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 2022 the digital universe will reach 44 zettabytes..."**

That's 44 trillion gigabytes!

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...

Machine Learning definition

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

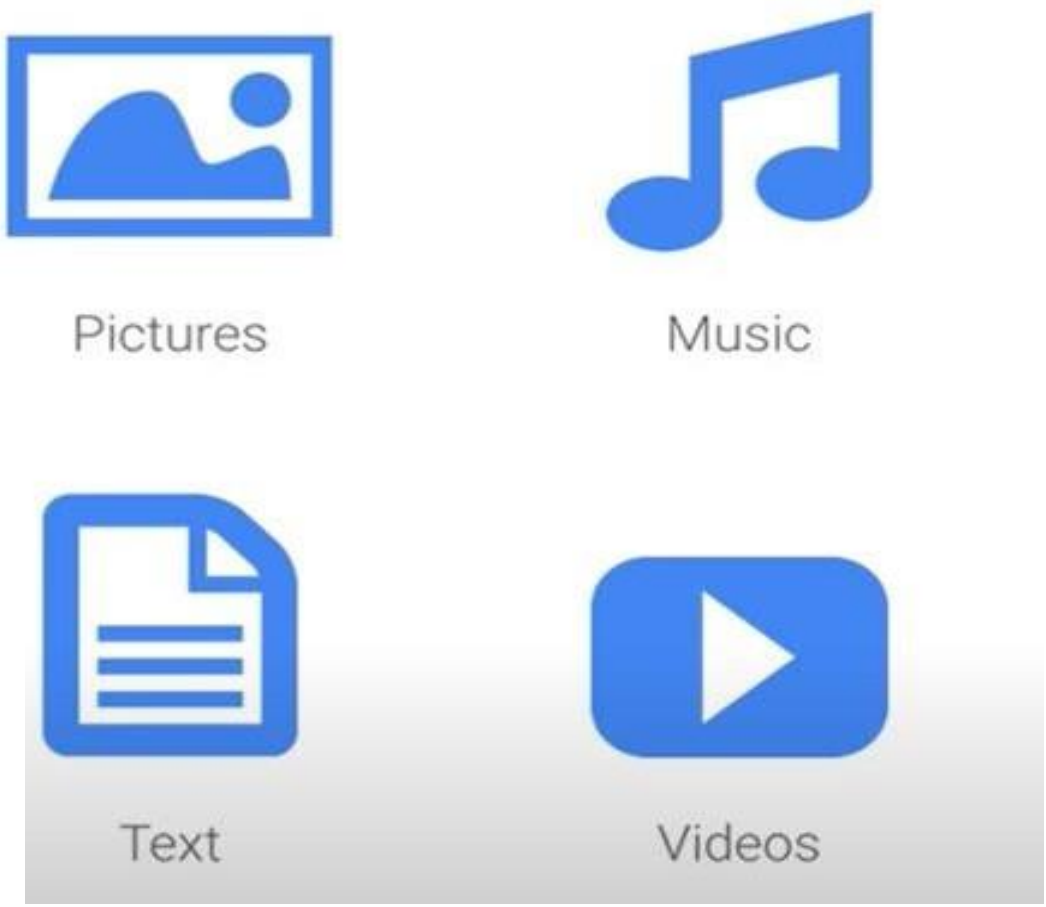
Tom Mitchell

http://videolectures.net/mlas06_mitchell_itm/

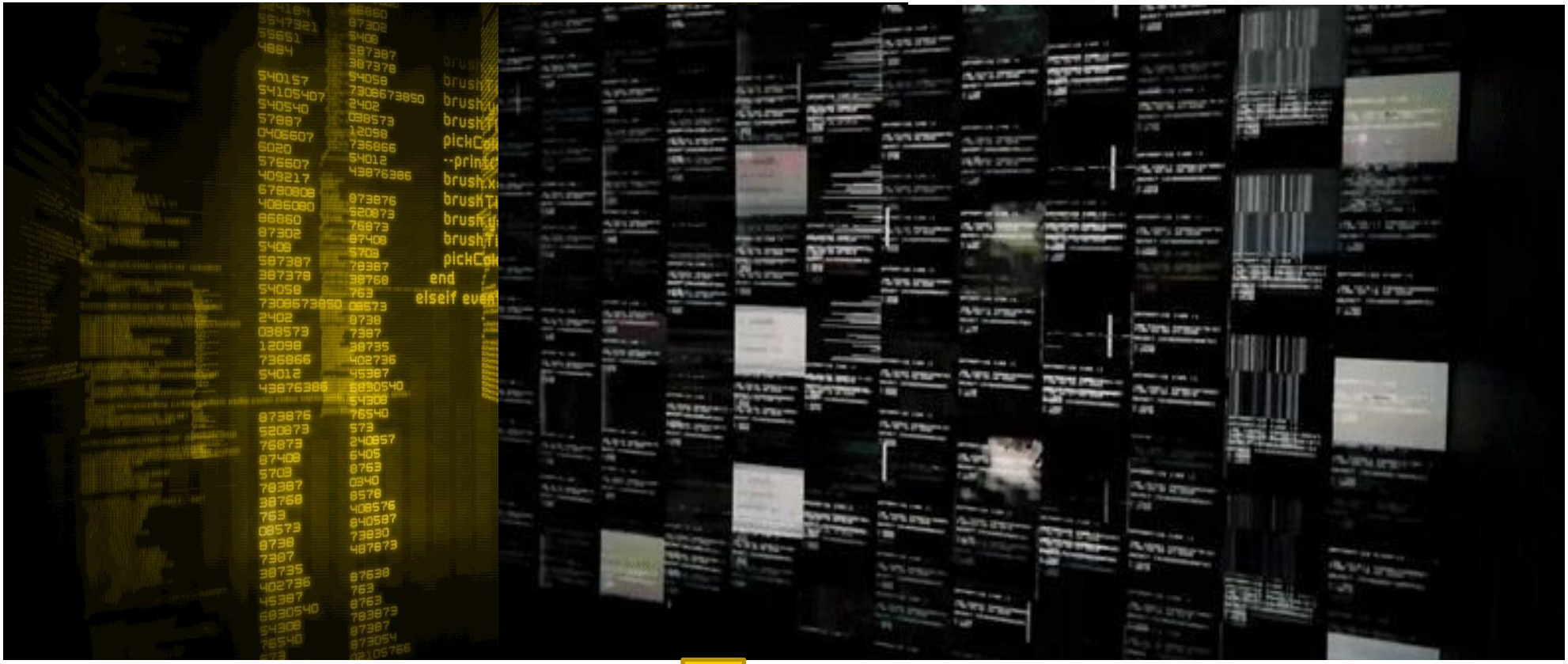
"**A computer program** is said to **learn** from **experience *E*** with respect to **some class of tasks *T*** and **performance measure *P*** , **if its performance *P* at tasks *T*, improves with experience *E*.**"

Tom Mitchell. Machine Learning 1997.

Why Machine Learning ?



Why Machine Learning ?



DATA

Identifying The Factors/ Opportunities

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

Why Machine Learning ?

- Machine Learning Systems **learn from data samples of solved cases.**
- They **do not require any expert knowledge,** since they infer such knowledge directly from the data.
- They are **useful in professional fields in which expertise is rare** and the codification of knowledge is limited.
- They are useful in domains where good tests and measurements are available, but methods of applying this information is insufficiently understood or systematized.
- They are useful in domains in which the information needs to be constantly updated, in order to maintain the system in routine use at high levels of performance

[From *Computer Systems that Learn*, Weiss & Kulikowski, Morgan Kaufmann, 1990]

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.

What is Machine Learning?

- **Machine Learning**
  - Study of algorithms that improve their performance at some task with experience
- Optimize a performance criterion using example data or past experience.
- **Role of Statistics:** Inference from a sample
- **Role of Computer science:** Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

Why "Learn"?

- **Machine Learning** is programming computers to optimize a performance criterion using example data or past experience.
- **Learning is used when**:
  - Human expertise does not exist (Navigating on Mars),
  - Humans are unable to explain their expertise (Speech Recognition)
  - Solution changes in time (Routing on a Computer Network)
  - 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:

- Introduction
- Types of Learning
- Mathematical Concepts
- Regression and Generalization
- Classification
- Clustering

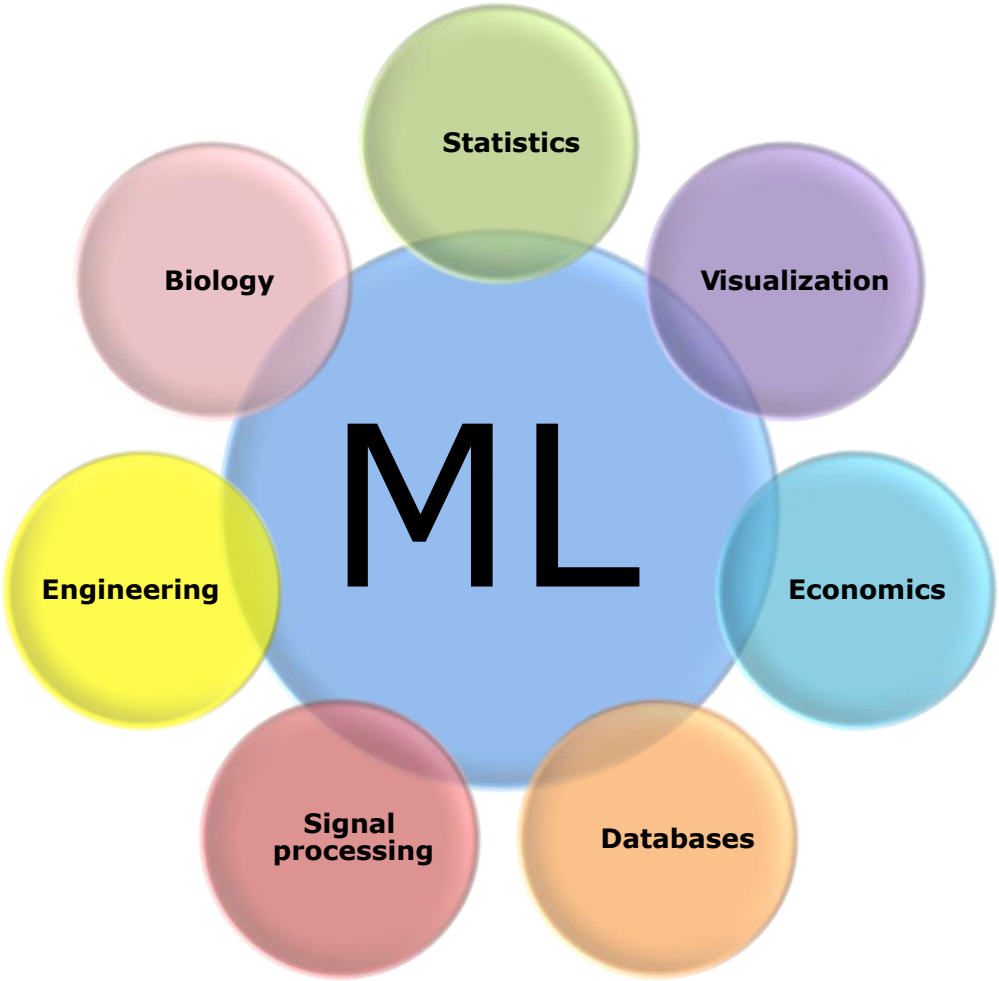
Section 2:

- Trends in Machine Learning
- Deep Learning
- Deep Learning Strategy

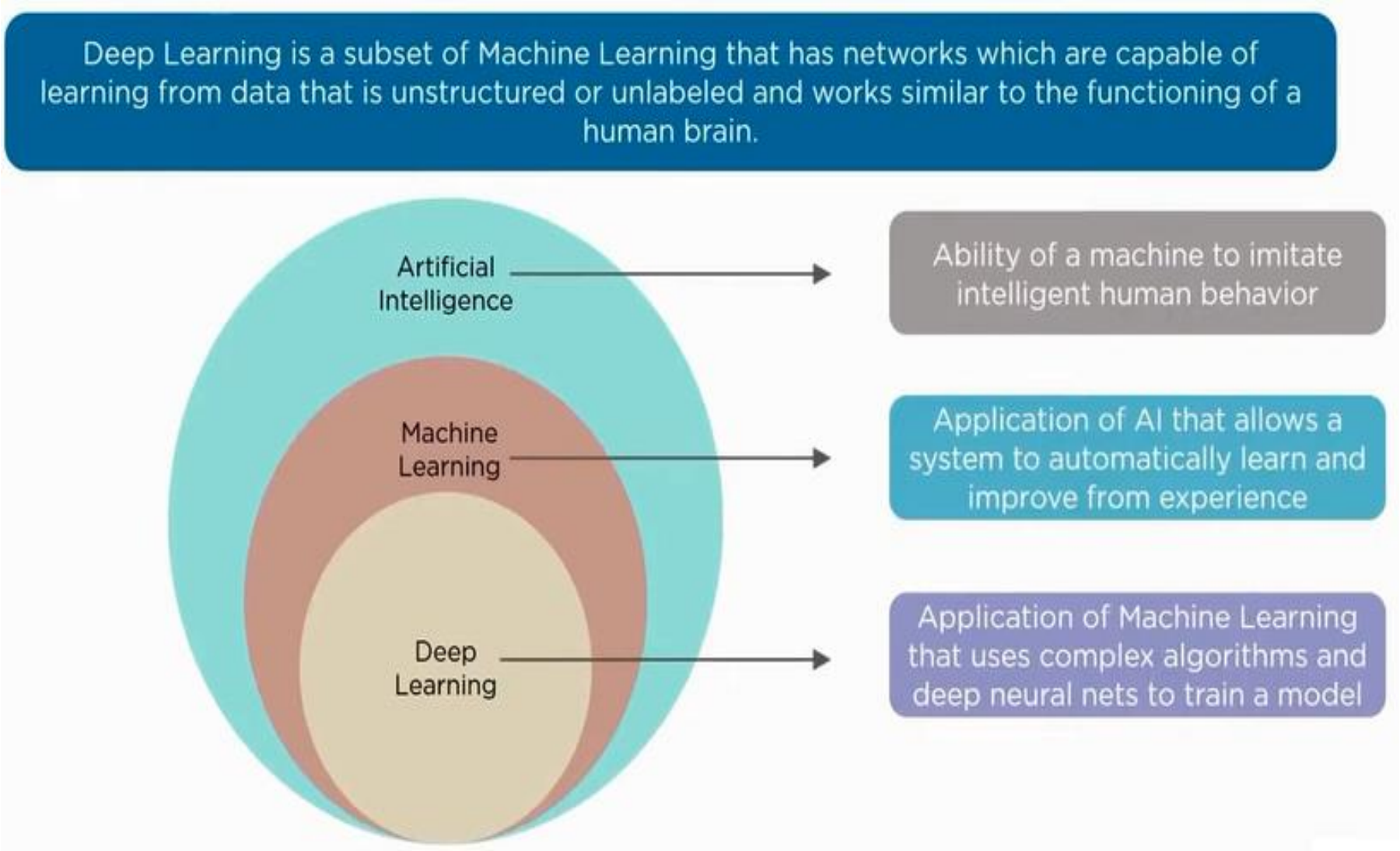
Machine Learning & Deep Learning Books

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

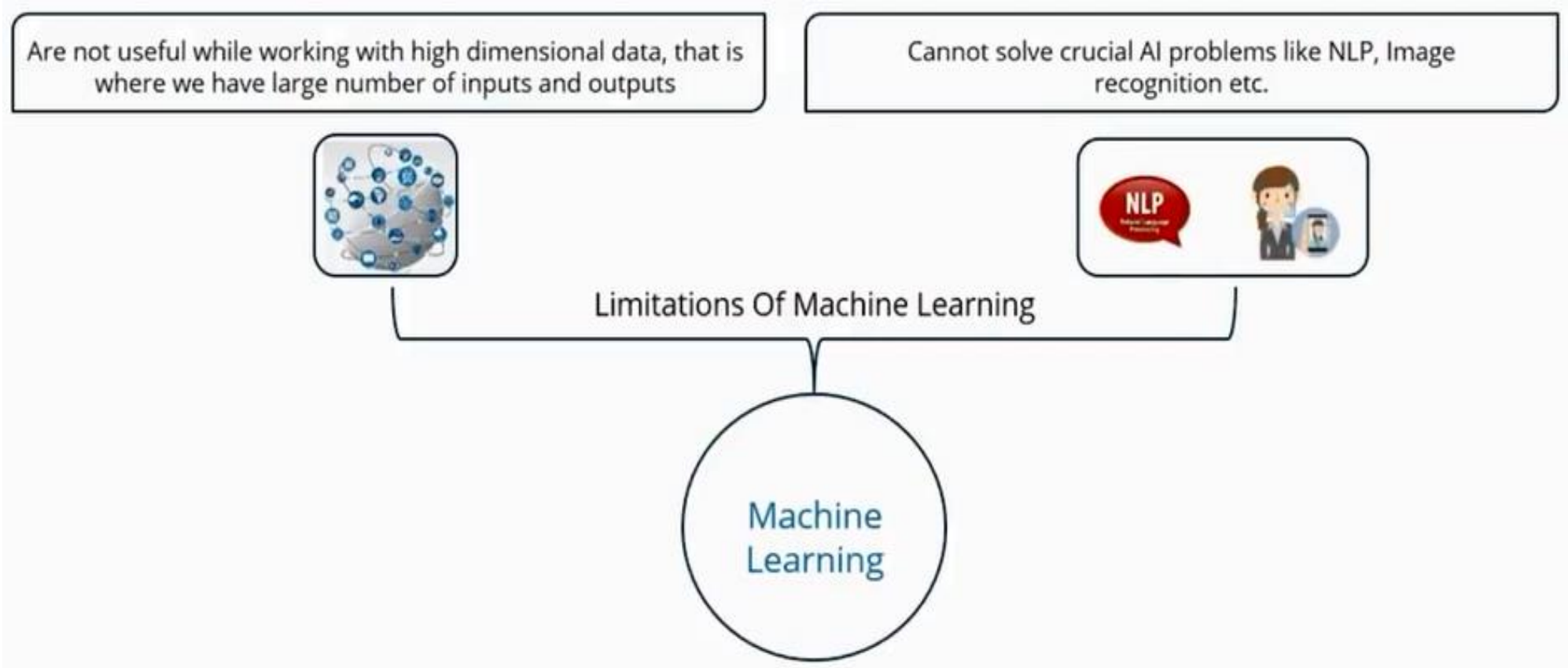


Deep Learning

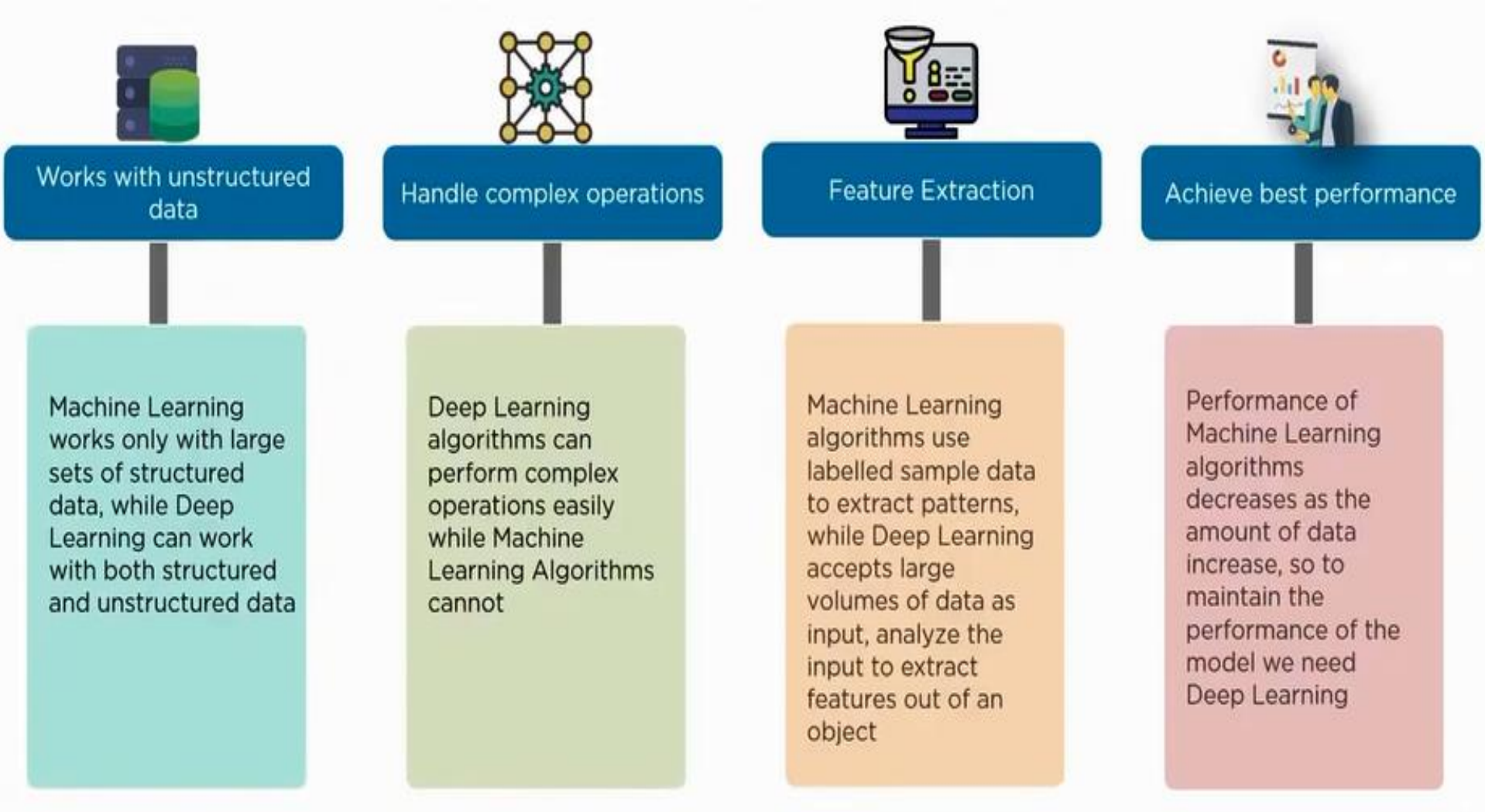


Deep Learning –Techniques that learns features and task from Data Directly
In Deep Learning, features can be learnt just from raw data.

Why do we need Deep Learning

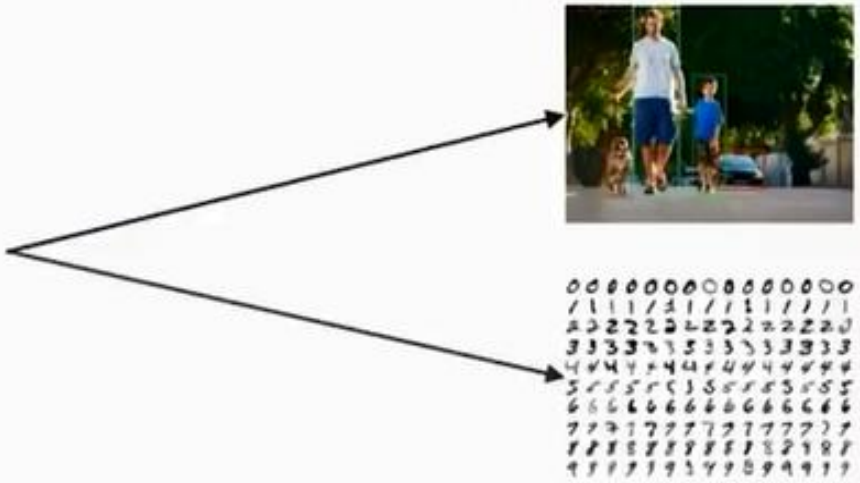


Why do we need Deep Learning



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.



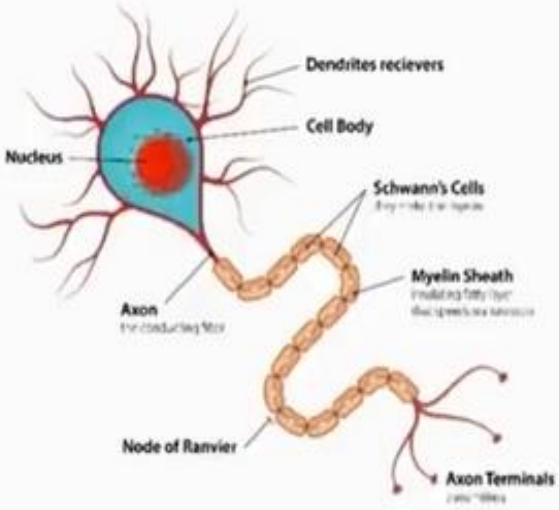
Why do we need Deep Learning

- 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.

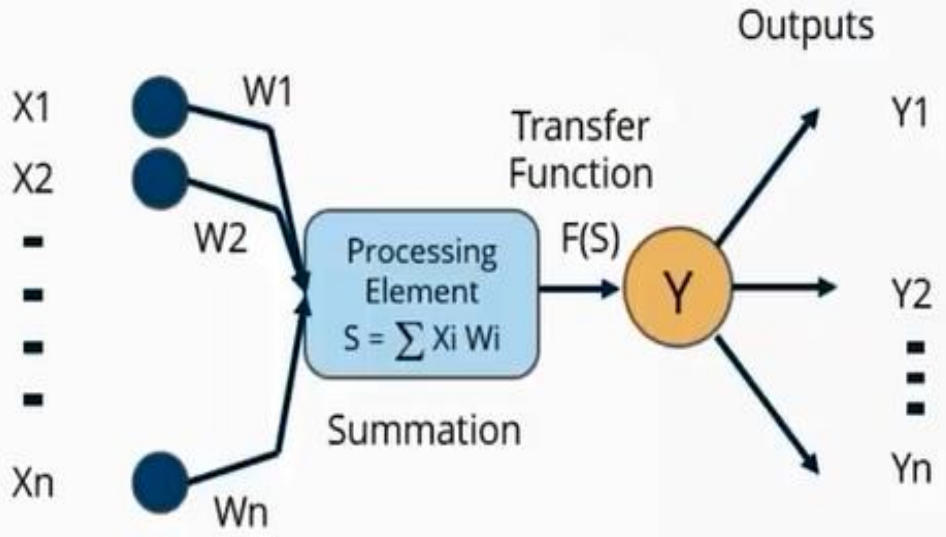
The idea behind Deep Learning is to build learning algorithms that mimic brain.

Why do we need Deep Learning

- Deep Learning is implemented through Neural Networks.
- Motivation behind Neural Networks is the biological Neuron.

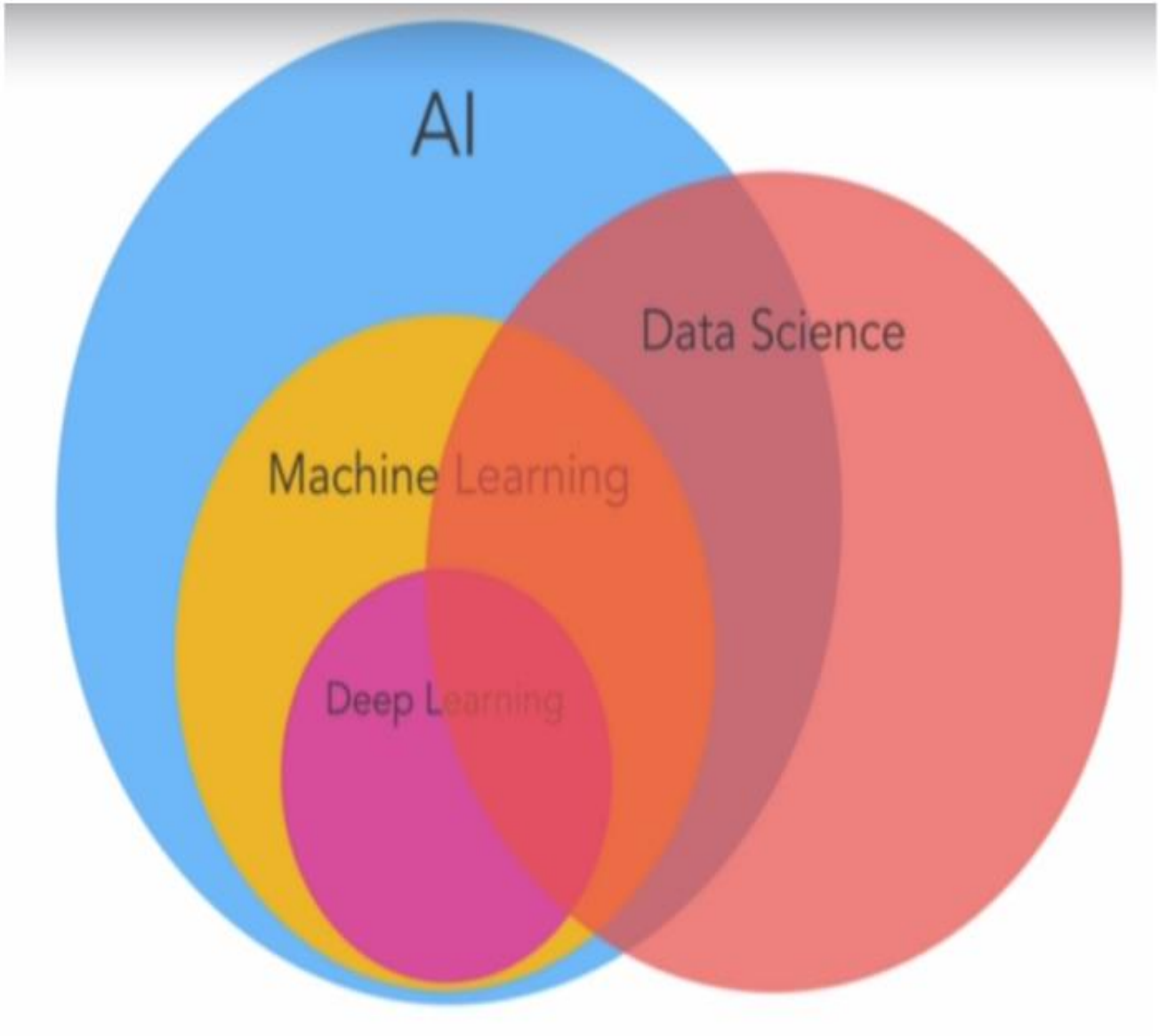


Neuron



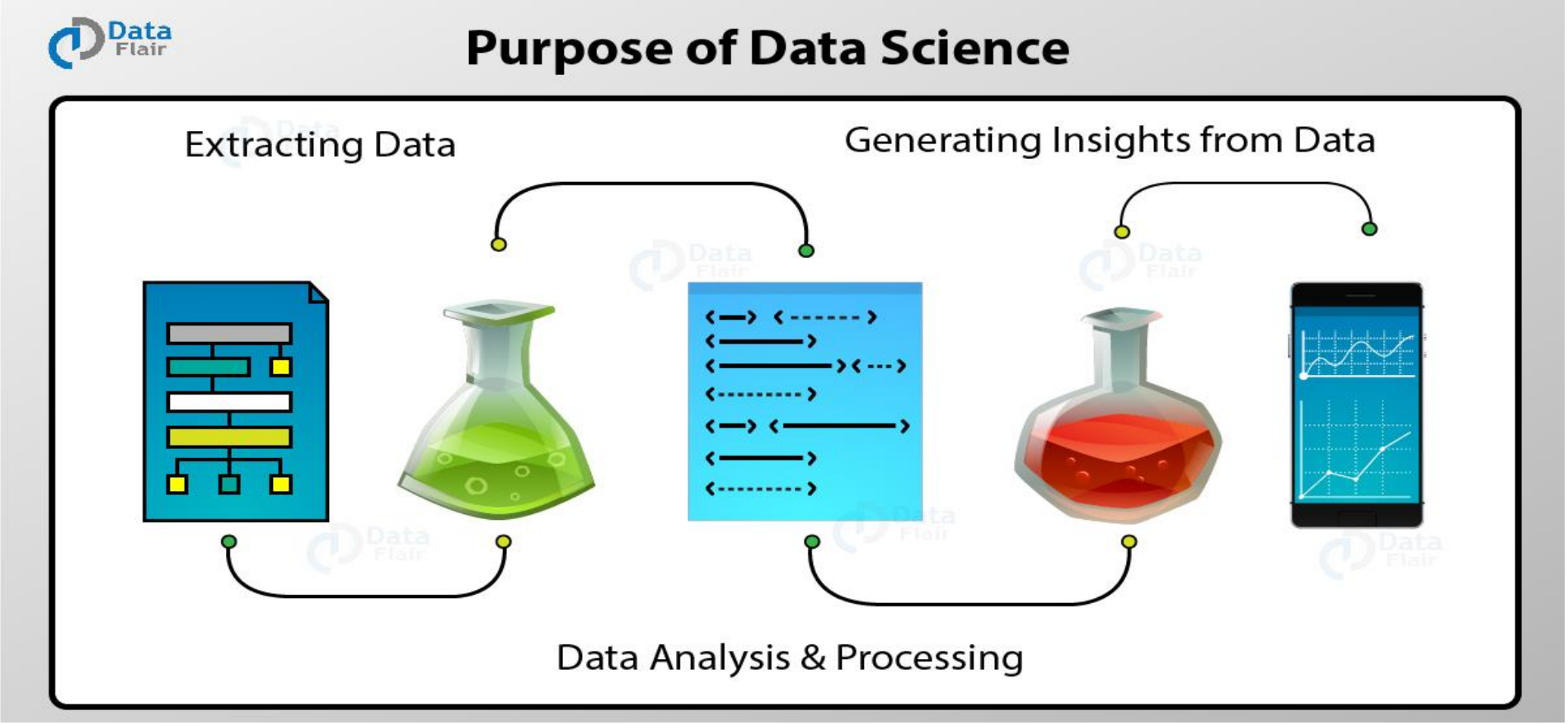
Artificial Neural Network

AI, ML and DS



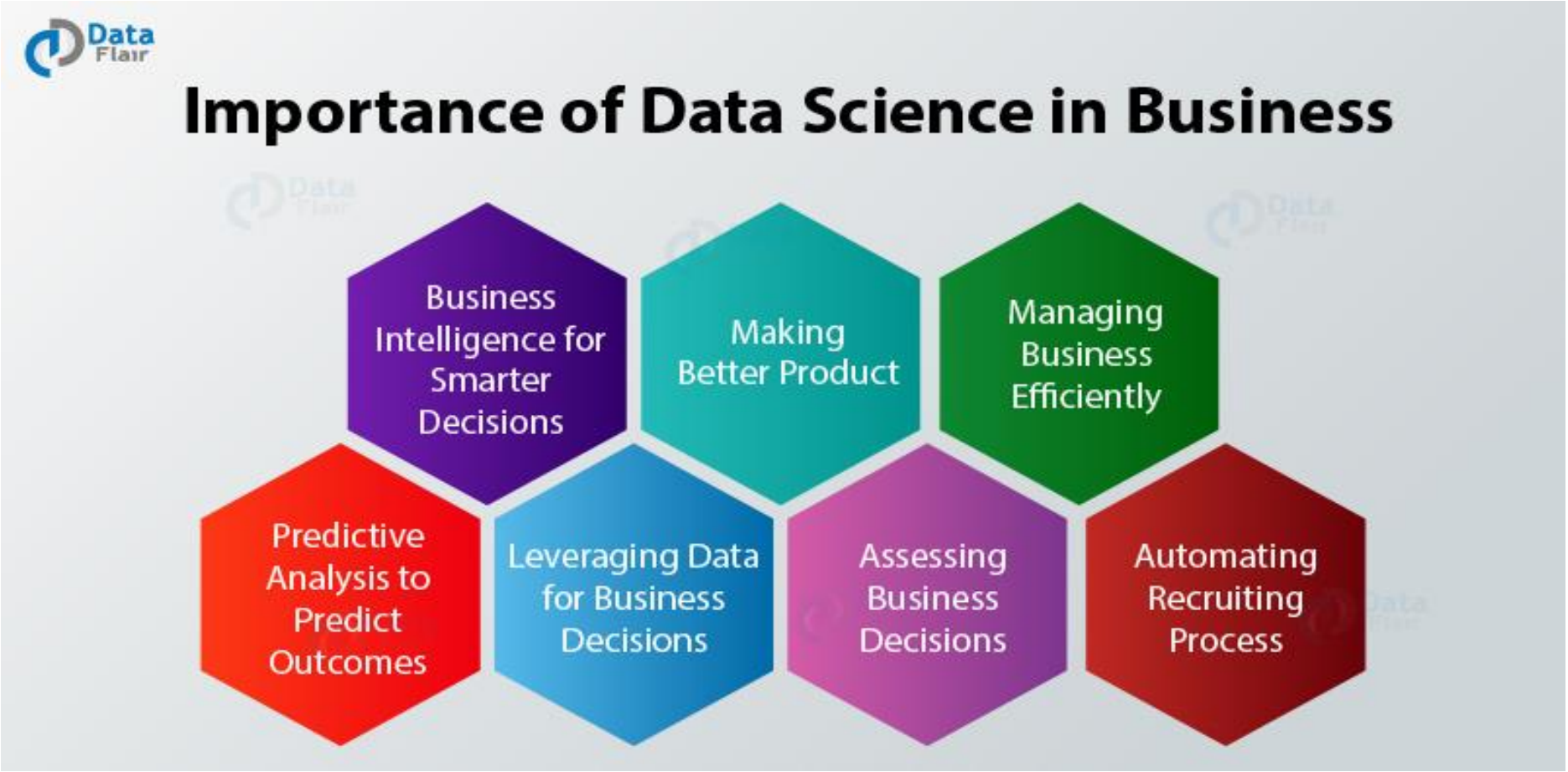
Why Data Science

- 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.**



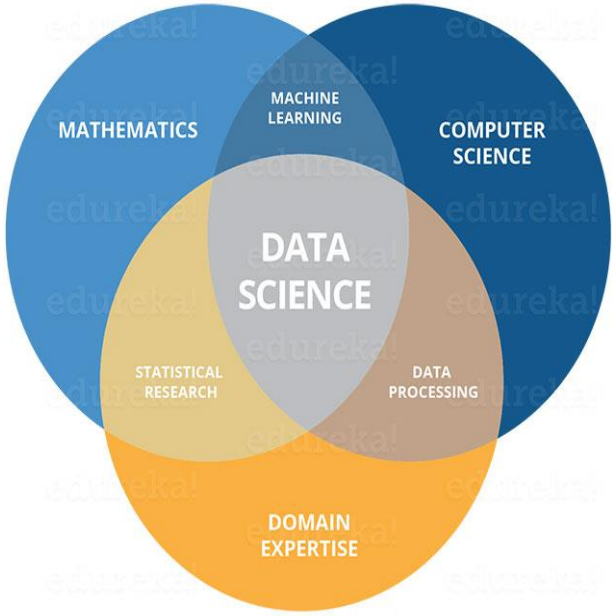
Introduction

- **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.



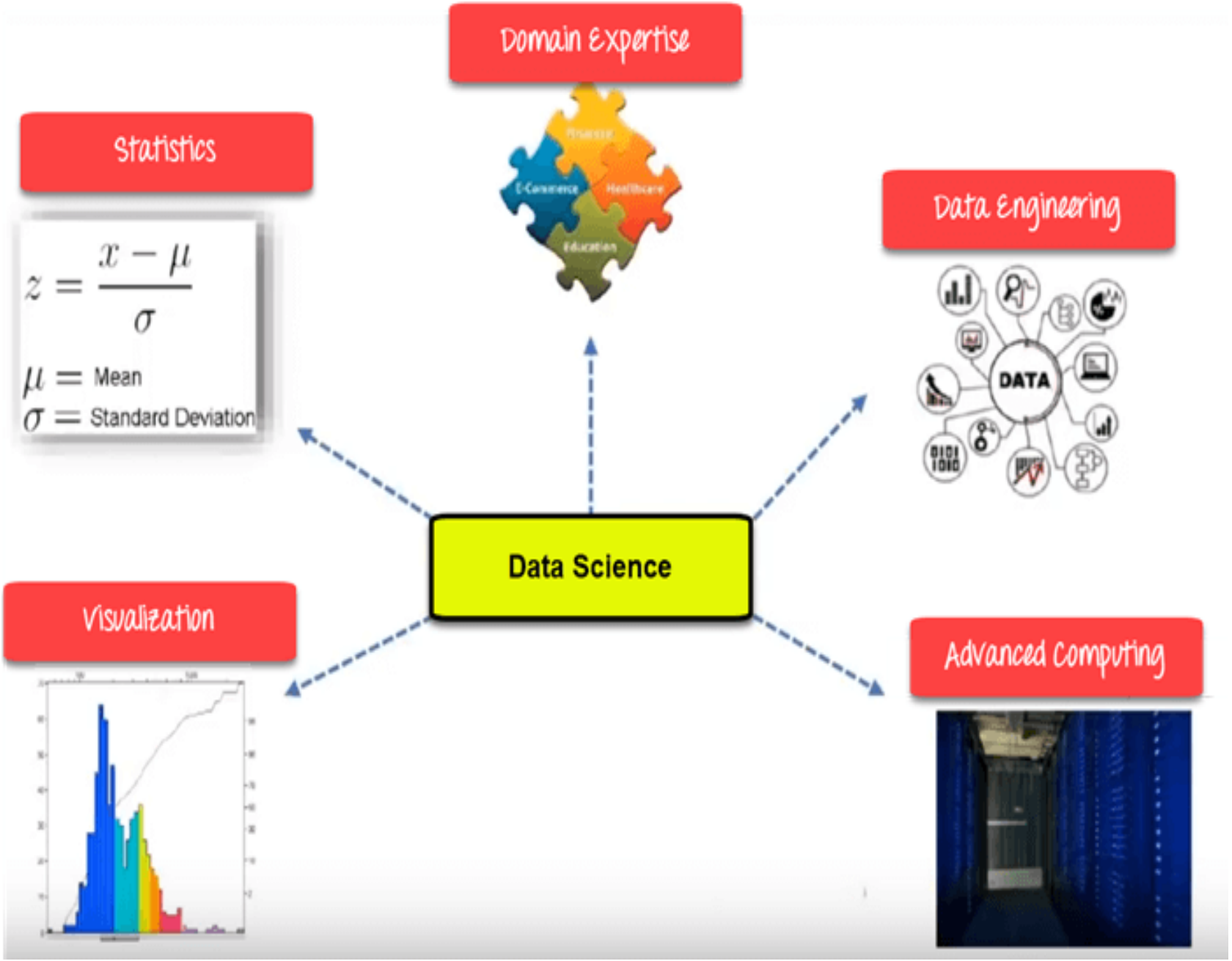
What is Data Science

- **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.**

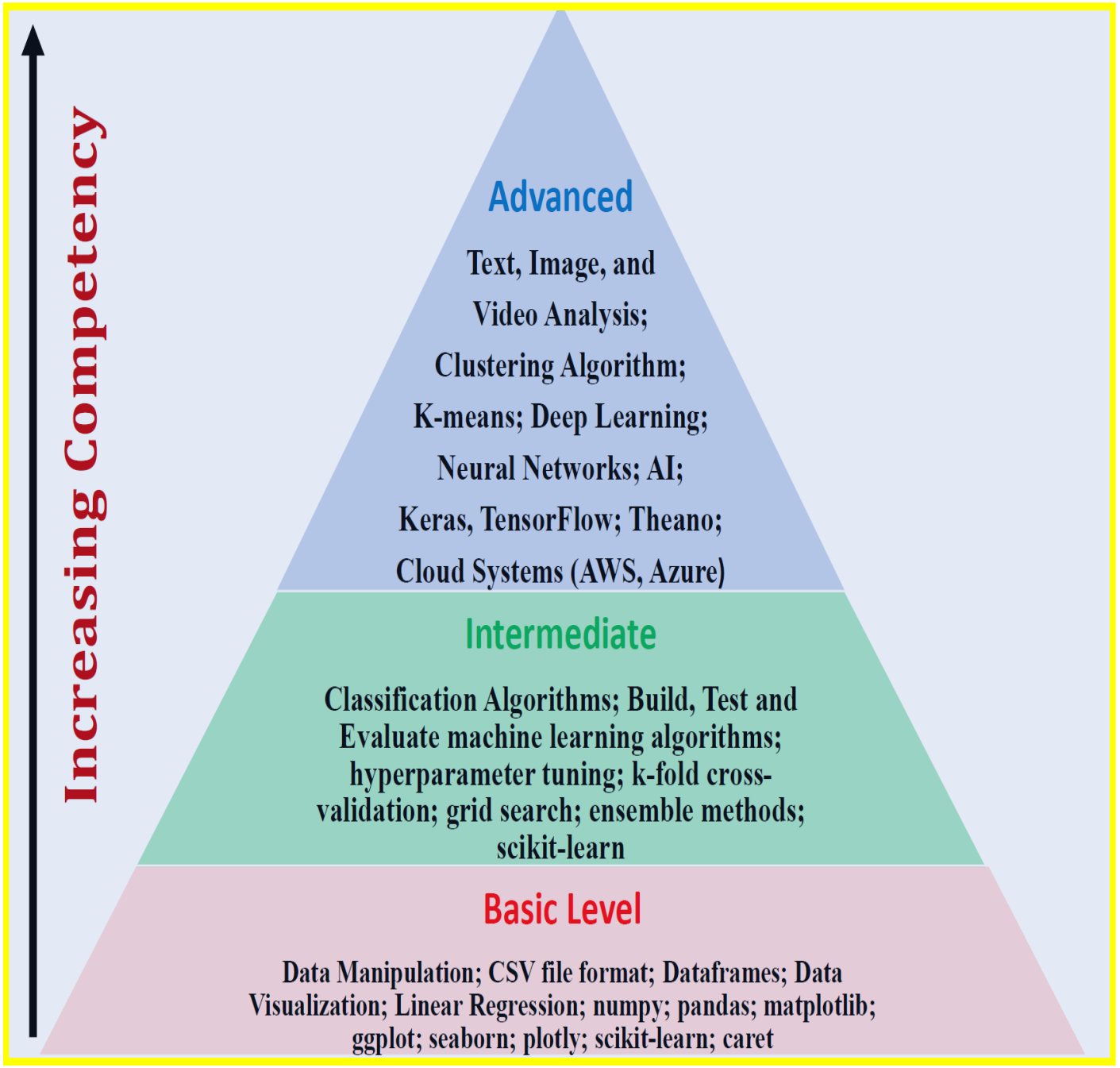


Goal of Data Science - Turn Data into Data Products

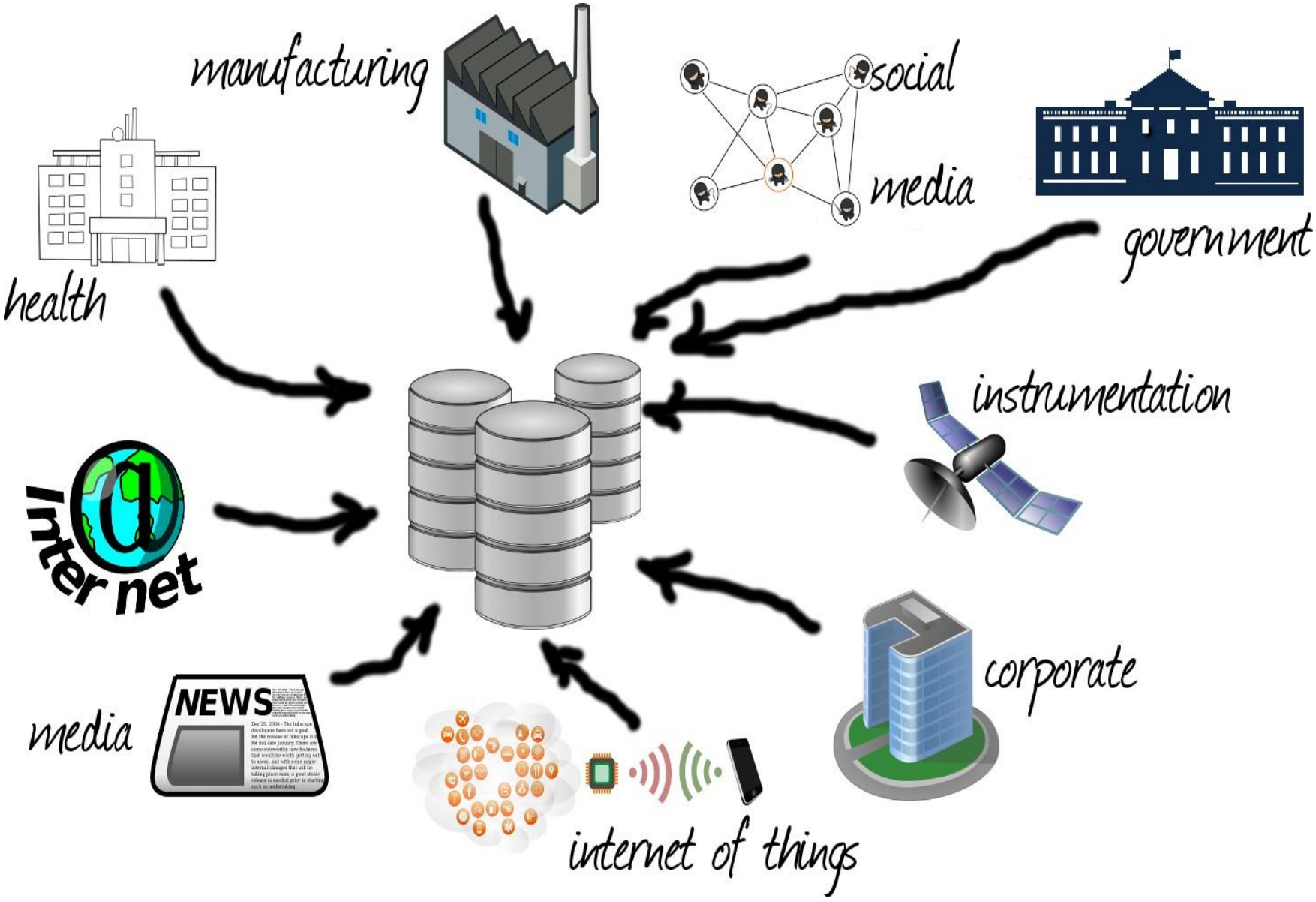
Data Science Components



Levels of Data Science



Data can be got from many sources



"Data is the New Oil"

Data

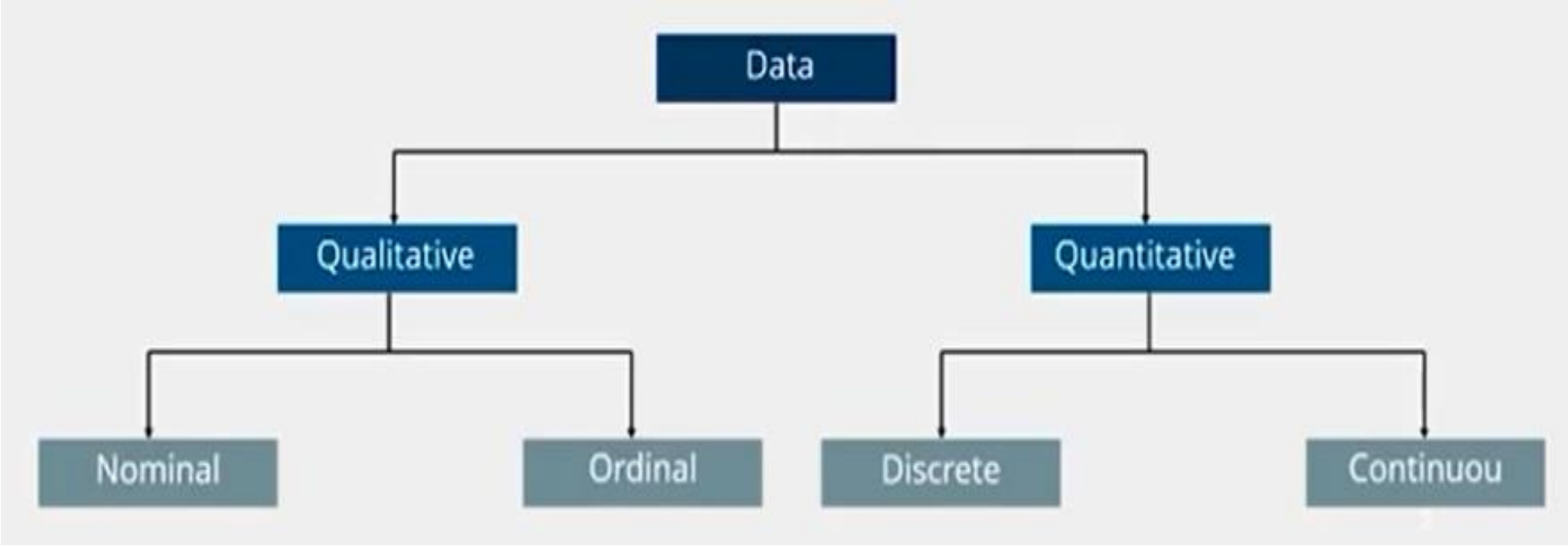
- Data refer to fact and Statistics collected together for reference and Analysis.
- Data can be collected/Stored
- Data can be Measured
- Data can be Analyzed
- Data can be visualized

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

Gender
Male
Female
Male
Male

Ordinal Data

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

Customer ID	Rating
001	Good
002	Average
003	Average
004	Bad

Quantitative Data

Quantitative data deals with numbers and things you can measure objectively

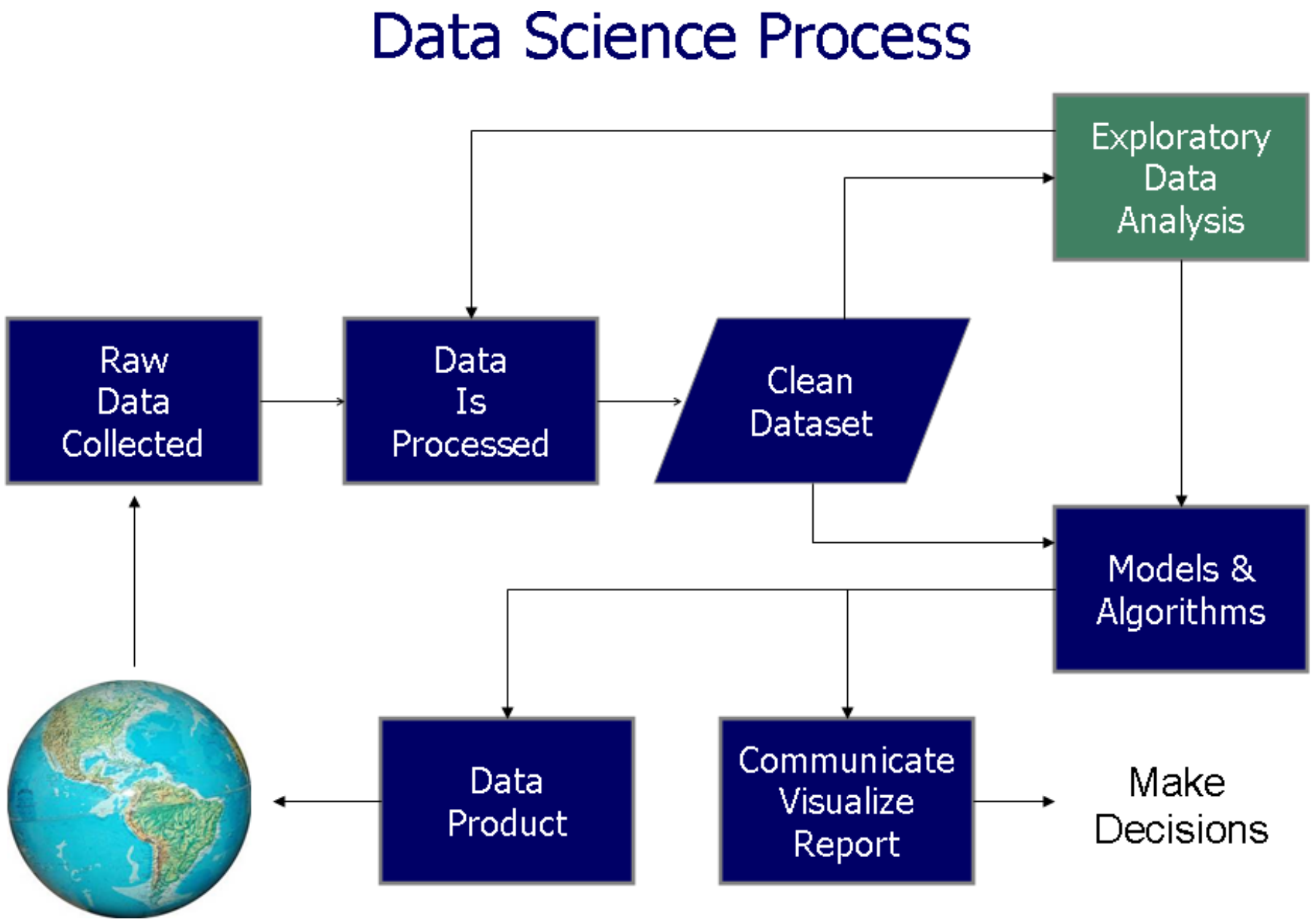
Discrete Data
Also known as categorical data, it can hold finite number of values

Example: Number of students in a class

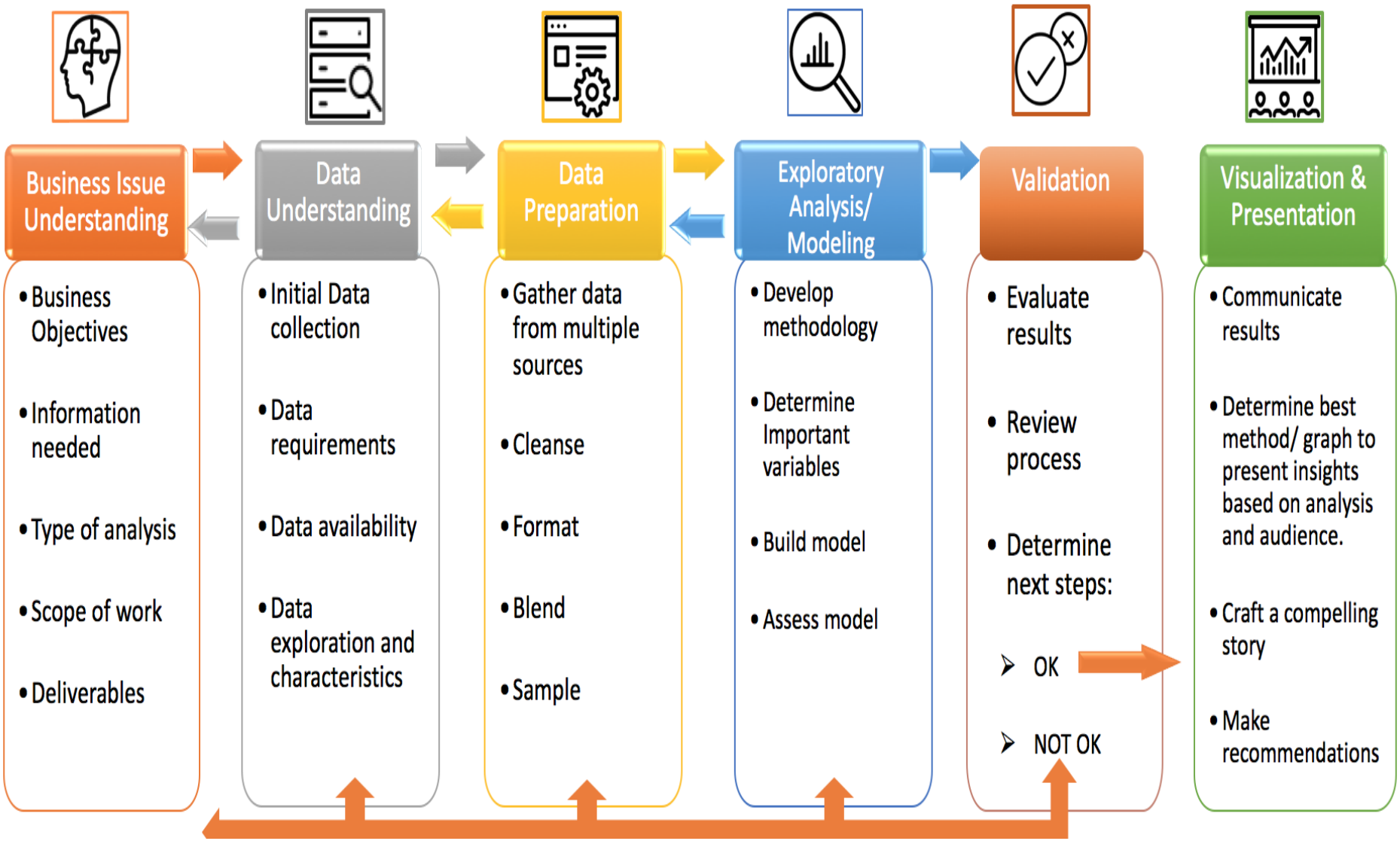
Continuous Data
Data that can hold infinite number of possible values

Example: Weight of a person

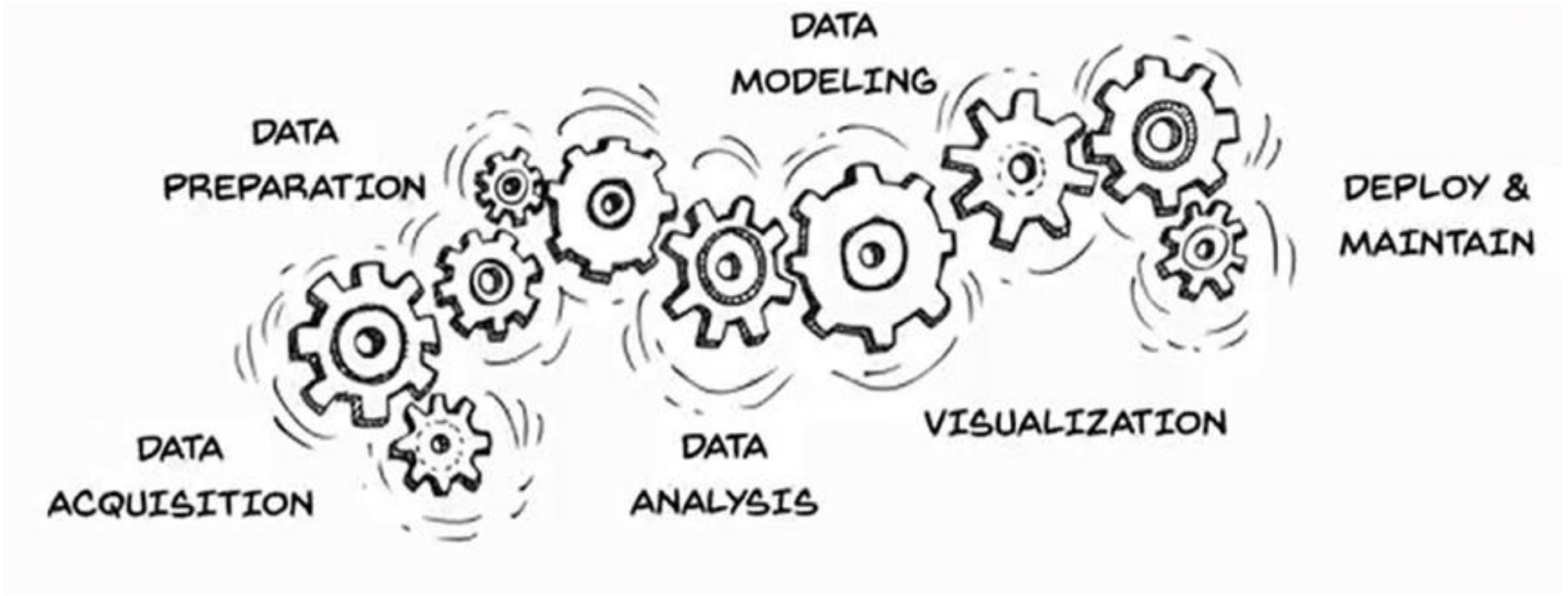
Data Science Process Flowchart



Life Cycle of Data Science Projects

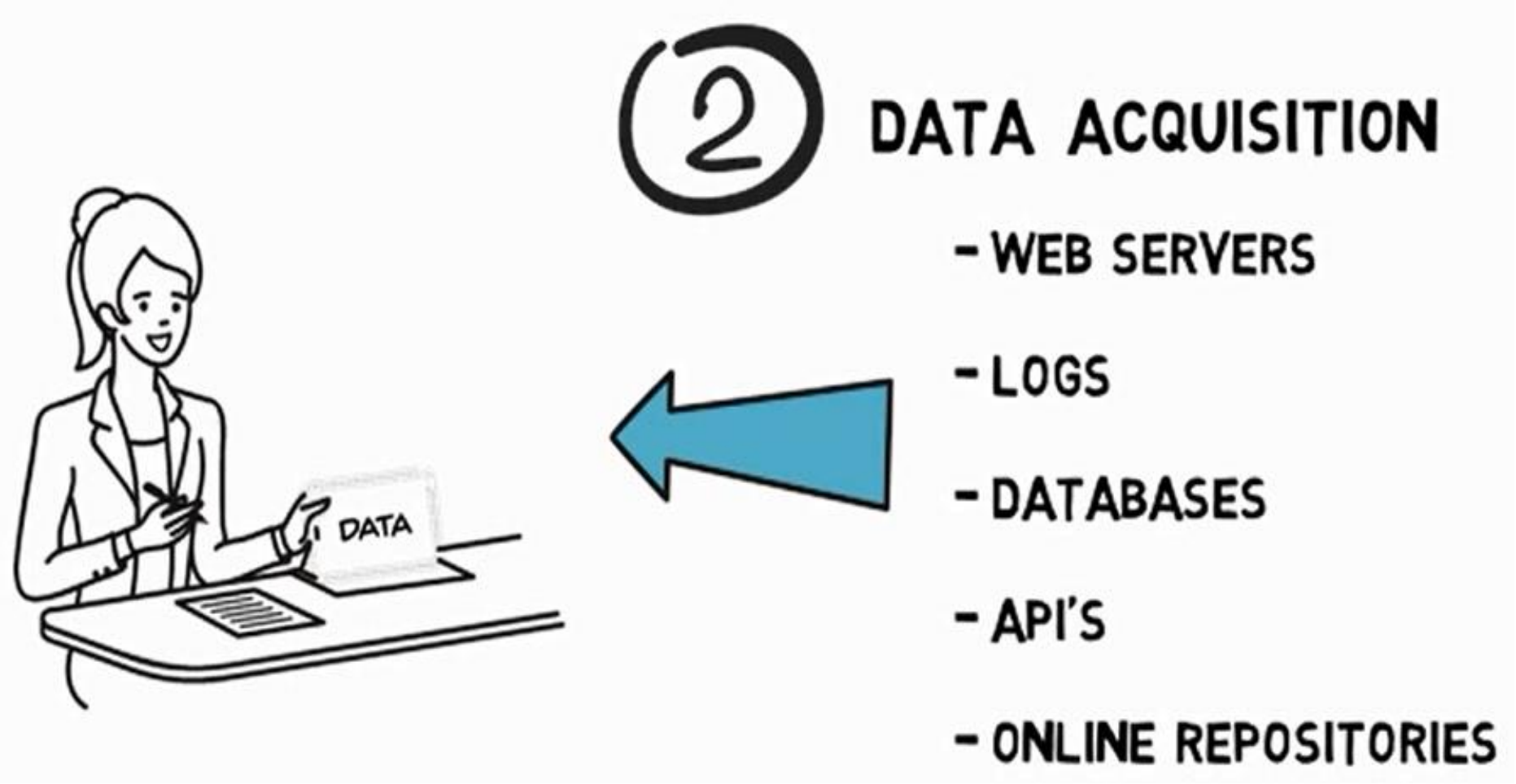


Data Science

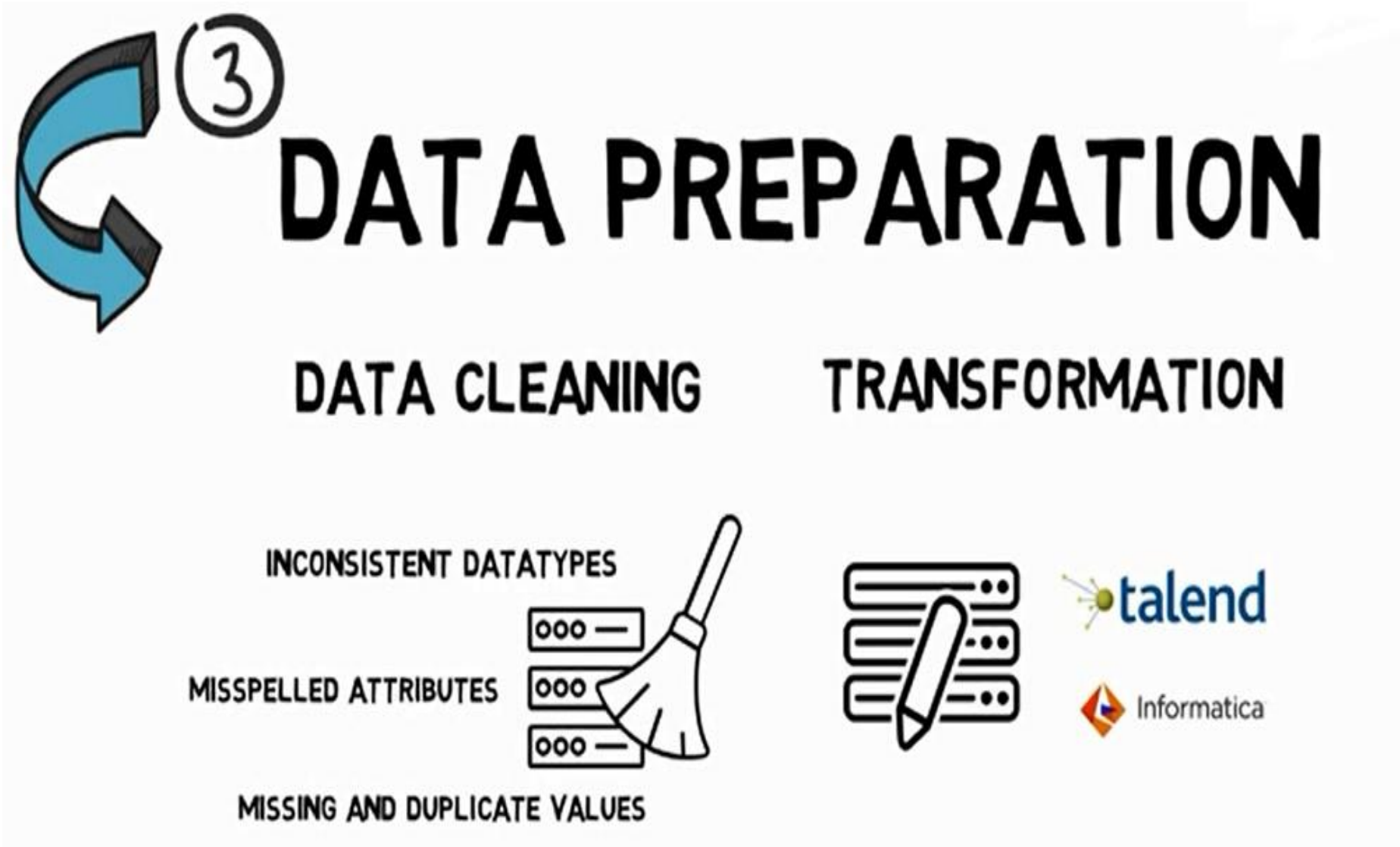


Business Problem Understanding

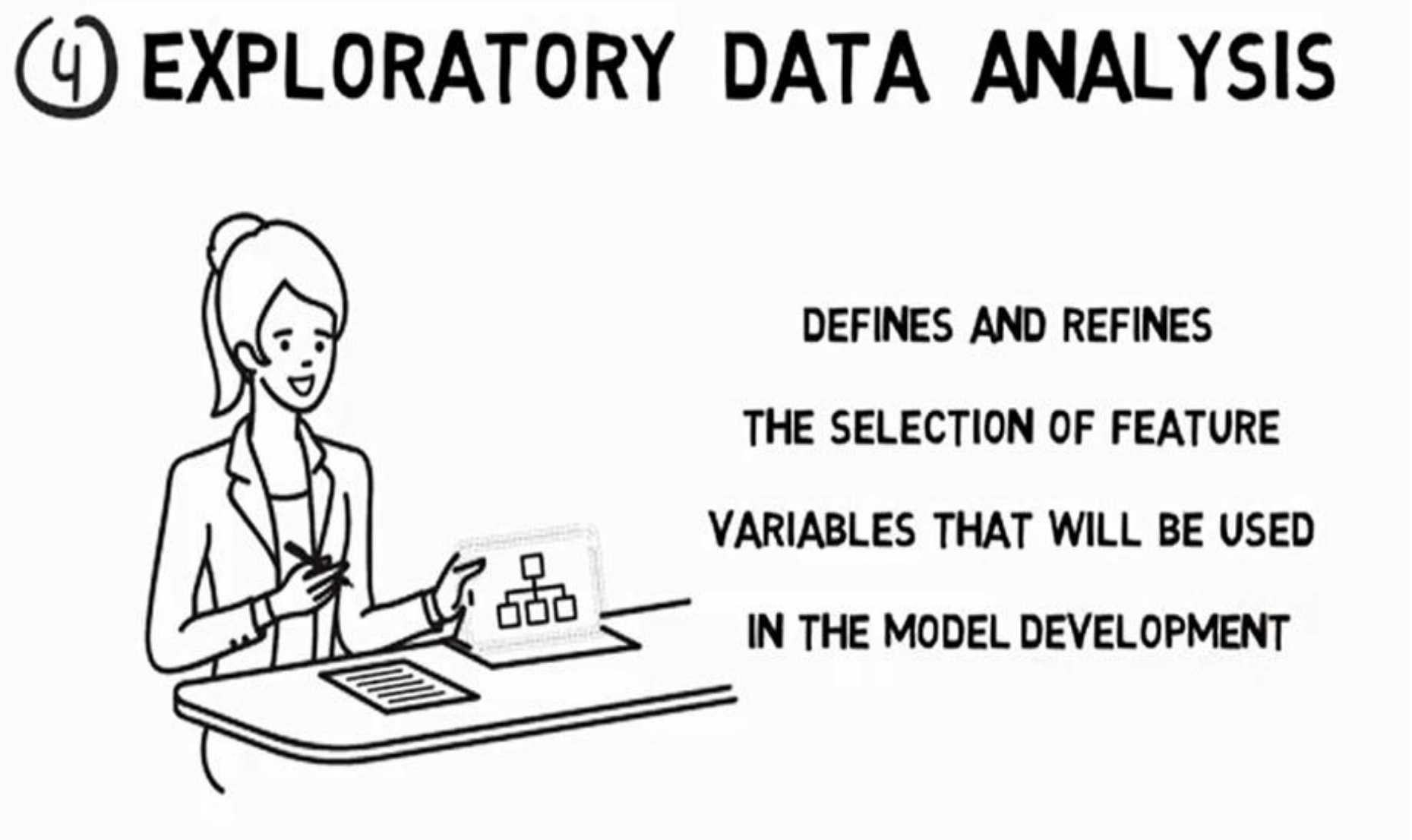
Data Acquisition



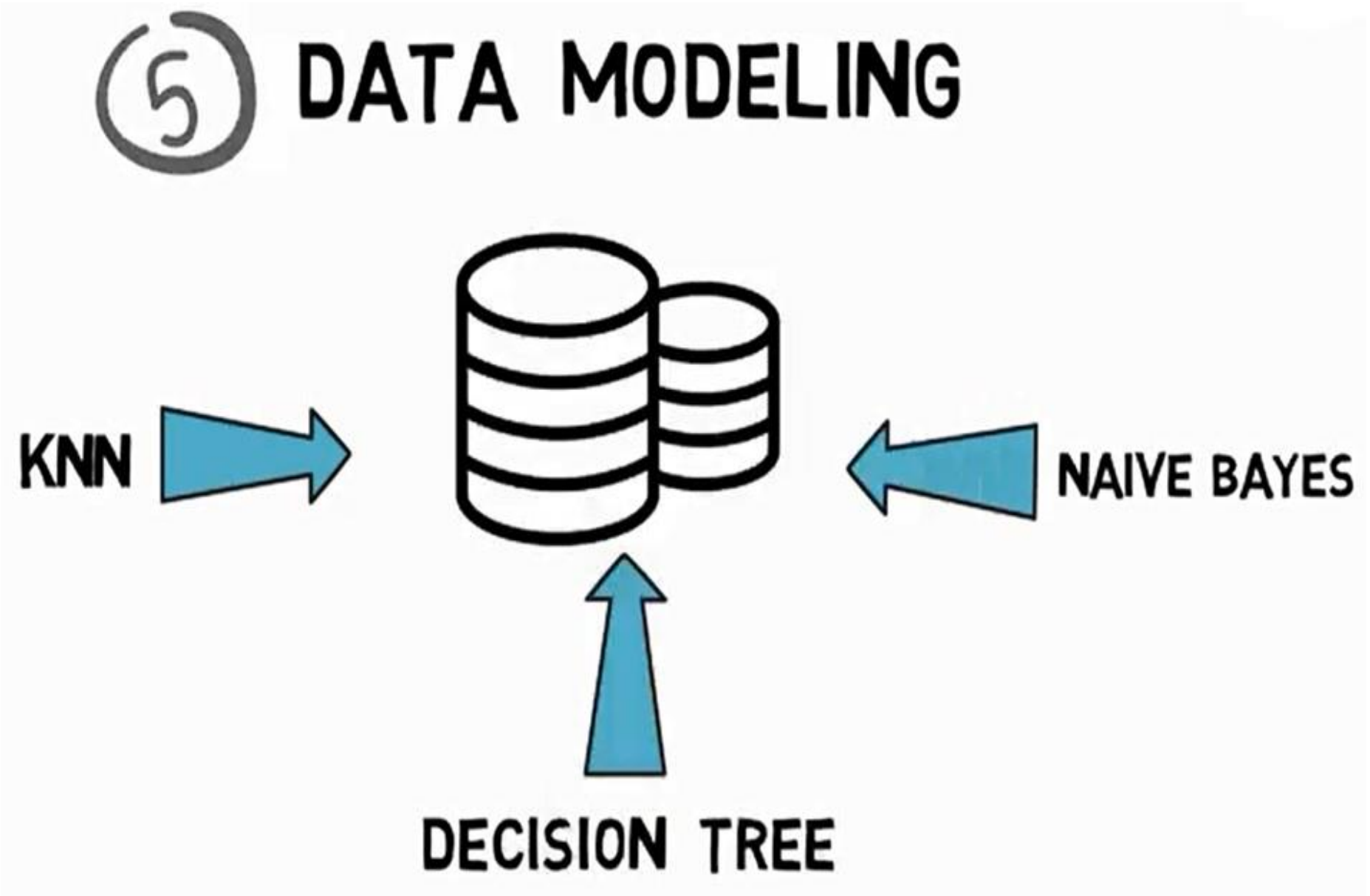
Data Preparation



Exploratory Data Analysis



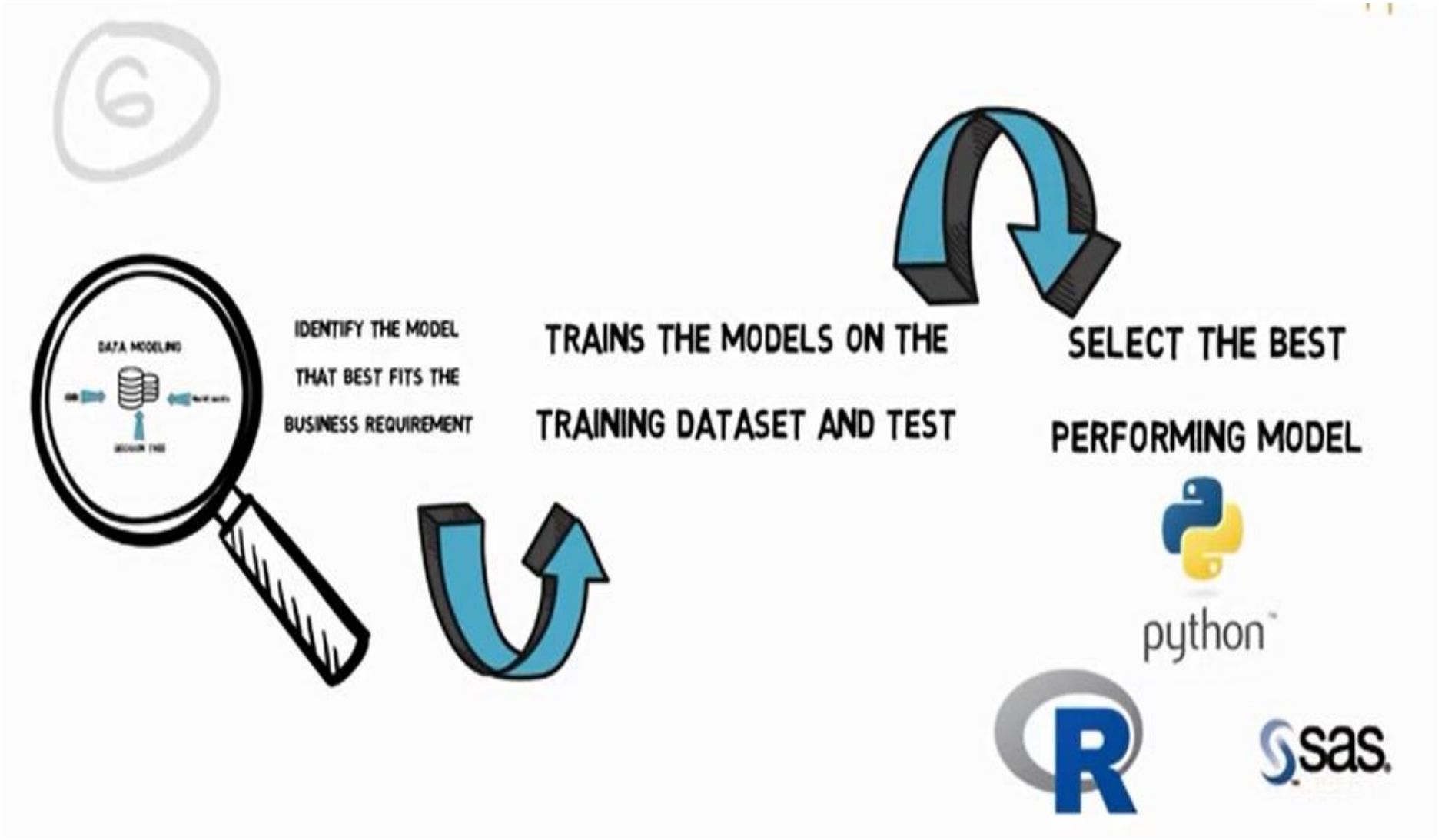
Data Modeling



Difference Between Data Science and M L

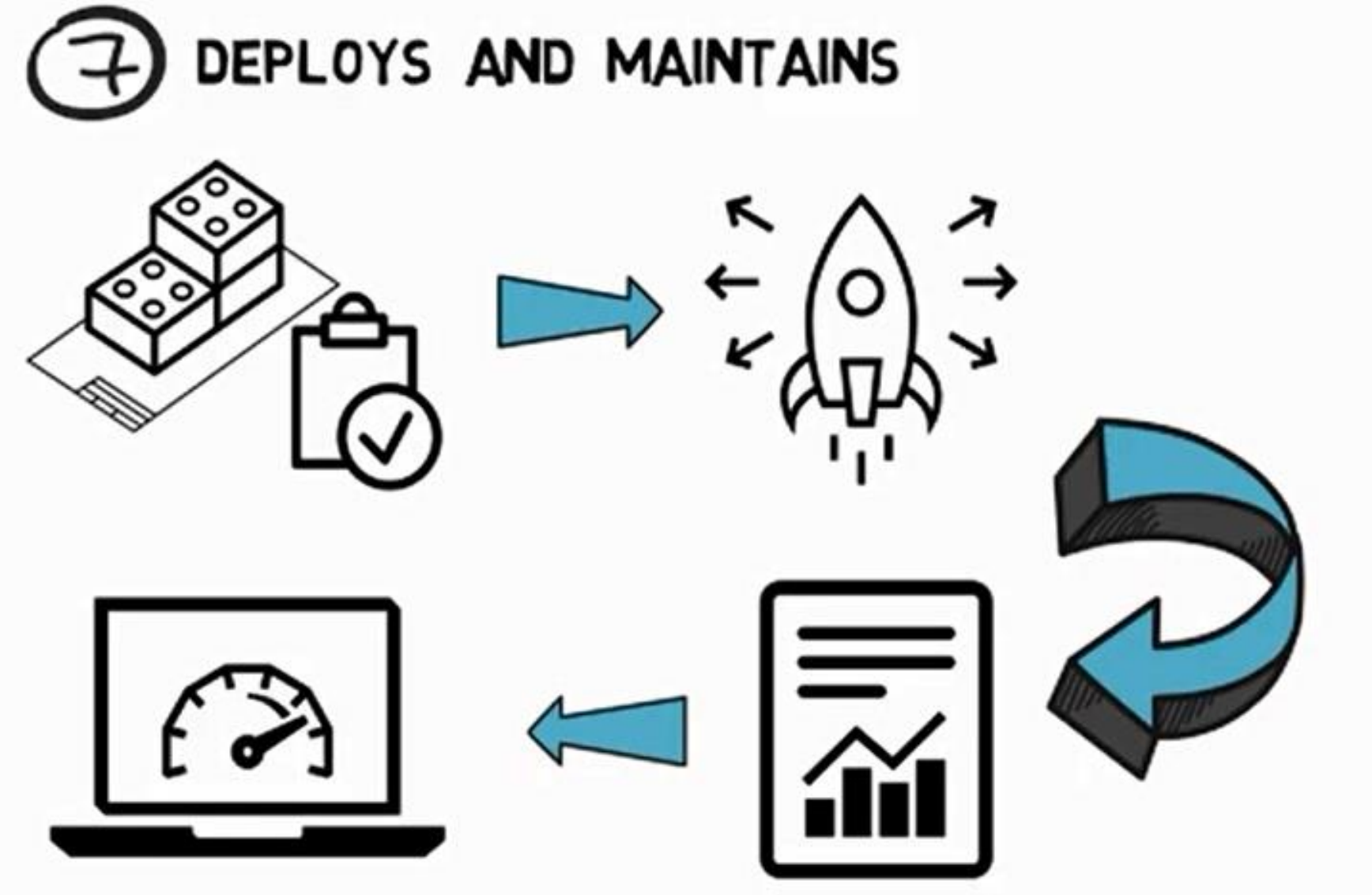
Data science	Machine Learning
Data Science technique helps you to create insights from Data dealing with all Real-World Complexities.	Machine Learning method helps you to predict from historical data with the help of Mathematical models.
Data science can work with manual methods as well, though they 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



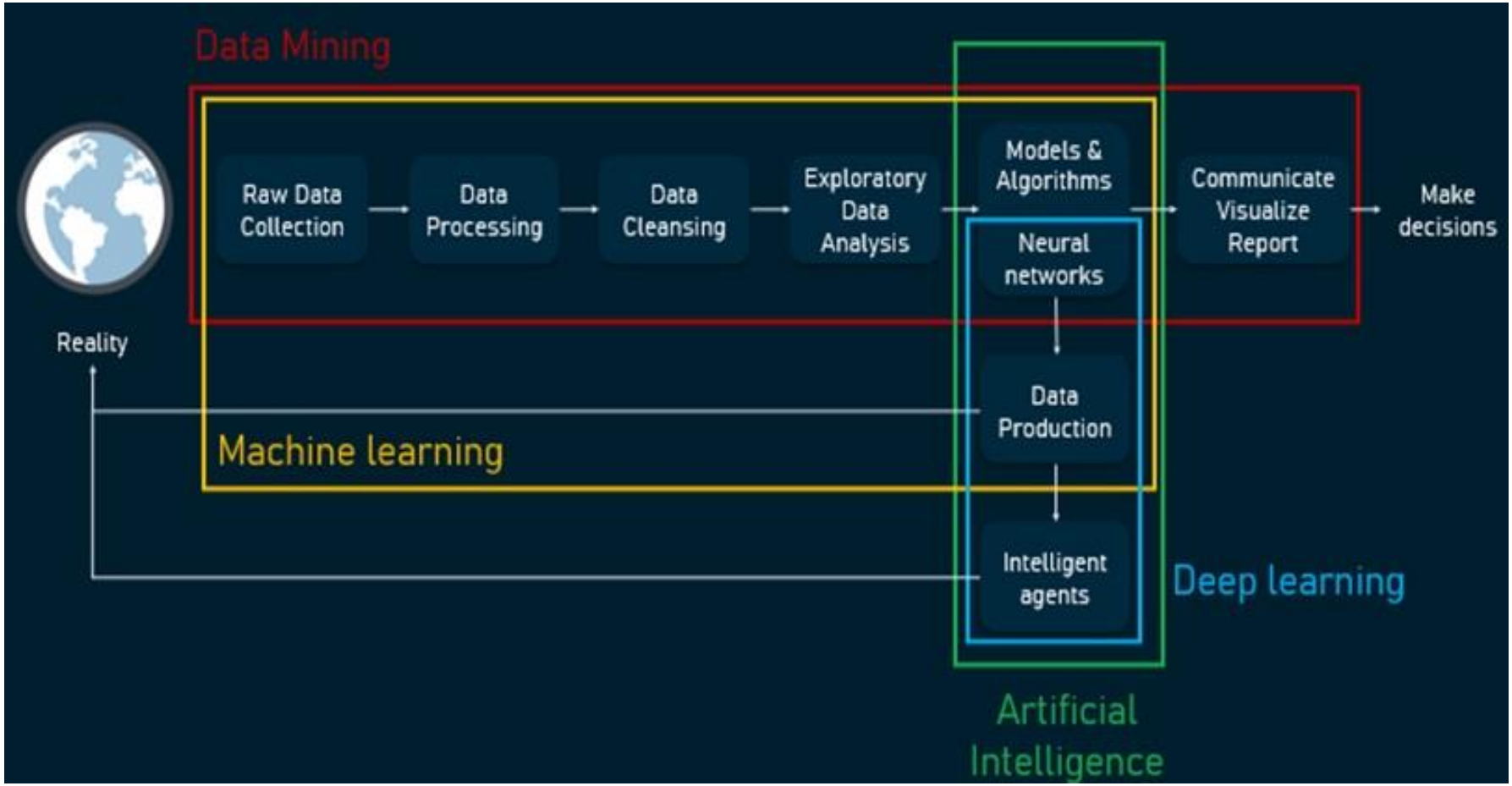
Visualization and Communication

Deploys and Maintains



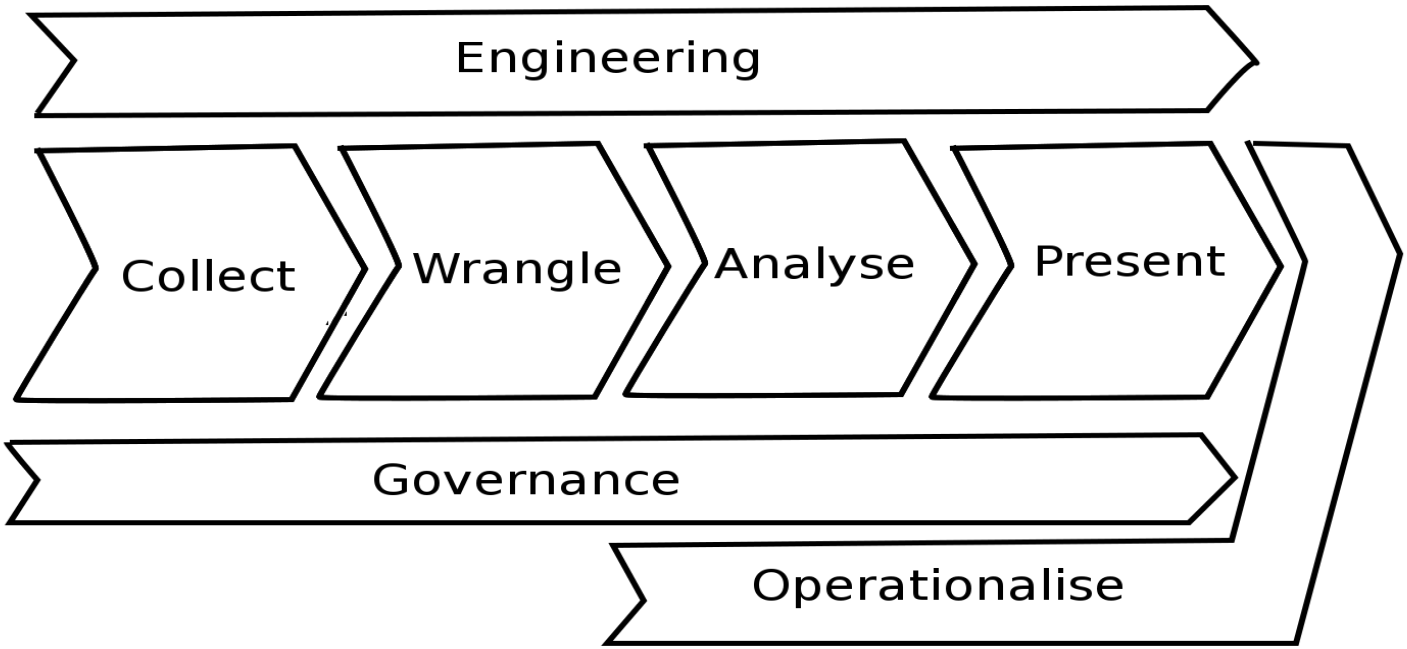
Relations between Data Science, AI, ML, DL, and DM

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

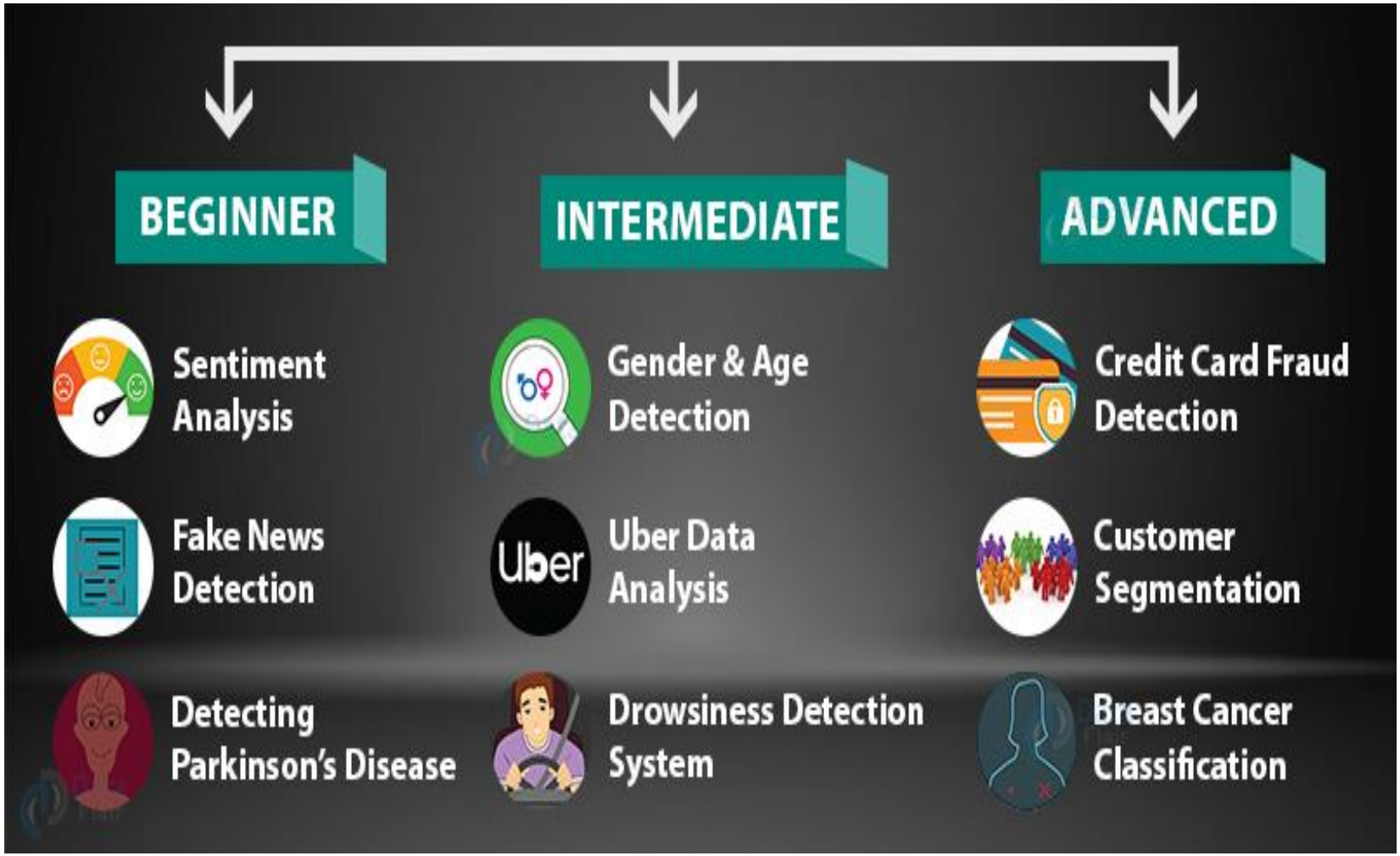
Multi-purpose: Python, R, SAS, KNIME, RapidMiner

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.

Project Ideas



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.

Driver Drowsiness Detection

- An excellent Data Science project idea for intermediate levels is the 'Keras & OpenCV Drowsiness Detection System'. Driving overnight is not only tough but a risky job too.
- We have heard of a lot of cases where accidents happen because the driver fell asleep while driving.
- Thus, this project can help prevent numerous road accidents that happen due to such cases.
- This project's main aim is to recognize whenever the driver may get drowsy & fall asleep while driving.

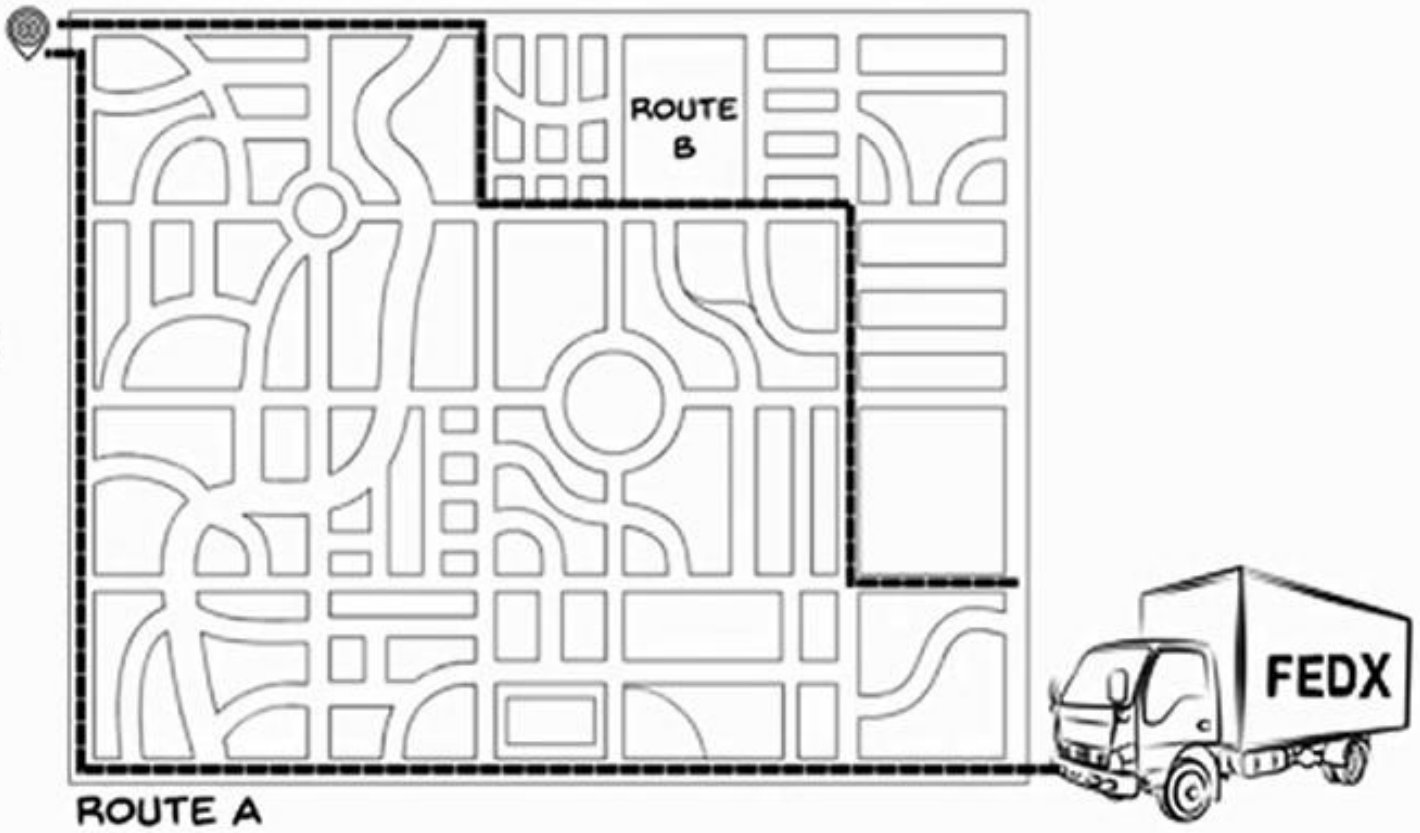
Logistics

LOGISTICS COMPANIES LIKE DHL, FEDEX HAVE

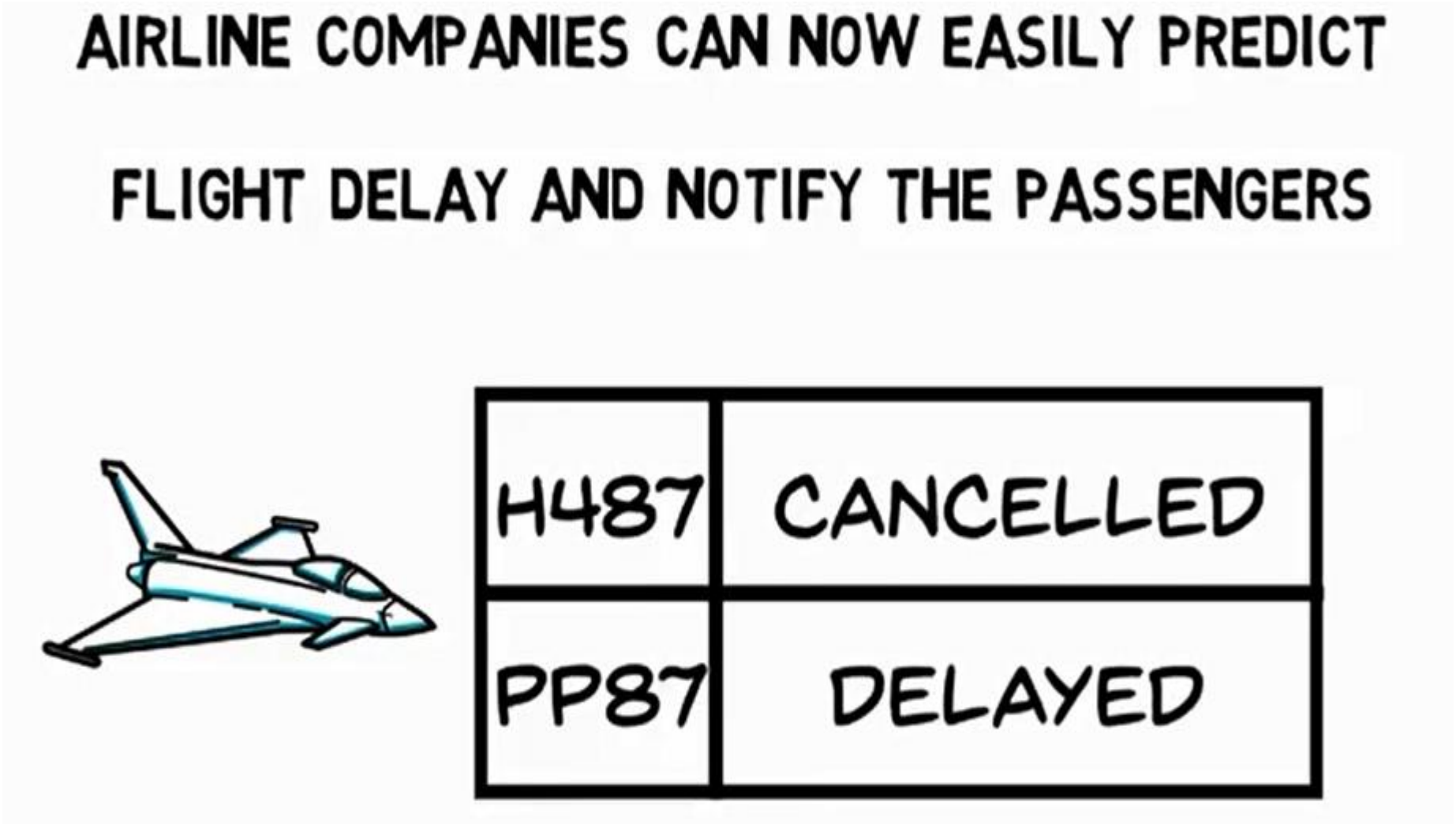
DISCOVERED THE BEST TIME AND ROUTES TO SHIP

BEST TIME:

12:00PM

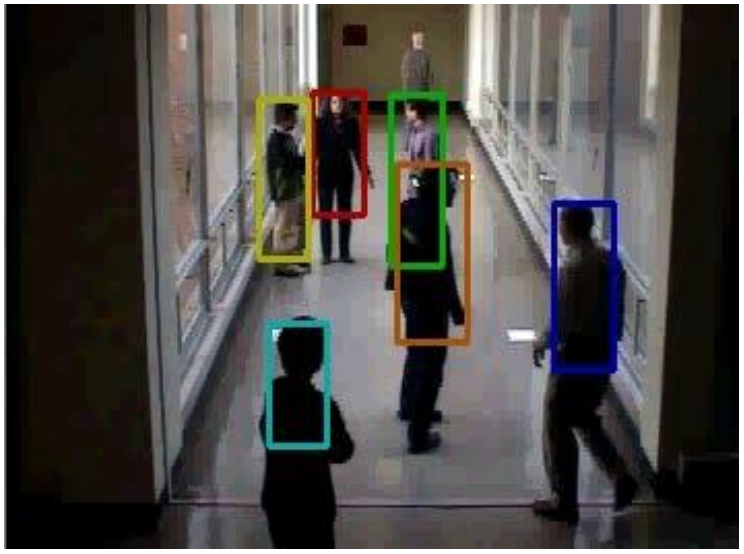


Airlines



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.**

Expected outcome

- Projects - Socially useful student projects, Research projects,
- Start-up
  - Commercialize product
  - Publish- Quality papers, Patent

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

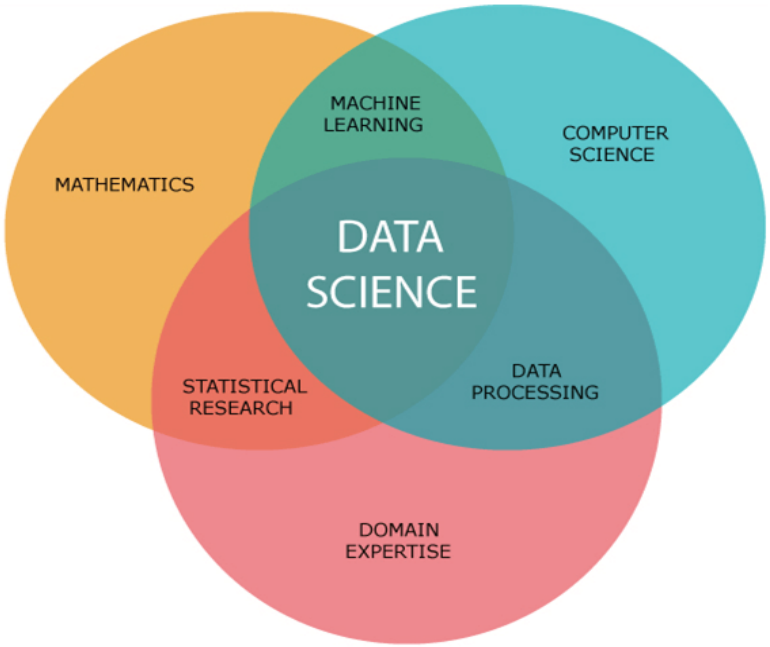
Impact of Data Science on Industries

- **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.

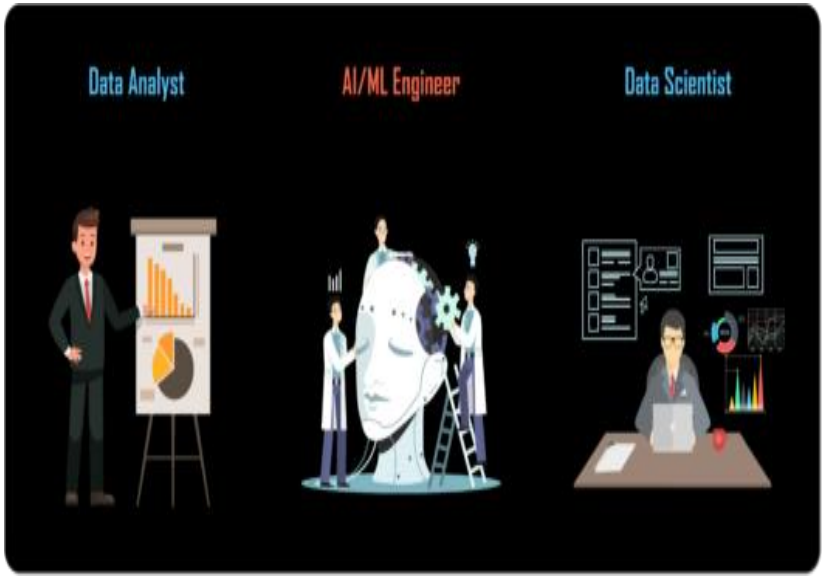
Data Science Engineer Skills

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



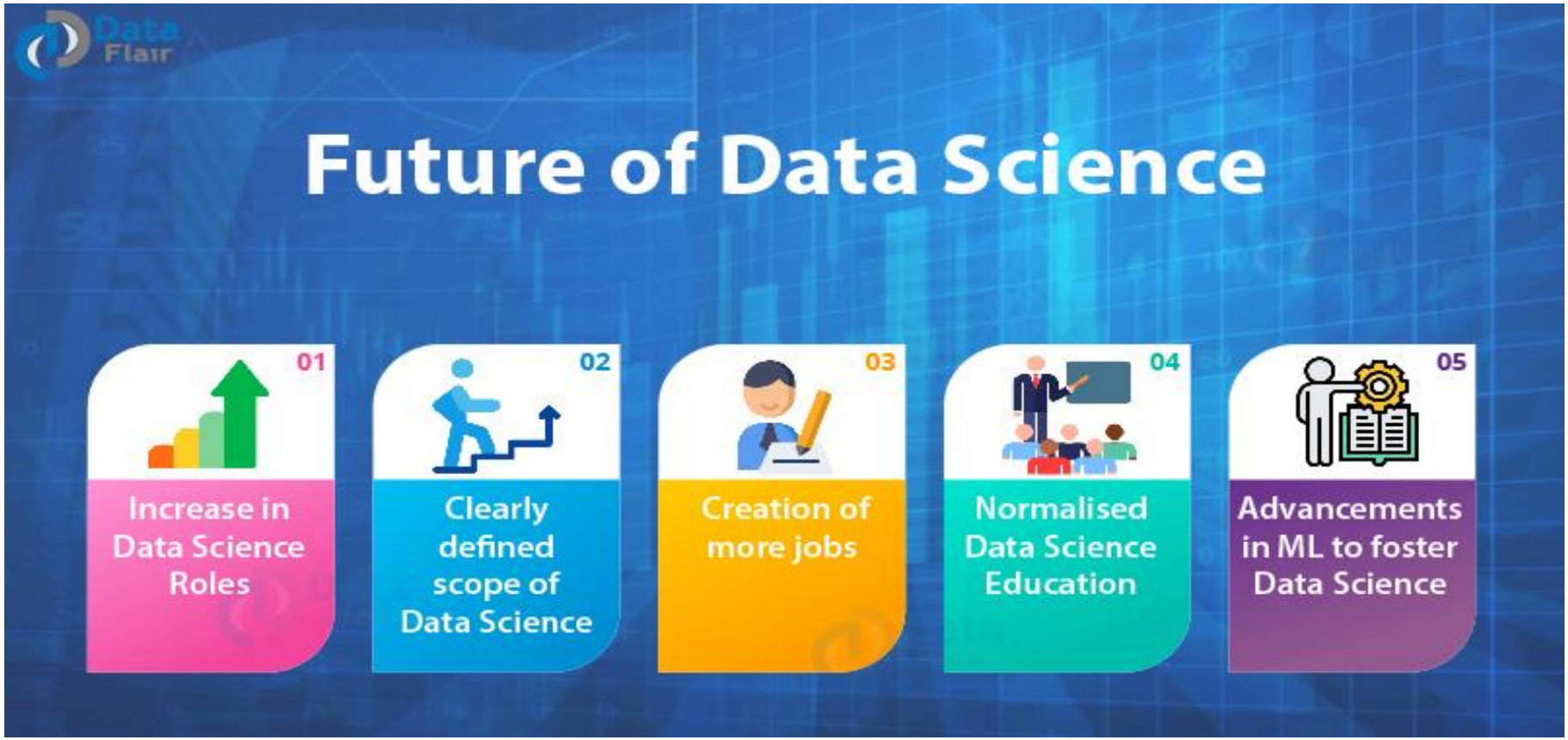
Job Roles & Opportunities

Average Salary 12 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.



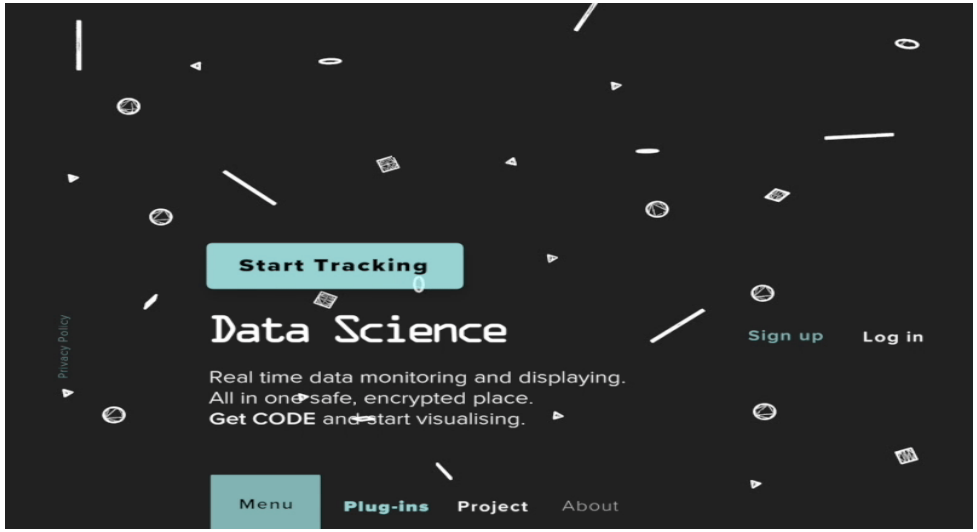
Market Trends and Projections

- **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.**




ThankYou
Merci
Kop
Salamat
Hvala
Tack
kun
Xie
Gra
Dankie
Kruthagnathalu
Arigatou
Dhonnobaad
Dhanyavad
Khopjai
Gomapsupnida
Terima
Danke
Enkosi
dank
Euxaristo
Kun
Shukriya
Shokriya
Mamnoon
Kasih
Todah
Sheun
umesc
Diolch
Kiitos
Shnorhakalutiun
Gamsahapnida
Dank
Takk
Waad
Daw
Dakujem
Dhanyavadaalu
krap
Grazzi
raibh
Gracias
Nandree
Blagodariya
Fyrir
Dekuju/Dekujeme
Te°ekkür
Ngiyabonga
Cam
Dziękuje
Shokrun
Spaas
Mul
Ačı
or
al
Dhanyavaad
Go
Grazie
Falemınderit
ederim
Hain
Asante
Dhan
daa
Or