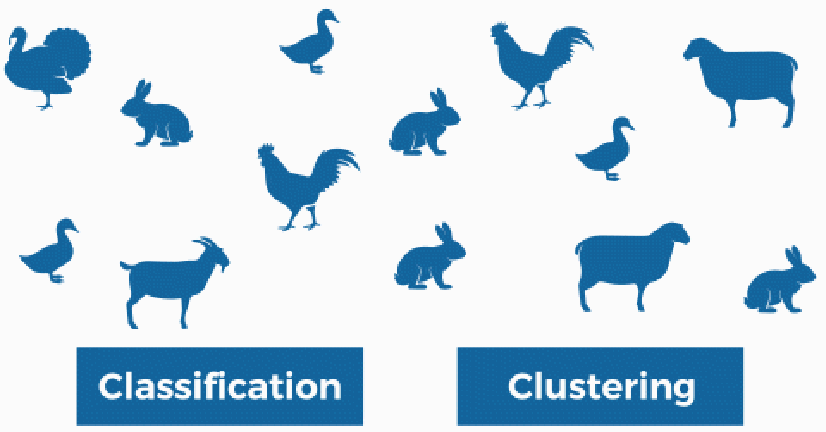
- The process of organizing data by relevant categories so that it may be used and protected more efficiently.
- Types of Data Classification
  - Content based
  - Context based
  - User based

Data Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.

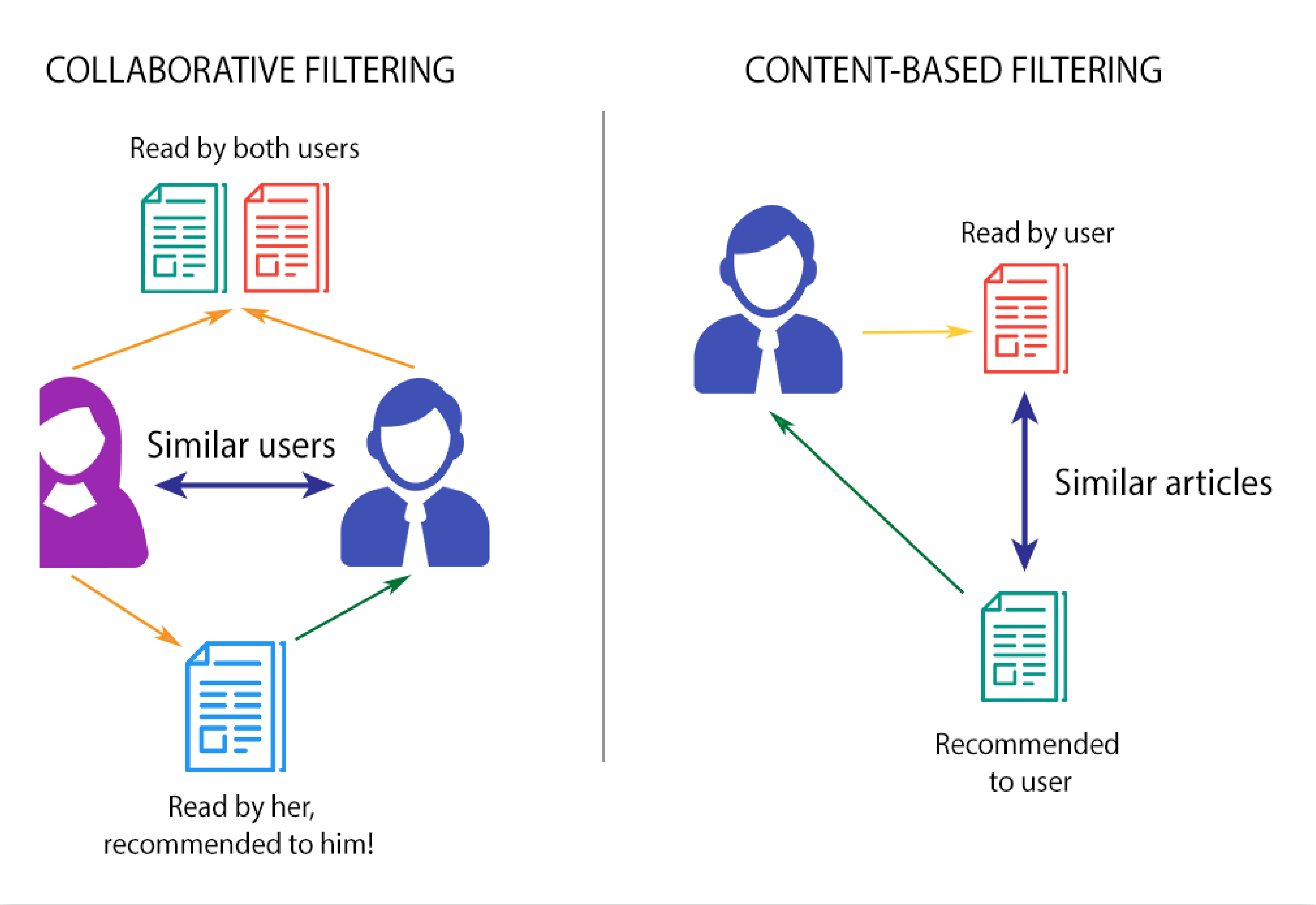
Data **Clustering** is a process which partitions a given **data** set into homogeneous groups based on given features such that similar objects are kept in a group whereas dissimilar objects are in different groups.

Difference between classification & clustering



दृष्टि... आने वाले समय का आभास

Vision... the shape of things to come



Recommender Systems

Seeks to predict the "rating" or "preference" a user would give to an item.