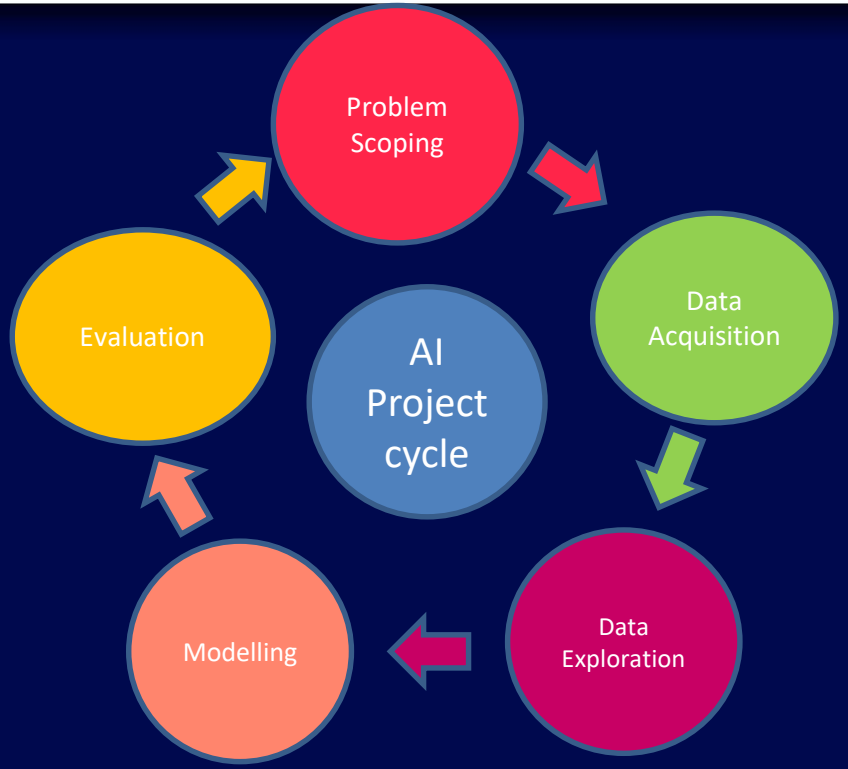
ARTIFICIAL INTELLIGENCE
CLASS – IX
UNIT 2

UNIT 2 - AI PROJECT CYCLE

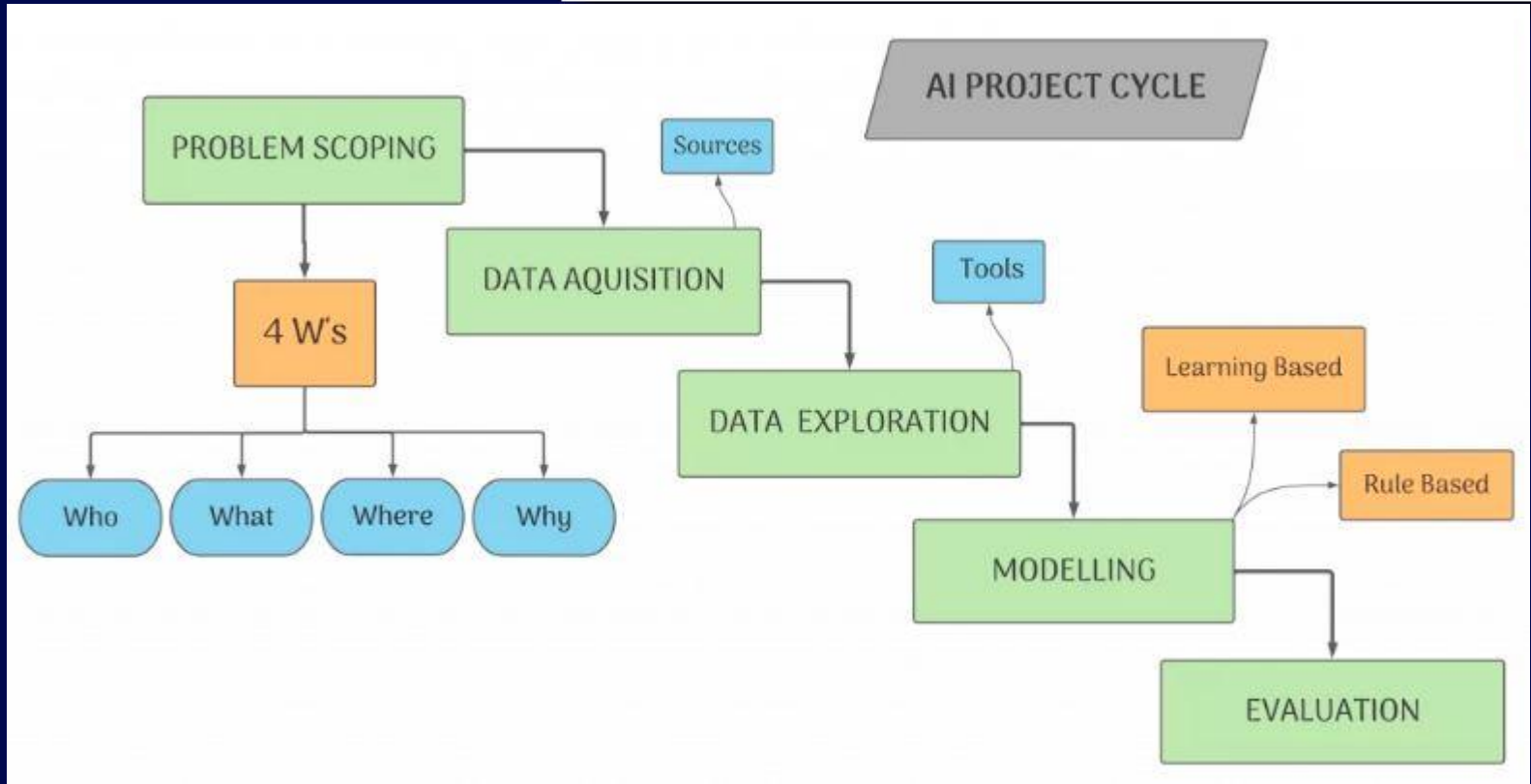
- Life cycle of AI project
- **Project Cycle** is a step-by-step **process to solve** problems using **proven scientific methods** and drawing inferences about them.



AI PROJECT CYCLE

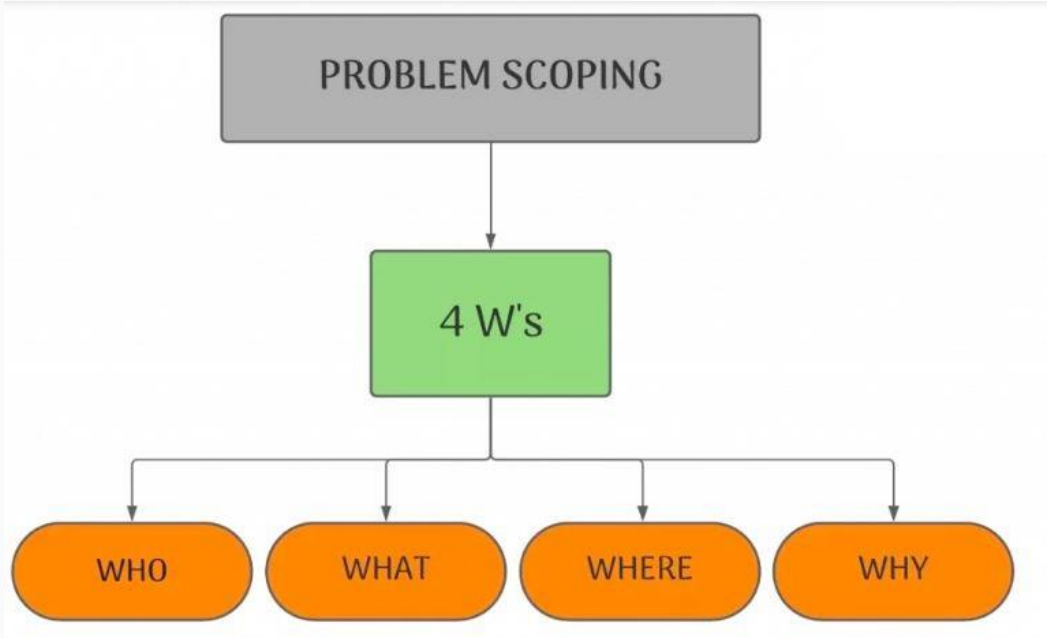
Problem Scoping	Understanding the problem
Data Acquisition	Collecting accurate and reliable data
Data Exploration	Arranging the data uniformly
Modelling	Creating Models from the data
Evaluation	Evaluating the project

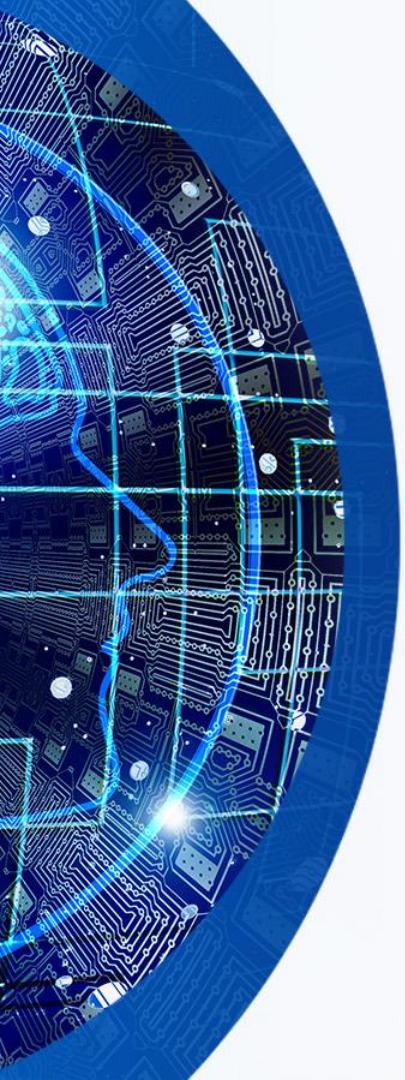
AI PROJECT CYCLE



Problem Scoping

- Problem identification and have a vision to solve it.

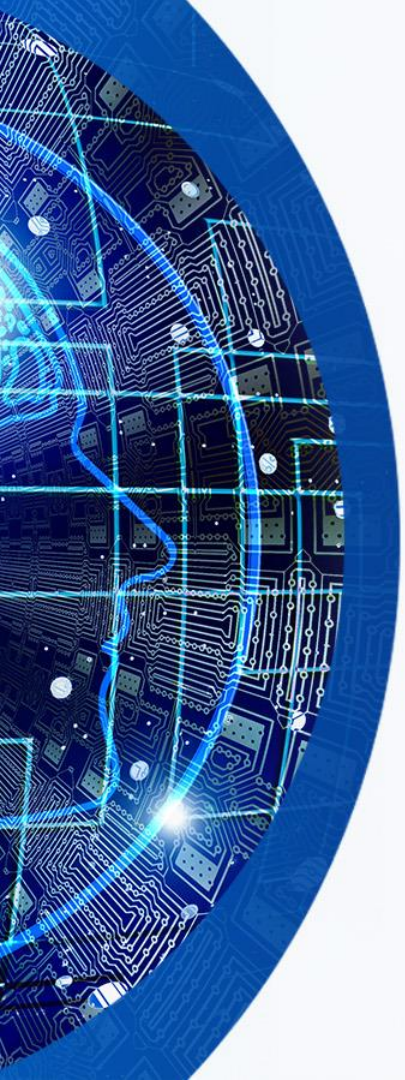




Problem Scoping

- Who? - Refers that who is facing a problem and who are the stakeholders of the problem.
- What? - Refers to what is the problem and how you know about the problem.
- Where? - It is related to the context or situation or location of the problem.
- Why? - Refers to why we need to solve the problem and what are the benefits to the stakeholders after solving the problem.

The final outcome of problem scoping is the problem statement template.

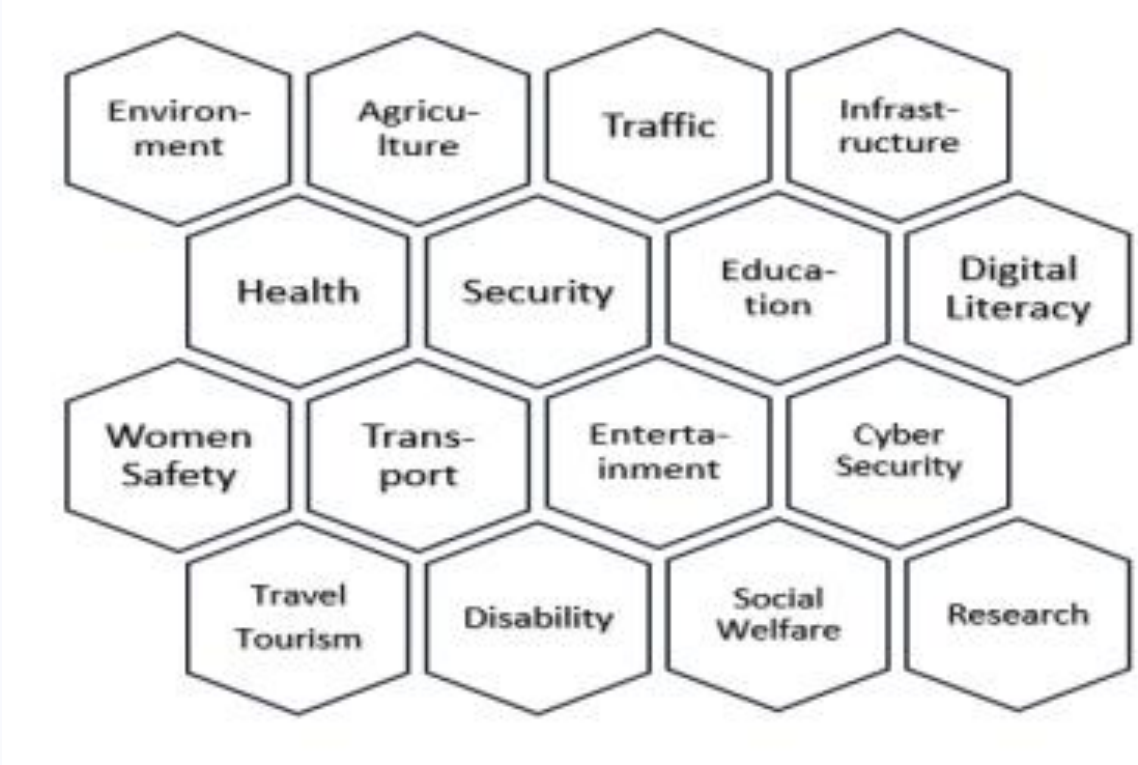


The problem statement template

- When the above 4Ws are completely filled you need to prepare a summary of these 4Ws.
- This summary is known as the problem statement template.
- This template explains all the key points in a single template.
- So if the same problem arises in the future this statement helps to resolve it easily.

Problem Scoping

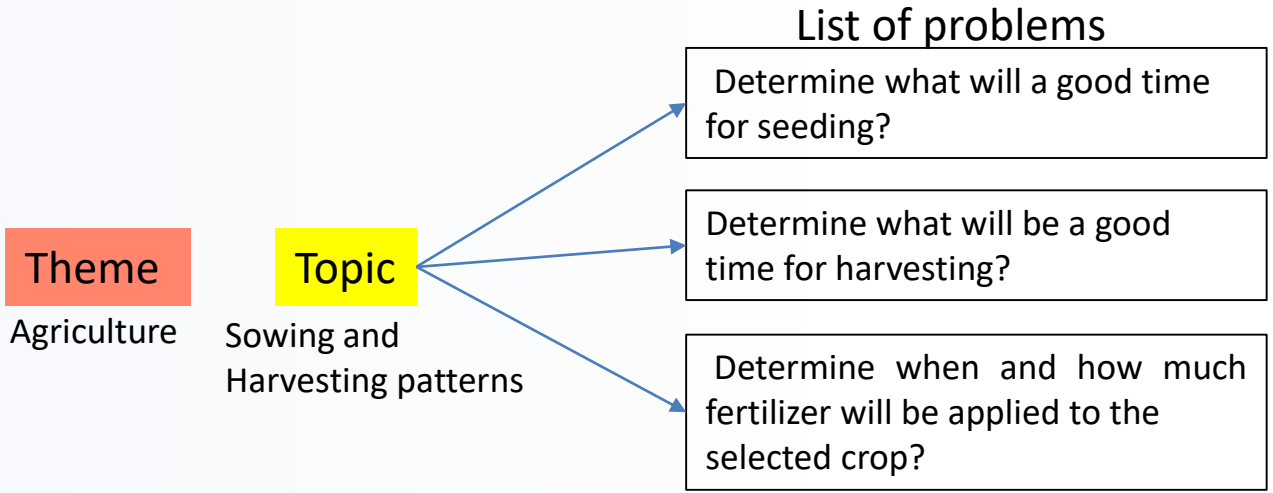
Activity - Brainstorm around the theme and set a goal for the AI project

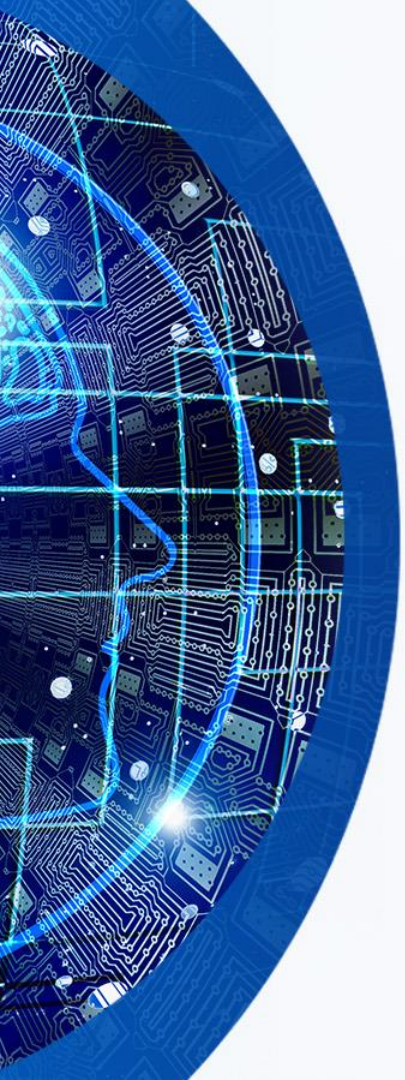


Problem Scoping

1. <u>Select a theme</u>

Eg: Agriculture: this theme includes different topics like pest issues, sowing and harvesting patterns, yield rates etc.



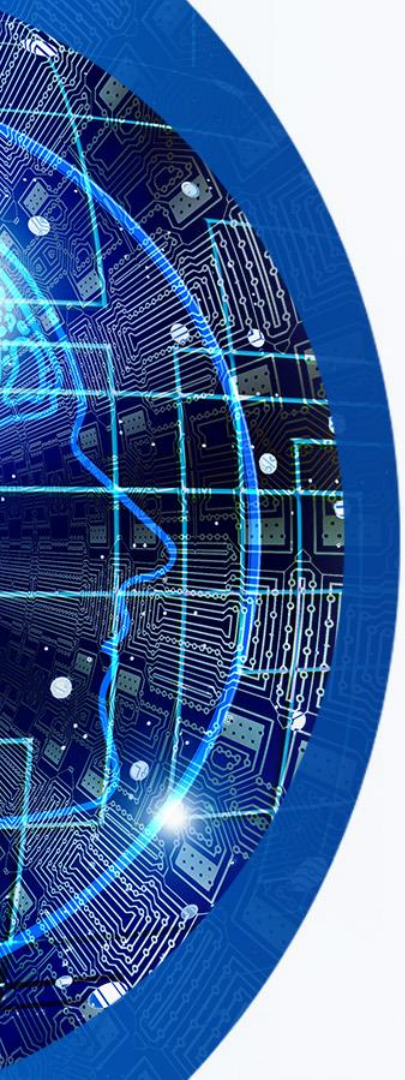


Problem Scoping

1. Select a goal

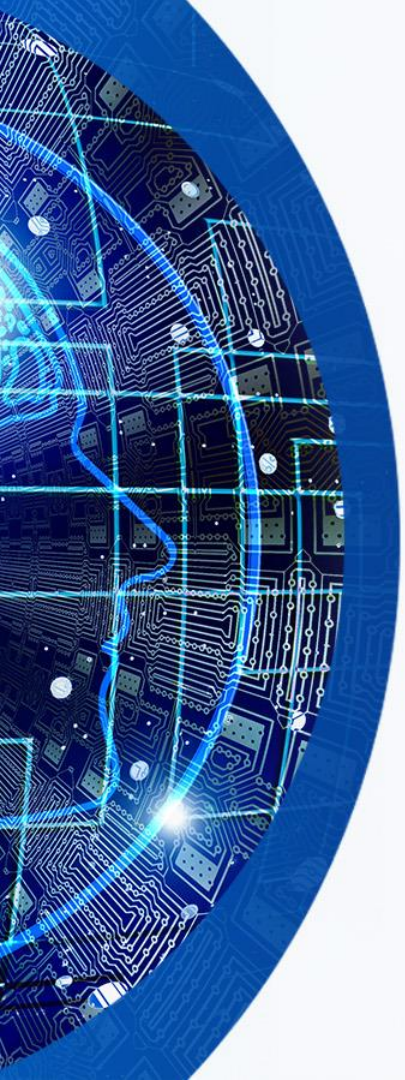
Select one problem out of the ones listed above to solve using AI knowledge. This Problem now becomes the target of AI project and helps for getting a clear vision of what is to be achieved.

Eg: Goal is - How might we help farmers determine the best times for seeding and for sowing their crops?



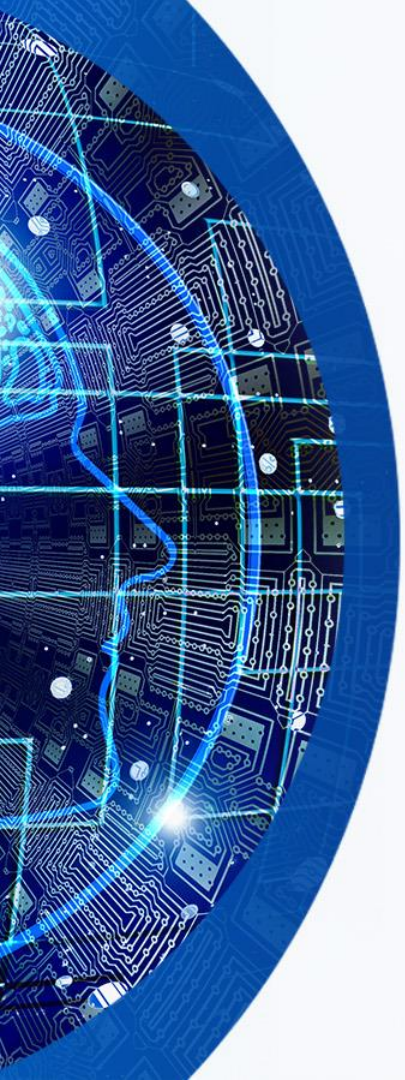
3. Final problem statement

	Stakeholders	Who
Our	Farmers, Fertilizer Producers, Labours, Tractor Companies	
has /have a problem that	The problem, Issue, Need	What
	Determine what will a good time for seeding or crop harvesting?	
When/while	Context/Situation	When
	Decide the mature age for the crop and determine its time	
An ideal solution would	Solution Benefits	Why
	Take the crop on time and supply against market demand on time	



Recap:
4 Ws

- Who?
  - a. Who are the stakeholders?
  - b. What do we know about them?
- 2. What?
  - a. What is the problem?
  - b. How do you that it is a problem? (is there an evidence?)
- 3. Where?
  - a. What is the context/situation the stakeholders experience this problem?
  - b. Where is the problem located?
- 4. Why?
  - a. What would hold value for the stakeholders?
  - b. How will the solution improve their situation?



Data Acquisition

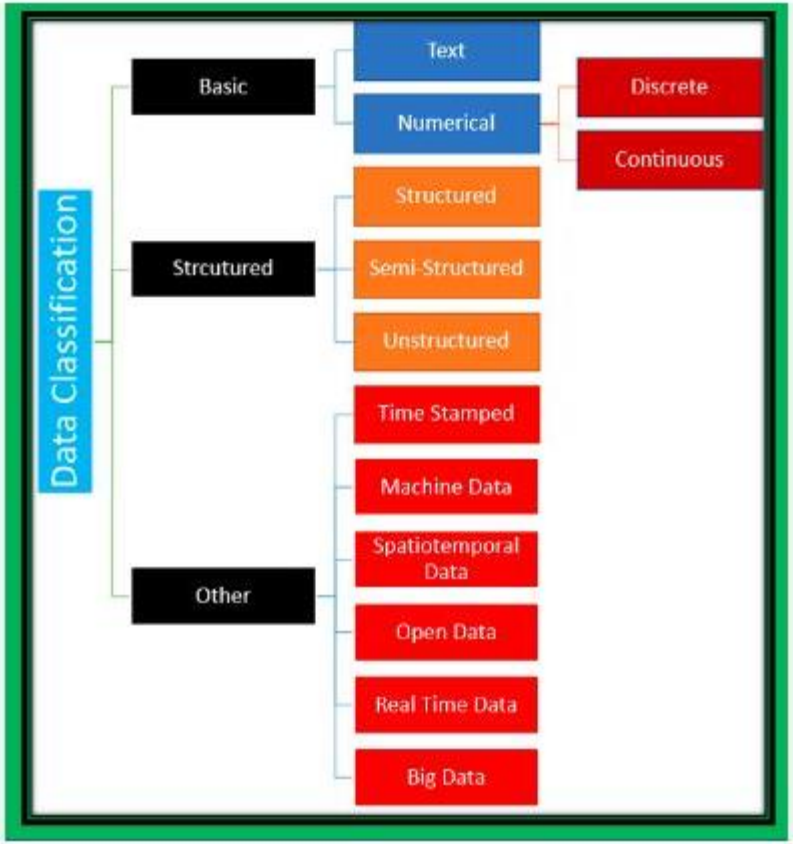
- **Acquiring data for the project.**

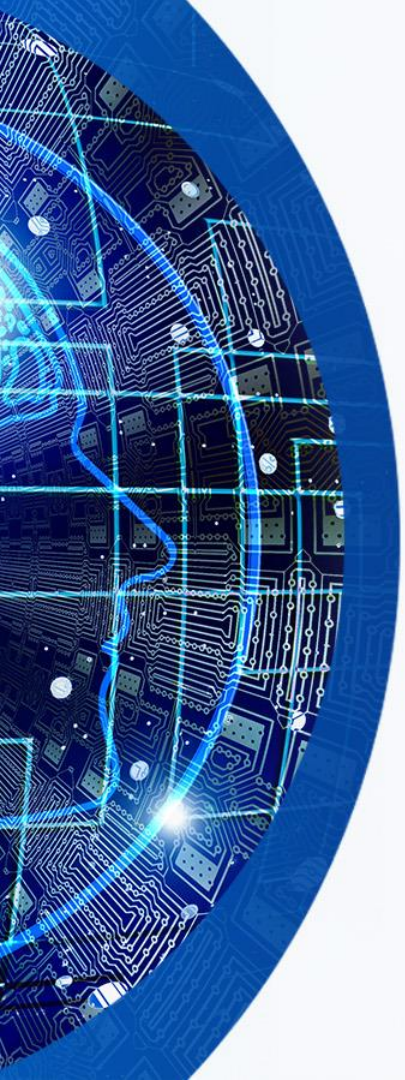
- What is data?
  - Data can be a piece of information or facts and statistics collected together for reference or analysis.
  - Whenever we want an AI project to be able to predict an output, we need to train it first using data.

Data Acquisition

- **Data features**
  - Data features refer to the type of data you want to collect.
  - data features would be salary amount, increment percentage, increment period, bonus, etc.

- **Types of Data**





Data Acquisition

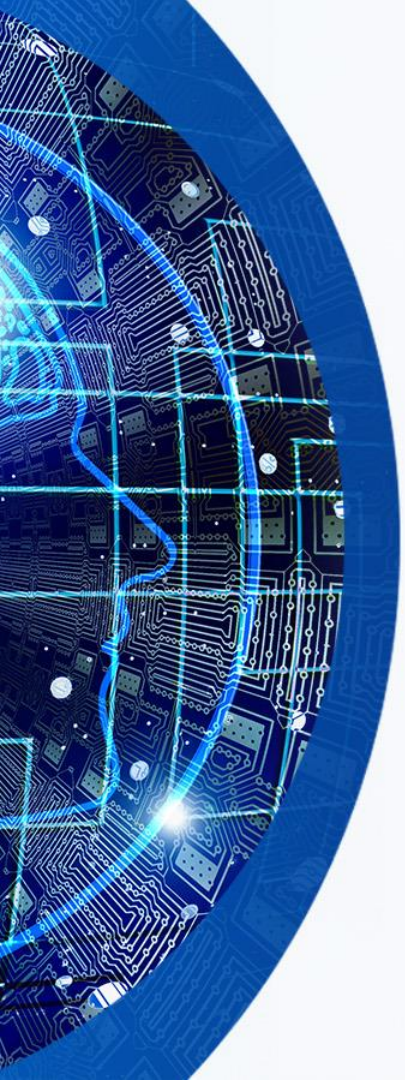
Types of Data :

a. Basic Data

Basic data is classified into two categories:

1. **Numeric Data:** Mainly used for computation.
   - **Discrete Data**: Discrete data only contains integer numeric data. It doesn't have any decimal or fractional value. The countable data can be considered as discrete data. Eg: 132 customers, 126 Students etc.
   - **Continuous Data**: It represents data with any range. The uncountable data can be represented in this category.Eg: 10.5 KGS, 100.50 Kms etc.

2. Text Data: mainly used to represent names, collection of words together, phrases, textual information etc.



Data Acquisition

Types of Data :

b. Structural Classification

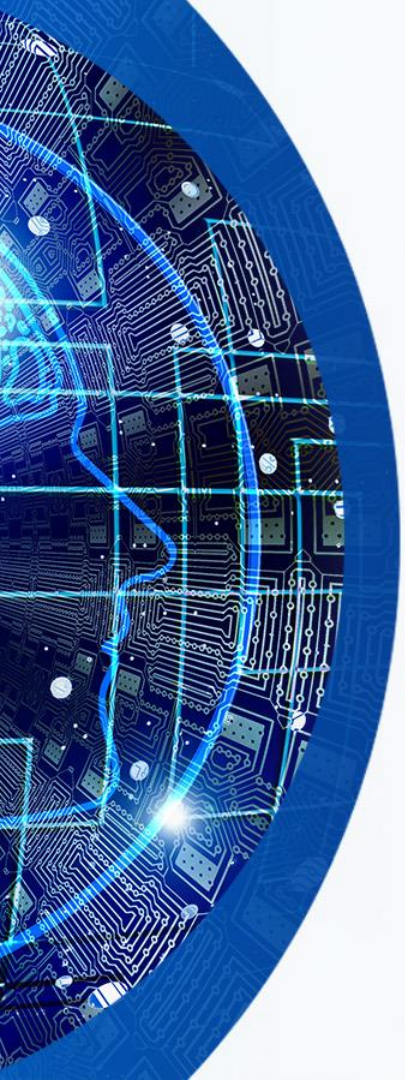
The data which is going to be feed in the system to train the model or already fed in the system can have a specific set of constraints or rules or unique pattern can be considered as structural data.

1. **Structured Data:** can have a specific pattern or set of rules.
    - These data have a simple structure and stores the data in specific forms such as tabular form. Eg: **The cricket scoreboard, school time table, Exam datasheet etc.**

2. Unstructured Data: doesn't have any specific pattern or constraints as well as can be stored in any form is known as unstructured data.

- Mostly the data that exists in the world is unstructured data.
- Eg: Videos, Facebook Photos, Dashboard data of any reporting tool.

3. Semi-Structured Data: combination of both structured and unstructured data. Some data can have a structure like a database whereas some data can have markers and tags to identify the structure of data.

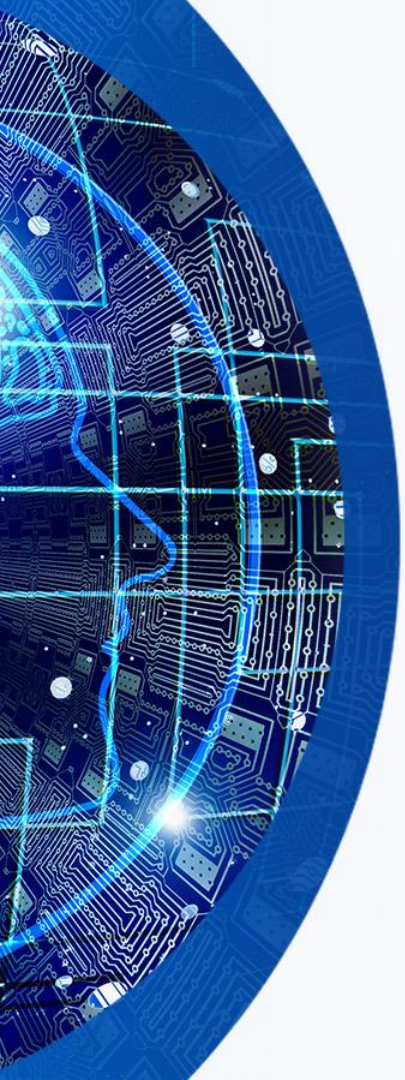


Data Acquisition

Types of Data :

Other Classification

1. **Time-Stamped Data:** This structure helps the system to predict the next best action. It is following a specific time-order to define the sequence. This time can be the time of data captured or processed or collected.
2. **Machine Data:** The result or output of a specific program, system or technology considered as machine data. It consists of data related to a user's interaction with the system like the user's logged-in session data. , specific search records, user engagement such as comments, likes and shares etc
3. **Spatiotemporal Data:** The data which contains information related to geographical location and time is considered as spatiotemporal data. It records the location through GPS and time-stamped data where the event is captured or data is collected.
4. **Open Data**: It is freely available data for everyone. Anyone can reuse this kind of data.
5. **Real-time Data**: The data which is available with the event is considered as real-time data.
6. **Big Data:** You may hear this word most often. The data which cannot be stored by any system or traditional data collection software like DBMS or RDBMS software can be considered as Big data.



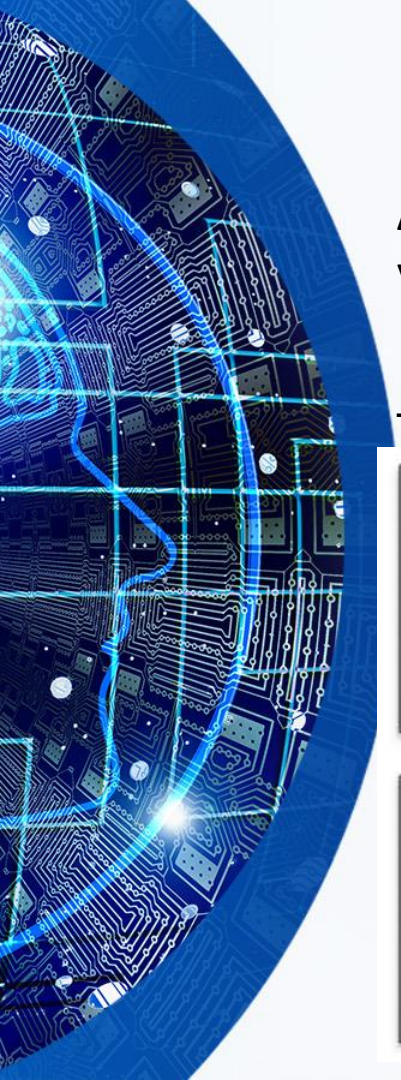
Data Acquisition

Data Features : type of data you want to collect.
Here two terms are associated with this:

1**. Training Data:** The collected data through the system is known as training data. In other words the input given by the user in the system can be considered as training data.

2. **Testing Data:** The result data set or processed data is known as testing data. In other words, the output of the data is known as testing data.

- Eg: to make an Artificially Intelligent system which can predict the salary of any employee based on his previous salaries.
- The previous salary data here is known as **Training Data** while the next salary prediction data set is known as the **Testing Data**
- For any AI project to be efficient, the training data should be authentic and relevant to the problem statement scoped. .

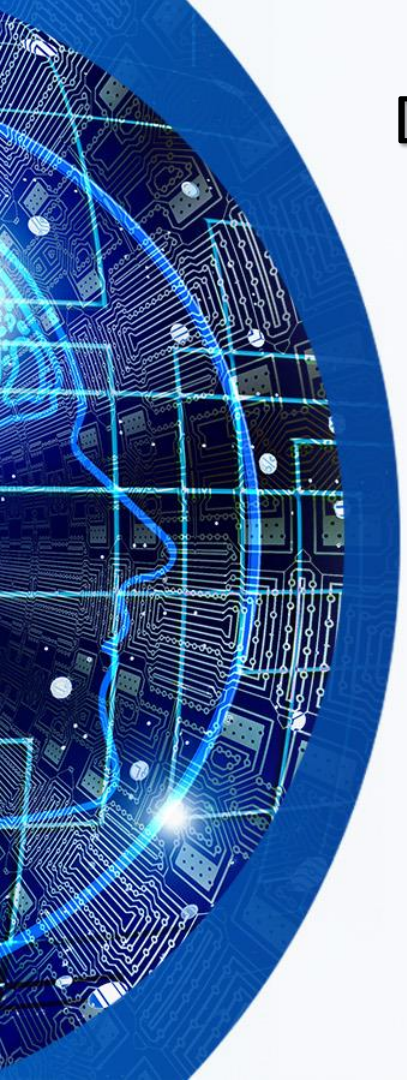


Data Acquisition

Acquiring Data from reliable sources: After mentioning the Data features, you get to know what sort of data is to be collected.

Different data sources are:

Surveys	Web Scraping	Sensors
Cameras	Observations	API(Application Program Interface)



Data sources

1. Web Scraping means collecting data from web using some technologies. We use it for monitoring prices, news and etc.
Example: Web Scrapping. using beautiful soup in python

2. Sensors are very Important but very simple to understand. Sensors are the part of IoT (Internet of things) Sensors collect the physical data and detect the changes.

3. Camera: captures the visual information and then that information which is called image is used as a source of data. Cameras are used to capture raw visual data.

4. Observations: When we observe something carefully we get some information For ex: Scientists Observe creatures to study them. Observations is a time consuming data source.

5. API: Application Programming interface. API is a messenger which takes requests and tells the system about requests and gives the response.
Ex: Twitter API, Google Search API

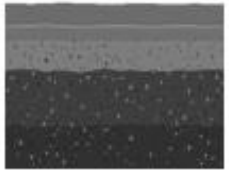
6 Surveys: The survey is a method of gathering specific information from a sample of people. Example, a census survey for analyzing the population.

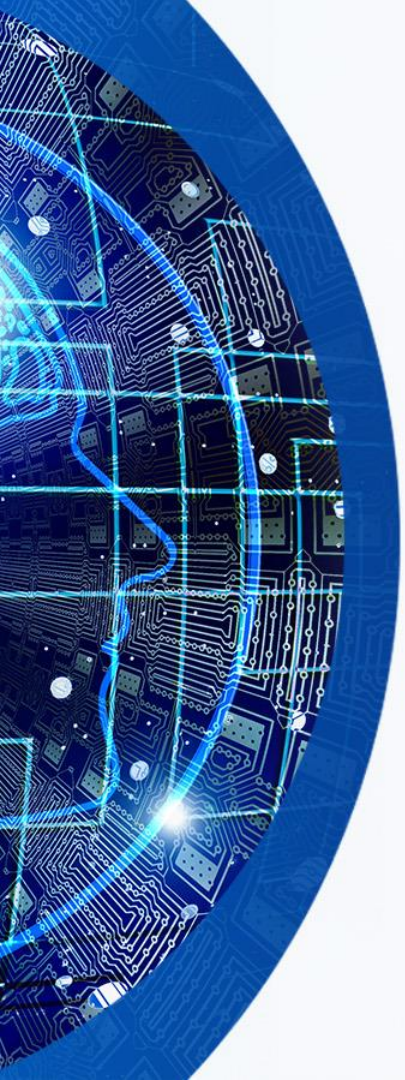
System Maps

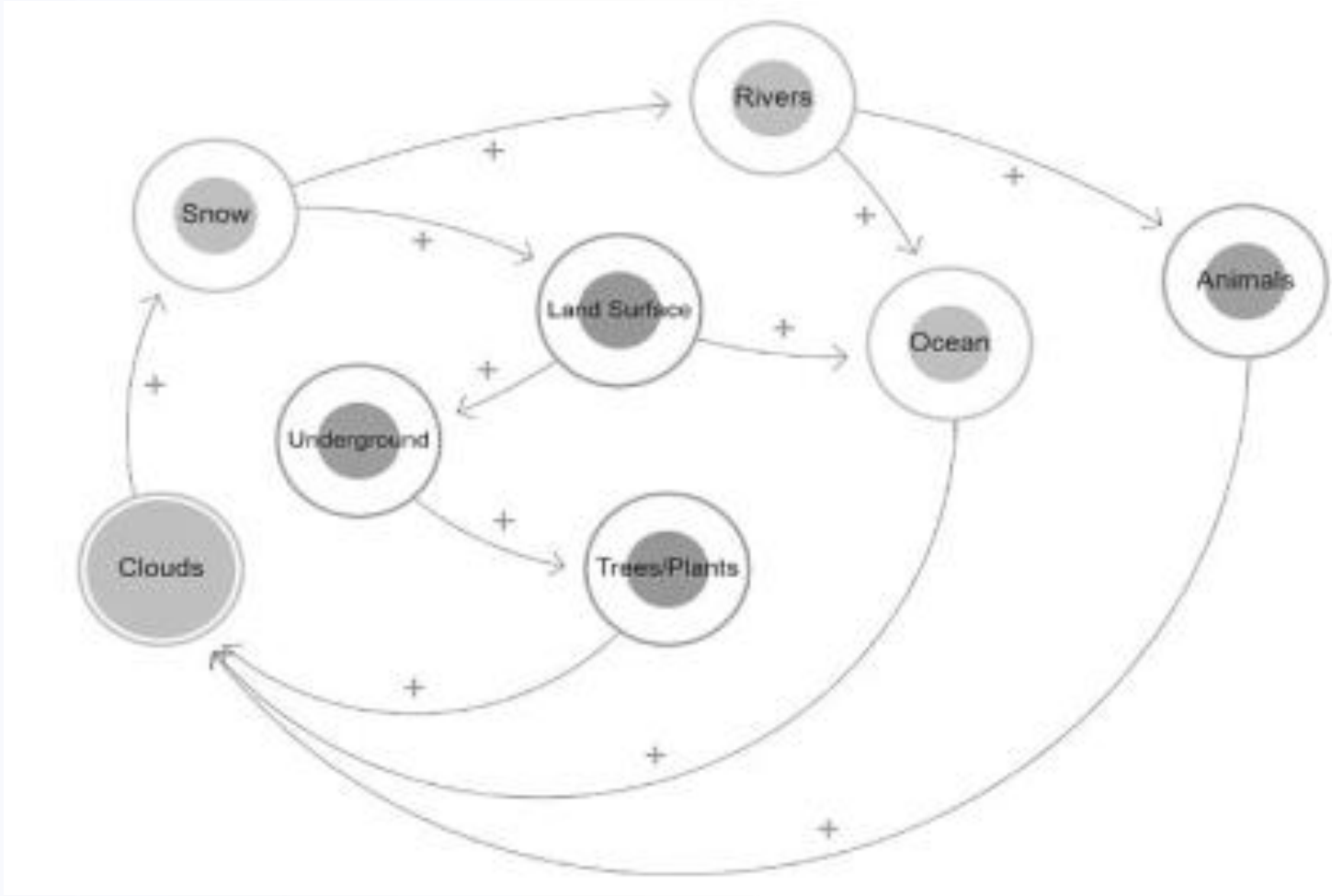
- System maps help us to find the relationship between the elements of the problem which we have scoped
- A system map shows the components and boundary of a system and the components of the environment at a point in time.
- The main use of a system map is to help structure a system and communicate the result to others.
- It helps us in strategizing the solution for achieving the goal of our project.
- help to understand complex issues with multiple factors that affect each other
- **Circles** represents the elements,
- **Arrows** represents the relationship between the elements. Length of arrow represents time for a change to happen. This is **time delay. The arrow- head** depicts the direction of the effect and the sign (+ or -) shows their relationship. **If the arrow goes from X to Y with a + sign, it means that both are directly related to each other. If the arrow goes from X to Y with a – sign, it means that both the elements are inversely related to each other .**
- **Loops** represent a specific chain of causes and effects.
- To change the outcome of a system, as a change maker, either change the elements in a system or change the relationships between elements.

System map for Water Cycle.

The major elements of this system are mentioned here.

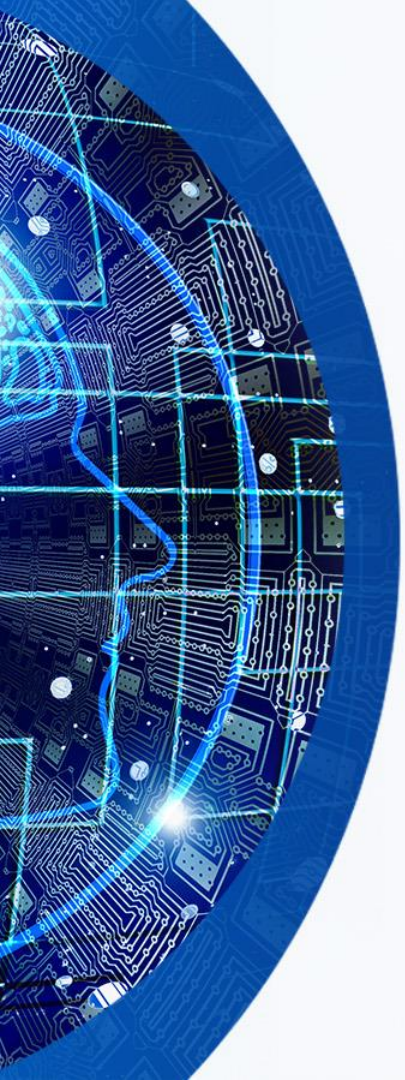
			
Clouds	Snow	Underground Soil	Rivers
			
Oceans	Trees	Land	Animals

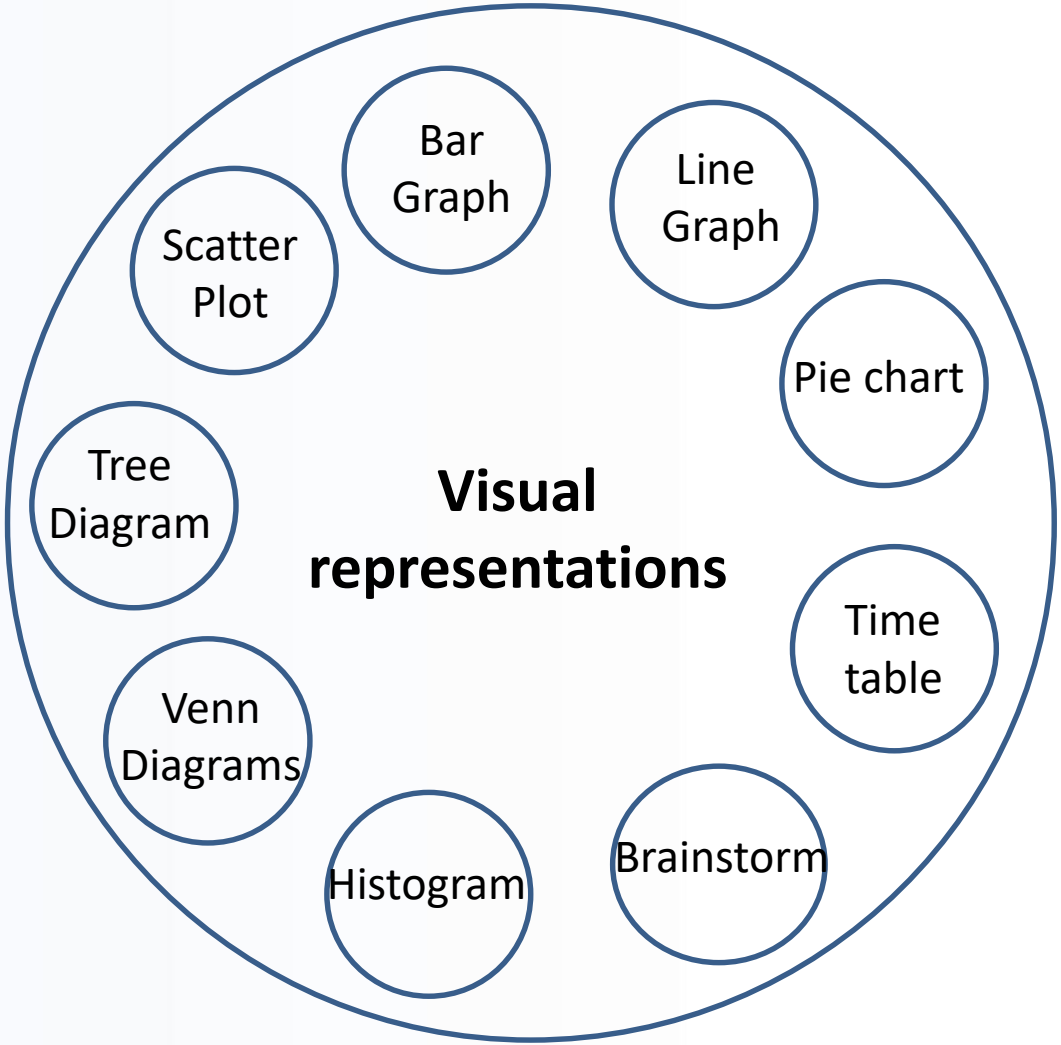


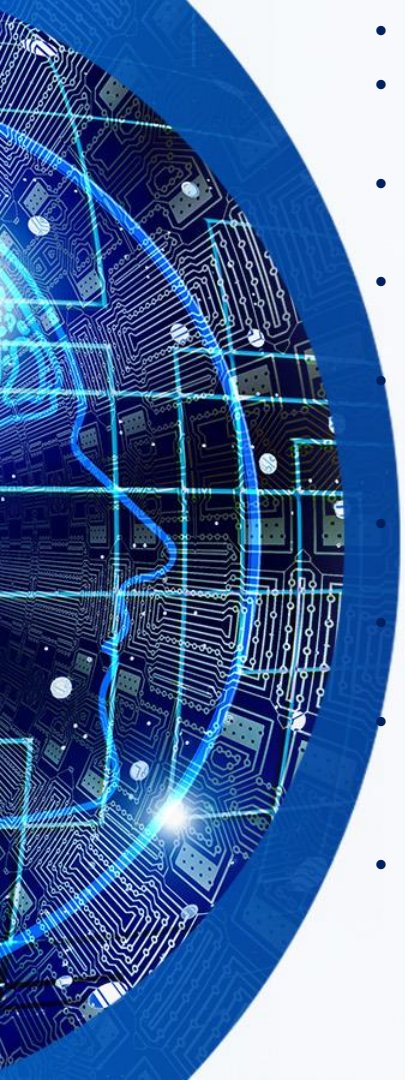

Rivers
Snow
Land Surface
Ocean
Animals
Underground
Clouds
Trees/Plants

Data Exploration

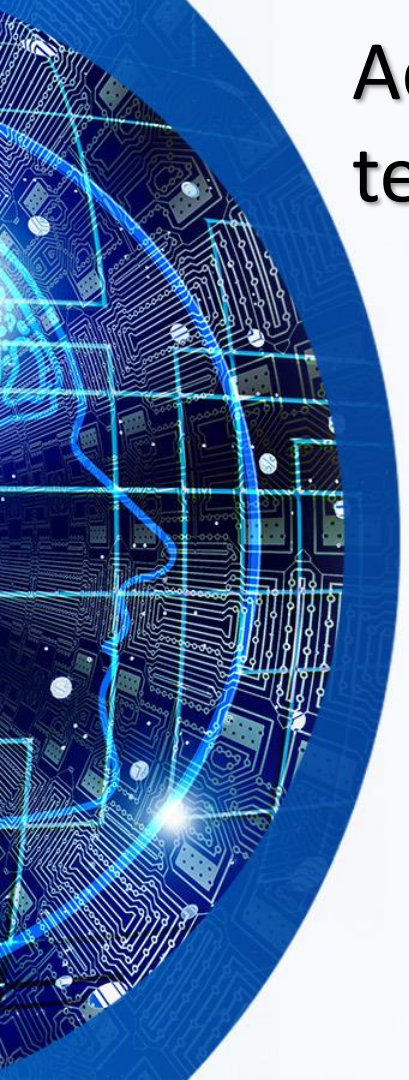
- Data Exploration refers to the techniques and tools used to visualize data through complex statistical methods.
- Advantages of Data Visualization
  - ❖ A better understanding of data
  - ❖ Provides insights into data
  - ❖ Allows user interaction
  - ❖ Provide real-time analysis
  - ❖ Help to make decisions
  - ❖ Reduces complexity of data
  - ❖ Provides the relationships and patterns contained within data
  - ❖ Define a strategy for your data model
  - ❖ Provides an effective way of communication among users




Visual representations
Bar Graph
Line Graph
Scatter Plot
Pie chart
Tree Diagram
Time table
Venn Diagrams
Brainstorm
Histogram

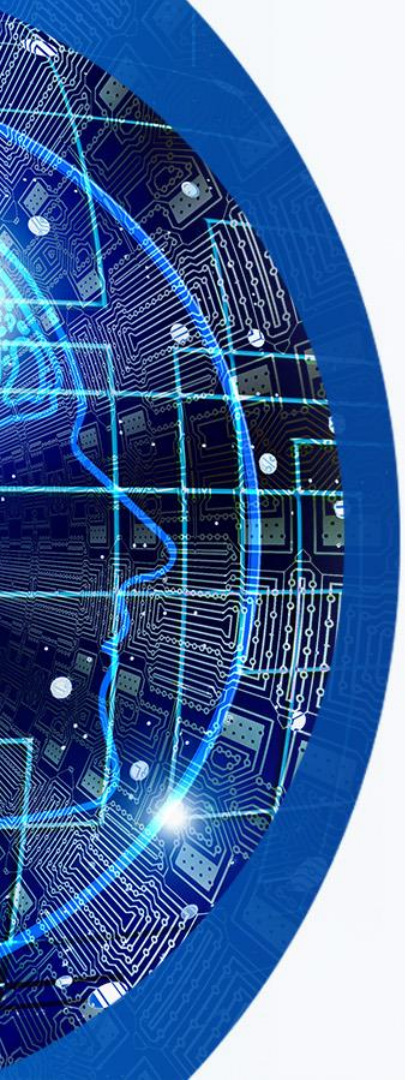


- Types of Graphical Representation
- Bar Graph: This uses either horizontal or vertical bars to categorize and compare different quantities. The length or height of the bar corresponds to the value it represents.
- Pie Chart: This represents data in the form of slices of a circle, where each slice represents a category. The size of each slice is proportional to the category it represents.
- Line Graph: This shows trends over time by connecting data points with straight lines. It's especially useful for showing changes over a period and trends.
- Pictograph: This type of graph uses pictures or icons to represent data values. Each image or icon represents a specific number or quantity, allowing for an easy-to-understand visual comparison.
- Histogram: This is a type of bar graph used to display frequency data. The bars are adjacent to each other, indicating that the data is in intervals.
- Frequency Distribution: This graphical representation (usually a table or a graph) shows how often each different value in a set of data occurs.
- Stem and Leaf Plot: This method of data representation organizes data into a semi-graphical representation that can display large amounts of data, and allows you to see the distribution and shape of the data set.
- Scatter Plot: This is used for displaying values for two variables for a set of data. The data is displayed as a collection of points, each representing the value variables plotted on a horizontal and vertical axes.



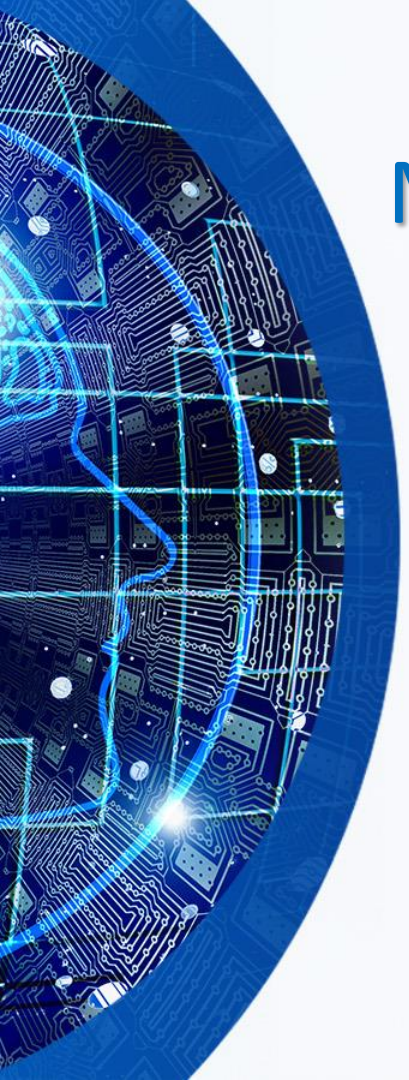
Activity: List down 5 new data visualization techniques which you learnt from

Data Visualisation Technique 1	
Name of the Representation	
One-line Description	
How to draw it	
Suitable for which data type?	

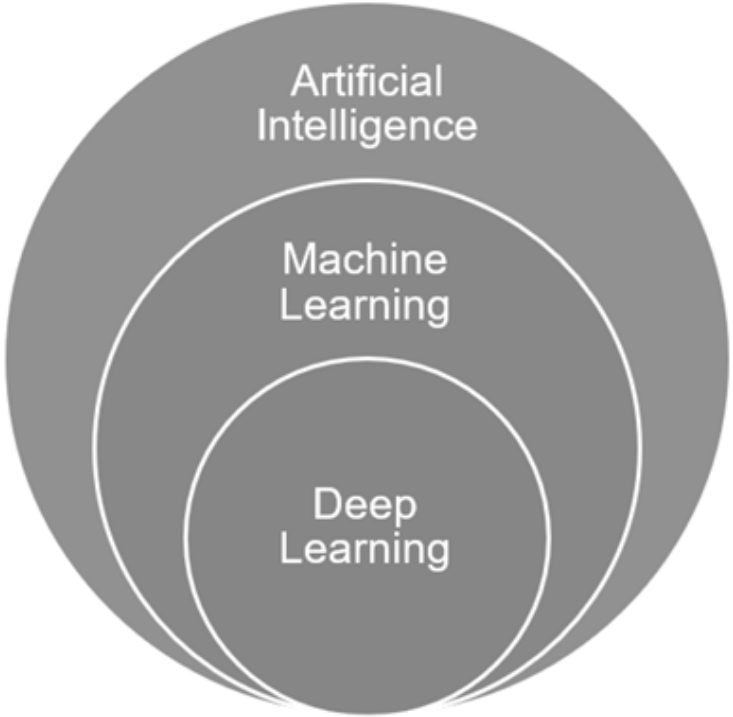


How to select a proper graph?

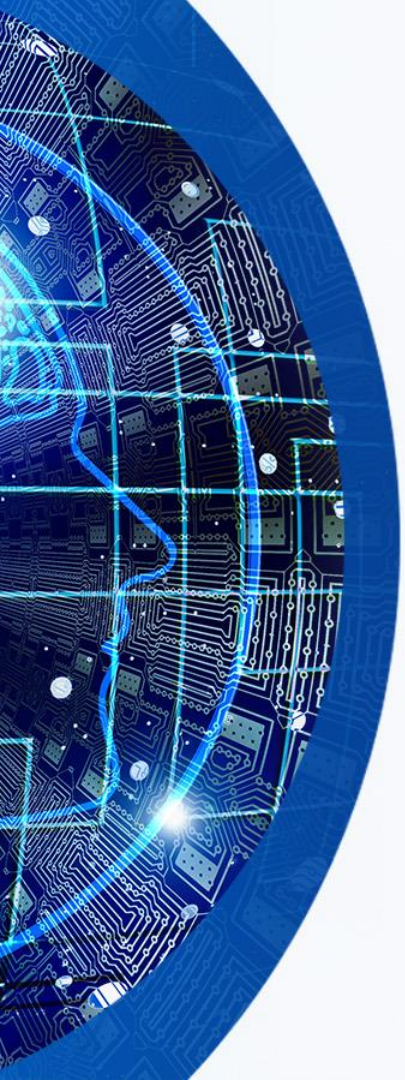
1. Comparison of Values - Show periodical changes i.e. Bar Chart
2. Comparison of Trends - Show changes over a period of time i.e. Line Chart
3. Distribution of Data according to categories - Show data according to category i.e. Histogram
4. Highlight a portion of a whole - Highlight data according to value i.e. Pie Chart
5. Show the relationship between data - Multiple charts can be used



Modelling



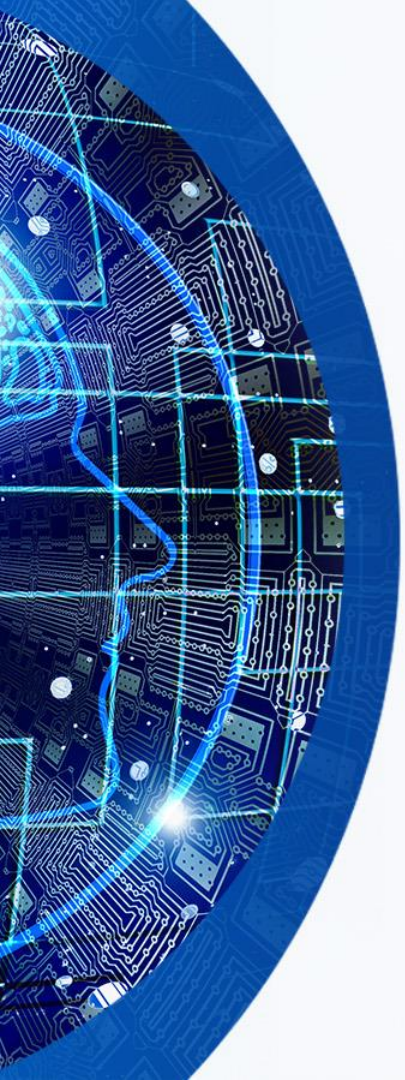
As you can see in the Venn Diagram, Artificial Intelligence is the umbrella terminology which covers machine and deep learning under it and Deep Learning comes under Machine Learning. It is a funnel type approach where there are a lot of applications of AI out of which few are those which come under ML out of which very few go into DL.



Modelling

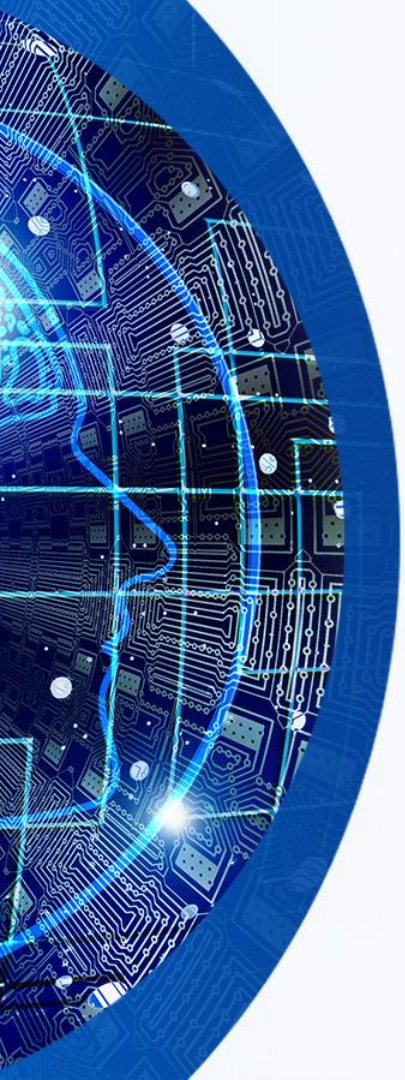
- **Artificial Intelligence,** or AI refers to any technique that enables computers to mimic human intelligence.
  - An artificially intelligent machine works on algorithms and data fed to it and gives the desired output.
- **Machine Learning,** or ML for short, enables machines to improve at tasks with experience.
  - The machine here learns from the new data fed to it while testing and uses it for the next iteration. The machine learns from its mistakes and takes them into consideration in the next execution.
  - It improvises itself using its own experiences.
- **Deep Learning,** or DL for short, enables software to train itself to perform tasks with vast amounts of data.
  - In deep learning, the machine is trained with huge amounts of data which helps it into training itself around the data. Such machines are intelligent enough to develop algorithms for themselves.

Deep Learning is the most advanced form of Artificial Intelligence out of these three. Then comes Machine Learning which is intermediately intelligent and Artificial Intelligence covers all the concepts and algorithms which, in some way or the other mimic human intelligence.



AI vs ML vs DL

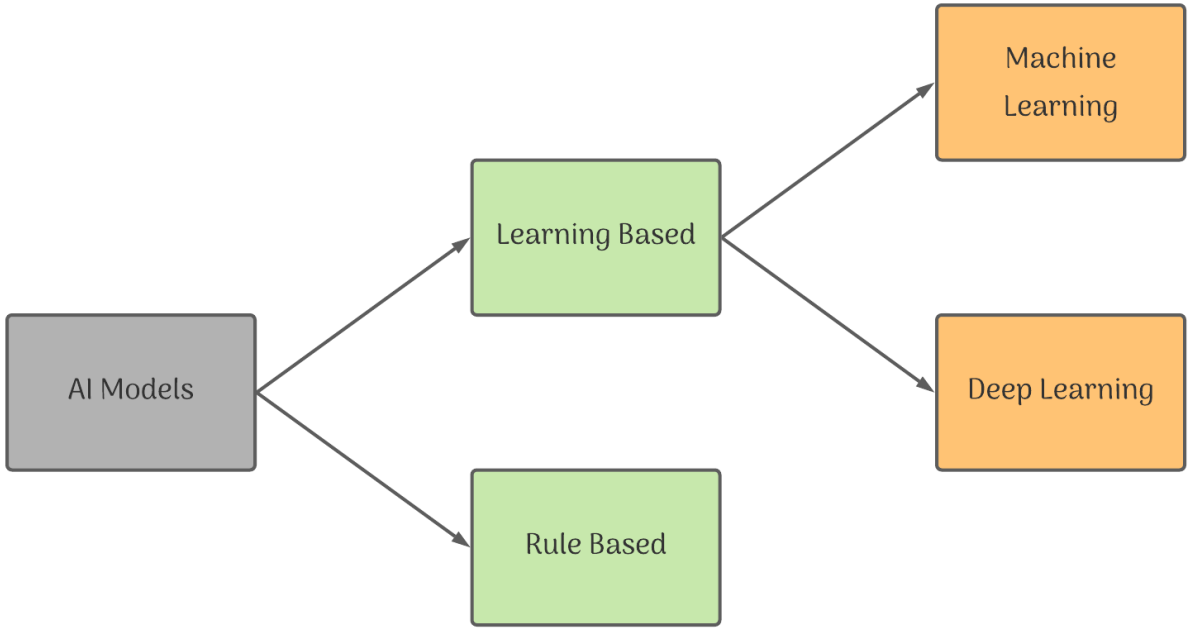
Artificial Intelligence	Machine learning	Deep Learning
AI represents stimulated intelligence in machines.	ML is the practice of getting machines to make decisions without being programmed.	It is artificial neural network to solve the complex problems.
AI is a subset of data science.	ML is the subset of AI and data science.	DL is the subset of AI , ML and data science.
AI aims toward building machines that are capable to think like humans	ML aims to learn through data to solve problems.	DL aim to build neural network that automatically discover patterns for feature detection



Modelling

- AI Modelling refers to developing algorithms, also called models which can be trained to get intelligent outputs. That is, writing codes to make a machine artificially intelligent.
- The graphical representation makes the data understandable for humans as we can discover trends and patterns out of it.
- But when it comes to machine accessing and analysing data, it needs the data in the most basic form of numbers (which is binary – 0s and 1s) and when it comes to discovering patterns and trends in data, the machine goes for mathematical representations of the same.
- The ability to mathematically describe the relationship between parameters is the heart of every AI model.
- Thus, whenever we talk about developing AI models, it is the mathematical approach towards analysing data which we refer to.

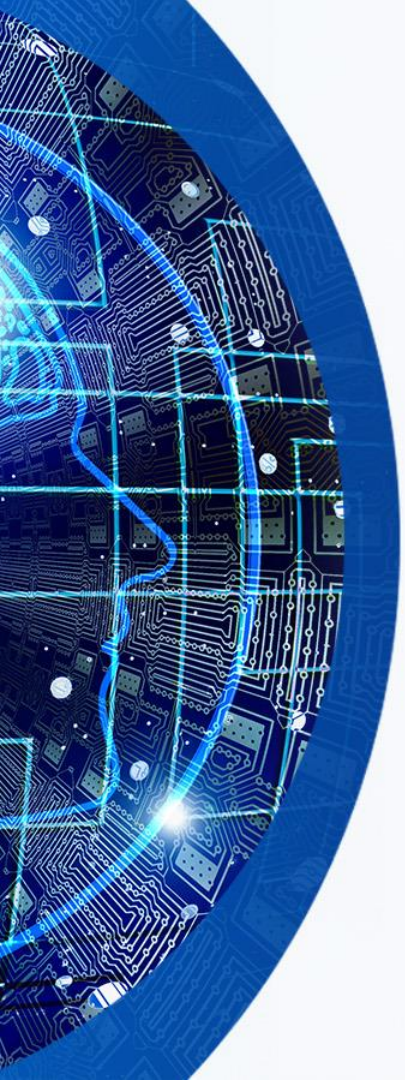
Modelling

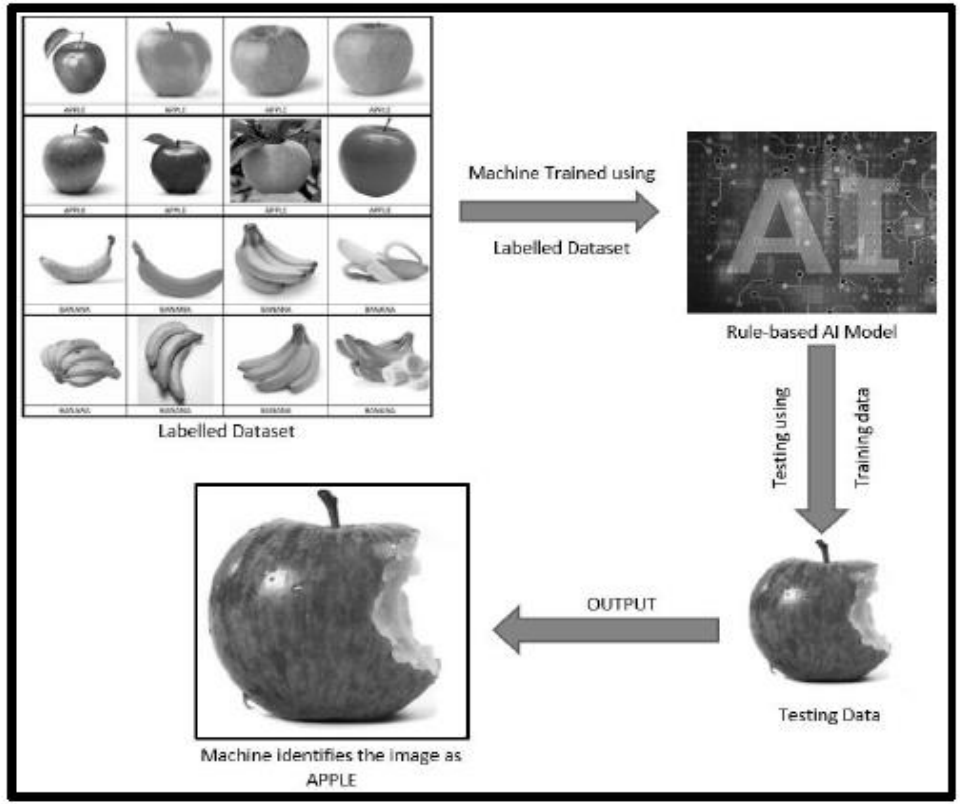

AI Models
Learning Based
Rule Based
Machine Learning
Deep Learning

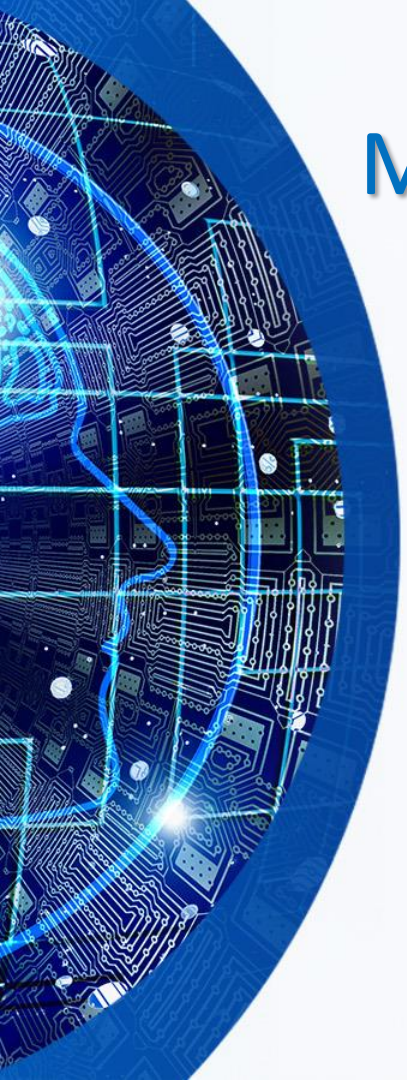
Modelling

Rule Based Approach :

- Rule Based Approach Refers to the AI modelling where the relationship or patterns in data are defined by the developer.
- The machine follows the rules or instructions mentioned by the developer, and performs its task accordingly.



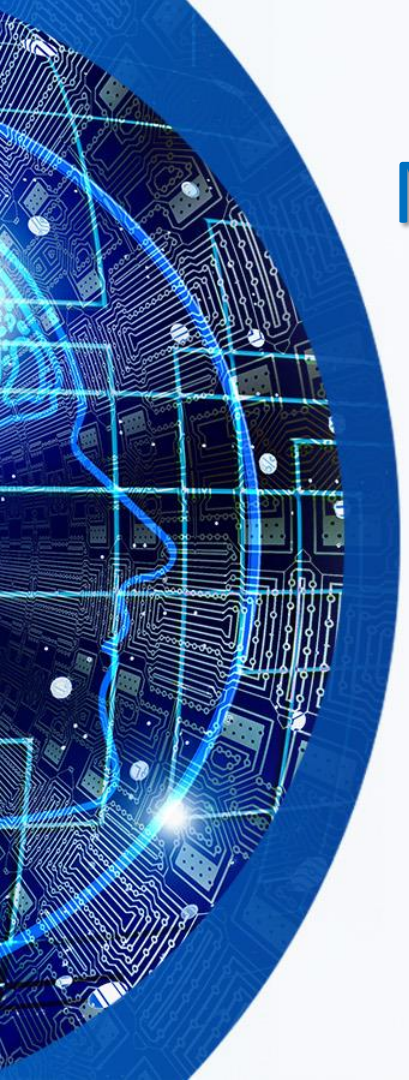

Machine Trained using
Labelled Dataset
AI
Rule-based AI Model
Testing using
Training data
Testing Data
OUTPUT
Machine identifies the image as
APPLE
Labelled Dataset



Modelling

Rule Based Approach :

- To train your machine, you feed this data into the machine and label each image as either apple or banana.
- Now if you test the machine with the image of an apple, it will compare the image with the trained data and according to the labels of trained images, it will identify the test image as an apple.
- This is known as Rule based approach. The rules given to the machine in this example are the labels given to the machine for each image in the training dataset.

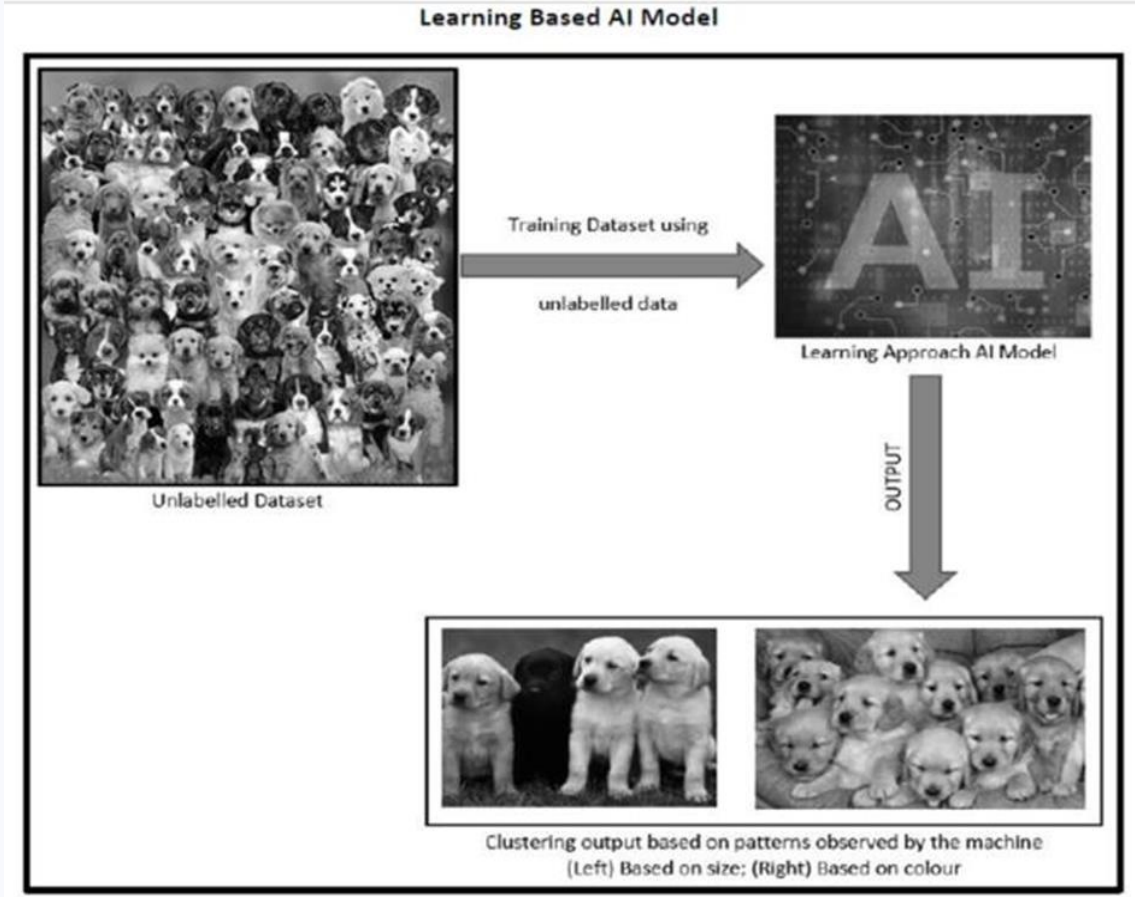


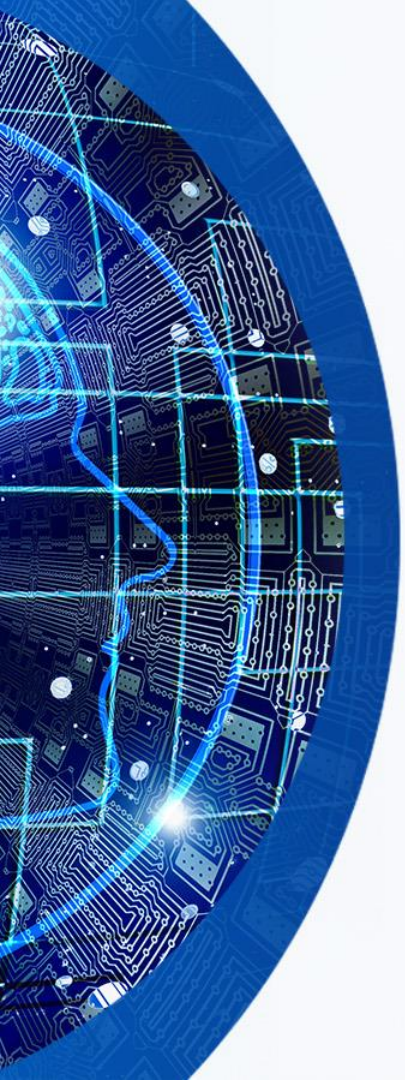
Modelling

Learning Based Approach

- AI modelling where the relationship or patterns in data are not defined by the developer.
- In this approach, random data is fed to the machine and it is left on the machine to figure out patterns and trends out of it.
- Generally this approach is followed when the data is unlabeled and too random for a human to make sense out of it. Thus, the machine looks at the data, tries to extract similar features out of it and clusters same datasets together.
- In the end as output, the machine tells us about the trends which it observed in the training data.

Modelling

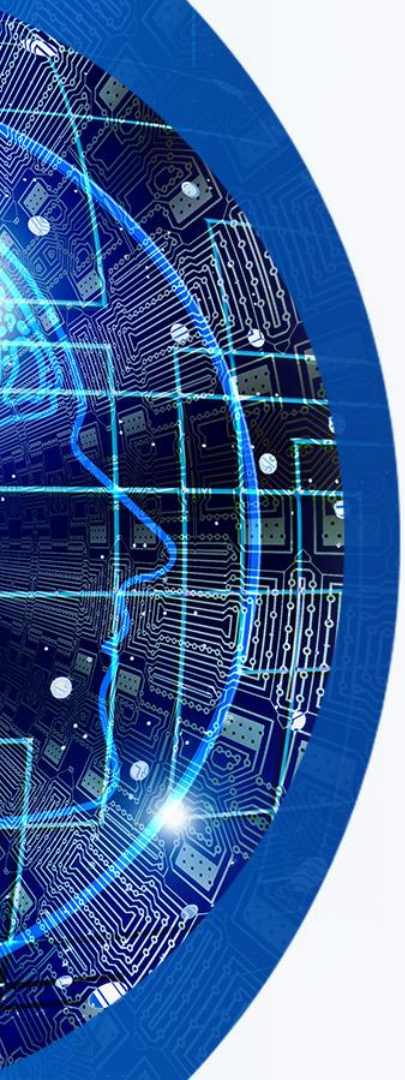

Learning Based AI Model
Training Dataset using
unlabelled data
AI
Learning Approach AI Model
Unlabelled Dataset
OUTPUT
Clustering output based on patterns observed by the machine
(Left) Based on size; (Right) Based on colour



Modelling

Learning Based Approach

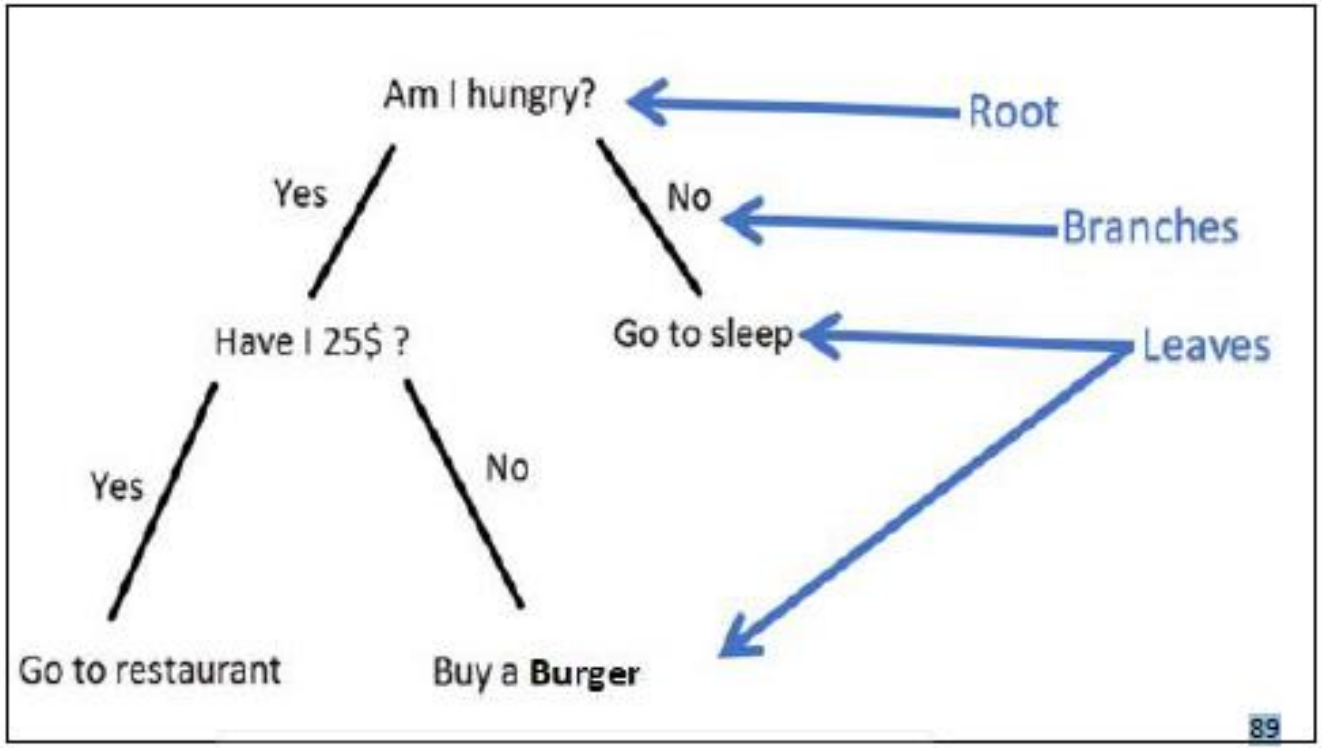
- For example, suppose you have a dataset of 1000 images of random stray dogs of your area.
- Now you do not have any clue as to what trend is being followed in this dataset as you don’t know their breed, or colour or any other feature.
- Thus, you would put this into a learning approach based AI machine and the machine would come up with various patterns it has observed in the features of these 1000 images.
- It might cluster the data on the basis of colour, size, fur style, etc. It might also come up with some very unusual clustering algorithm which you might not have even thought of!



Decision tree

- Decision tree is the most powerful and popular tool for classification and prediction.
- It made up of several nodes with top-down approach
- A Decision tree is a flowchart like tree structure.
- The basic structure of a Decision Tree starts from the root which the point where the decision tree starts. From there, the tree diverges into multiple directions with the help of arrows called branches.
- Each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

Decision Tree



THANK YOU