Introduction to

Machine Learning and Deep Learning

By

Prof(Dr) Premanand P Ghadekar

Outline

- ❖ Introduction to ML & DL
- ❖ Syllabus
- ❖ Data Science Process Flow chart
- ❖ Best Practices for Data Science Projects
- ❖ Data Science based Project ideas
- ❖ Impact of Data Science on Industries
- ❖ Data Science Engineers Skill Set
- ❖ Job Roles & Opportunities for Students
- ❖ Research Opportunities etc.

Terminology

Machine Learning, Deep Learning, Data Science, Data Mining, Data Analysis, Statistical Learning, Knowledge Discovery in Databases, Pattern Discovery, AI.

Smile, we are 'DATAFIED'!

- ❖ Wherever we go, we are "datafied".

- ❖ Smartphones are tracking our locations.

- ❖ We leave a data trail in our web browsing.

- ❖ Interaction in social networks.

- ❖ Privacy is an important issue.

Data everywhere!

1. **Google:** processes 24 peta bytes of data per day.

2. **Facebook:** 10 million photos uploaded every hour.

3. **Youtube:** 1 hour of video uploaded every second.

4. **Twitter:** 400 million tweets per day.

5. **Astronomy:** Satellite data is in hundreds of PB.

6. . . .

7. **"By 2025 the digital universe will reach 50+ zettabytes..."**

GTPEZY Bytes

Data types

Data comes in different sizes and also flavors (types):

☒ **Texts**

☒ **Numbers**

☒ **Clickstreams**

☒ **Graphs**

☒ **Tables**

☒ **Images**

☒ **Transactions**

☒ **Videos**

☒ **Some or all of the above!**

Applications of ML

- We all use it on a daily basis. Examples:

Machine Learning

❖ Spam filtering

❖ Credit card fraud detection

❖ Digit recognition on checks, zip codes

❖ Detecting faces in images

❖ MRI image analysis

❖ Recommendation system

❖ Search engines

❖ Handwriting recognition

❖ Scene classification

❖ etc...

What is Machine Learning?

"How do we create computer programs that improve with experience?"

- Tom Mitchell
- http://videolectures.net/mlas06_mitchell_itm/

❖ **Machine Learning** is the Science of making computers learn and act like humans by feeding data and information without being explicitly programmed.

❖ A branch of **artificial intelligence,** concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.

What is Machine Learning?

❖ **Machine Learning** is the Science of making computers learn and act like humans by feeding data and information without being explicitly programmed.

❖ A branch of **artificial intelligence,** concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.

❖ As intelligence requires knowledge, it is necessary for the computers to acquire knowledge.

❖ What machine learning can do is perform predictive analytics far faster than any human can.

❖ As a result, machine learning can help humans work more efficiently.

Why Machine Learning ?

Identifying The Factors/ Opportunities

Which Can Be Further Used For Decision Making And Value Creation.

Why "Learn"?

❖ **Machine Learning** is programming computers to optimize a performance criterion using example data or past experience.

❖ **Learning is used when**:

o Human expertise does not exist (Navigating on Mars),

o Humans are unable to explain their expertise (Speech Recognition)

o Solution changes in time (Routing on a Computer Network)

o Solution needs to be adapted to particular cases (user Biometrics).

Machine Learning & Deep Learning

Prerequisite- Linear Algebra, Statistics, Probability, Calculus, and Programming Languages

Section 1:

o Introduction

o Types of Learning

o Mathematical Concepts

o Regression and Generalization

o Classification

o Clustering

Section 2:

o Trends in Machine Learning

o Deep Learning

o Deep Learning Strategy

ML & DL Syllabus

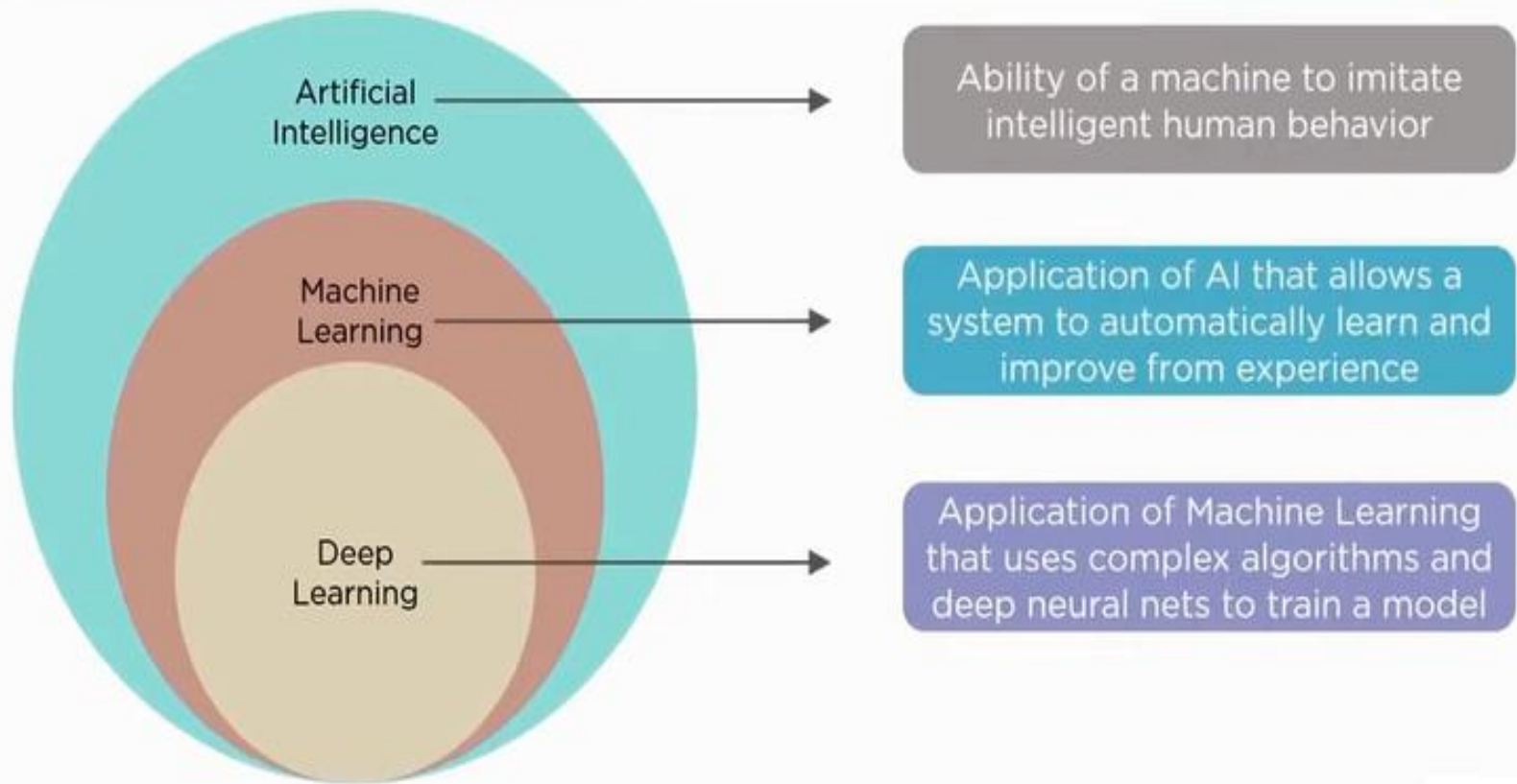# Machine Learning & Deep Learning Books

1. Tom Mitchell, Machine Learning.

2. Abu-Mostafa, Yaser S. and Magdon-Ismail, Malik and Lin, Hsuan-Tien, Learning From Data, AMLBook.

3. The elements of statistical learning Data mining, inference, and prediction T. Hastie, R. Tibshirani, J. Friedman.

4. Christopher Bishop. Pattern Recognition and Machine Learning.

5. Richard O. Duda, Peter E. Hart, David G. Stork. Pattern Classification. Wiley.

6. Deep Learning with Python by François Chollet, Manning Publications Co, ISBN: 9781617294433.

7. Deep Learning - A Practical Approach by Rajiv Chopra, Khana Publications, ISBN: 9789386173416.

8. Deep Learning by Ian Goodfellow and Yoshua Bengio and Aaron CourvillePublished by An MIT Press book.

Interdisciplinary field
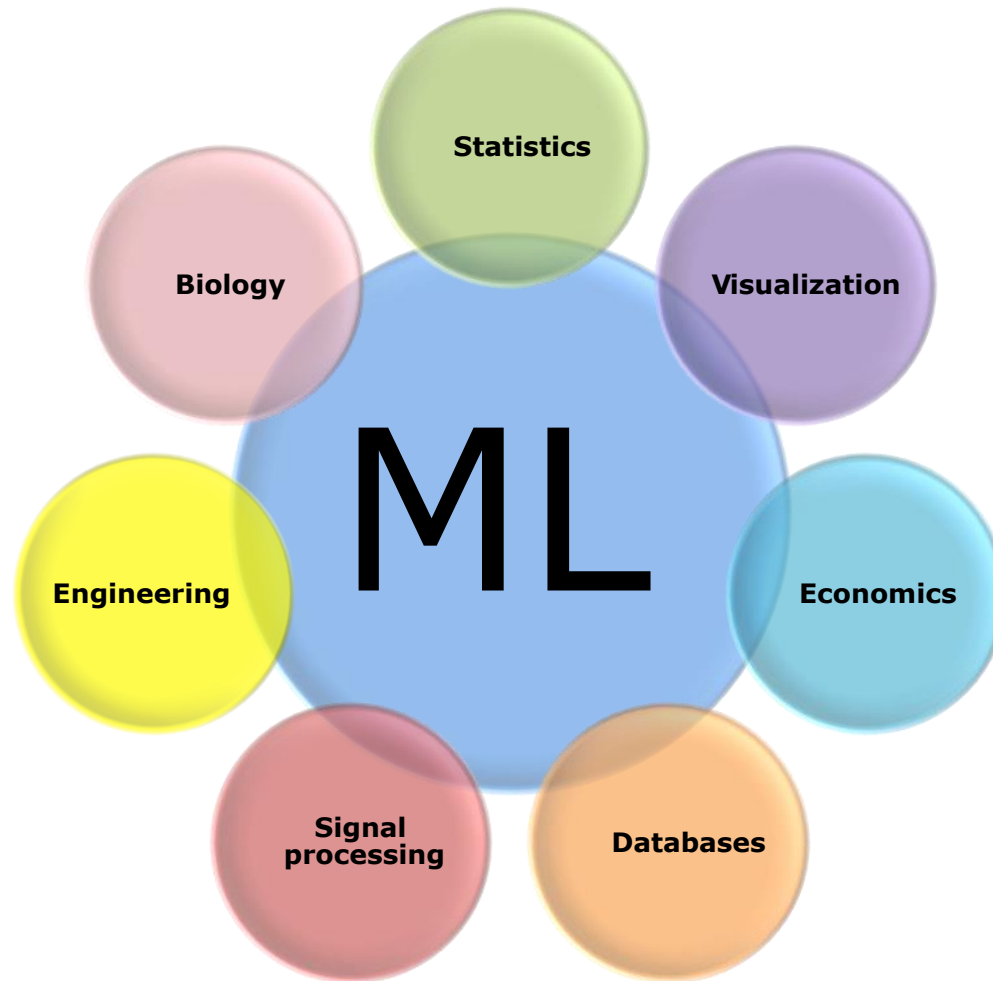
Deep Learning

Deep Learning is a subset of Machine Learning that has networks which are capable of learning from data that is unstructured or unlabeled and works similar to the functioning of a human brain.

Artificial Intelligence → Ability of a machine to imitate intelligent human behavior

Machine Learning → Application of AI that allows a system to automatically learn and improve from experience

Deep Learning → Application of Machine Learning that uses complex algorithms and deep neural nets to train a model

Deep Learning –Techniques that learns features and task from Data Directly In Deep Learning, features can be learnt just from raw data.
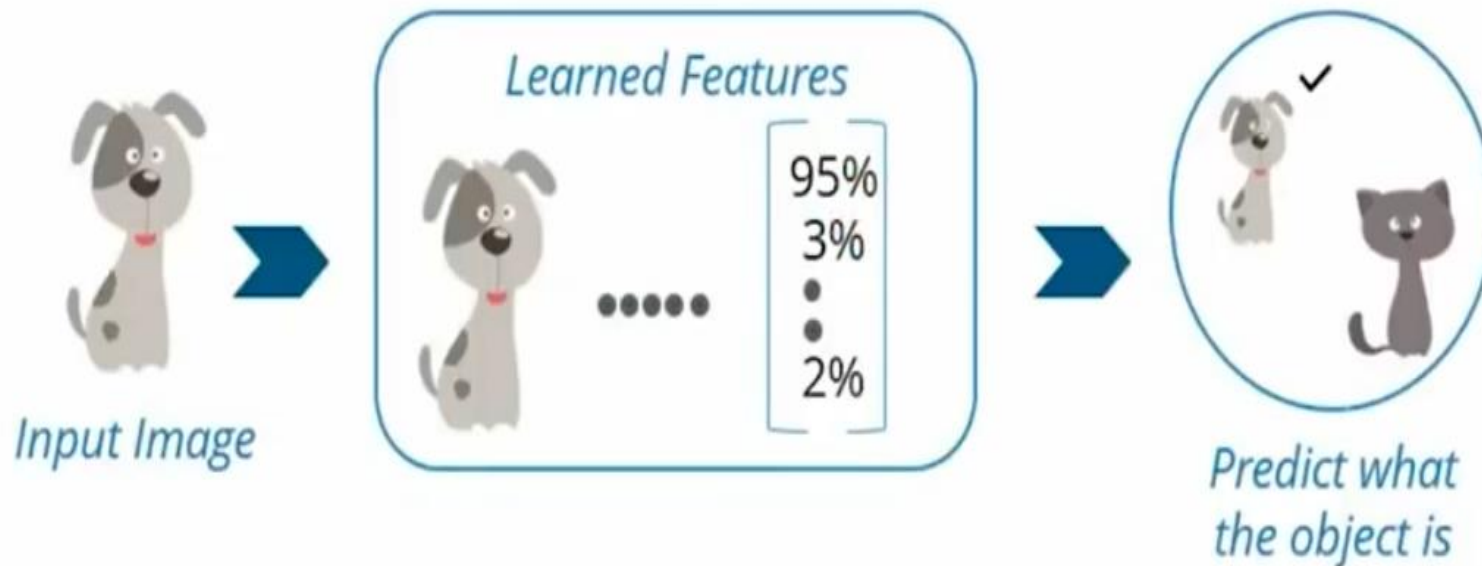
Deep Learning

Training

Input Image → Feature extraction → Machine Learning Model → Predict what the object is
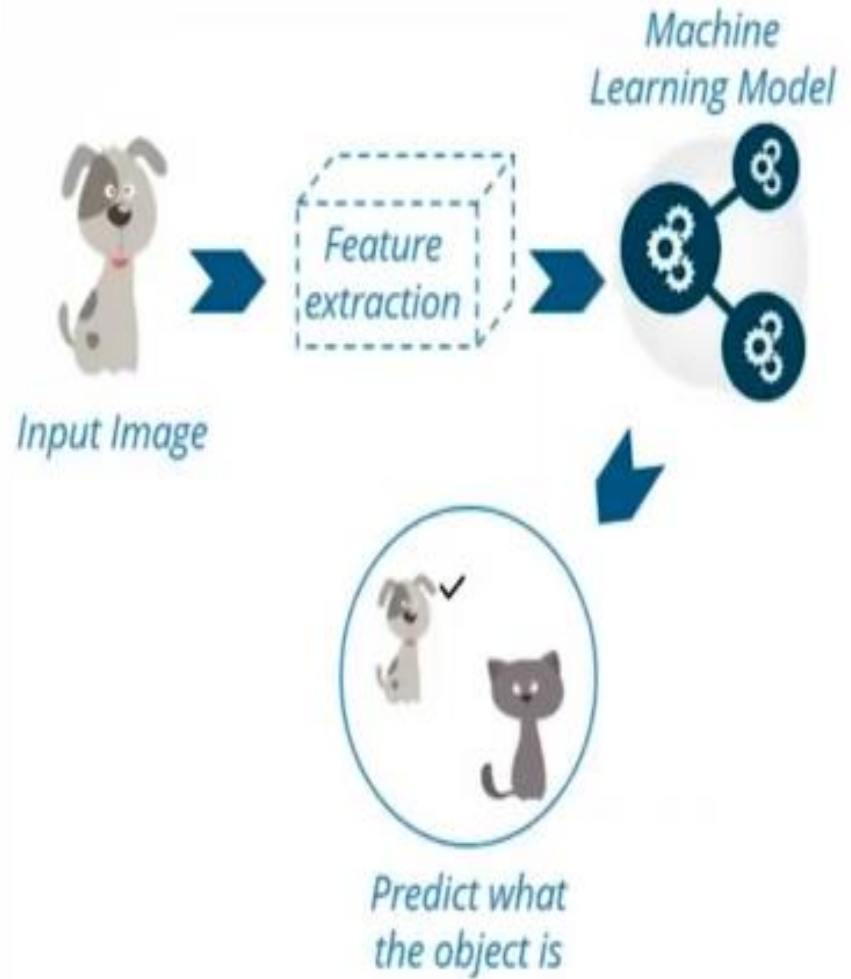
What is Deep Learning

Deep Learning Skips the manual steps of extracting features, you can directly feed images to the Deep learning algorithm, which then predicts the object.

Why do we need Deep Learning

Are not useful while working with high dimensional data, that is where we have large number of inputs and outputs
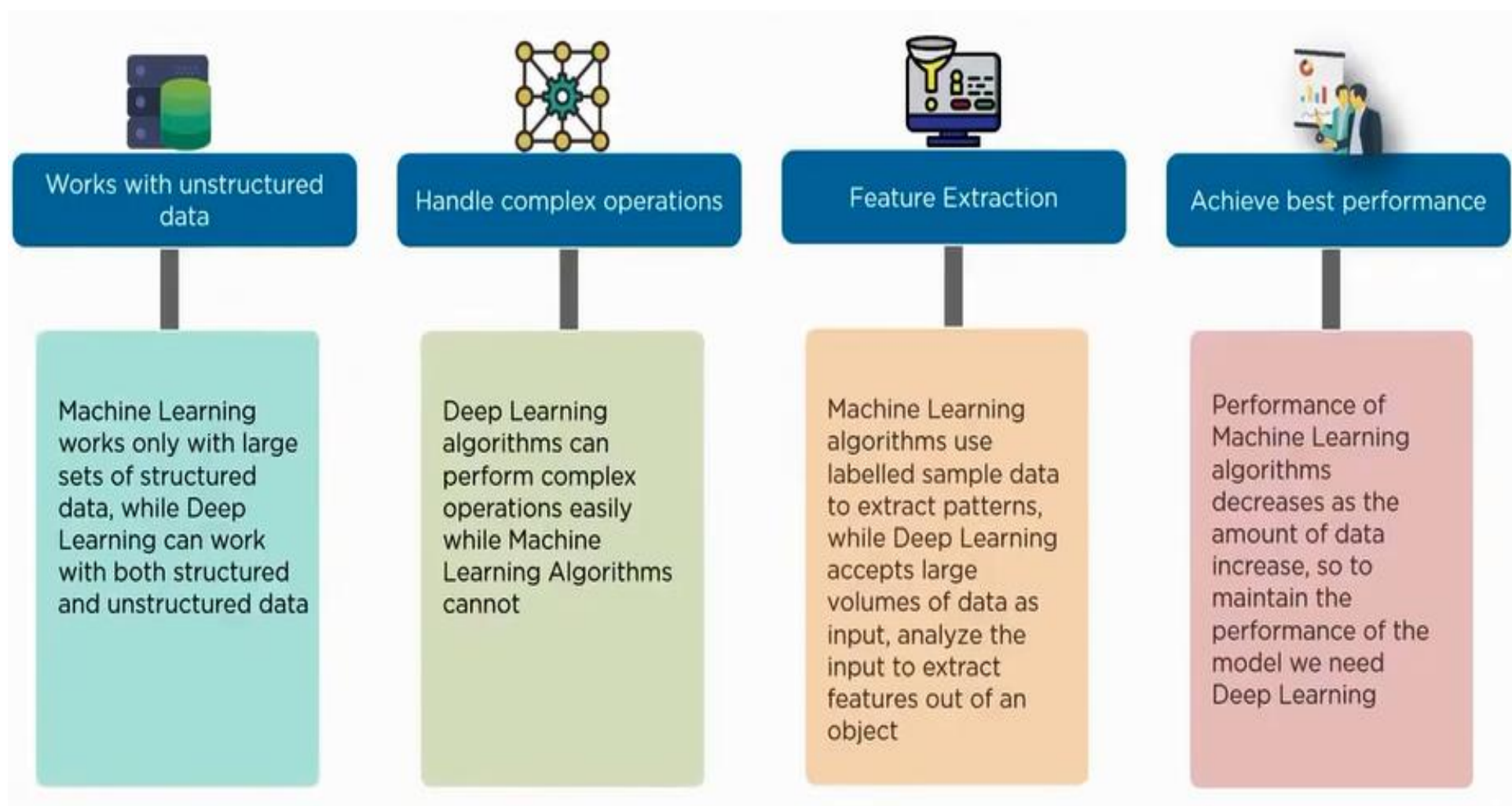
Cannot solve crucial AI problems like NLP, Image recognition etc.

Limitations Of Machine Learning

Machine Learning

Why do we need Deep Learning

Works with unstructured data

Machine Learning works only with large sets of structured data, while Deep Learning can work with both structured and unstructured data

Handle complex operations

Deep Learning algorithms can perform complex operations easily while Machine Learning Algorithms cannot

Feature Extraction

Machine Learning algorithms use labelled sample data to extract patterns, while Deep Learning accepts large volumes of data as input, analyze the input to extract features out of an object

Achieve best performance

Performance of Machine Learning algorithms decreases as the amount of data increase, so to maintain the performance of the model we need Deep Learning

Why do we need Deep Learning

❑ One of the big challenges with traditional Machine Learning models is a process called feature extraction.

❑ For complex problems such as object recognition or handwriting recognition, this is a huge challenge.
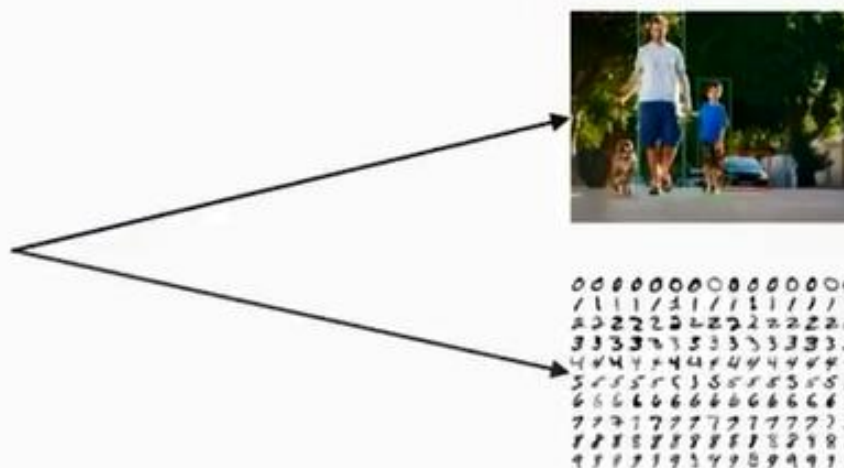
Why do we need Deep Learning

- Deep Learning models are capable to focus on the right features by themselves, requiring little guidance from the programmer.
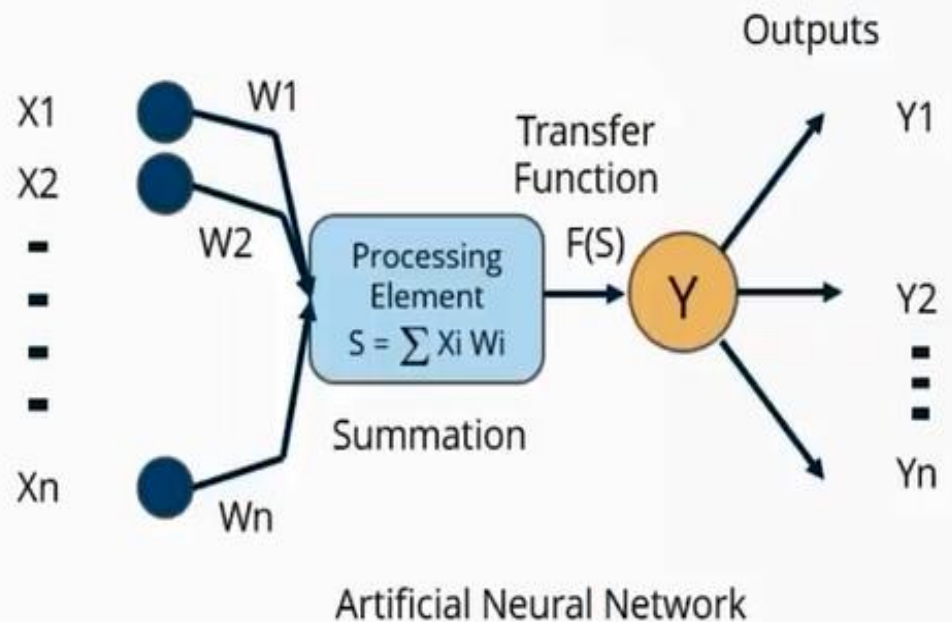- These models also partially solve the dimensionality problem.

What is Deep Learning?

The idea behind Deep Learning is to build learning algorithms that mimic brain.

Why do we need Deep Learning

□ Deep Learning is implemented through Neural Networks.

□ Motivation behind Neural Networks is the biological Neuron.

Neuron

Artificial Neural Network
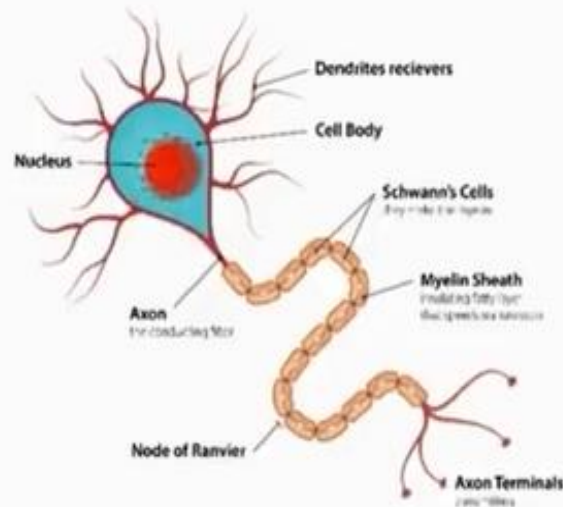
AI, ML and DS
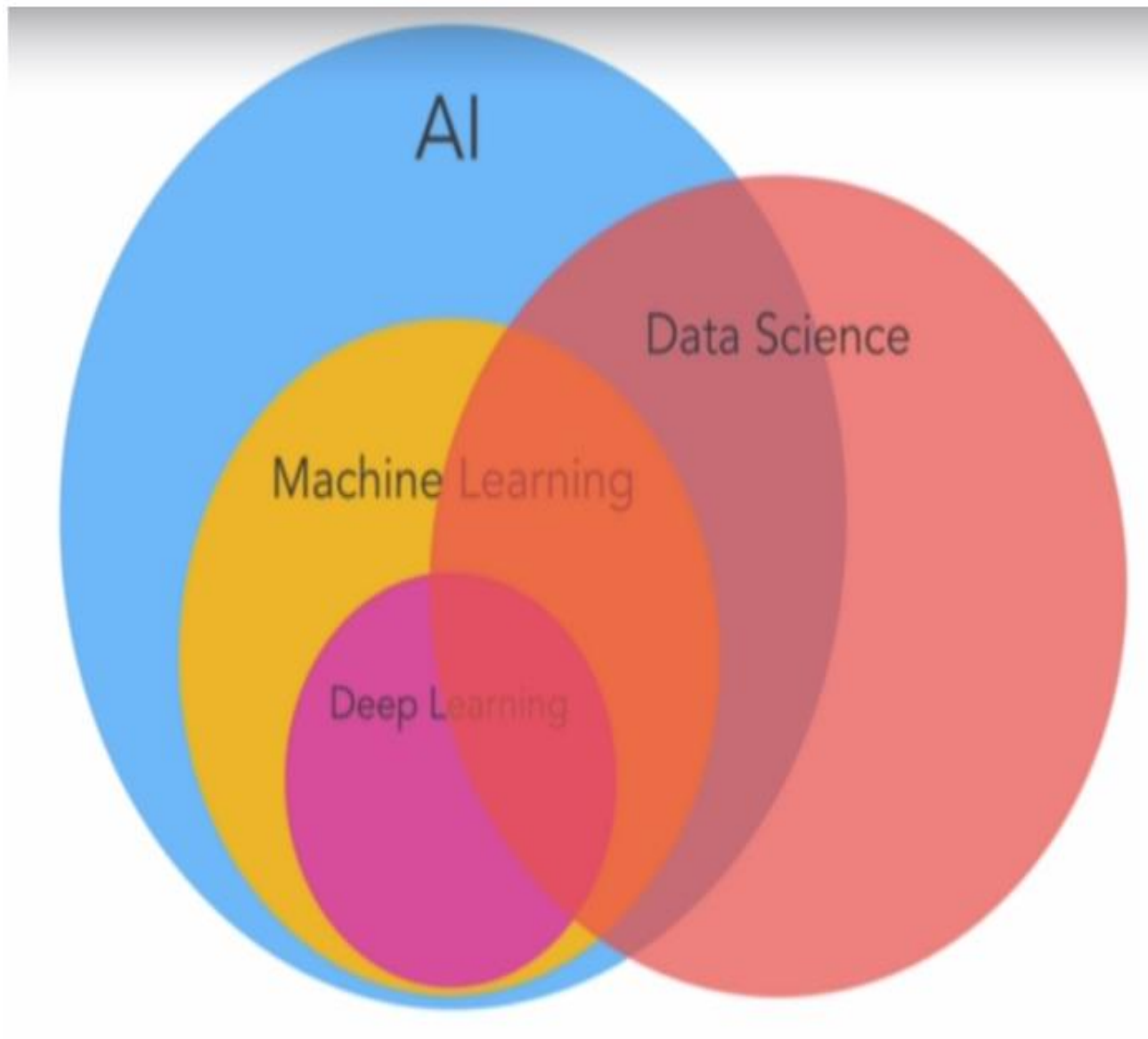
Why Data Science

- ❖ In the last two years alone, **90% of the world's data** has been created.
- ❖ **2.5 Quintillion bytes of data** are produced by humans every day.
- ❖ **Huge Amount of data** and **High Performance Computing** facilities are available.
- ❖ **Data Science** is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics **to extract meaningful insights from Data.**

Purpose of Data Science — Extracting Data → Data Analysis & Processing → Generating Insights from Data
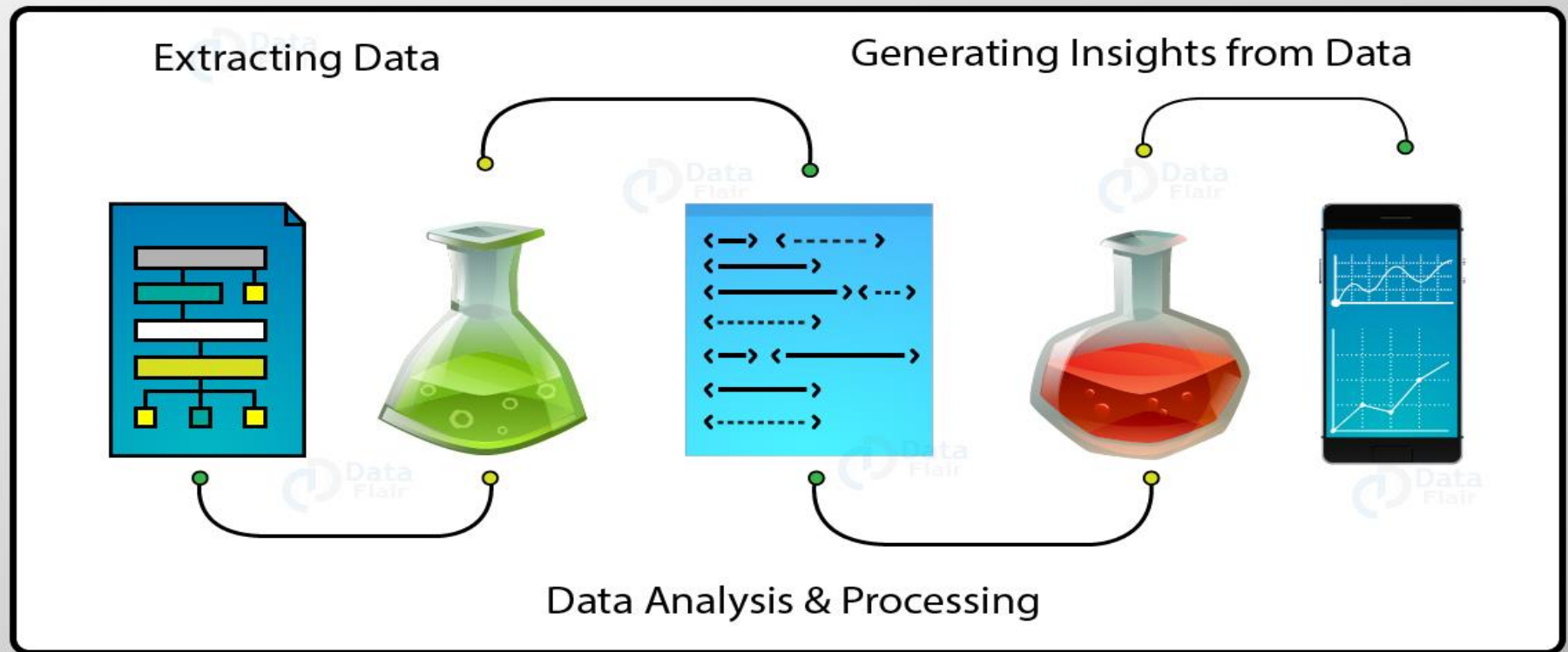
Introduction

❖ **Data Science** is rapidly growing to occupy all the industries of the world.
❖ The **future is Automation** with tremendous scope for Data Science
❖ **Data creates magic.**
❖ **Industries** need data to help them make **Careful Decisions.**
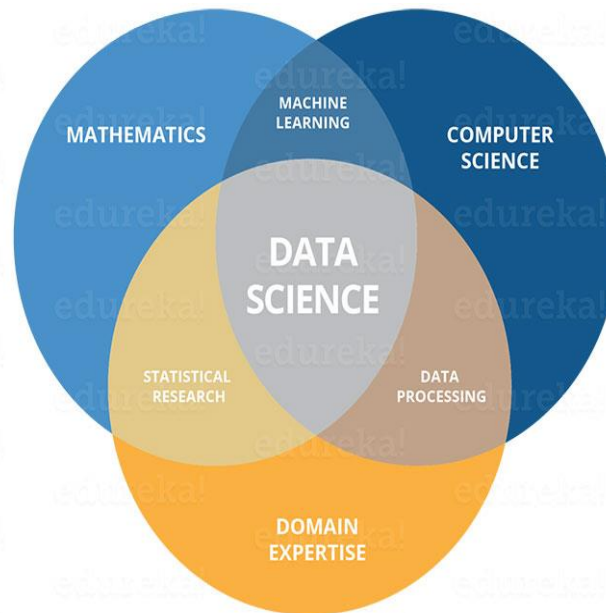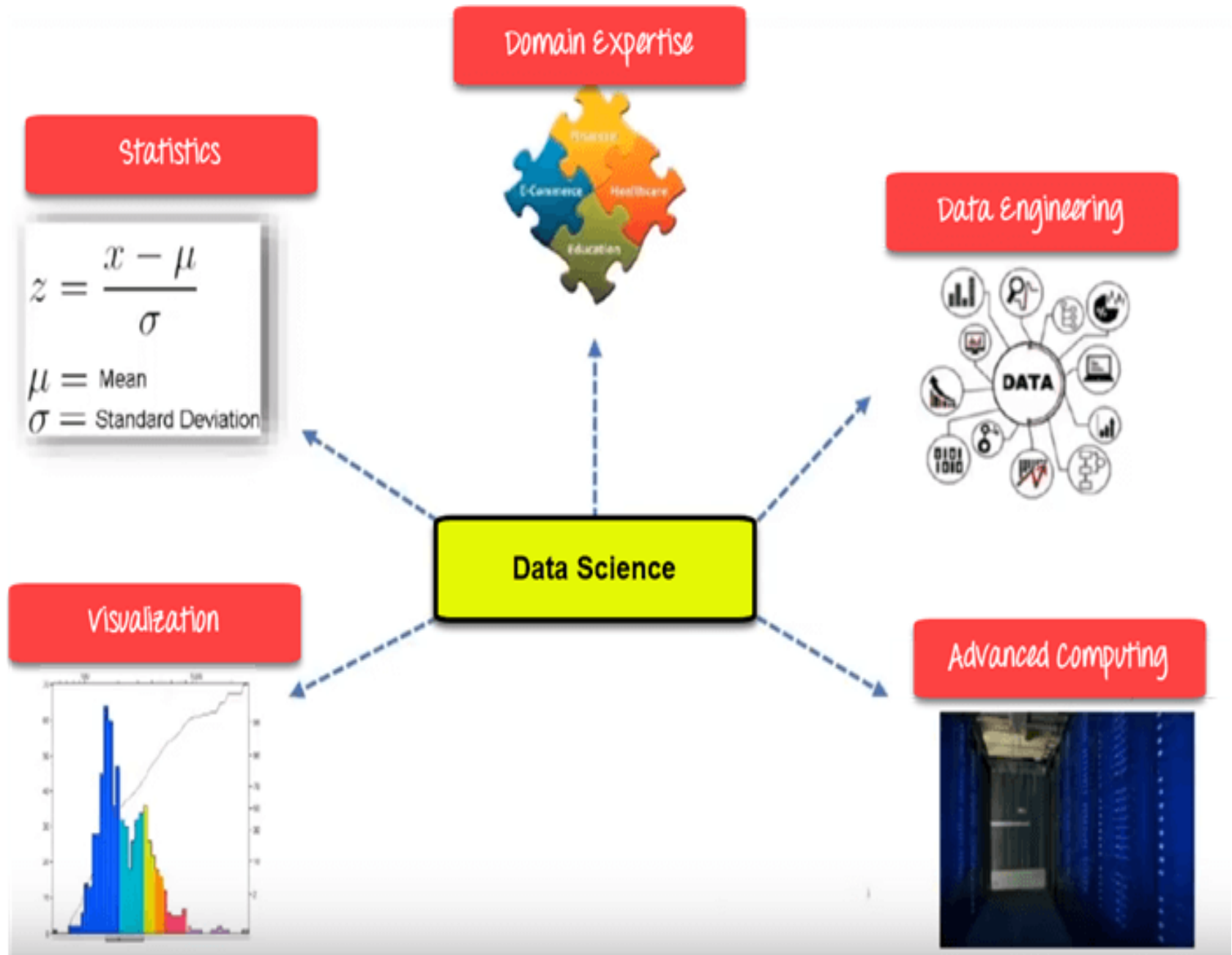❖ Data Science churns raw data into meaningful insights. Therefore, industries need data science.

What is Data Science

- **Data Science** is the science which uses **Computer science, Statistics, Visualization and Machine Learning Principles** to **collect, clean, integrate, analyze, visualize, and to interact with data** to **create data products.**
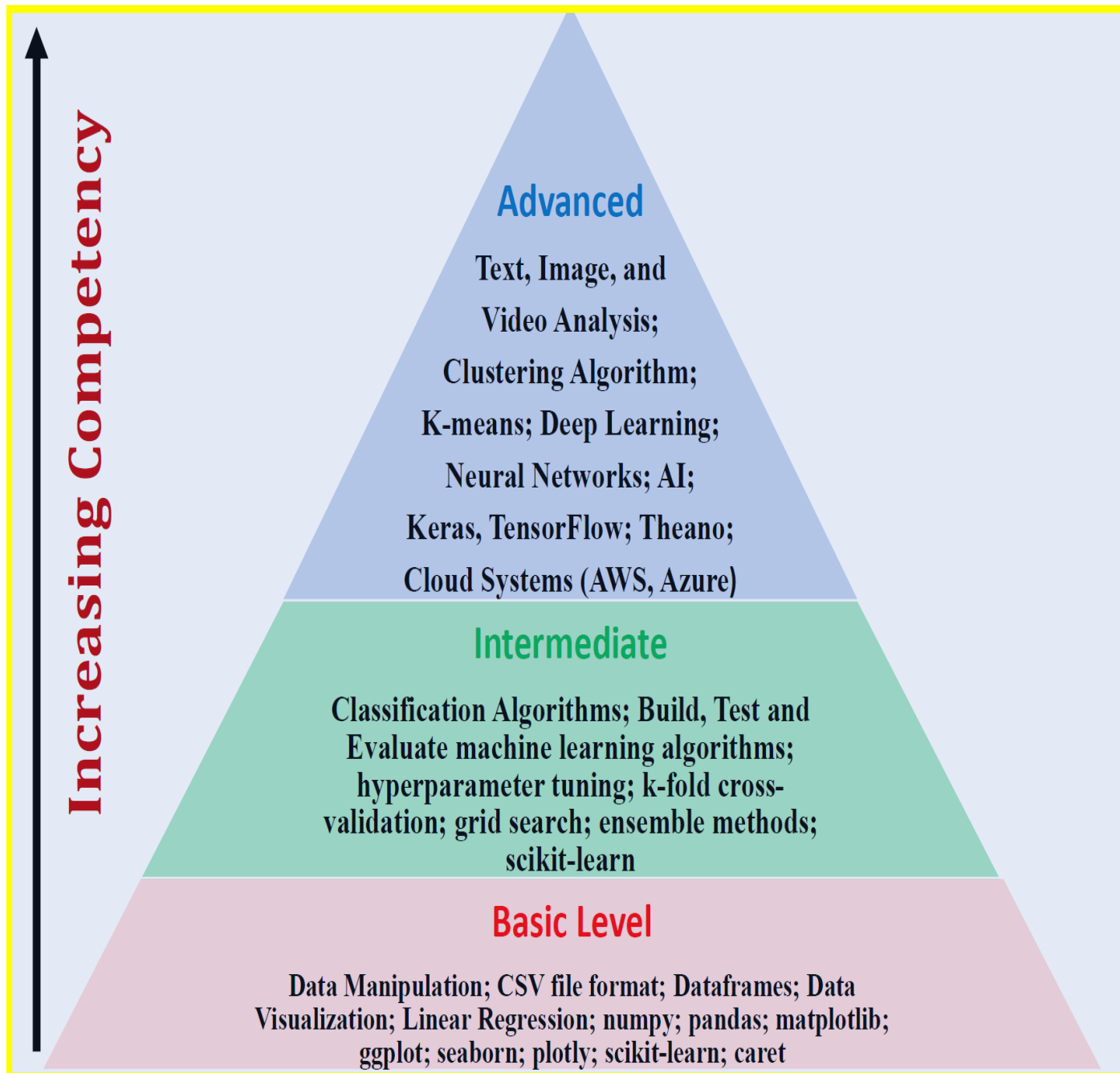

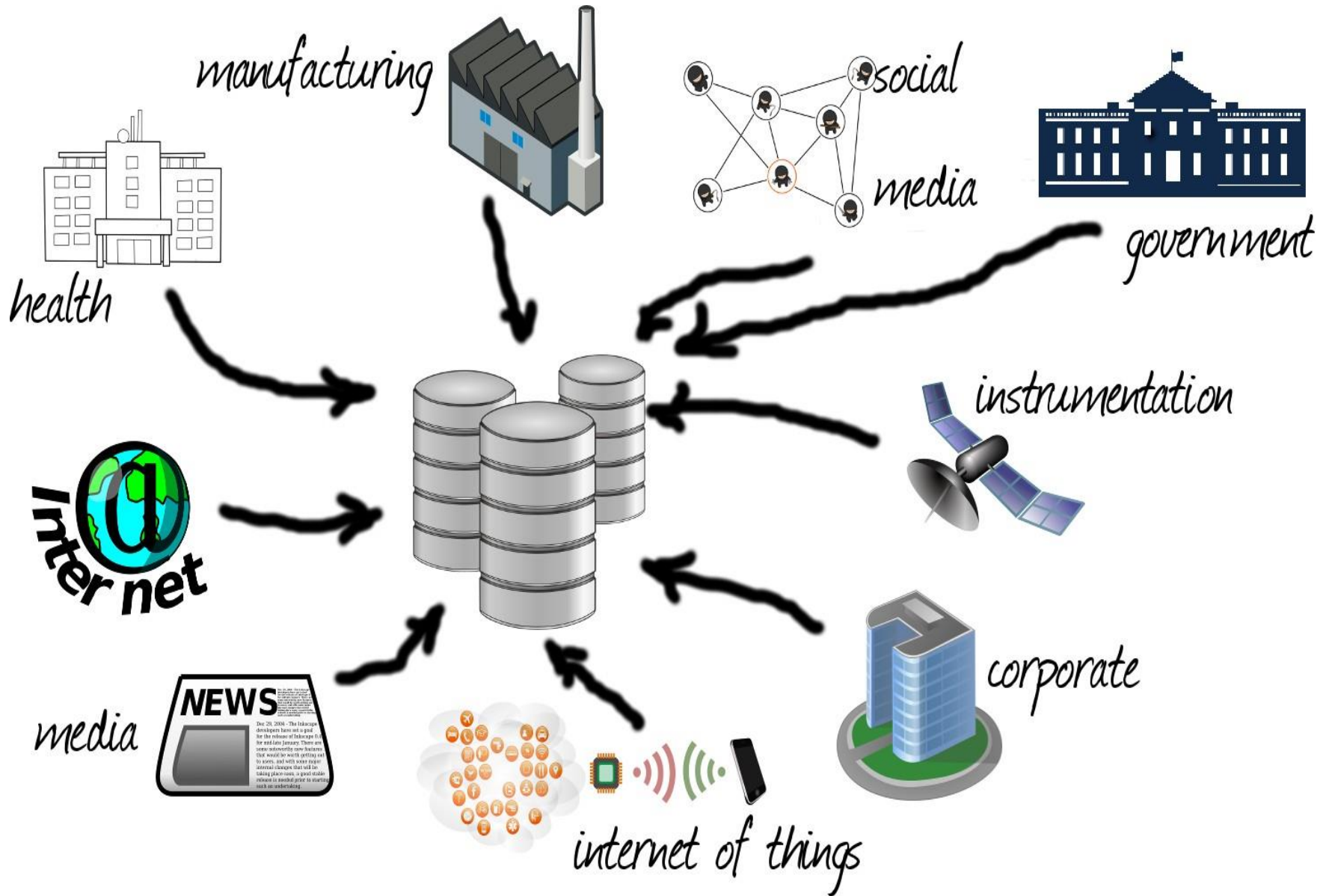**Goal of Data Science - Turn Data into Data Products**

Data Science Components

Levels of Data Science

Increasing Competency

Advanced

Text, Image, and
Video Analysis;
Clustering Algorithm;
K-means; Deep Learning;
Neural Networks; AI;
Keras, TensorFlow; Theano;
Cloud Systems (AWS, Azure)

Intermediate

Classification Algorithms; Build, Test and
Evaluate machine learning algorithms;
hyperparameter tuning; k-fold cross-
validation; grid search; ensemble methods;
scikit-learn

Basic Level

Data Manipulation; CSV file format; Dataframes; Data
Visualization; Linear Regression; numpy; pandas; matplotlib;
ggplot; seaborn; plotly; scikit-learn; caret

Data can be got from many sources

"Data is the New Oil"

Data

- ❖ Data refer to fact and Statistics collected together for reference and Analysis.

- ❖ Data can be collected/Stored

- ❖ Data can be Measured

- ❖ Data can be Analyzed
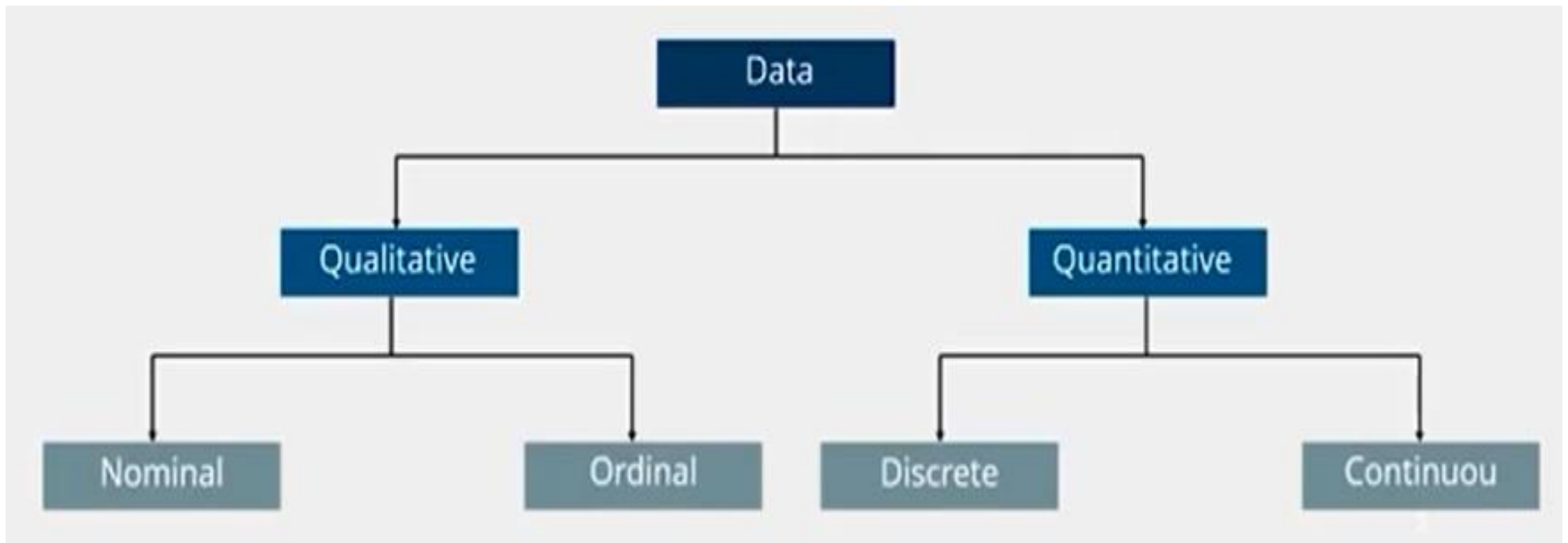
- ❖ Data can be visualized

What is Data

Data refers to facts and Statistics collected together for reference or Analysis

Categories of Data

Data is categorized in to **Qualitative** and **Quantitative**

Qualitative Data

Qualitative data deals with characteristics and descriptors that can't be easily measured, but can be observed subjectively.

Nominal Data
Data with no inherent order or ranking such as gender or race, such kind of data is called Nominal data

Ordinal Data
Data with an ordered series, such as shown in the table, such kind of data is called Ordinal Data

Gender
Male
Female
Male
Male

Customer ID	Rating
001	Good
002	Average
003	Average
004	Bad

Quantitative Data

Quantitative data deals with numbers and things you can measure objectively

Discrete Data
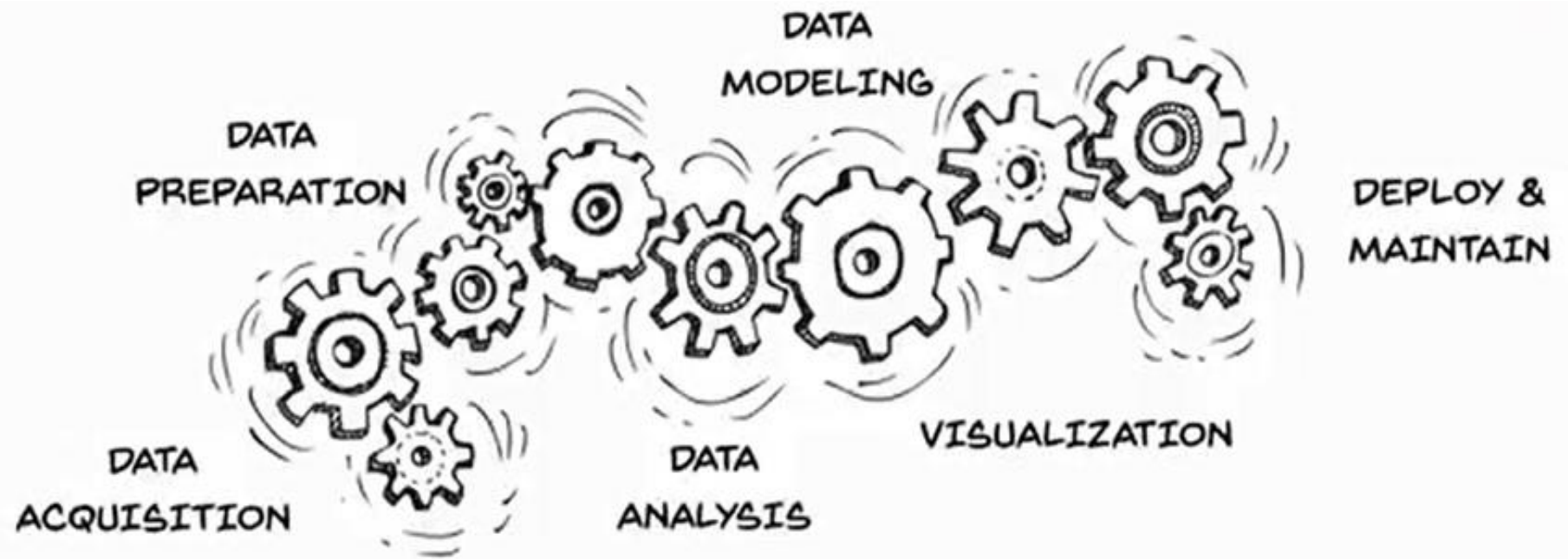Also known as categorical data, it can hold finite number of values

Continuous Data
Data that can hold infinite number of possible values

Example: Number of students in a class

Example: Weight of a person

Data Science

Business Problem Understanding

Data Acquisition
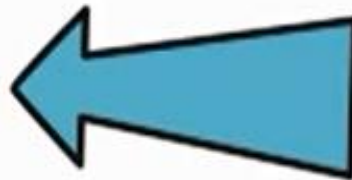
Data Preparation
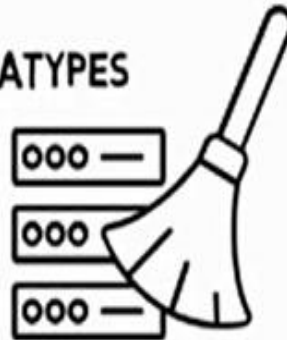
Exploratory Data Analysis

④ EXPLORATORY DATA ANALYSIS

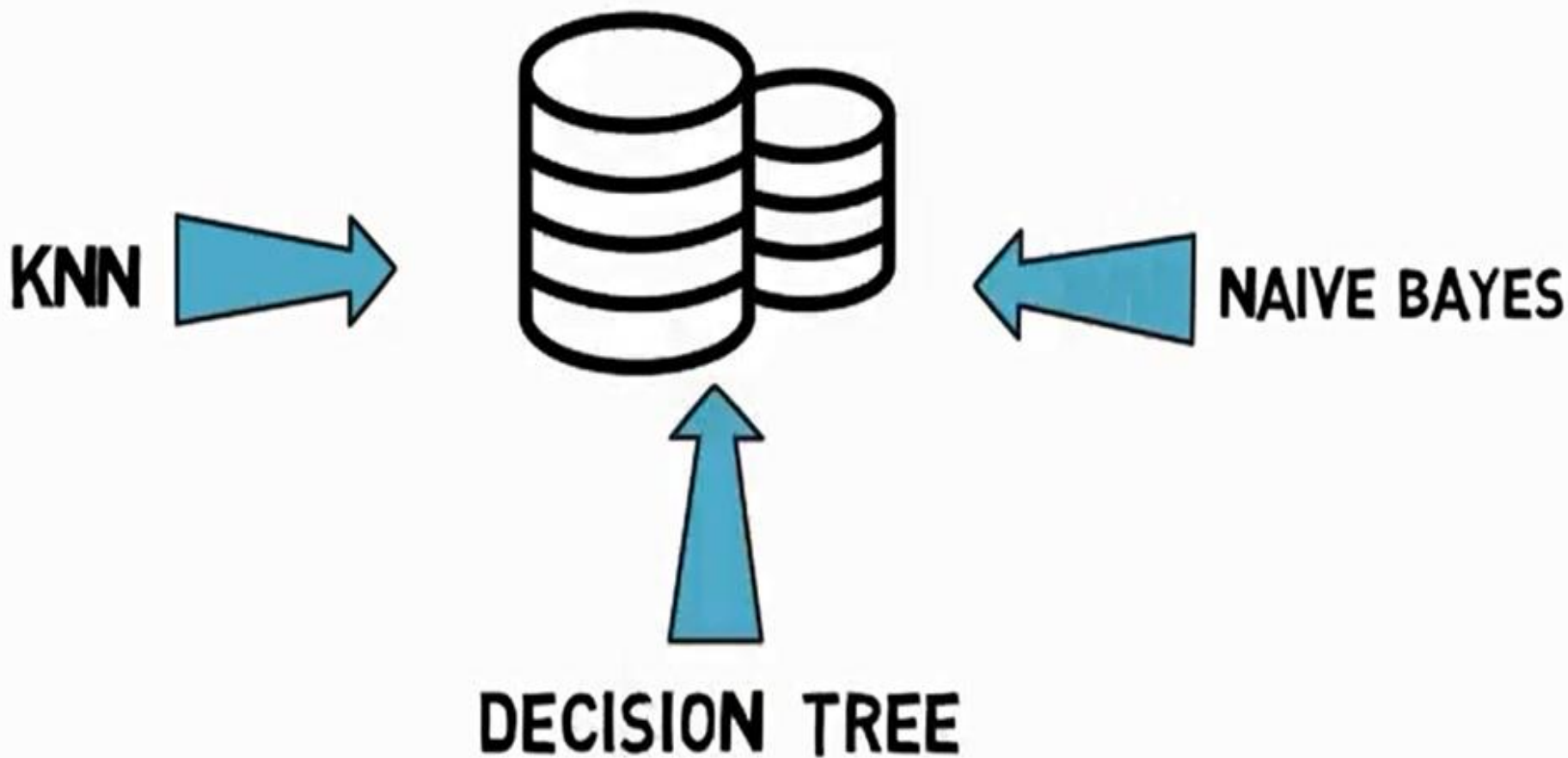**Exploratory Data Analysis (EDA)** is an analysis approach that identifies general patterns in the data.

DEFINES AND REFINES

THE SELECTION OF FEATURE

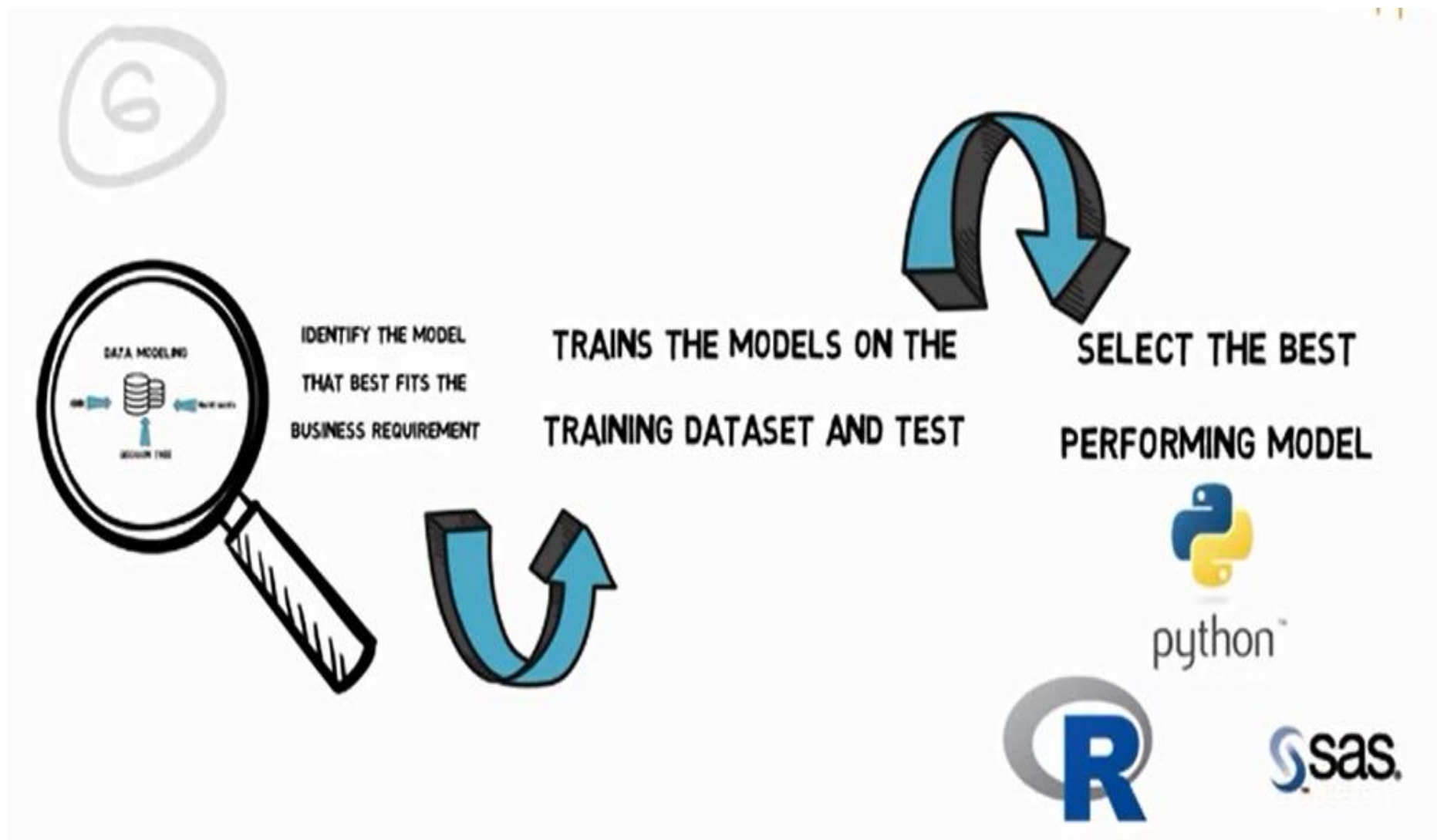VARIABLES THAT WILL BE USED

IN THE MODEL DEVELOPMENT

Difference Between Data Science and M L

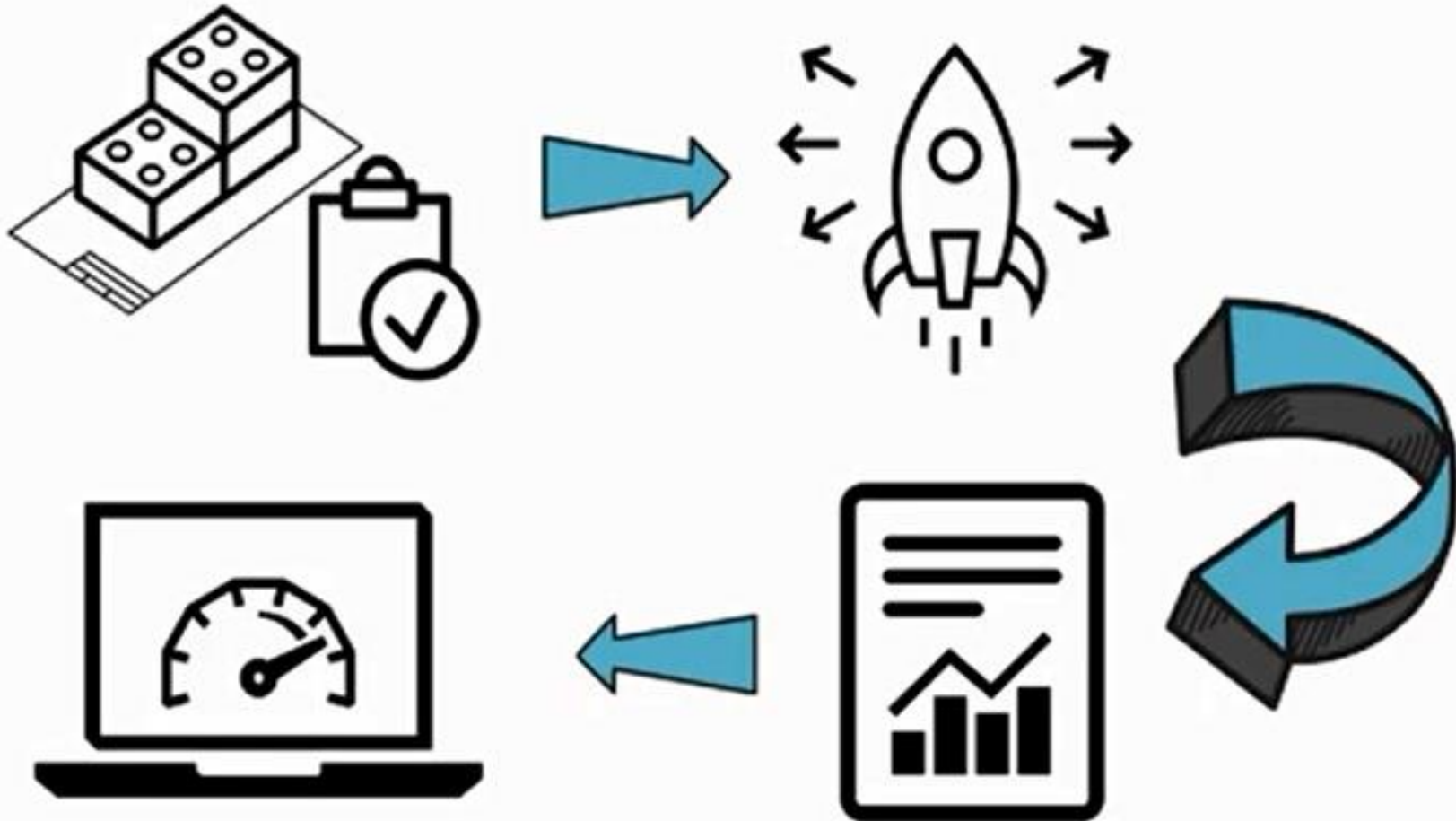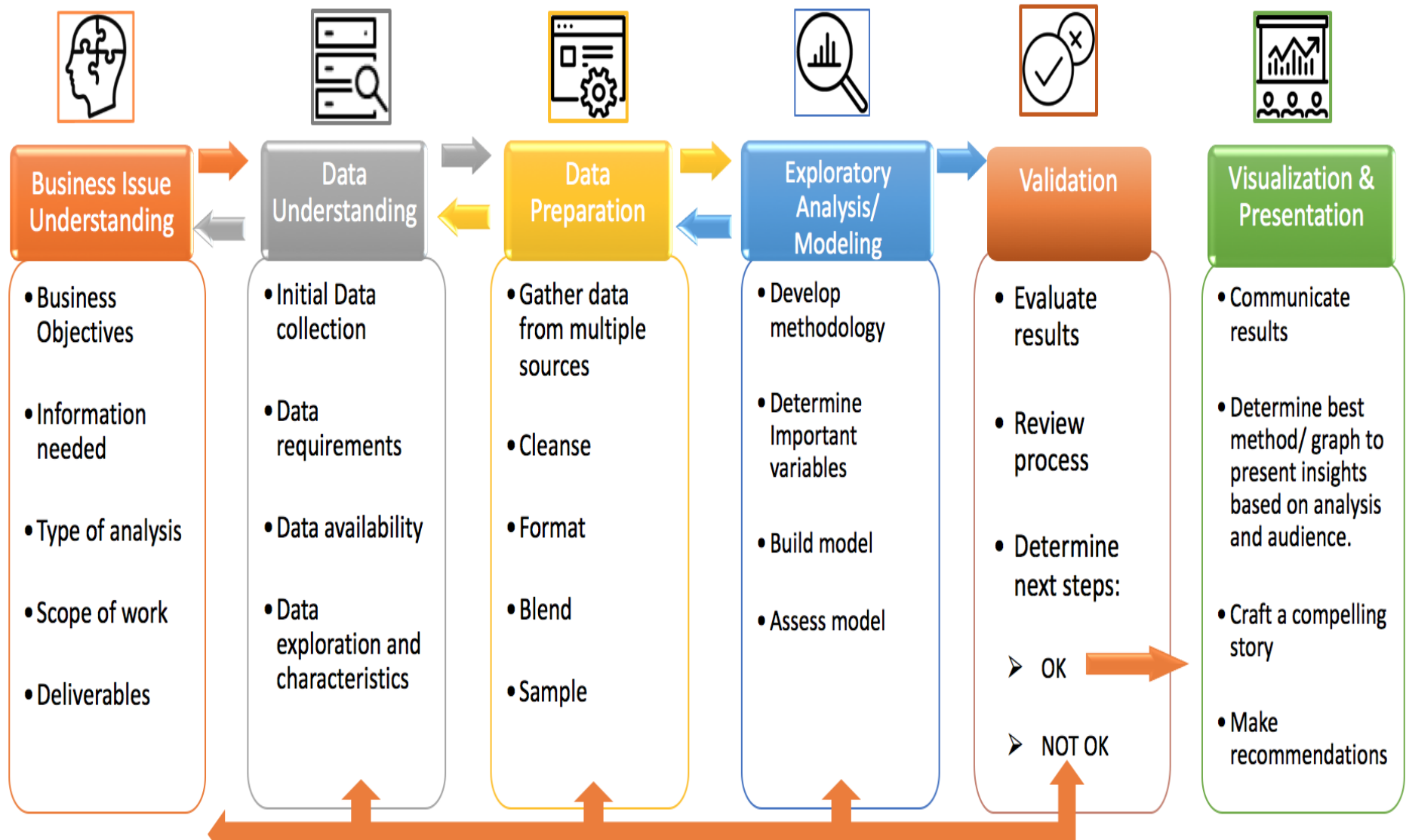| Data science	Machine Learning
Data Science technique helps you to create insights from Data dealing with all Real-World Complexities.	Machine Learning method helps you to predict from historical data with the help of Mathematical models.
Data science can work with manual methods as well, though they are not very useful.	Machine learning algorithms hard to implement manually.
Data science is a complete process.	Machine Learning is a single step in the entire Data Science process.
Data Science is not a subset of Artificial Intelligence (AI).	Machine Learning is a subset of Artificial Intelligence (AI).

Data Modeling

Visualization and Communication

[Tableau](#)

Life Cycle of Data Science Projects

Business Issue Understanding
- Business Objectives
- Information needed
- Type of analysis
- Scope of work
- Deliverables

Data Understanding
- Initial Data collection
- Data requirements
- Data availability
- Data exploration and characteristics

Data Preparation
- Gather data from multiple sources
- Cleanse
- Format
- Blend
- Sample

Exploratory Analysis/ Modeling
- Develop methodology
- Determine Important variables
- Build model
- Assess model

Validation
- Evaluate results
- Review process
- Determine next steps:
  ➢ OK
  ➢ NOT OK

Visualization & Presentation
- Communicate results
- Determine best method/ graph to present insights based on analysis and audience.
- Craft a compelling story
- Make recommendations

Data Science

Parts of a Data Science Project

Collection: getting the data
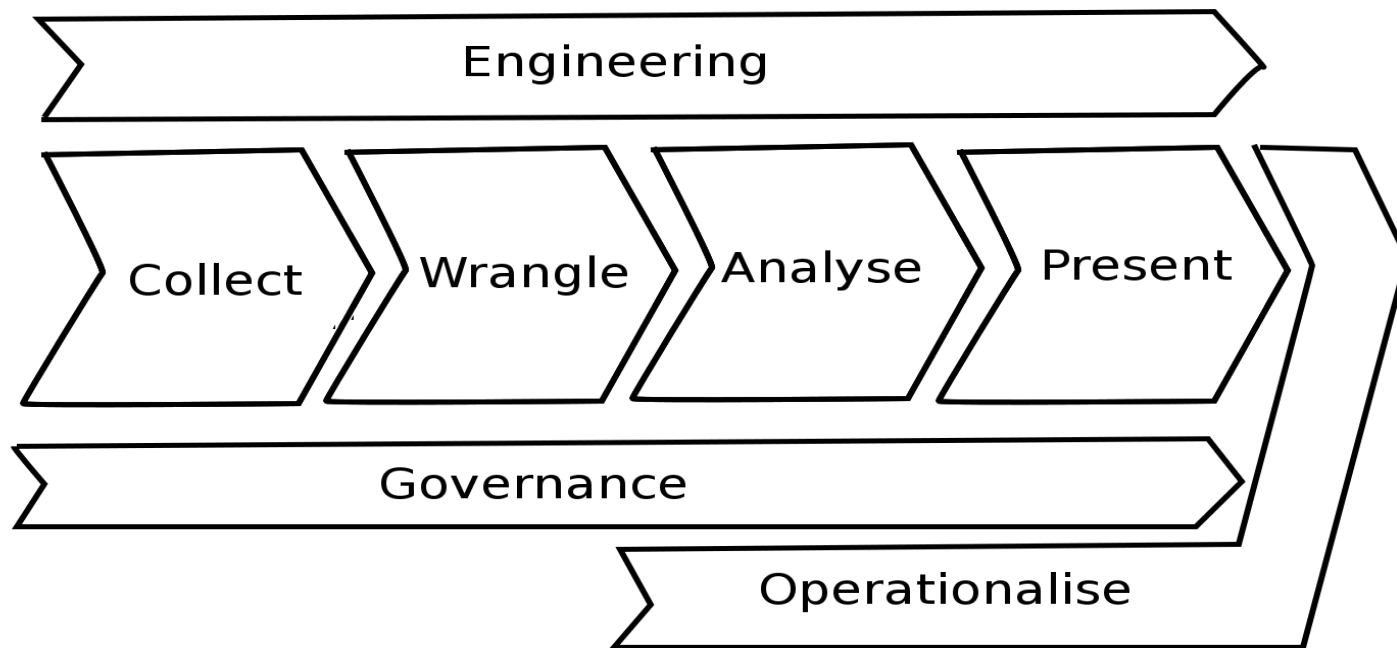**Engineering:** storage and computational resources
Governance: overall management of data
Wrangling: data preprocessing, cleaning
Analysis: discovery (learning, visualisation, *etc.*)
Presentation: arguing that results are significant and useful
Operationalisation: putting the results to work

Wrangling

If customer records are in 4 different databases in different formats; you want a single standardised set of customer names and addresses.

► Convert addresses in your customer database into geographic latitude and longitude.

► Convert free text dates to standard format, *e.g.* "next Tuesday", "2nd January 15", "January 3 next year", "3rd   Friday in the month", "03/31/15", "31/03/15"

► Recognize what values in your data are **"unknown" or "illegal"**

► Want the **core news text, title, date**, *etc.* off the following page
  Apple's iPhone loses top spot to Android in Australia

► Want the **text plus details from the PDF file**
  "Data Wrangling: The Challenging Journey from the Wild to the L

► Want to digitize the text off a *scanned letter*

Tools for the Data Science-Common Softwares

Access: SQL, Hadoop, MS SQL Server, PIG, Spark

Data Analysis: R, Spark, Python and SAS

Wrangling: Common Scripting Languages (Python, Perl)

Visualization: Tableau, Matlab, Javascript+D3.js

Statistical Analysis: Weka, SAS, R

Multi-purpose: Python, R, SAS, KNIME, RapidMiner

Cloud-based: Azure ML (Microsoft), AWS ML (Amazon)

Python versus R

▶ Both are free

▶ R is better for Stand-alone analysis and exploration

▶ Python lets you integrate easier with other systems

▶ Python easier to learn and extend than R (better language)

▶ R has vectors and arrays as first class objects; similar to Matlab

▶ R currently less Scalable.

Fake News Detection

❖ With this data science project idea, you can use Python language to develop a specific model that can precisely detect whether the news is real journalism or false information.

❖ For this, 'Passive Aggressive Classifier' to classify the news into either a "Real" and "Fake" segmentations.

❖ The main idea of this Data Science project is to develop a real-time machine learning model that can correctly detect social media news authenticity.

Forest Fire Prediction

❖ One of the alarming & common disasters happening in today's world is forest fires. These disasters are highly damaging to the ecosystem.

❖ To deal with such a disaster, a lot of money on infrastructure & controlling and handling is required.

❖ We can build a Data Science project using 'k-means clustering'- it can identify any forest fires hotspots along with the severity of the fire at that particular spot.
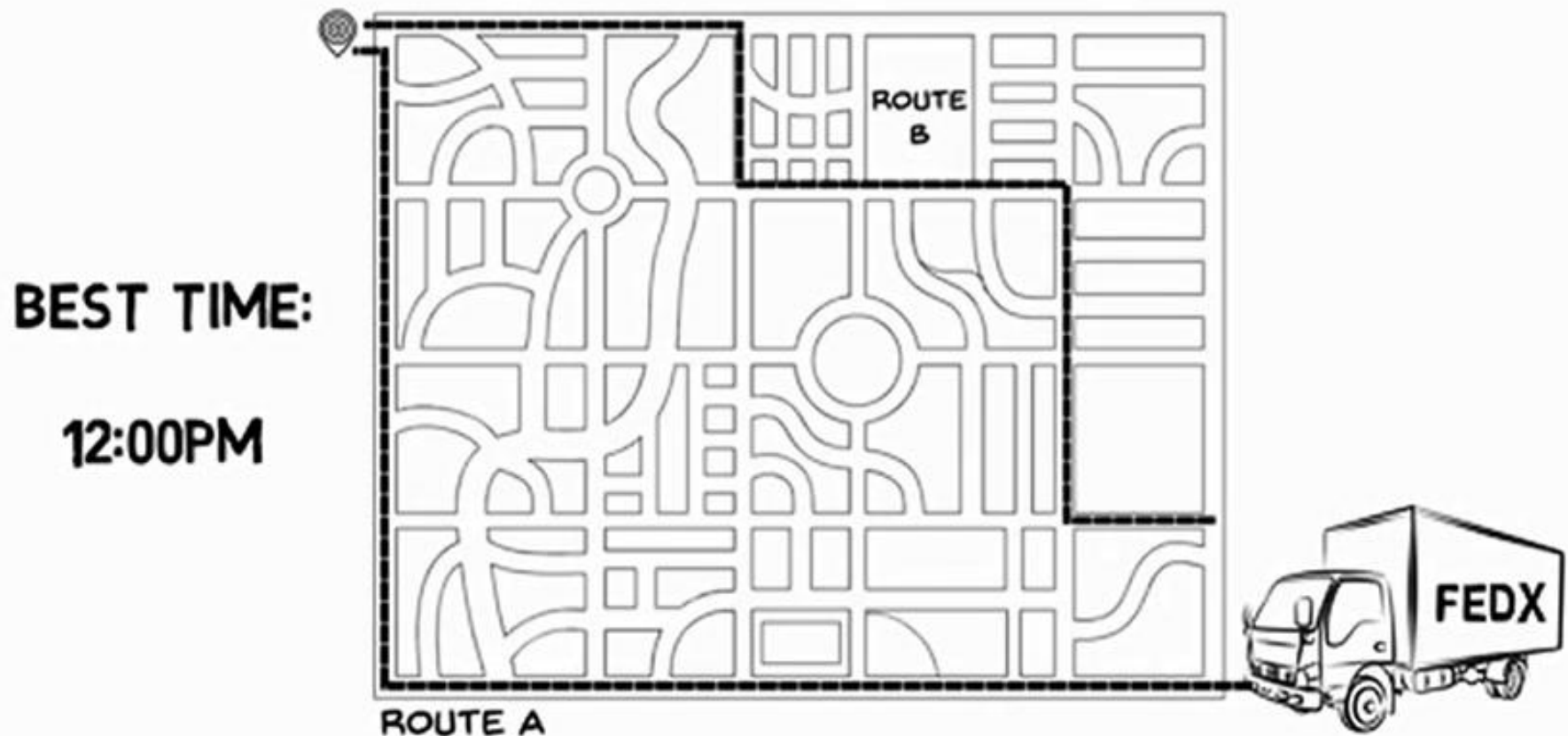
Airlines

AIRLINE COMPANIES CAN NOW EASILY PREDICT

FLIGHT DELAY AND NOTIFY THE PASSENGERS

H487	CANCELLED
PP87	DELAYED

Revenue Management And Route Planning, Fuel Consumption And Optimization, Boarding And Checking Bags With Facial Recognition, Preparing Plane For Next Flight

Video Surveillance: Tracking
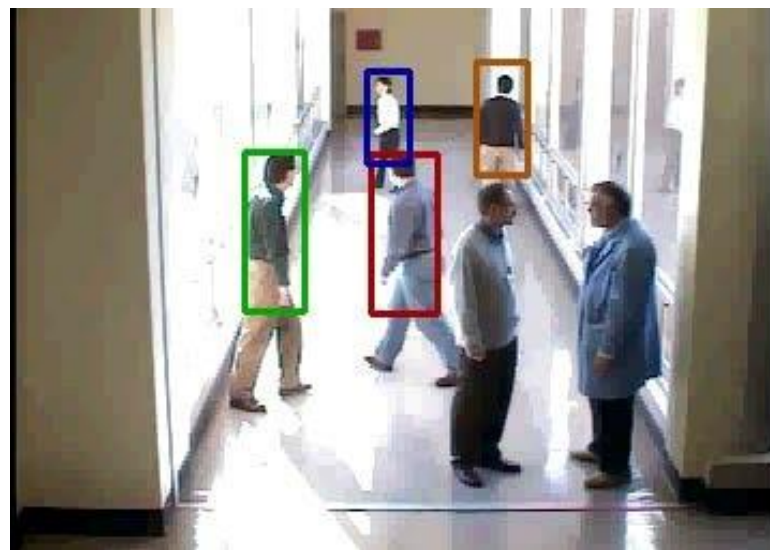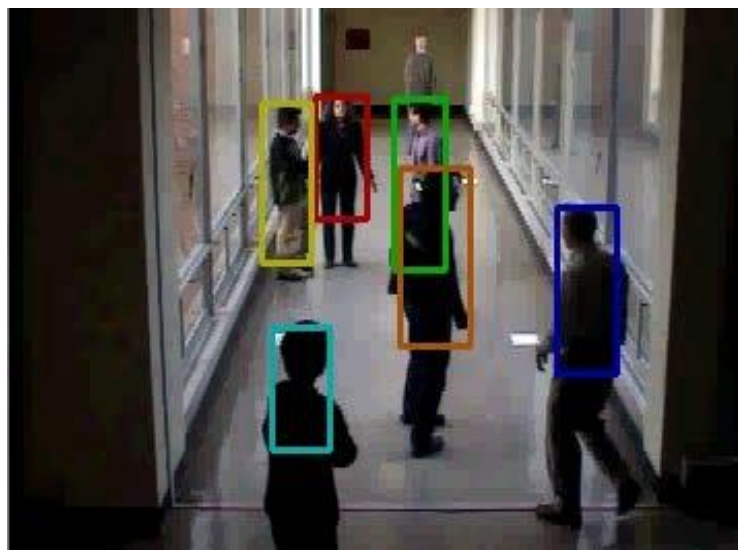
Waste Management and Healthcare Using Data Science

❖ The world produces over **2 billion tons of municipal solid waste** every year.

❖ The World Bank warns that global waste will increase up to 70 percent on current levels by 2050 unless **urgent actions are undertaken.**

❖ **Data science** can change the health care sector in so many ways.

❖ This, in turn, allows to **optimize the workforce and throughput, improves care recipients' satisfaction, and balances the supply.**

Top 5 cities in India which generate the highest municipal waste

3.3	2.7	1.6	1.4	1.1
Delhi	Mumbai	Chennai	Hyderabad	Kolkata

Total Municipal Waste Generated (Million Tonnes Per Year)

Expected outcome

o Projects - Socially useful student projects, Research projects,

o Start-up

o Commercialize product

o Publish- Quality papers, Patent

O Explore research areas in Waste management, Health Care using Data Analytics, Transportation, Logistics, Finance, Robotics etc.

Impact of Data Science on Industries

❖ **Data Science Impact on Various Sectors**

❑ Healthcare
❑ Automotive
❑ Financial Services
❑ Retail
❑ Technology
❑ Manufacturing
❑ Energy
❑ Transport and Logistics
❑ Robotics
❑ Analytics
❑ Education etc.

No sector or business is in any way immune from the impact of Data Science.

Data Science Engineer Skills

1. Analytical skills

2. Creativity

3. Data Driven Problem-solving skills

4. Critical thinking skills

5. Communication skills

6. Leadership skills

7. Self Learning

8. Mathematics and Algorithms

9. Cognitive Computing
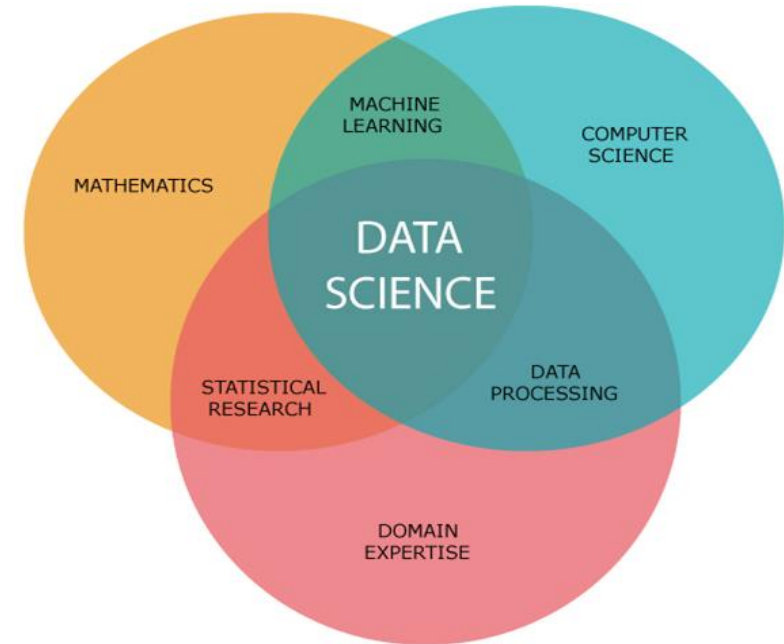
Job Roles & Opportunities

Average Salary 15 Lakhs

Principal Data Scientist – Rs. 17,11,180

Computer Scientist – Rs. 16,24,615

Machine Learning Engineer – Rs. 9,50,000

Data Engineer/Data Architect – Rs. 8,35,755

Data Scientist – Rs. 6,99,928

Research Engineer – Rs. 6,52,230

Algorithm Engineer – Rs. 5,40,220

Data Analyst – Rs. 4,97,550

Computer Vision Engineer – Rs. 4,50,000

Machine Learning Analyst
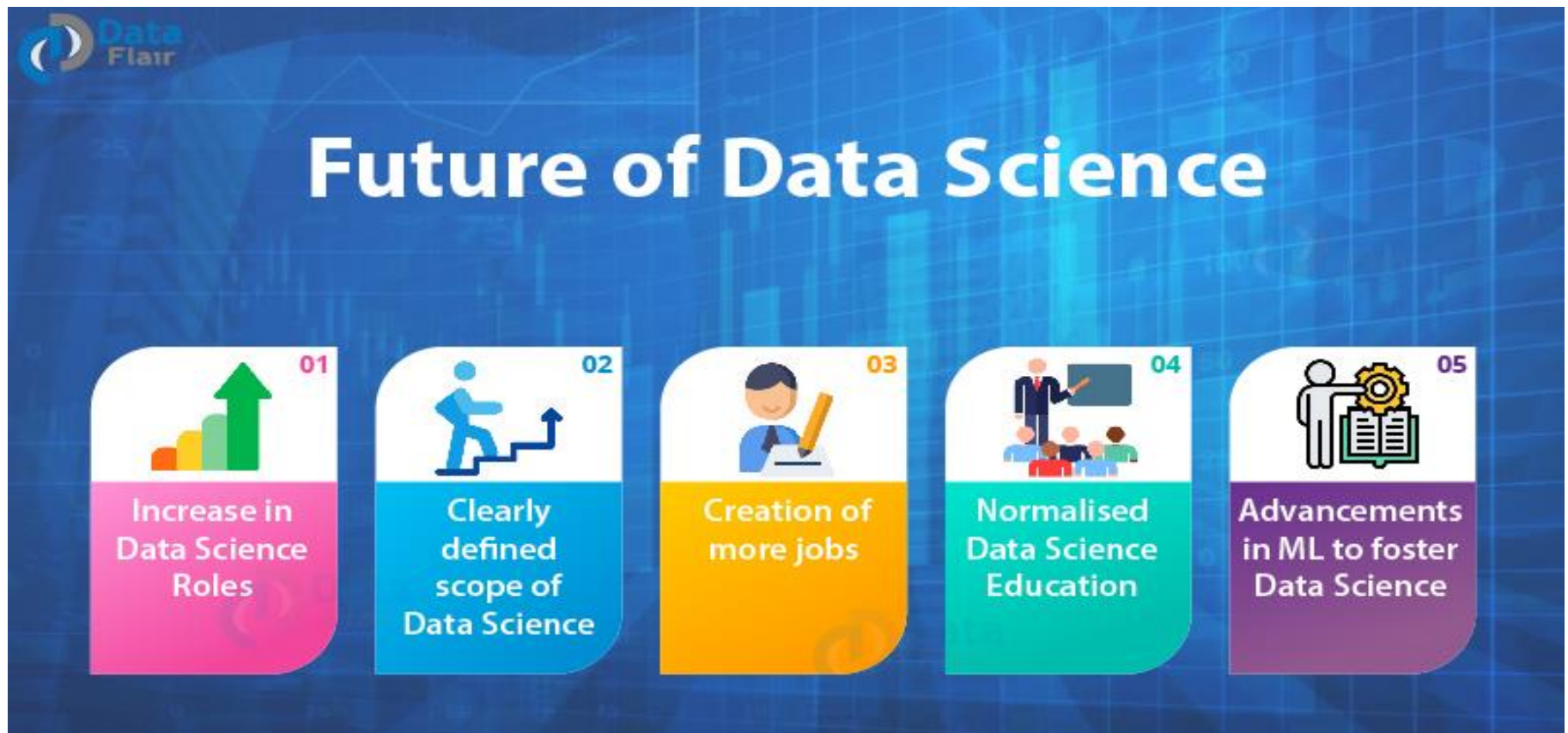
NLP Data Scientist, Risk Analyst

Data Sciences Lead, Data Manager

Machine Learning Scientist etc.

Market Trends and Projections

- **2.3 Millions of Jobs in Data Science in Future (Gartner Report)**.

- Business Analyzing data are estimated to be around **$430 billions** by over 2021 over their rivals.

- **39% increase** in demands for **Data Scientist** and **Data Engineers** by 2021.

Research Opportunities

- Waste Management and Health Analytics using Data Science.

- Handling synchronized video analytics in a distributed cloud.

- Multiple and Heterogeneous information sources

- Predictive Business Analytics

- Computerizing front-end phases of the data life cycle

- Building domain-sensitive large scale frameworks

- Predictive Maintenance, Autonomous Car, & Surveillance etc.

- **Research Projects, Collaborative Projects** with Industry

- **Research Publications & Patents** etc.

- **Post covid-19 Situation**, there is **lot of scope for Start-ups** in **Data Science domain.**