ARTIFICIAL INTELLIGENCE

Class XI

TEACHER HANDBOOK

Subject Code:843

ARTIFICIAL INTELLIGENCE CURRICULUM

Teacher Handbook for Class XI

Acknowledgments

Patrons
Mr. Rahul Singh, IAS, Chairperson, Central Board of Secondary Education

Strategic Guidance
Dr. Biswajit Saha, Director (Skill Education), Central Board of Secondary Education
Sh. Ravinder Pal Singh, Joint Secretary, Department of Skill Education, Central Board of Secondary Education

Strategic Advisory
Ms. Shipra Sharma, CSR Leader, India/South Asia, IBM
Ms. Joyeeta Das, Lead & Strategist, Global Education and Workforce Development, IBM
Dr. Mani Madhukar, Program Lead - SkillsBuild, IBM

Curriculum Planning Team
Mr. Manav Subodh, Founder & Chief Mentor, 1M1B
Mr. Saffin Mathew, Programs Director, 1M1B

Lead Curriculum Curator
Ms. Mehreen M Shamim, AI Curriculum Manager, 1M1B

AI Teacher Advisory for Curriculum Revision
Mr. Akhil R, TGT, DAV Public School Pushpanjali Enclave, Delhi
Ms. Anni Kumar, PGT, Vikas Bharati Public School, Delhi
Ms. Harmeet Kaur, PGT, Vasant Valley School, New Delhi
Ms. Jyoti P B, PGT, Jyothis Central School, Thiruvananthapuram, Kerala
Mr. Naveen Gupta, PGT, St. Mark's Sr. Sec Public School, Meera Bagh, Delhi
Ms. Neeru Mittal, PGT, Shaheed Rajpal DAV Public School, Delhi
Ms. Rani Kumari, PGT, DLF Public School, Ghaziabad, Uttar Pradesh
Ms. Shelly Batra, TGT, Mount Carmel School, Dwarka, Delhi
Ms. Smitha R Athreya, PGT, Delhi Public School Rourkela, Odisha
Ms. Soumya Iyer, PGT, Sanskriti School, Pune, Maharashtra
Ms. Swati Sharma, TGT, Heritage Xperiential Learning School, Gurugram, Haryana
Mr. Tushar Upadhyay, TGT, Navrachana Higher Secondary School, Vadodara Gujarat
Ms. Varsha Vijay K, TGT, Delhi Public School Bangalore North, Bangalore, Karnataka
Ms. Vineeta Garg, PGT, Shaheed Rajpal DAV Public School, Delhi

Foreword

The world around us is undergoing a dramatic transformation, driven by the relentless advancement of Artificial Intelligence (AI). From self-driving cars navigating city streets to virtual assistants understanding complex inquiries, AI is rapidly reshaping industries, societies, and the very way we interact with technology.

This revised textbook, designed for students in Classes XI and XII, dives into the captivating world of AI, offering a comprehensive exploration of its core concepts, applications, and potential impact. As you embark on this journey, you will not only delve into the fascinating algorithms that power AI systems, but also examine its ethical considerations and its profound implications for the future.

This is no longer science fiction. AI is here, and it holds immense potential to improve our lives in countless ways. This textbook equips you, the future generation, with the knowledge and critical thinking skills necessary to navigate this rapidly evolving landscape. Through engaging exercises and thought-provoking questions, you will be challenged to not only understand AI but also to consider its role in your own future.

The Central Board of Secondary Education (CBSE) recognizes the transformative power of Artificial Intelligence (AI) and its impact on the future. Building upon this successful introduction, CBSE extended the AI subject to Class XI, starting in the 2020-2021 academic session. Thus, allowing students to delve deeper into the world of AI and develop a more comprehensive understanding.

This AI Curriculum has been created with the help of teacher advisors managed by 1M1B and supported by IBM. This curriculum aligns with industry standards as set forth by the National Skills Qualification Framework (NSQF) at Levels 3 & 4.

CBSE acknowledges and appreciates the valuable contribution of IBM India in developing the AI curriculum and conducting training programs. This collaborative effort ensures educators are well-equipped to deliver the AI curriculum effectively.

By working together, CBSE and its partners aim to empower students to embrace the future. By incorporating AI into their learning experience, students gain the knowledge and skills necessary to not only understand AI but also leverage its potential to enhance their learning and future prospects.

The future is full of possibilities, and AI is poised to play a pivotal role. Are you ready to be a part of it?

Embrace the challenge. Explore the potential. Shape the future with Artificial Intelligence.

Contents

UNIT 1: Introduction: Artificial Intelligence for Everyone

Title: Introduction: AI for Everyone	**Approach**: Example based learning**,** Hands- on activities, Discussion

Summary:

This unit covers various aspects of Artificial Intelligence (AI), including its definition, evolution, types, domains, terminologies, and applications. It explains the fundamental concepts of AI, such as supervised learning, natural language processing (NLP), computer vision etc. Additionally, it delves into machine learning (ML) and deep learning (DL) and discusses their differences, types, and applications. The content also outlines the benefits and limitations of AI, addressing concerns such as job displacement, ethical considerations, explainability, and data privacy.

Learning Objectives:

1. Understand the basic concepts and principles of Artificial Intelligence.
2. Explore evolution of AI and identify the different types of AI.
3. Learn about the domains of AI, such as statistical data, natural language processing, and computer vision.
4. Understand the terminologies associated with AI, including machine learning, deep learning, and reinforcement learning.

Key Concepts:

1. What is Artificial Intelligence?
2. Evolution of AI
3. Types of AI
4. Domains of AI
5. AI Terminologies
6. Benefits and limitations of AI

Learning Outcomes:

Students will be able to -

1. Communicate effectively about AI concepts and applications in written and oral formats.
2. Describe the historical development of AI.
3. Differentiate between various types and domains of AI, including their applications.
4. Recognize the key terminologies and concepts related to machine learning and deep learning.
5. Formulate informed opinions on the potential benefits and limitations of AI in various contexts.

Pre-requisites: Reasonable fluency in English language and basic computer skills

Sparking Curiosity: A Teacher's Guide to an Engaging AI Curriculum Warm-Up

This lesson equips you to launch your AI curriculum with a captivating activity that ignites student enthusiasm.

1. Technology Habits - A Revealing Look:

- **Engaging Warm-Up:** Begin with an interactive activity. Ask students a series of questions about their daily technology habits:
    - How often do they use smartphones?
    - Do they utilize voice assistants like Siri or Alexa?
    - Have they played AI-powered games?

This discussion sparks curiosity and reveals the prevalence of AI in everyday life.

2. A Journey Through AI History:

- **Time Travel Through AI Evolution:** Provide a captivating overview of AI's journey:
    - Discuss key milestones in AI history, from early philosophical concepts to modern advancements.
    - Introduce influential figures who shaped the field of AI.
    - Highlight groundbreaking achievements that propelled AI forward.

Encourage critical thinking by prompting students to analyze:

- How AI has impacted society throughout history.
- Possible future implications of AI's continued development.

3. Demystifying AI Categories:

- **Unlocking the AI Landscape:** Introduce the different categories of AI:
    - Explain the characteristics and functionalities of each category (e.g., Narrow AI, General AI, Artificial Superintelligence).
    - Discuss real-world applications of each type to solidify understanding.

Spark a discussion by posing questions about:

- The potential future development of AI and its potential impact on various sectors.
- Ethical considerations surrounding different AI categories.

4. Unveiling the Three AI Domains:

- **The AI Ecosystem:** Introduce the three fundamental domains of AI:
    - Natural Language Processing (NLP): How machines understand and process human language.
    - Computer Vision: How machines extract meaning from visual data (images, videos).

    - Statistical Data: The crucial fuel powering AI systems for learning and development.

Engage students by discussing:

- Real-world examples from each domain (e.g., NLP in chatbots, computer vision in self-driving cars, using data for healthcare AI).
- How these domains work together to create sophisticated AI systems.

5. A Hierarchy of Intelligence:

- **AI, Machine Learning, Deep Learning: A Ladder of Complexity:** Introduce the concepts of AI, Machine Learning (ML), and Deep Learning (DL) as successive subsets.
- Explain how each level builds upon the previous one, ultimately aiming to replicate human-like intelligence in machines.

6. Machine Learning Demystified:

- **Exploring Machine Learning Types:** Guide students through various types of Machine Learning:
    - **Supervised Learning:** Machines learn from labeled data to make predictions (e.g., spam filtering).
    - **Unsupervised Learning:** Machines identify patterns and relationships in unlabeled data (e.g., customer segmentation).
    - **Reinforcement Learning:** Machines learn through trial and error, receiving rewards for desired actions (e.g., playing games against themselves).

Encourage exploration by asking students to identify real-world applications for each type of Machine Learning.

7. A Balanced Perspective: Benefits and Challenges of AI:

- **Weighing the Impact:** Discuss the advantages AI offers:
    - Increased efficiency in various tasks.
    - Enhanced decision-making capabilities.
    - Fueling innovation across diverse fields.

However, foster balanced thinking by addressing AI limitations:

- Potential job displacement due to automation.
- Ethical considerations surrounding bias and fairness in AI systems.
- Challenges in explainability and understanding how AI reaches decisions.
- Data privacy concerns and potential misuse of personal data.

Encourage critical analysis by facilitating a discussion on these benefits and challenges, prompting students to consider the future of AI from a responsible and balanced perspective.

What is Artificial Intelligence (AI)?

Teachers can ask the following questions:

- Think about your daily life. Can you name any technologies you use that might involve AI?
- Imagine a machine that can learn and improve on its own. What kind of tasks or problems do you think such a machine could help us with?

Artificial Intelligence (AI), has evolved drastically over the years, touching various aspects of our lives. It is a technology that has not only fascinated us but also significantly impacted how we live, work, and interact with the world around us. Within the vast landscape of AI, there exist several distinct Domains of Artificial Intelligence, each with its unique characteristics and applications. According to Statista, the global AI market, with a value of billion 113.60 GBP in 2023, is on a continuous growth trajectory, primarily fueled by substantial investments.

Artificial intelligence (AI) refers to the ability of a machine to learn patterns and make predictions.

In its simplest form, Artificial Intelligence is a field that combines computer science and robust datasets to enable problem-solving. AI does not replace human decisions; instead, AI adds value to human judgment. Think of AI as a smart helper that can understand things, learn from examples, and do tasks on its own without needing to be told exactly what to do each time. For example, AI can:

- **Understand Language:** AI can understand and respond to what you say, like virtual assistants such as Siri or Alexa.
- **Recognize Images:** AI can look at pictures and recognize what is in them, like identifying animals in photos.
- **Make Predictions:** AI can analyze data to make predictions, like predicting the weather or suggesting what movie you might like to watch next.
- **Play Games:** AI can play games and learn to get better at them, like playing chess or video games.
- **Drive Cars:** AI can help cars drive themselves by sensing the road and making decisions to stay safe.

What is not AI?

When we talk about machines, not all of them are considered Artificial Intelligence (AI). Here are some examples:

- **Traditional Rule-Based Systems**: These machines follow set rules without learning from data.
- **Simple Automation Tools**: Basic tools like timers or calculators do specific tasks but do not think or learn.
- **Mechanical Devices**: Machines like pulleys or gears work based on physics but do not learn or think.
- **Fixed-Function Hardware**: Devices like microwave ovens perform tasks without learning or thinking.
- **Non-Interactive Systems**: Machines that do not change based on new information, like a basic electric fan.
- **Basic Sensors**: Sensors collect data but do not analyze or understand it.

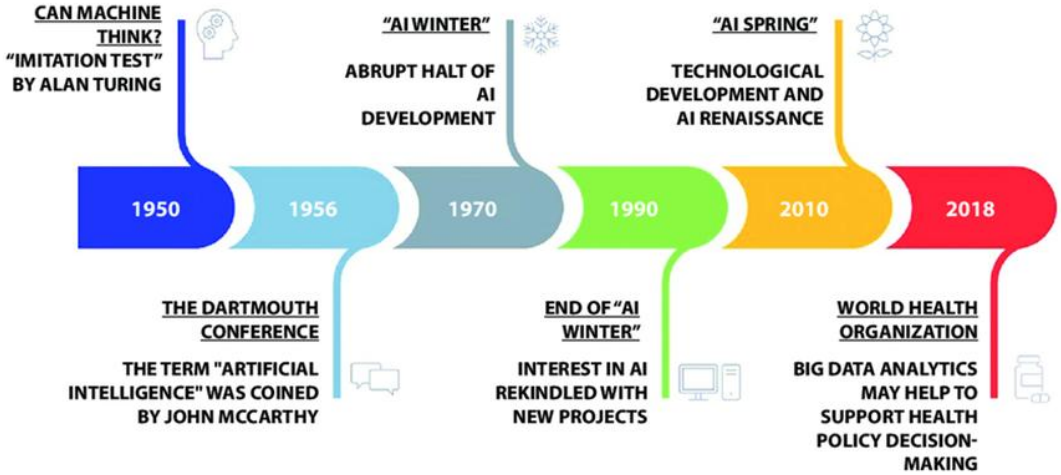
Artificial Intelligence machines are different. They learn from data and can make decisions on their own. For example, a smart washing machine can adjust its settings based on what it is washing. AI goes beyond just following rules; it can learn, adapt, and make decisions based on data and context.

Evolution of AI

Teachers can ask the following questions:

- **Can you think of any stories or myths from history that involve machines or artificial beings acting intelligently?** (This question prompts students to make connections between historical ideas of artificial intelligence and the modern concept.)
- **Imagine you could create a machine that could do one thing really well. What task would you choose and why?** (This question taps into students' own ideas about what intelligence means and how it could be applied in technology.)

The history of AI can be traced back to ancient times, with philosophical discussions about the nature of intelligence and the possibility of creating artificial beings. However, the modern era of AI began in the mid-20th century with significant developments and milestones:



Source:https://www.researchgate.net/figure/Timeline-diagram-showing-the-history-of-artificial-intelligence_fig1_364826401

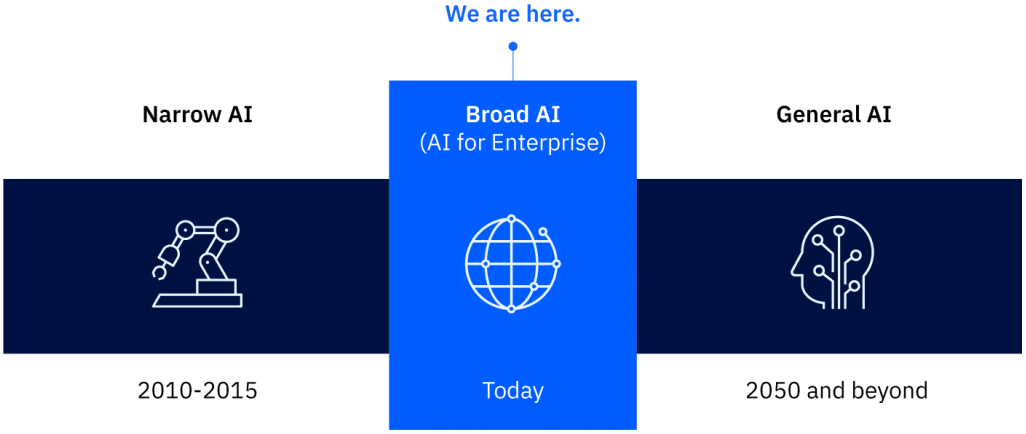
Time Period	Key Events and Developments
1950	1950 was a landmark year for the question of machine intelligence because of Alan Turing's famous paper "Computing Machinery and Intelligence." In this paper, Turing proposed a thought experiment called the "imitation game" (later known as the Turing test).
1956	Dartmouth Conference was organized by McCarthy that marked birthplace of AI as a field. The term "Artificial Intelligence" was coined by John McCarthy. McCarthy, along with Turing, Minsky, and Simon, laid foundation for AI.
1960-1970	Significant progress in AI research that led to the development of expert systems, early neural networks, exploration of symbolic reasoning and problem-solving techniques.
1980-1990	Mixed optimism and skepticism about AI with breakthroughs in machine learning, neural networks led to "AI winter".
21st Century	Resurgence of interest and progress in AI with advancements in computing power, data availability, and algorithmic innovation. Also, there were breakthroughs in machine learning, deep learning, reinforcement learning. That led to transformative applications of AI in healthcare, finance, transportation, and entertainment.

Types of AI

Teachers can ask the following questions:

- **Imagine you have a robot helper at home. What kind of tasks do you think it would be most helpful with? Why?** (This question gets students thinking about specific tasks AI can perform, priming them for the concept of Narrow AI).
- **Can you think of any examples of machines that seem to get "smarter" over time? What makes them seem that way?** (This question encourages students to consider the development and growth of AI capabilities, introducing the idea of different levels of AI).

Computer scientists have identified three levels of AI based on predicted growth in its ability to analyze data and make predictions.



1. Narrow AI:
    - Focuses on single tasks like predicting purchases or planning schedules.
    - Rapidly growing in consumer applications, such as voice-based shopping and virtual assistants like Siri.
    - Capable of handling specific tasks effectively, but lacks broader understanding.
2. Broad AI:
    - Acts as a midpoint between Narrow and General AI.
    - More versatile than Narrow AI, capable of handling a wider range of related tasks.
    - Often used in businesses to integrate AI into specific processes, requiring domain-specific knowledge and data.
3. General AI:
    - Refers to machines that can perform any intellectual task a human can.
    - Currently, AI lacks abstract thinking, strategizing, and creativity like humans.
    - Artificial Superintelligence (ASI) may emerge in the future, potentially leading to self-aware machines, but this is far from current capabilities.

Types of Data

Teachers can ask the following questions:

- **Can you think of all the different ways we share information every day? For example, when you send a message, take a photo, or post on social media, what kind of information are you creating? (**This question encourages students to think about the variety of data they generate daily and sets the stage for discussing how different types of data (structured, unstructured, and semi-structured) play roles in our digital lives.)
- **Imagine you are organizing your digital photos. Would it be easier if they were automatically sorted by date, location, or who is in the picture? Why might this kind of organization be helpful?** (This question helps students understand the concept of structured versus unstructured data by drawing a parallel between organized (structured) and unorganized (unstructured) information. It prepares them for learning how different types of data require different methods for sorting and analysis.)

Data might be facts, statistics, opinions, or any kind of content that is recorded in some format. This could include voices, photos, names, and even dance moves! It surrounds us and shapes our experiences, decisions, and interactions.

It is often referred to as the "**new oil**" of the 21st century. Did you know? 90% of the world's data has been created in just the last 2 years, compared to the previous 6 million years of human existence.

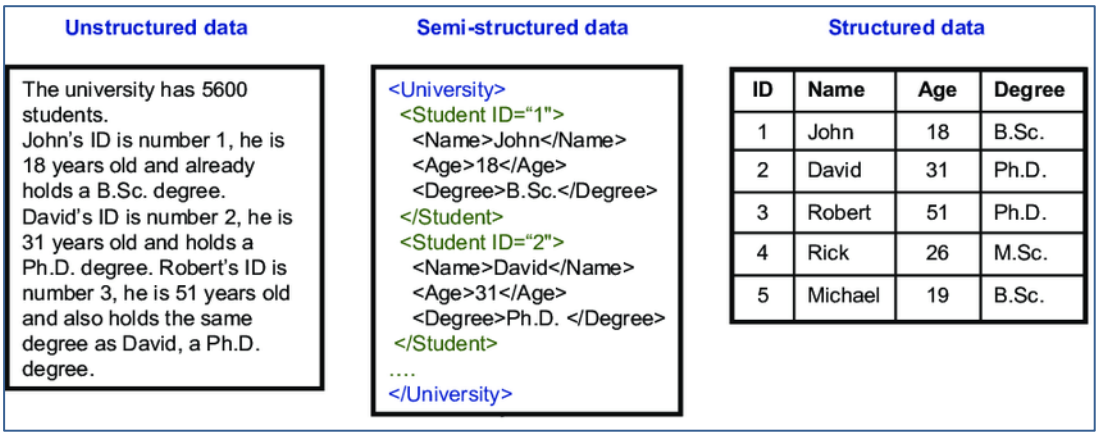
Data can be classified into three different types which are as follows:

- Structured Data
- Unstructured Data
- Semi-structured Data

Structured data is like a neatly arranged table, with rows and columns that make it easy to understand and work with. It includes information such as names, dates, addresses, and stock prices. Because of its organized nature, it is straightforward to analyze and manipulate, making it a preferred format for many data-related tasks.

On the other hand, **unstructured data** lacks any specific organization, making it more challenging to analyze compared to structured data. Examples of unstructured data include images, text documents, customer comments, and song lyrics. Since unstructured data does not follow a predefined format, extracting meaningful insights from it requires specialized tools and techniques.

Semi-structured data falls somewhere between structured and unstructured data. While not as organized as structured data, it is easier to handle than unstructured data. Semi-structured data uses metadata to identify certain characteristics and organize data into fields, allowing for some level of organization and analysis. An example of semi-structured data is a social media video with hashtags used for categorization, blending structured elements like hashtags with unstructured content like the video itself.



Source: https://www.researchgate.net/figure/Unstructured-semi-structured-and-structured-data_fig4_236860222

Domains of AI

Teachers can ask the following questions:

- **Have you ever wondered how your phone can suggest the next word while you're texting or how streaming platforms recommend movies you might like? What kind of 'intelligence' do you think is behind these features?** (This question prompts students to think about everyday AI applications, setting the stage to introduce the concepts of Statistical Data and Natural Language Processing (NLP) in AI.)

- **Imagine a computer that can 'see' what's in a picture or 'hear' and understand your voice commands. What kinds of tasks could this ability help with, and where do you think it's already being used?** (This question helps students envision the possibilities of Computer Vision and NLP, leading naturally into a discussion of how AI uses visual and language data to perform complex tasks like facial recognition, autonomous driving, and virtual assistants.)

Artificial Intelligence (AI) encompasses various fields, each focusing on different aspects of replicating human intelligence and performing tasks traditionally requiring human intellect. These fields are classified based on the type of data input they handle:

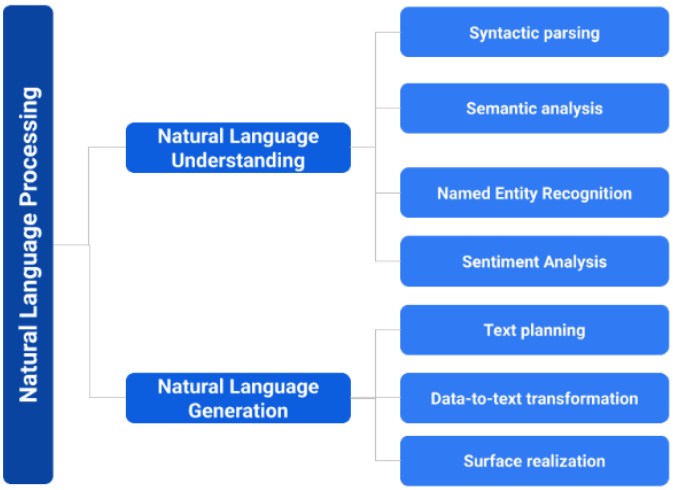
a) **Statistical Data**: Statistical data includes numerical, categorical, and alphanumeric inputs. This domain involves the collection, analysis, and interpretation of large datasets to extract insights and patterns using statistical methods, machine learning algorithms, and data visualization techniques. Examples include:

- **Search recommendations and Google Maps history**, which are based on your previous interactions and location data.
- **Amazon's personalized recommendations**, which are influenced by your shopping habits and browsing history.
- **Social media activity, cloud storage, and digital textbooks**, all of which generate and store vast amounts of data that can be analyzed for insights.

b) **Natural Language Processing (NLP):** NLP focuses on processing text and speech inputs to enable computers to understand, interpret, and generate human language. It involves tasks such as language translation, sentiment analysis, text summarization, and speech recognition, facilitating communication between humans and machines through natural language interfaces.

The goal of an NLP-Trained model is to be capable of “understanding” the contents of documents, including the slangs, sarcasm, inner meaning, and contextual definitions of the language in which the text was written.

Differences Between NLP, NLU, and NLG:



Source: https://www.baeldung.com/cs/natural-language-processing-understanding-generation

Natural Language Processing (NLP): This is the broad umbrella term encompassing everything related to how computers interact with human language. Think of it as the "what" - what computers can do with human language. *It is like the whole library - filled with different tools and techniques for working with language data.*

Natural Language Understanding (NLU): This is a subfield of NLP that focuses on understanding the meaning of human language. It analyzes text and speech,

extracting information, intent, and sentiment. NLU helps computers understand the language and what it means. *Imagine finding a specific book in the library.*

Natural Language Generation (NLG): This is another subfield of NLP, but instead of understanding, it focuses on generating human language. It takes structured data as input and turns it into coherent and readable text or speech. *Think of this as writing a new book based on the information gathered in the library.*

c) **Computer Vision:** Computer Vision deals with visual data inputs, primarily images and videos. It enables computers to interpret and understand visual information, and perform tasks such as object detection, image classification, facial recognition, and scene understanding, enabling applications such as autonomous vehicles, medical imaging, and augmented reality.

Computer Vision is like giving computers the ability to see and understand the world through digital images and videos, much like how humans use their eyes to perceive their surroundings. In this domain, computers analyze visual information from images and videos to recognize objects, understand scenes, and make decisions based on what they "see."

When we take a digital image, it is essentially a grid of tiny coloured dots called **pixels**. Each pixel represents a tiny portion of the image and contains information about its colour and intensity.

Resolution is expressed as the total number of pixels along the width and height of the image. For example, an image with a resolution of 1920x1080 pixels has 1920 pixels horizontally and 1080 pixels vertically. Higher resolution images have more pixels, providing more detail.

Now, here's where AI comes in. To make sense of these images, computers convert them into numbers. They break down the image into a series of numbers that represent the colour and intensity of each pixel. This numerical representation allows AI algorithms to process the image mathematically and extract meaningful information from it. For instance, AI algorithms might learn to recognize patterns in these numbers that correspond to specific objects, like cars or faces. By analyzing large amounts of labeled image data, AI systems can "learn" to identify objects accurately.

Activity:

Divide the students into groups and provide them with a list of real-world applications without specifying which domain each application belongs to. Ask each group to categorize the applications into the three domains: Statistical Data, Natural Language Processing (NLP), and Computer Vision.

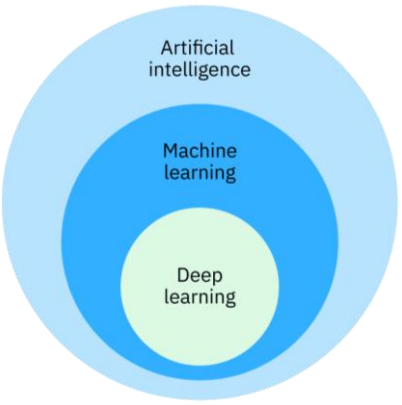
1. *Gesture recognition for human-computer interaction*
2. *Chatbots for customer service*
3. *Spam email detection*
4. *Autonomous drones for surveillance*
5. *Google Translate*

6. *Fraud detection in financial transactions*
7. *Augmented reality applications (e.g., Snapchat filters)*
8. *Sports analytics for performance optimization*
9. *Object detection in autonomous vehicles*
10. *Recommendation systems for e-commerce platforms*
11. *Customer segmentation for targeted marketing*
12. *Text summarization for news articles*
13. *Automated subtitles for videos*
14. *Medical image diagnosis*
15. *Stock prediction*

Statistical Data	Natural Language Processing	Computer Vision
6. Fraud detection in financial transactions	3. Spam email detection	1.Gesture recognition for human-computer interaction
8.Sports analytics for performance optimization	2.Chatbots for customer service	4.Autonomous drones for surveillance
10.Recommendation systems for e-commerce platforms	5. Google Translate	7.Augmented reality applications (e.g., Snapchat filters)
11.Customer segmentation for targeted marketing	12. Text summarization for news articles	9.Object detection in autonomous vehicles
15.Stock prediction	13. Automated subtitles for videos	14.Medical image diagnosis

AI Terminologies

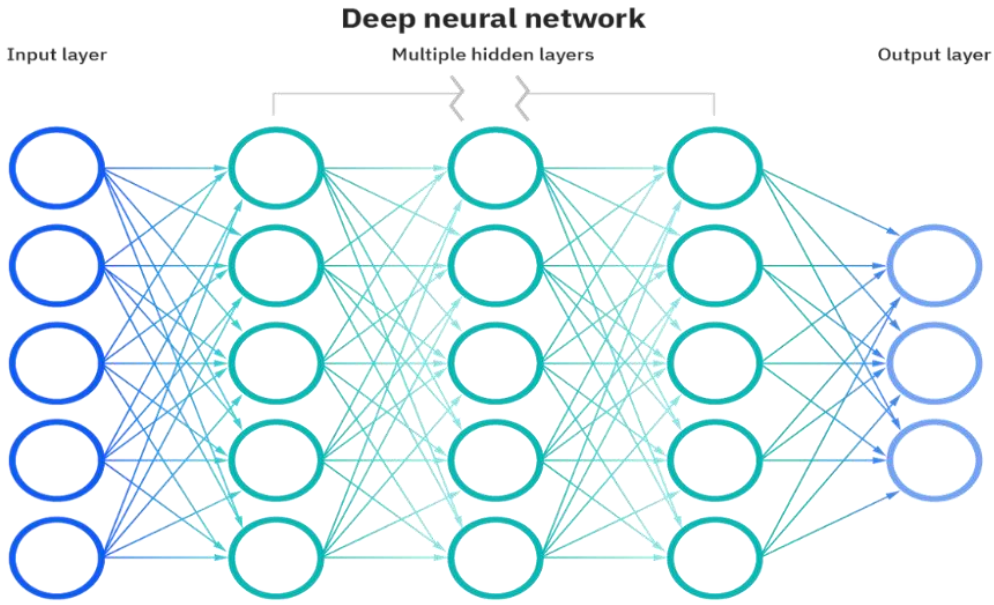
Teachers can ask the following questions:



- **Can you think of any machines that seem really smart, even though they aren't alive? For example, a self-driving car or a program that recommends movies you might like.** (This question activates prior knowledge and gets students thinking about machines that might be considered "intelligent")
- **How do humans learn new things? Do we always need someone to tell us exactly what to do, or can we also figure things out on our own by looking at examples?** (This question helps connect the concept of machine learning to familiar human experiences)
- **Artificial intelligence** machines don't think. They calculate. They represent some of the newest, most sophisticated calculating machines in human history. It is a computer system that can perform tasks that ordinarily require human intelligence or human interference.
- Some can perform what is called **machine learning** as they acquire new data. Machine learning is a subset of artificial intelligence (AI) that focuses on developing algorithms

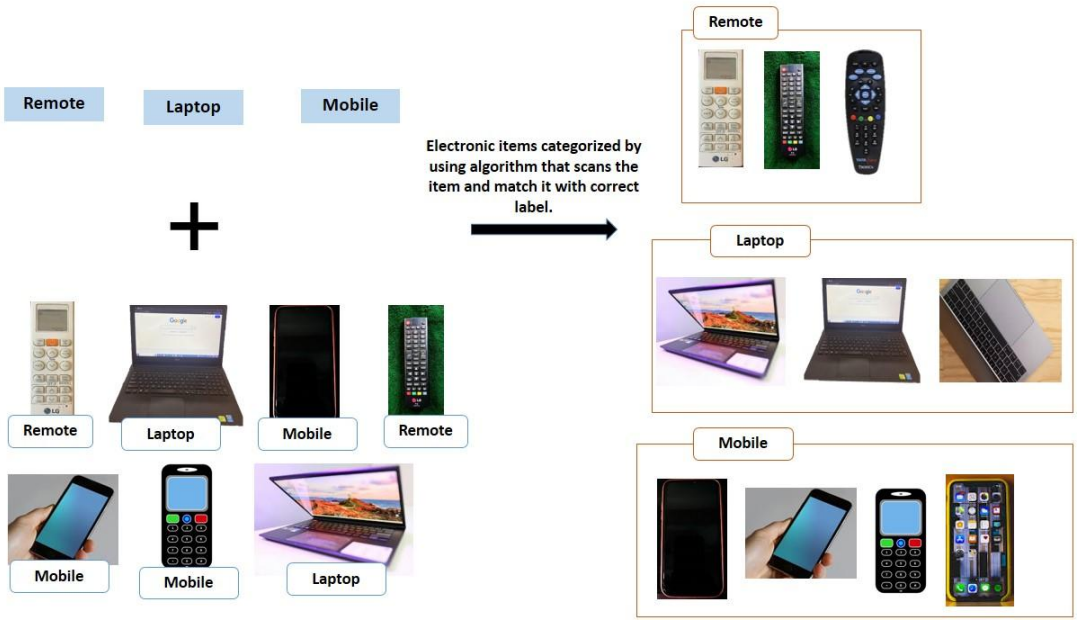
and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed.

- Others, using calculations arranged in ways inspired by neurons in the human brain, can even perform **deep learning** with multiple levels of calculations. Deep learning is an AI function that imitates the working of the human brain in processing data and creating patterns for use in decision making.
    - The structure of Deep Learning is inspired by the structure of the neurons and neuron connections in the human brain.
    - Neural networks, also known as Artificial Neural Networks (ANNs), are a subset of Machine Learning and the core heart and concept of Machine Learning.
    - They comprise of node layers, containing an input layer, one or multiple hidden layers, and an output layer.
    - If the output of any node is above a specified threshold, that node is activated, sending data to the next layer of the network.
    - Otherwise, no data is passed along to the next layer of the network.
    - If the number of Layers including the Input and Output Layer is more than three, then it is called a Deep Neural Network.

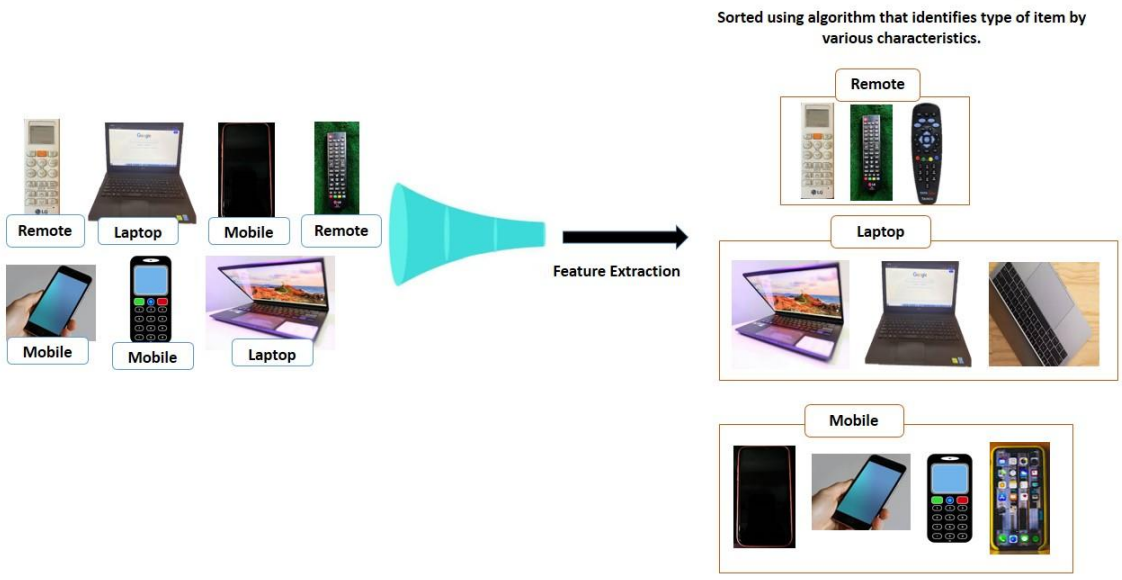


MACHINE LEARNING	DEEP LEARNING
1. Works on small dataset for accuracy	1. Works on Large dataset
2. Dependent on Low-end machine	2. Heavily dependent on high-end machine
3. Divides the tasks into sub-tasks, solves them individually and finally combine the results	3. Solves problem end to end
4. Takes less time to train	4. Takes longer time to train
5. Testing time may increase	5. Less time to test the data

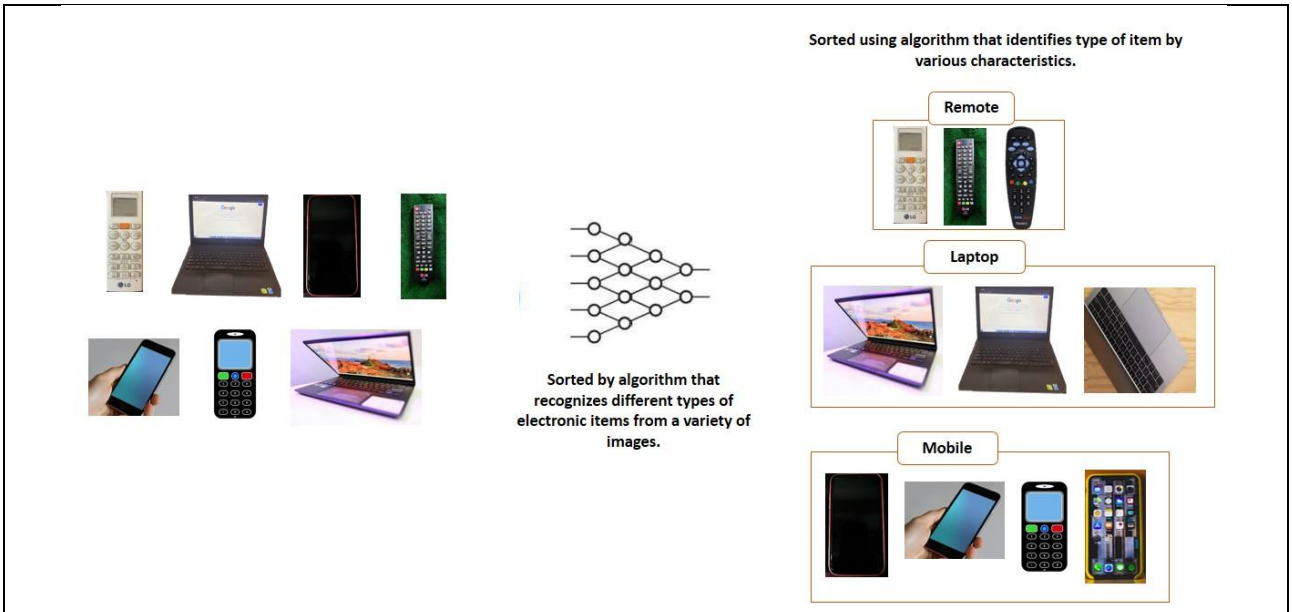
Example: Imagine you are given the job to sort items in the electronic items department at an electronic store. You realize that there are dozens of products and very less time to sort them manually. How will you use artificial intelligence, machine learning, and deep learning to help with your work?



By training the model with labeled data, AI can match the visual characteristics of the electronic items with the correct category, significantly speeding up the sorting process and reducing human error.

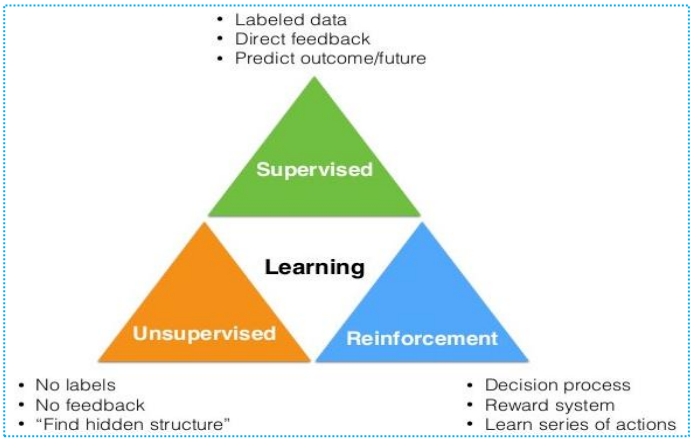


To improve the performance of the machine, you expose it to more data to ensure that the machine is trained on numerous characteristics of each type of electronic items, such as size, shape, and colour. The more data you provide for the algorithm, the better the model gets. By providing more data and adjusting parameters, the machine minimizes errors by repetitive guess work.

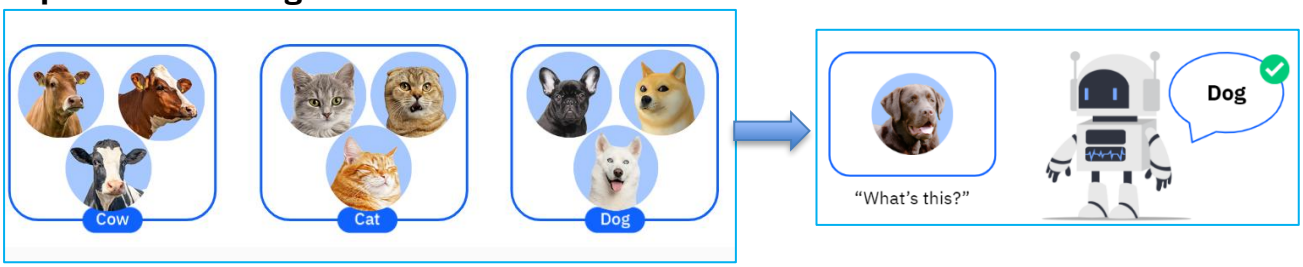


Deep learning models eliminate the need for feature extractions. Decide the algorithms based on deep learning to sort electronic items by removing the need to define what each product looks like. Feature extraction is built into the process without human input. Once you have provided the deep learning model with dozens of electronic items pictures, it processes the images through different layers of neural networks. The layers can then learn an implicit representation of the raw data on their own.

Types of Machine Learning



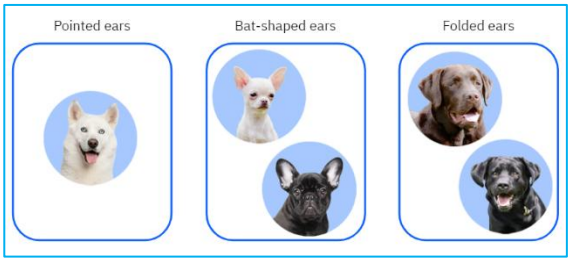
Supervised learning

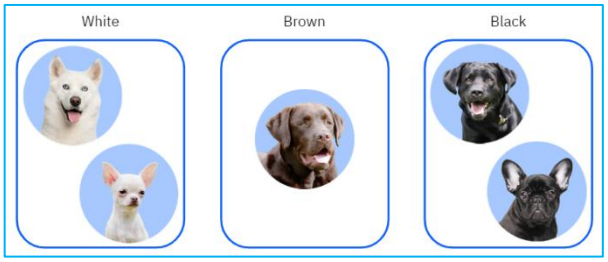


- Supervised learning is a type of machine learning where the model learns from labelled data, which means that the input data is accompanied by the correct output.
- In supervised learning, the algorithm learns to map input data to output labels based on example input-output pairs provided during the training phase.

- The goal of supervised learning is to learn a mapping function from input variables to output variables so that the model can make predictions on unseen data.
- Examples of supervised learning algorithms include linear regression, logistic regression, decision trees, support vector machines (SVM), and neural networks.

Unsupervised Learning:

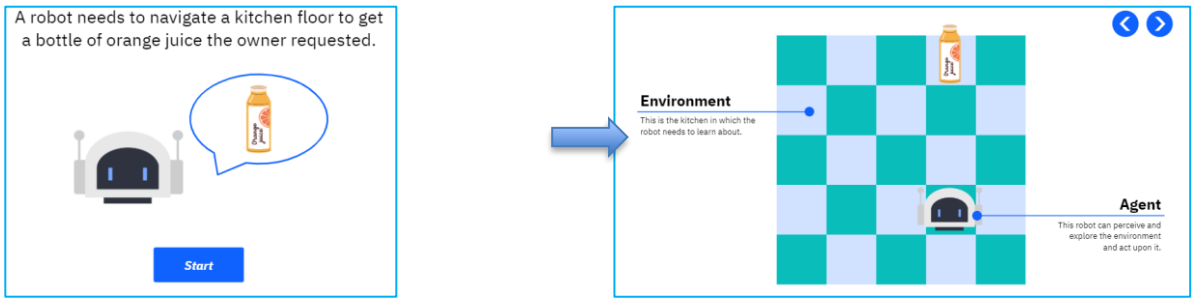




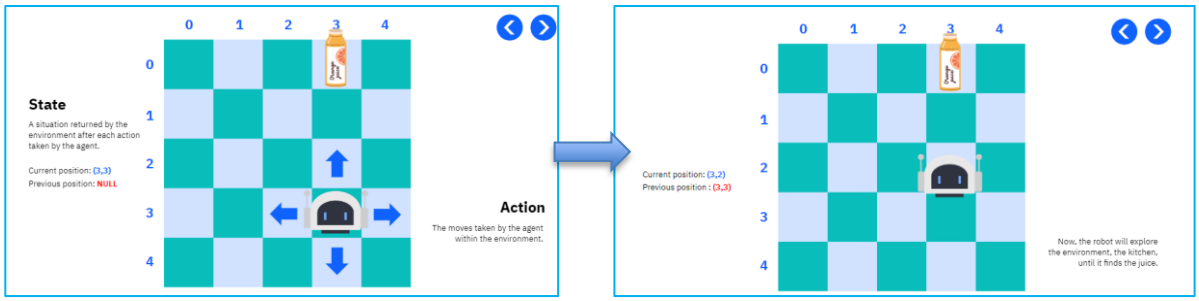
- Unsupervised learning is a type of machine learning where the model learns from unlabelled data, which means that the input data is not accompanied by the correct output.
- In unsupervised learning, the algorithm tries to find hidden patterns or structure in the input data without explicit guidance.
- The goal of unsupervised learning is to explore and discover inherent structures or relationships within the data, such as clusters, associations, or anomalies.
- Examples of unsupervised learning algorithms include k-means clustering, hierarchical clustering, principal component analysis (PCA), and autoencoders.

Reinforcement Learning:

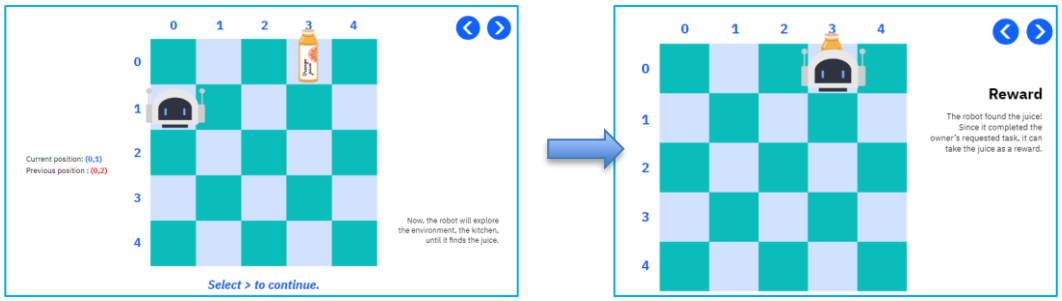
1.



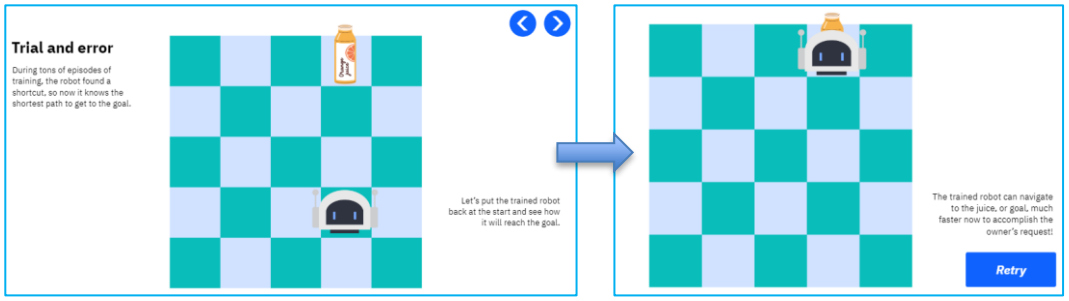
2.



3.



4.



- Reinforcement learning is a type of machine learning where an agent learns to make decisions by *interacting with an environment* to maximize cumulative rewards.
- In reinforcement learning, the agent learns through *trial and error* by taking actions and receiving feedback from the environment in the form of rewards or penalties.
- The goal of reinforcement learning is to learn a policy or strategy that guides the agent to take actions that lead to the highest cumulative reward over time.
- Reinforcement learning is commonly used in scenarios where the agent must make a sequence of decisions over time, such as playing games, controlling robots, or managing financial portfolios.
- Examples of reinforcement learning algorithms include Q-learning, deep Q-networks (DQN), policy gradients, and actor-critic methods.

Benefits and limitations of AI

Teachers can ask the following questions:

Benefits of AI:

- Imagine you're a scientist researching a new disease. How do you think AI could help you analyze massive amounts of data to find patterns and potential cures?

Limitations of AI:

- Self-driving cars are becoming a reality. However, some people worry that AI might not be able to handle unexpected situations. What are some challenges you see with AI making decisions in critical situations?

Benefits:

- **Increased efficiency and productivity:** AI automates tasks, analyzes data faster, and optimizes processes, leading to increased efficiency and productivity across various sectors.
- **Improved decision-making:** AI analyzes vast amounts of data and identifies patterns that humans might miss, assisting in data-driven decision-making and potentially leading to better outcomes.

- **Enhanced innovation and creativity:** AI tools can generate new ideas, explore possibilities, and automate repetitive tasks, freeing up human resources for more creative pursuits and innovation.
- **Progress in science and healthcare:** AI aids in drug discovery, medical diagnosis, and personalized medicine, contributing to advancements in healthcare and scientific research.

Limitations:

- **Job displacement:** Automation through AI raises concerns about job displacement and the need for workforce retraining and upskilling.
- **Ethical considerations:** Concerns exist around bias in AI algorithms, potential misuse for surveillance or manipulation, and the need for ethical guidelines and regulations.
- **Lack of explainability:** Some AI models, particularly complex ones, lack transparency in their decision-making, making it difficult to understand how they arrive at their outputs.
- **Data privacy and security:** Large-scale data collection and use for AI development raise concerns about data privacy and security vulnerabilities.

Earn a credential on IBM SkillsBuild on the topic Artificial Intelligence Fundamentals using the link:

https://students.yourlearning.ibm.com/activity/PLAN-CC702B39D429

- *Semantris*, is an NLP-Based game by Google based on Word association powered by semantic search.
  https://experiments.withgoogle.com/semantris
- This is a game built with machine learning. We **draw**, and a neural network tries to guess what you're **drawing**.
  https://quickdraw.withgoogle.com/
- *The experiment based on the computer vision domain of AI. It identifies that you draw and suggests the related images. To play the game, visit the following link on any computing device with speakers.*
  https://www.autodraw.com/

Extension Activities:

These activities provide opportunities for students to explore various aspects of artificial intelligence, develop critical thinking skills, and engage in hands-on learning experiences in the classroom.

1. **AI in the News:** Have students research recent news articles or stories related to artificial intelligence? They can explore topics such as AI advancements, ethical dilemmas, or AI applications in various industries. Students can then present their findings to the class and facilitate discussions on the implications of these

developments.

2. **AI Applications Showcase:** Divide students into small groups and assign each group a specific AI application or technology (e.g., virtual assistants, self-driving cars, healthcare diagnostics). Ask students to research and create presentations or posters showcasing how their assigned AI technology works, its benefits, potential drawbacks, and real-world examples of its use.
3. **AI Coding Projects:** Introduce students to basic coding concepts and tools used in AI development, such as Python programming language and machine learning libraries like TensorFlow or scikit-learn. Guide students through hands-on coding projects where they can build simple AI models, such as image classifiers or chatbots. Encourage experimentation and creativity in designing and training their AI systems.
4. **AI Film Analysis:** Screen and analyze films or documentaries that explore themes related to artificial intelligence, such as "Ex Machina," "Her," "I, Robot," or "The Social Dilemma." After watching the films, facilitate discussions on how AI is portrayed, its potential impact on society, and ethical considerations raised in the narratives.

EXERCISE

A. Multiple-choice questions (MCQs):

1. Who is often referred to as the "Father of AI"?
    a. Alan Turing
    b. John McCarthy
    c. Marvin Minsky
    d. Herbert A. Simon
2. In which year was the term "Artificial Intelligence" first used by John McCarthy?
    a. 1930
    b. 1955
    c. 1970
    d. 2000
3. What does the term "Data is the new oil" imply?
    a. Data is as valuable as oil.
    b. Data is used as fuel for machines.
    c. Data is a non-renewable resource.
    d. Data and oil are unrelated.
4. Divya was learning neural networks. She understood that there were three layers in a neural network. Help her identify the layer that does processing in the neural network.
    a. Output layer
    b. Hidden layer
    c. Input layer
    d. Data layer

5. Which category of machine learning occurs in the presence of a supervisor or teacher?
    a. Unsupervised Learning
    b. Reinforcement Learning
    c. Supervised Learning
    d. Deep Learning
6. What does Deep Learning primarily rely on to mimic the human brain?
    a. Traditional Programming
    b. Artificial Neural Networks
    c. Machine Learning Algorithms
    d. Random Decision Making
7. What is the role of reinforcement learning in machine learning?
    a. Creating rules automatically
    b. Recognizing patterns in untagged data
    c. Rewarding desired behaviors and/or penalizing undesirable ones
    d. Mimicking human conversation through voice or text
8. Which AI application is responsible for automatically separating emails into "Spam" and "Not Spam" categories?
    a. Gmail
    b. YouTube
    c. Flipkart
    d. Watson

B. Fill in the Blanks:

1. To determine if a machine or application is AI-based, consider its ability to perform tasks that typically require _____ intelligence.
2. Artificial intelligence (AI) enables a machine to carry out cognitive tasks typically performed by _____.
3. Supervised, unsupervised, and reinforcement learning are three categories of _____.
4. _____ is a subset of artificial intelligence that is entirely based on artificial neural networks.
5. Machine learning can be used for online fraud detection to make cyberspace a _____ place.

C. True or False:

1. Chatbots like Alexa and Siri are examples of virtual assistants.
2. Supervised learning involves training a computer system without labeled input data.
3. Unstructured data can be easily analyzed using traditional relational database techniques.
4. Deep learning typically requires less time to train compared to machine learning.
5. Machine learning is not used in everyday applications like virtual personal assistants and fraud detection.

D. Short Answer Questions:

1. How is machine learning related to AI?
2. Define machine learning.
3. What do you mean by Reinforcement Learning? Write any two applications of Reinforcement Learning at School.
4. How do you understand whether a machine/application is AI based or not? Explain with the help of an example.

E. Long Answer Questions:

1. Define Data. Briefly explain the types of data.
2. What is deep learning, and how does it differ from traditional machine learning?

F. Case-study/Application Oriented Questions:

A hospital implemented an AI system to assist doctors in diagnosing diseases based on medical images such as X-rays and MRI scans. However, some patients expressed concerns about the accuracy and reliability of the AI diagnoses. How can the hospital address these concerns?

ANSWERS

A. Multiple-choice questions (MCQs):

1. b. John McCarthy 2. b. 1955 3. a. Data is as valuable as oil
4. b. Hidden layer 5. c. Supervised Learning 6. b. Artificial Neural Networks
7. c. Rewarding desired behaviors and/or penalizing undesirable ones
8. a. Gmail

B. Fill in the Blanks:

1. human-like 2. Humans 3. Machine Learning 4. Deep Learning 5. Secure

C. True or False:

1. True 2. False 3. False 4. False 5. False

D. Short Answer Questions:

1. How is machine learning related to AI?

Ans. Machine learning enables machines to learn, forecast, and improve on their own, contributing to the broader field of AI.

2. Define machine learning.

Ans. Machine learning is defined as the science of getting computers to act without being explicitly programmed, and its primary categories include supervised, unsupervised, and reinforcement learning.

3. What is deep learning, and how does it differ from traditional machine learning?

Ans. Deep learning is a subset of machine learning entirely based on artificial neural networks, distinguished by its ability to solve end-to-end problems and its heavy reliance on high-end machines for computation.

4. What do you mean by Reinforcement Learning? Write any two applications of Reinforcement Learning at School.

Ans. Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. Two applications in schools include adaptive learning systems that personalize content and educational games/simulations that engage students in interactive learning experiences.

5. How do you understand whether a machine/application is AI based or not? Explain with the help of an example.

Ans. To understand whether a machine/application is AI based or not, we need to check if it learns with data and whether it's able to decide/predict.

E. Long Answer Questions:

1. Define Data. List the types of data.

Ans. Data refers to raw facts, figures, or information collected from various sources, which can be processed, analyzed, and used to draw conclusions or make decisions. Data can be in various forms and can be processed by both humans and machines.

Types of Data:

1. Structured Data: Data that is highly organized, following a predefined schema, and stored in a tabular format (e.g., rows and columns in databases).
   Examples: Spreadsheets, databases, names, ages, addresses, sales data.
2. Unstructured Data: Data that does not have a specific format or structure, making it difficult to organize in traditional databases.
   Examples: Text files, videos, audio recordings, social media content, emails, images.
3. Semi-Structured Data: Data that does not conform to a rigid structure but still contains tags or markers to separate elements and enforce hierarchies.
   JSON files, XML files, emails (with headers and body), social media posts with metadata.

2. What is Deep Learning, and How Does It Differ from Traditional Machine Learning?

Deep learning is an AI function that imitates the working of the human brain in processing data and creating patterns for use in decision-making. It is based on artificial neural networks (ANNs) and involves multiple layers of processing to extract features and learn from large datasets. Deep learning can handle complex data such as images, audio, and unstructured text, allowing for end-to-end learning without the need for manual feature extraction.

Differences Between Deep Learning and Traditional Machine Learning:

Feature	Traditional Machine Learning	Deep Learning
Dataset Size	Works well with small datasets	Requires large datasets
Computational Power	Can run on low-end machines	Heavily dependent on high-end machines (GPUs/TPUs)
Approach	Divides tasks into sub-tasks and combines results	Solves problems end-to-end
Training Time	Takes less time to train	Takes longer to train
Testing Time	May increase testing time	Generally faster testing once trained

In traditional machine learning, models often rely on manual feature extraction, where human experts define which features of the data are important for the model to learn. In contrast, deep learning models automatically learn these features through layered neural networks, which makes them particularly effective for complex data types such as images and natural language

F. Case-study/Application Oriented Questions:

A hospital implemented an AI system to assist doctors in diagnosing diseases based on medical images such as X-rays and MRI scans. However, some patients expressed concerns about the accuracy and reliability of the AI diagnoses. How can the hospital address these concerns?

Ans. The hospital can address these concerns by conducting thorough validation studies to assess the accuracy and reliability of the AI system compared to human diagnoses. They can also ensure transparency by providing detailed information about how the AI system works and how it complements the expertise of human doctors.

G. Competency Based Questions

1.Rahul is an architect. He has designed and built a beautiful home for his client in Pune. He has installed these systems/appliances/gadgets at the newly constructed home. Identify which of these are AI systems and which of these are not AI systems.

Solar water heater, Smart TV, Security cameras, rainwater harvesting system, cleaning robots, smart lighting, automatic door, Siri, automatic washing machine.

After separating the AI systems, mention some parameters on which you choose these appliances/systems as AI systems.

Ans: The AI systems are: Smart TV, cleaning robots, smart lighting, Siri.

The systems which are not AI systems are: Solar water heater, security cameras, rainwater harvesting systems, automatic door, automatic washing machine.

The parameters are: ability to make decisions, problem-solving, recommendations, adapt to new situations, and learn from past experiences. (any other AI feature can be included).

2.If you were designing a robot to sort recyclable items like glass, plastic, and paper, which type of learning would be used to help the robot.
Ans: Supervised learning

3.Can you think of a scenario where you have a bunch of different fruits mixed together and you want the computer to organize them into groups based on similarities?
Ans: Unsupervised learning

4.Mr. Shankar owns a company that deals with services to customers related to financial investments. Lately, he has been using AI technology in his company due to which his employees are facing less job responsibility, customers are feeling insecure about their data. What is this scenario known as?
Ans: Potential impact of AI on society

5.Jatin is a student who has just enrolled in a course in AI. He attended a few introductory classes and learned that systems can learn from the data using algorithms to perform a task without explicitly programming it. In some situations, the system mimics the human brain's learning process. Identify the concepts in this scenario.
Ans: Machine learning and Deep learning

REFERENCES
Videos to watch

- Understand about AI : https://youtu.be/ad79nYk2keg?feature=shared
- History of AI : https://www.youtube.com/watch?v=L-9ZqkeNNJE
- Natural Language Processing: https://www.youtube.com/watch?v=CMrHM8a3hqw
- Introduction to Computer Vision: https://www.youtube.com/watch?v=puB-4LuRNys

UNIT 2: Unlocking your Future in AI

Title: Unlocking your Future in AI	**Approach:** Team Discussion, Web search

Summary: This lesson explores the global demand for artificial intelligence (AI) professionals, highlighting the diverse career opportunities available across various industries. It discusses common job roles in AI, essential skills and tools for prospective AI careers, and opportunities for AI professionals in different sectors. Additionally, it provides a curated list of resources for individuals interested in exploring AI further and staying updated with the latest developments in the field.

Learning Objectives:

1. Understand the increasing demand for AI professionals in today's global market.
2. Identify common job roles in the field of artificial intelligence and their respective responsibilities.
3. Recognize the essential skills and tools required for a successful career in AI.
4. Explore the diverse opportunities for AI professionals across various industries.
5. Discover curated resources for further learning and staying updated in the field of AI.

Key Concepts:

1. The Global Demand
2. Some Common Job Roles In AI
3. Essential Skills and Tools for Prospective AI Careers
4. Opportunities in AI Across Various Industries

Learning Outcomes:

Students will be able to:

1. Articulate the demand for AI professionals and the diverse career opportunities available in the field.
2. Identify the requisite skills and tools needed to pursue a career in artificial intelligence.
3. Understand the potential roles and responsibilities of AI professionals across different industries.
4. Explore resources for further learning and skill development in the field of AI.
5. Evaluate their own interests and skills to determine potential pathways for a career in AI.

Pre-Requisites:

Basic understanding of artificial intelligence concepts and applications, familiarity with programming languages such as Python, and interest in exploring career opportunities in the field of artificial intelligence.

Integrating the Future: A Teacher's Guide to AI and the Evolving Workforce

This lesson equips you to guide students through the dynamic relationship between AI and the future of work.

1. Staying Ahead of the Curve:

- **Continuous Learning:** Emphasize the importance for teachers themself to stay updated on the evolving landscape of AI techniques. This ensures you can provide students with the most current information. Explore online resources, attend workshops, or collaborate with industry professionals to stay informed.

2. AI and Jobs: A Shifting Landscape:

- **AI's Impact on Employment:** Discuss how AI may transform the job market:
    - Identify existing jobs that may be automated or significantly impacted by AI.
    - Explore how AI might create new job opportunities requiring different skill sets.

Focus on a Positive Outlook:

- Frame the discussion to emphasize that AI is more likely to create a job transformation than widespread unemployment.
- Encourage students to develop skills that complement AI, such as:
    - Critical thinking and problem-solving.
    - Creativity and innovation.
    - Collaboration and interpersonal skills.
    - Adaptability and lifelong learning.

3. AI Professionals in Action:

- **Bringing AI to Life:** Organize group discussions or role-plays to showcase the diverse roles of AI professionals in various departments:
    - **AI Engineers:** Design, develop, and maintain AI systems. (Role-play: A team brainstorming the design of an AI-powered customer service chatbot.)
    - **Data Scientists:** Collect, analyse, and interpret data to train AI models. (Group discussion: How data scientists ensure data quality and address potential bias in AI systems.)
    - **AI Ethicists:** Ensure the responsible and ethical development and use of AI. (Role-play: A team discussing ethical considerations for using AI in the healthcare industry.)

These activities provide students with a deeper understanding of the various career paths within the AI field and how they contribute to the overall success of AI projects.

Additional Tips:

- Invite guest speakers from the AI industry to share their experiences and insights.
- Encourage students to research specific AI applications in their areas of interest.
- Guide students in exploring online resources and tools for learning more about AI.

By incorporating these elements, you can equip students with the knowledge and skills to navigate the evolving job market and thrive in a future shaped by AI.

THE GLOBAL DEMAND

Teachers can ask the following questions:

1. **What are some examples of artificial intelligence (AI) that you encounter in your daily lives?** This gets the students thinking about AI and its real-world applications, making the passage more relevant.
2. **Imagine a future where AI is even more integrated into society. What kind of jobs do you think will be in high demand?** This question activates their imagination and taps into their existing understanding of jobs. It also creates a nice bridge into the passage which highlights the variety of AI career paths.

Artificial intelligence (AI) was once confined to the realms of science fiction, but today, it permeates our daily lives in ways we often take for granted. From personalized recommendations on streaming platforms to the algorithms powering autonomous vehicles, AI has become an indispensable part of the modern society. As the field continues to evolve and expand, so do the opportunities it presents for career growth and development.

Amidst the concerns about automation and job displacement, it is essential to recognize the significant demand for AI professionals across various industries. While it is true that AI technologies may replace some traditional roles, they also create a multitude of new and exciting career paths. *Rather than viewing AI as a threat, individuals should embrace it as an opportunity for advancement and innovation*.

Image Source: https://media.licdn.com

The global demand for AI talent is skyrocketing, driven by the rapid pace of technological advancements and the increasing integration of AI solutions into diverse sectors. From healthcare and finance to transportation and retail, organizations are harnessing the power of AI to streamline operations, optimize processes, and deliver enhanced services to consumers. One of the most significant advantages of pursuing a career in AI is the sheer breadth of opportunities it offers. Whether you are passionate about machine learning, natural language processing, robotics, or data analytics, there is a niche within the AI field suited to your interests and skills. Moreover, as AI technologies continue to mature, new specialties and job roles are emerging, creating avenues for specialization and expertise.

SOME COMMON JOB ROLES IN AI:

In today's market, there is a wide range of job roles within the field of artificial intelligence (AI) that are in high demand. Some common job roles include:

- **Machine Learning Engineer**: Machine learning engineers bridge software engineering and data science, utilizing big data tools and programming frameworks

to develop scalable data science models capable of handling vast volumes of real-time data. Strong mathematical skills, experience in machine learning and deep learning, and proficiency in programming languages like Java, Python, and Scala are essential for success in this role.

- **Data Scientist**: Data scientists leverage machine learning and predictive analytics to extract insights from large datasets, to take proper business decisions. Proficiency in big data platforms like Hadoop, Pig, and Spark, fluency in programming languages such as SQL, Python, and Scala, and a solid understanding of descriptive and inferential statistics are the key requirements for this role.
- **Business Intelligence Developer**: Business intelligence (BI) developers design, model, and maintain complex data sets to analyse business and market trends, enhance organizational profitability and efficiency. Strong technical and analytical skills, along with expertise in data warehouse design and BI technologies, are essential for success in this role.
- **Robotics Engineer**: They design and maintain AI-powered robots, develop mechanical devices capable of performing tasks with human commands. Proficiency in programming, along with expertise in disciplines like robotic engineering, mechanical engineering, and electrical engineering, is crucial for success in this field.
- **Software Engineer**: AI software engineers build and maintain software products for AI applications, staying updated on the latest artificial intelligence technologies. Proficiency in software engineering, programming languages, and statistical analysis is essential, typically requiring a bachelor's degree in computer science, engineering, or related fields.

- **Natural Language Processing (NLP) Engineer**: NLP engineers specialize in human language processing, working on voice assistants, speech recognition, and document processing. A specialized degree in computational linguistics or a combination of computer science, mathematics, and statistics is typically required for this role.
- **Computer Vision Engineer**: Computer vision engineers specialize in developing algorithms and systems that enable computers to analyse and interpret visual information from images or videos. Their expertise lies in creating software solutions that can understand and process visual data, requiring proficiency in image processing techniques and programming languages such as Python and C++.
- **AI Ethicist**: AI ethicists address ethical considerations and implications related to the development and deployment of AI technologies, ensuring that they are used

responsibly and ethically. They provide guidance on ethical frameworks, policies, and practices to promote fairness, transparency, and accountability in AI systems, often requiring a background in ethics, philosophy, or law, combined with expertise in AI technology.

- **AI Consultant**: AI consultants offer expert guidance and advice to organizations on how to leverage AI technologies to solve business challenges and drive innovation. They assess business needs, identify opportunities for AI integration, and develop strategic AI initiatives, requiring a deep understanding of AI technologies, business processes, and industry trends, along with strong communication & analytical skills.

Activity 1:

Divide the class into small groups and distribute the list of AI job roles to each group. Using the roles written in the chit, the teams will identify ten companies currently hiring employees for those specific AI positions.

AI Consultant	NLP Engineer	AI Ethicist
IBM, Deloitte, Accenture,	Google, Amazon, Microsoft	EFF, Amnesty International
Ernst & Young (EY), KPMG	Facebook, Apple, IBM	MIT Media Lab, EARL, CDT
Cognizant, Capgemini	NVIDIA, OpenAI, Twitter	IBM, Google, OpenAI

ESSENTIAL SKILLS AND TOOLS FOR PROSPECTIVE AI CAREERS

Teachers can ask the following questions:

1. **Based on the different AI job roles we discussed, what technical skills do you think might be important across most of them?** This question helps bridge the gap between the different job roles and the specific skills needed. It gets them thinking about the practical application of the knowledge they just learned.
2. **Can you think of any tools or software you already use that might be helpful in an AI career?** This taps into their existing knowledge and encourages them to make connections between familiar tools and the world of AI. It can also lead to an interesting discussion about the accessibility of AI tools.

A successful career in artificial intelligence requires a diverse set of skills that encompass both technical expertise and soft skills. According to industry leaders, here are some of the top skills that AI professionals need:

Technical Skills:

- Expertise in neural networks, machine learning, and deep learning is essential for developing advanced AI applications.
- Knowledge of big data technologies and techniques for handling and analysing large datasets is crucial in AI applications.

- Understanding of frameworks and libraries like TensorFlow, SciPy, and NumPy is vital for building and deploying AI solutions.
- Familiarity with programming languages such as Python, R, Java, and C++ is necessary for developing AI models and algorithms.
- Proficiency in linear algebra, probability, statistics, and signal processing is essential for understanding the mathematical principles underlying AI algorithms.

Soft Skills:

- Effective communication skills are crucial for conveying complex technical concepts to non-technical stakeholders and collaborating with multidisciplinary teams.
- Strong teamwork and collaboration abilities are essential for working effectively in cross-functional teams to develop AI products and solutions.
- Problem-solving, decision-making, and analytical thinking skills are critical for identifying and addressing challenges in AI projects.
- Time management and organizational skills are essential for managing multiple projects and meeting deadlines.
- Business intelligence and critical thinking skills are valuable for understanding business requirements and translating them into AI solutions that deliver tangible value.

Your Professional Toolkit

In addition to acquiring the necessary skills, it is essential for AI professionals to familiarize themselves with popular AI tools, platforms, and programming languages. Here are some essential tools and their purposes:

- **Python**: A versatile programming language with pre-made libraries for advanced computing and scientific computation.
- **R**: A programming language for data collection, organization, and analysis, particularly useful for machine learning and statistical functions.
- **Java**: Widely used in AI for implementing intelligence programming, neural networks, and machine learning solutions.
- **C++**: Known for its flexibility and object-oriented functions, used for procedural programming and hardware manipulation in AI.
- **TensorFlow**: An open-source machine learning platform with tools and libraries for developing sophisticated AI applications.
- **SciPy and NumPy**: Python libraries for scientific computing and mathematical operations, ideal for manipulating and visualizing data.

By acquiring the right combination of technical skills and tools, aspiring AI professionals can position themselves for success in this dynamic and rapidly growing field. Whether you are interested in developing AI algorithms or implementing AI solutions in real-world applications, building a strong foundation of skills and expertise is the key to unlocking exciting career opportunities in artificial intelligence.

Activity 2:

In continuation with the previous activity, list the technical skills and soft skills listed by any two companies for the specific AI position.

Technical Skills	Soft Skills	Link to the website
1. Proficiency in programming languages such as Python, Java, or C++. 2. Experience with machine learning frameworks such as TensorFlow, PyTorch, or scikit-learn. 3. Knowledge of natural language processing (NLP) techniques and libraries. 4. Familiarity with deep learning architectures and algorithms.	1. Strong problem-solving abilities and analytical thinking. 2. Excellent communication skills, both verbal and written. 3. Collaboration and teamwork, including the ability to work effectively in interdisciplinary teams. 4. Adaptability to changing requirements and environments.	(https://www.ibm.com/employment/)
1. Proficiency in programming languages such as Python, Java, or R. 2. Strong understanding of machine learning algorithms and statistical modeling techniques. 3. Experience with machine learning frameworks such as TensorFlow, PyTorch, or scikit-learn. 4. Knowledge of natural language processing (NLP) techniques and libraries like NLTK or SpaCy.	1. Excellent problem-solving abilities and analytical thinking. 2. Strong communication skills, both verbal and written. 3. Collaboration and teamwork, including the ability to work effectively in interdisciplinary teams. 4. Attention to detail and commitment to quality.	https://careers.cognizant.com/global/en)

OPPORTUNITIES IN AI ACROSS VARIOUS INDUSTRIES

Teachers can ask the following questions:

1. **Think about the different industries we discussed earlier. Can you imagine how AI might be used to improve or revolutionize any of them?** This question gets them thinking about the broad applicability of AI and encourages them to make connections between the technology and the real world.

2. **What are your interests or what subjects do you enjoy in school? How do you think AI could be used in those fields in the future?** This question personalizes the topic and gets them thinking about potential career paths at the intersection of their interests and AI.

Artificial intelligence professionals design and develop AI systems that use machine learning and neural networks to predict trends, provide better customer experiences and recommendations, and offer solutions to difficult problems. While some AI professionals work towards the goal of General AI—systems interconnected and able to be nearly as creative as human beings—others focus on narrower applications. This following table gives you a variety of opportunities to choose from depending upon your choice of subject.

Industry	Employment Opportunities	Some Existing/Expected Job Roles	Relevant Subjects in School
Automobile	Design, manufacturing, and sale of motor vehicles.	1. Autonomous Vehicle Engineer: Develops AI algorithms for self-driving cars. 2. Simulation Engineer: Creates virtual environments for testing autonomous vehicle technologies. 3. Robotics Engineer: Designs AI-powered robots for automotive tasks.	Mathematics, Physics, Computer Science/Artificial Intelligence
Agriculture	Monitoring crop health, optimizing irrigation, and maximizing yields.	1. Precision Agriculture Specialist: Uses AI-powered drones and sensors for monitoring crops. 2. Crop Yield Prediction Analyst: Forecasts crop yields using AI models. 3. Livestock Monitoring Specialist: Tracks the health and productivity of farm animals.	Biology, Mathematics, Computer Science/Artificial Intelligence
Retail	Optimizing inventory, sales forecasting, and enhancing customer experience.	1. Inventory Management Specialist: Optimizes inventory levels using AI algorithms. 2. Sales Forecasting Analyst: Forecasts sales using AI models. 3. Customer Experience Designer: Enhances customer experience using AI-driven insights.	Business Studies, Mathematics, Computer Science/Artificial Intelligence
Media	Creating visual effects, content generation, and audience analysis.	1. Visual Effects Artist: Uses AI tools for creating visual effects. 2. Content Creator: Generates content using AI-generated insights. 3. Audience Analyst: Analyzes audience behavior using AI algorithms.	Fine Arts, Media Studies, Computer Science/Artificial Intelligence
Information Technology	Developing AI algorithms, systems, and	1. Machine Learning Engineer: Develops AI algorithms and systems.	Computer Science/Artificial Intelligence,

	infrastructure for various applications.	2. AI Software Developer: Builds AI-powered applications. 3. AI Infrastructure Specialist: Maintains and optimizes AI infrastructure.	Mathematics, Physics
Healthcare	Medical imaging analysis, personalized healthcare, and drug discovery.	1. Medical Imaging Analyst: Analyzes medical images using AI algorithms. 2. Virtual Nurse Assistant: Provides personalized healthcare recommendations. 3. Drug Discovery Researcher: Identifies potential drug candidates using AI.	Biology, Chemistry, Computer Science/Artificial Intelligence
Finance	Market analysis, fraud detection, risk management, and investment recommendations.	1. Quantitative Analyst: Analyzes market trends using AI algorithms. 2. Fraud Detection Analyst: Identifies fraudulent activities using AI models. 3. Financial Advisor: Offers personalized investment recommendations using AI-driven analytics.	Economics, Mathematics, Computer Science/Artificial Intelligence
Government & Military	Surveillance, predictive analytics, citizen services, and military technologies.	1. National Security Analyst: Uses AI-powered surveillance systems. 2. Defense Contractor: Develops AI-enabled military technologies. 3. Government AI Specialist: Implements AI for citizen services and regulatory compliance.	Political Science, Computer Science/Artificial Intelligence, Mathematics
Tourism	Personalized travel recommendations, customer service, and itinerary planning.	1. Travel Recommendation Engine Developer: Provides personalized travel recommendations. 2. Chatbot for Customer Service: Assists travelers with booking and inquiries. 3. Smart Travel Itinerary Planner: Optimizes travel routes and schedules.	Geography, Business Studies, Computer Science/Artificial Intelligence
Beauty & Wellness	Skincare analysis, virtual styling, and wellness guidance.	1. AI-powered Skincare Assistant: Provides personalized skincare recommendations. 2. Virtual Hair Stylist: Simulates different hairstyles using AI. 3. Wellness Chatbot: Offers guidance on nutrition and fitness.	Chemistry, Biology, Computer Science/Artificial Intelligence
Banking	Loan approval automation, fraud detection, and personalized financial advice.	1. Loan Approval Specialist: Automates loan approval process using AI. 2. Fraud Detection Analyst: Identifies fraudulent transactions using AI algorithms. 3. Financial Advisor: Offers personalized financial advice using AI-driven analytics.	Economics, Mathematics, Computer Science/Artificial Intelligence

Geospatial	Spatial data analysis, remote sensing, and mapping technologies.	1. Geographic Information Systems (GIS) Specialist: Analyzes spatial data using AI. 2. Remote Sensing Analyst: Interprets satellite imagery using AI algorithms. 3. Mapping Technician: Uses AI-enabled drones for mapping.	Geography, Geology, Computer Science/Artificial Intelligence
Textile	Fabric design, quality control, and inventory management.	1. AI-powered Fabric Design Specialist: Creates innovative textile patterns using AI. 2. Textile Quality Control Inspector: Ensures product quality using AI-enabled systems. 3. Smart Inventory Management Specialist: Optimizes inventory levels using AI algorithms.	Chemistry, Art & Design, Computer Science/Artificial Intelligence
Design	Design optimization, user experience enhancement, and content creation.	1. Generative Design Assistant: Optimizes design solutions using AI algorithms. 2. AI-powered UX Designer: Enhances user experience using AI-driven insights. 3. AI-powered Content Creator: Generates content using AI tools.	Art & Design, Computer Science/Artificial Intelligence, Mathematics
Sales & Marketing	Campaign automation, customer segmentation, and sales forecasting.	1. Marketing Campaign Automation Specialist: Automates marketing campaigns using AI. 2. Customer Segmentation Analyst: Segments customers based on behavior using AI. 3. Sales Forecasting Analyst: Forecasts sales using AI models.	Business Studies, Mathematics, Computer Science/Artificial Intelligence
Fashion	Personalized fashion recommendations, trend analysis, and virtual try-on.	1.AI-powered Fashion Stylist: Recommends personalized clothing combinations using AI. 2. Trend Analyst: Analyzes fashion trends using AI algorithms. 3. Virtual Clothing Try-on Specialist: Allows virtual try-on of clothing using AI and AR.	Fashion Design, Mathematics, Computer Science/Artificial Intelligence

ADDITIONAL LEARNING RESOURCES:

Here are some resources you can explore, bookmark, and keep in mind if you would like to explore more about AI and stay in touch with the latest developments in the field. This is a curated listing. There are many organizations and websites to explore, depending on your interests.

News and blogs to stay current

- Analytics Insight offers insights, latest news, and a magazine featuring opinions and views of top industry leaders and executives who share their journeys, experiences, success stories, and knowledge to grow profitable businesses.
- Towards Data Science is an online publication in which independent authors who follow their rules and guidelines can publish their work, share their knowledge and expertise, and engage a wide audience on Medium.
- KDnuggets is a leading site on data science, machine learning, AI, and analytics. It contains excellent tutorial materials, courses, webinars, online events.
- Data Science Central is a leading online resource for data practitioners. From statistics and analytics to machine learning and AI, Data Science Central provides a community experience that includes a rich editorial platform, social interaction, forum-based support, and the latest information on technology, tools, trends, and careers.
- Datanami is a news portal dedicated to providing insight, analysis, and up-to-the-minute information about emerging trends and solutions in big data.

Free learning opportunities to build skills

Note: *You will need to sign up for a free account for the following online learning opportunities.*

- You can take advantage of IBM SkillsBuild to power your future in tech with job skills, courses, digital credentials, and more.
- Kaggle offers free online micro courses to help you gain the skills you need to do independent data science projects. Kaggle also allows you to grow your data science and machine learning skills by competing in Kaggle competitions. Find help in the documentation or learn about Community Competitions.
- Udemy offers a variety of free video-based courses on artificial intelligence, including a short, practical hands-on course on artificial intelligence, called Kickstart Artificial Intelligence. Udemy also offers a course called Artificial Intelligence: Preparing Your Career for AI, which covers what you should be doing now to prepare for the coming of AI.
- freeCodeCamp.org offers a rundown of All the Math You Need to Know in Artificial Intelligence. Jason Dsouza gives you an overview of the core math principles you need to focus on to work in AI.
- DataCamp offers a free, two-hour Machine Learning for Everyone course which introduces machine learning without coding involved.
- W3Schools is the world's largest web developer site that offers a variety of free online tutorials with hands-on practice. The site includes tutorials on some popular data science programming languages, such as Python, R, and SQL.
- Codecademy offers free coding classes on 12 different programming languages including Python, Java, and C++.

Additional information regarding colleges offering professional course in AI.

- ✓ IIT Madras – Four year Bachelor of Science Degree in Data Science and Applications https://study.iitm.ac.in/ds/
- ✓ AICTE – All India Council for Technical Education's unique website – "Digital Skilling". Explore this site for a wide variety of course and internships. https://1crore.aicte-india.org/
- ✓ Most of the top colleges in India now offer B.Tech courses in AI and Ml, Data Science, Robotics and Computer Science with specializations. Students can visit college websites to know more about these courses. Also, some colleges are offering BSc in AI and ML.

REFERENCES : IBM Skills Build

EXERCISES:

A. Multiple Choice Questions

1. Which of the following is a job role in AI related to the automobile industry?
   A. Robotics Engineer
   B. Virtual Nurse Assistant
   C. Sales Forecasting Analyst
   D. Autonomous Vehicle Engineer
2. Identify the important soft skill required for AI professionals.
   A. Expertise in neural networks
   B. Effective communication
   C. Proficiency in Python
   D. Knowledge of big data technologies
3. Which industry uses AI for personalized travel recommendations?
   A. Tourism
   B. Banking
   C. Healthcare
   D. Geospatial
4. What is the purpose of the website "Data Science Central"?
   A. Providing a community experience for data practitioners
   B. Offering free video-based courses on AI
   C. Analyzing market trends using AI algorithms
   D. Providing insight into emerging trends in big data
5. Which industry uses AI for market analysis and fraud detection?
   A. Finance
   B. Media
   C. Textile
   D. Design

ANSWERS

1. D. Autonomous Vehicle Engineer
2. B. Effective communication
3. A. Tourism
4. A. Providing a community experience for data practitioners
5. A. Finance

B. Short answer questions:

1. Name some common job roles in the field of artificial intelligence (AI).

Some common Job roles in AI are Machine Learning Engineer, Data Scientist, Business Intelligence Developer, Robotics Engineer, Software Engineer, NLP Engineer, Computer Vision Engineer, AI ethicist, AI consultant

2. What are some essential technical skills required for a successful career in AI, and why are they important?

Some important technical skills required are expertise in neural networks, machine learning, and deep learning, knowledge of big data technologies and techniques, understanding of frameworks and libraries like TensorFlow, SciPy, and NumPy, familiarity with programming languages such as Python, R, Java, and C++, proficiency in linear algebra, probability, statistics, and signal processing.

By acquiring the right combination of technical skills and tools, aspiring AI professionals can position themselves for success in this dynamic and rapidly growing field.

3. What is the role of AI professionals in healthcare, finance, and retail industry?

In Healthcare:

1. Medical Imaging Analyst: Analyzes medical images using AI algorithms.
2. Virtual Nurse Assistant: Provides personalized healthcare recommendations.
3. Drug Discovery Researcher: Identifies potential drug candidates using AI.

In Finance

1. Quantitative Analyst: Analyzes market trends using AI algorithms.
2. Fraud Detection Analyst: Identifies fraudulent activities using AI models.
3. Financial Advisor: Offers personalized investment recommendations using AI-driven analytics.

In Retail Industry

1. Inventory Management Specialist: Optimizes inventory levels using AI algorithms.
2. Sales Forecasting Analyst: Forecasts sales using AI models.
3. Customer Experience Designer: Enhances customer experience using AI-driven insights.

4. List some popular AI tools and programming languages used by AI professionals.

Some of the popular AI tools are Tensor Flow, SciPy and NumPy. Programming Languages used by AI professionals are Python, R, Java, C++.

5. What soft skills do AI professionals need, and how do they help them succeed?

Soft skills needed by AI professionals are:

- Effective communication skills
- Strong teamwork and collaboration
- Problem-solving, decision-making, and analytical thinking skills
- Time management and organizational skills
- Business intelligence and critical thinking skills

Acquiring soft skills helps professional develop the personality and thus work better

6. Why is continuous learning crucial in AI, and how do professionals stay updated with the latest advancements?

Continuous learning is crucial in AI due to the rapid pace of advancements and innovations in the field. Evolution of techniques, new research findings, addressing ethical and social implications have made continuous AI learning crucial.

Professionals stay updated with the latest advancement through continuous education, research and publications, online resources, community involvement, Handson projects.

C. Long answer questions.

1. How does the global demand for AI professionals affect career opportunities in the field?

The global demand for AI talent is skyrocketing, driven by the rapid pace of technological advancements and the increasing integration of AI solutions into diverse sectors. From healthcare and finance to transportation and retail, organizations are harnessing the power of AI to streamline operations, optimize processes, and deliver enhanced services to consumers.

One of the most significant advantages of pursuing a career in AI is the sheer breadth of opportunities it offers. Whether you are passionate about machine learning, natural language processing, robotics, or data analytics, there is a niche within the AI field suited to your interests and skills.

Moreover, as AI technologies continue to mature, new specialties and job roles are emerging, creating avenues for specialization and expertise.

2. What are some common job roles in the field of AI, and how do they contribute to the development and implementation of AI solutions across various industries?

Some common job roles include:

- Machine Learning Engineer: Machine learning engineers bridge software engineering and data science, utilizing big data tools and programming frameworks to develop scalable data science models capable of handling vast volumes of real-time data.
- Data Scientist: Data scientists leverage machine learning and predictive analytics to extract insights from large datasets, informing business decisions.

- Business Intelligence Developer: Business intelligence (BI) developers design, model, and maintain complex data sets to analyse business and market trends, enhancing organizational profitability and efficiency.
- Robotics Engineer: They design and maintain AI-powered robots, develop mechanical devices capable of performing tasks with human commands.
- Software Engineer: AI software engineers build and maintain software products for AI applications, staying updated on the latest artificial intelligence technologies.
- Natural Language Processing (NLP) Engineer: NLP engineers specialize in human language processing, working on voice assistants, speech recognition, and document processing.
- Computer Vision Engineer: Computer vision engineers specialize in developing algorithms and systems that enable computers to analyse and interpret visual information from images or videos.
- AI Ethicist: AI ethicists address ethical considerations and implications related to the development and deployment of AI technologies, ensuring that they are used responsibly and ethically.
- AI Consultant: AI consultants offer expert guidance and advice to organizations on how to leverage AI technologies to solve business challenges and drive innovation.

Competency Based Questions

1.Medha just finished her Class X exams and dreams of a career in AI for Finance. She's unsure what subjects to choose in higher studies (Class XI & XII) to prepare for this future goal. What subjects would you recommend for Medha to take in school, and are there any additional courses she can consider to strengthen her profile?
Ans- Medha can combine her AI aspirations with a finance career. In school, she should focus on Math, Computer Science, and Commerce subjects (Accountancy, Economics, Business Studies). Additionally, online courses on platforms like Kaggle, Udemy, and Data Science programs can provide valuable technical skills and bridge the gap between finance and AI. This combination will equip her for a successful future in AI for Finance.

2.Akash, a recent graduate, just joined the IT and AI startup company as a software developer. He's brilliant at coding but feels a bit nervous about collaborating with his new team. He feels explaining technical concepts to non-technical team members and clearly documenting code could be a challenge. Also, he’s not sure how to effectively collaborate with others and handle situations where there might be disagreements about the best approach. List some soft skills that can help him excel in this environment:

Ans:

- Effective Communication: Being able to explain complex technical ideas in clear and concise language, both verbally and in writing, is crucial.
- Teamwork and Collaboration: Success in an AI startup often depends on working effectively with others. Actively listen to your colleagues, share ideas openly, and be willing to compromise when necessary.
- Problem-solving: AI projects often involve unexpected hurdles. Develop a logical approach to problem-solving, considering different solutions and collaborating with your team to find the best approach.

3. Saloni, a career counselor, is visiting a Class XII assembly to talk about exciting career opportunities in Artificial Intelligence (AI). After explaining various AI fields, a student named Aditya raises his hand and asks, "This all sounds fascinating, but how is AI being used in healthcare specifically?" How will Saloni address Aditya's query?
Ans- Saloni addressed Aditya's query by showcasing how AI is revolutionizing healthcare through tasks like medical image analysis, drug discovery, virtual assistants, robotic surgery, predicting health risks, and automating administrative tasks. This highlights the vast potential of AI careers at the intersection of technology and medicine.

4. Naveen is a student. Help him to differentiate the commonly used AI tools, platforms and programming languages used for the development of AI applications.
Python, GW-BASIC, R, MS-Word, IBM Watson, TensorFlow, MS-Paint, SciPy, Perl.
Ans: Python, R, IBM Watson, TensorFlow and SciPy.

5.Mr. Sourav Shukla is a journalist and owns a media house. He wants to leverage AI in his field. List some employment opportunities and some job roles related to the field of media.
Ans: Visual effects, content generation, and audience analysis. Job roles are: Visual Effect Artist, Content Creator, Audience or Data Analyst, and targeted advertising.

UNIT 3: Python Programming

Title: Python Programming	**Approach**: Group Discussion, Hands on Practice using the software

Summary: This unit will introduce students to the fundamentals/ basics of Python programming language its history, evolution, operators, variables, constants, lists, strings, iterative and select statements. Students will explore three essential Python libraries: NumPy, Pandas, and Scikit-learn. Students will learn how Python is used to create programs. They will also learn how to use NumPy for numerical computing, Pandas for data manipulation and analysis, and Scikit-learn for implementing machine learning algorithms.

Learning Objectives:

Students will be able to

1. Understand the basics of python programming language- tokens, data types, lists, string manipulation, iterative and decision statements.
2. Learn how to use NumPy for mathematical operations and numerical computing.
3. Explore Pandas for data manipulation, analysis, and exploration of structured data.
4. Gain proficiency in using Scikit-learn for implementing machine learning algorithms, including classification.
5. Develop the skills necessary to use Python libraries effectively in Data Science and machine learning projects.

Key concepts:

1. Basics of Python programming language
2. Understanding of character sets, tokens, modes, operators and data types.
3. Control Statements
4. CSV Files
5. Libraries – NumPy, Pandas, Scikit-learn

Learning Outcomes:

Students will be able to –

1. Explain the basics of Python programming language and write programs with basic concepts of tokens
2. Use selective and iterative statements effectively
3. Gain practical knowledge on how to use the libraries efficiently.

Pre-requisites: Reasonable fluency in English language and basic computer skills

Unveiling the Power of Python: A Teacher's Guide

This lesson equips you to introduce students to the wonders of Python, a powerful and versatile programming language.

1. Demystifying Python:

- **Comprehensive Introduction:** Begin with a thorough explanation of the Python language, highlighting its key features:
    - **Easy to Read and Learn:** Python's syntax is clear and readable, making it approachable for beginners.
    - **Versatile:** Python can be used for various tasks, including data science, web development, scripting, and more.
    - **Large Community:** Python boasts a vast and supportive community, offering resources and assistance.

2. Jupyter Notebook: Your Interactive Playground:

- **Showcase Jupyter Notebook:** While you can choose any code editor, demonstrate Jupyter Notebook as an interactive environment for students to learn and experiment with Python code and visualizations.

3. Exploring Development Tools:

- **Introduce Options:** Introduce students to popular development tools like Anaconda, Github, and Colab. These tools offer features for project management, version control, and cloud-based coding environments respectively. However, emphasize their choice of using any preferred editor.

4. Leveraging Learning Resources:

- **Website and Course Links:** Provide a curated list of websites and online courses at the chapter's end to encourage further exploration. Be transparent about potential sign-up requirements for specific resources.

5. Sample Programs for AI Applications:

- **Real-World Focus:** Present sample Python programs that demonstrate its capabilities in AI applications. Explain how these programs can be used as building blocks for future AI projects.

6. Deepening Understanding with Practice:

- **Hands-on Exercises:** Supplement the curriculum with in-lab practical exercises tailored to the chapter's concepts. This reinforces learning and boosts confidence in using Python.
- **Extra Challenges:** Offer additional programs for students to practice independently, pushing their skills further.

7. Expanding the Toolkit:

- **While Loops:** While not explicitly mentioned, you can incorporate While Loops into the curriculum if deemed beneficial for a more comprehensive understanding of control flow.

8. Tuples: A Note on Immutability:

- **Focus and Simplicity:** For a concise lesson, consider omitting Tuples as they share similarities with Lists (but are immutable). You can introduce them later or as an optional topic for advanced students.

9. Dive Deeper with Built-in Functions:

- **Seminar or Activity:** Consider offering a seminar or activity where students explore built-in functions for basic data types (e.g., strings, integers, lists). This encourages independent learning and fosters curiosity.

10. Project-Based Learning:

- **Building with Knowledge:** Assign small projects that utilize Python programs and libraries like Scikit-learn. This project-based approach allows students to apply their knowledge and gain valuable experience.

Remember, this is a guideline. Tailor it to your classroom dynamics and adjust the pace based on students' progress. Encourage exploration, experimentation, and a spirit of inquiry to make learning Python an engaging and rewarding journey for your students.

Introduction to Python

Python is a general-purpose, high level programming language. It was created by Guido van Rossum, and released in 1991. Python got its name from a BBC comedy series – "Monty Python's Flying Circus"

Features of Python

- High Level language
- Interpreted Language
- Free and Open Source
- Platform Independent (Cross-Platform) – runs virtually in every platform if a compatible python interpreter is installed.
- Easy to use and learn – simple syntax similar to human language.
- Variety of Python Editors – Python IDLE, PyCharm, Jupyter, Spyder
- Python can process all characters of ASCII and UNICODE.
- Widely used in many different domains and industries.

Python Editors

There are various editors and Integrated Development Environments (IDEs) that you can use to work with Python. Some popular options are PyCharm, Spyder, Jupyter Notebook, IDLE, Google Colab etc. Let us look at how we can work with Jupyter Notebook.

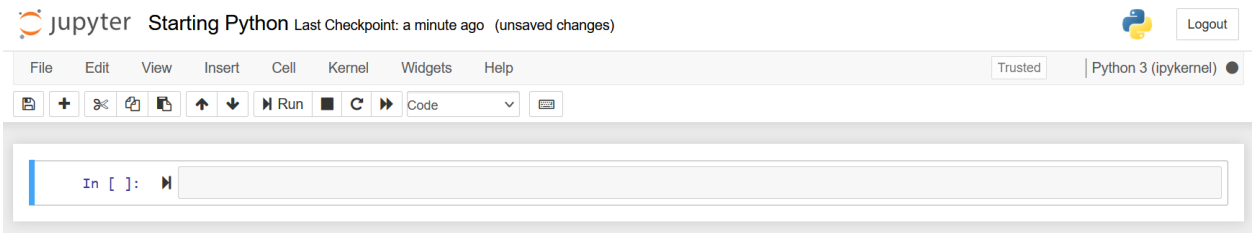
Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text. It's widely used in data science and research. It can be installed using Anaconda or with pip. For more details of installation use the link
https://docs.jupyter.org/en/latest/install/notebook-classic.html
Those who are familiar with Python, open the command prompt in administrative mode and type

```
pip install notebook
```

To run the notebook, Open the command prompt and type

```
jupyter  notebook
```

Following window will open



You can type the code in the cell provided. Then click Run to see the output just below it.

```
In [1]: print ( " Hai all this is how to display a text")

Hai all this is how to display a text

In [4]: #To Add two numbers
num1=10
num2=20
print (num1 + num2)

30
```

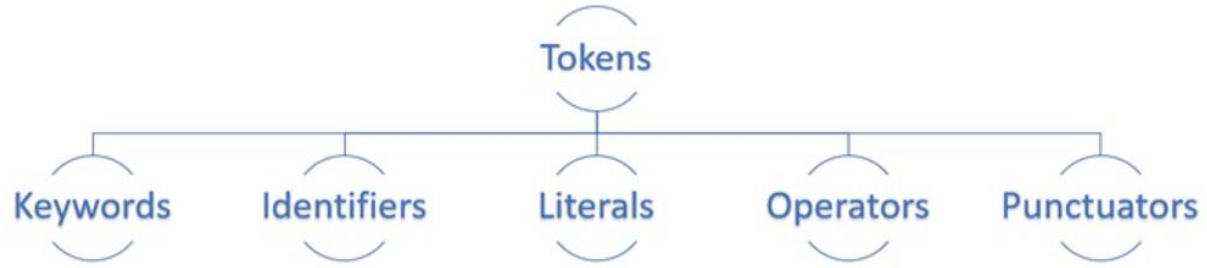
Output – We can work on multiple programs in same window

Point to be Noted

Another way to install and start using Jupyter Notebook is through Anaconda. It is the most widely used Python distribution for data science and comes pre-loaded with all the most popular libraries and tools. Jupyter notebook can easily be accessed using the Anaconda Prompt with the help of a local host.

Getting Started with Python Programs

Python program consists of Tokens. It is the smallest unit of a program that the interpreter or compiler recognizes. Tokens consist of identifiers, keywords, literals, operators, and punctuators. They serve as the building blocks of Python code, forming the syntactic structure that the interpreter parses and executes. During lexical analysis, the Python interpreter breaks down the source code into tokens, facilitating subsequent parsing and interpretation processes.



https://www.studytrigger.com/wp-content/uploads/2022/08/Tokens-in-Python.jpg

Keywords

Reserved words are used for special purposes. List of keywords are given below.

False	None	True	for	in	or	while
and	class	elif	from	is	pass	with
as	continue	else	global	lambda	raise	yield
assert	def	except	if	nonlocal	return	async
break	del	finally	import	not	try	await

Identifier

An identifier is a name used to identify a variable, function, class, module or other object. Generally, keywords (list given above) are not used as variables. Identifiers cannot start with digit and it can't contain any special characters except underscore.

Literals:

Literals are the raw data values that are explicitly specified in a program. Different types of Literals in Python are String Literal, Numeric Literal (Numbers), Boolean Literal (True & False), Special Literal (None) and Literal Collections.

Operators:

Operators are symbols or keywords that perform operations on operands to produce a result. Python supports a wide range of operators:

- Arithmetic operators (+, -, *, /, %)
- Relational operators (==, !=, <, >, <=, >=)
- Assignment operators (=, +=, -=)
- Logical operators (and, or, not)
- Bitwise operators (&, |, ^, <<, >>)
- Identity operators (is, is not)
- Membership operators (in, not in)

Punctuators:

Common punctuators in Python include

: ( ) [ ] { } , ; . ` ' ' " " / \ & @ ! ? | ~ etc.

Example

```
In [2]: #Finding Square root of a number with the function sqrt() of math library
        import math
        num = 625
        root = math.sqrt(num)
        print("Square root=  ", root)
```

output

```
Square root=   25.0
```

Tokens in the above program are given below

- Keyword - import
- Identifier - num , root (Here it can be said as variables also)
- Literal - 625
- Operator - =
- Punctuator - " " , ( ) .

Point to be Noted

In the above program

- print () is used to display the output on the screen
- # symbol is used to write comments which are used to increase readability and will not be executed
- import statement is used to load the functions from the library (math)
- Variables – Named labels whose value can be used and processed during the execution of the program .

Sample Program-1

Display the string “National Animal-Tiger” on the screen

```
In [1]: print("National animal - Tiger")
```

output

```
National animal - Tiger
```

Sample Program-2

Write a program to calculate the area of a rectangle given the length and breadth are 50 and 20 respectively.

```
In [1]: length = 12
        breadth =7
        area = length * breadth
        print("Area of Rectangle=", area)
```

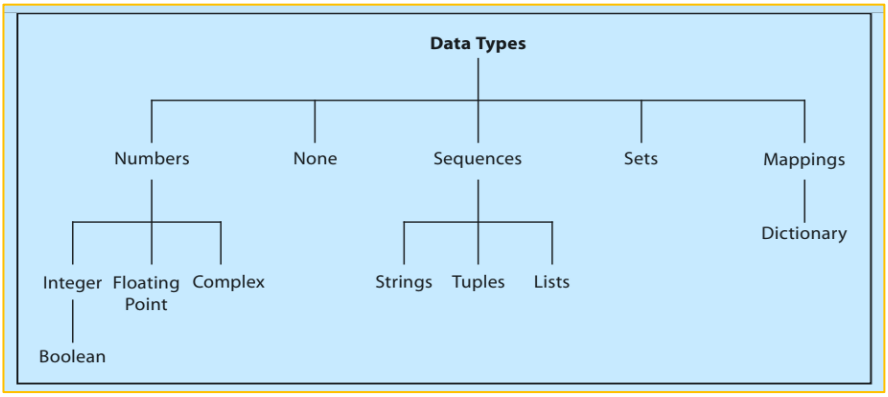
output

```
Area of Rectangle= 84
```

Data Types:

Data types are the classification or categorization of data items. It represents the kind of value that tells what operations can be performed on a particular data. Python supports Dynamic Typing. Python automatically determines the variable's data type based on its value. It infers the data type from the value, and this is why Python is often called a "**dynamically-typed**" language.

The following are the standard or built-in data types in Python:



Data Type Description

Integer	Stores whole number	a=10
Boolean	Boolean is used to represent the True and False values of the expressions.	Result = True
Floating point	Stores numbers with fractional part	x=5.5
Complex	Stores a number having real and imaginary part	num=a+bj
String	immutable sequences (After creation values cannot be changed in-place) Stores text enclosed in single or double quotes	name= “Ria”
List	mutable sequences (After creation values can be changed in-place) Stores list of comma separated values of any data type between square [ ]	lst=[25, 15.6, “car”, “XY”]

Tuple	Immutable sequence (After creation values cannot be changed in-place) Stores list of comma separated values of any data type between parentheses ( )	tup=(11, 12.3, “abc”)
Set	Set is an unordered collection of values, of any type, with no duplicate entry.	s = { 25, 3, 3.5}
Dictionary	Unordered set of comma-separated key:value pairs within braces {}	dict= { 1 : “One”, 2: “Two”, 3: “Three”}

Accepting values from the user

The input() function retrieves text from the user by prompting them with a string argument. For instance:

```
name = input("What is your name?")
```

Return type of input function is string. So, to receive values of other types we have to use conversion functions together with input function.

Sample Program-3

Write a program to read name and marks of a student and display the total mark.

```
In [3]: name=input("Enter Student's Name")
        m1=float(input("Enter the Mark of English"))
        m2=float(input("Enter the Mark of Artificial Intelligence"))
        m3=float (input("Enter the Mark of Maths"))
        Total=m1+m2+m3
        print("Name : ", name)
        print("Total Marks : ", Total)
```

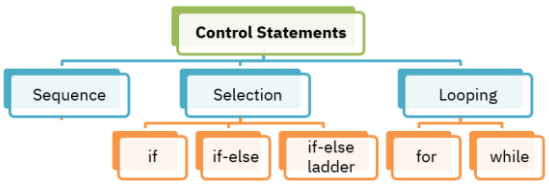
output

```
Enter Student's Name  M J Anakha
Enter the Mark of English  99
Enter the Mark of Artificial Intelligence  100
Enter the Mark of Maths  96
Name :    M J Anakha
Total Marks :  295.0
```

In the above example float( ) is used to convert the datatype into floating point. The explicit conversion of an operand to a specific type is called **type casting.**

Control flow statements in Python

Till now, the programs you've created have followed a basic, step-by-step progression, where each statement executes in sequence, every time. However, there are many practical programs where we have to selectively execute specific sections of the code or iterate over parts of the program. This capability is achieved through selective statements and looping statements.



Selection Statement

The if/ if..else statement evaluates test expression and the statements written below will execute if the condition is true otherwise the statements below else will get executed. Indentation is used to separate the blocks.

Syntax:

```
if <test expression>:
    Statements
```

```
if <test expression>:
    Body of if
else:
    Body of else
```

```
if <test expression>:
    Body of if
elif <test expression>:
    Body of elif
else:
    Body of else
```

Let's check out different examples to see the working of if and if-else statements

Sample Program-4

Asmita with her family went to a restaurant. Determine the choice of food according to the options she chooses from the main menu.

Case 1: All Members are vegetarians. They prefer to have veg food. No other options. (menu-veg)

Program & Output

```
choice=input("Enter the choice of food-Veg/Nonveg")
if choice == "Veg":
    print("Welcome to Vegetarian Food House")
```

```
Enter the choice of food-Veg/NonvegVeg
Welcome to Vegetarian Food House
```

Case 2: Family members may choose non-vegetarian foods also if veg foods are not available. (menu-veg/Nonveg)

```
choice=input("Enter the choice of food-Veg/Nonveg")
if choice == "Veg":
    print("Welcome to Vegetarian Food House")
elif choice == "Nonveg":
    print("Welcome to Non-vegetarian Foods")
```

```
Enter the choice of food-Veg/NonvegNonveg
Welcome to Non-vegetarian Foods
```

Case 3: Family members can choose from variety of options

```
print("Menu: Veg | Nonveg | Mixed")
choice=input("Enter the choice of food")
if choice == "Veg":
    print("Welcome to Vegetarian Food House")
elif choice == "Nonveg":
    print("Welcome to Non-vegetarian Foods")
else:
    print("Welcome to your Favourite Choice of Foods")
```

```
Menu: Veg | Nonveg | Mixed
Enter the choice of foodMixed
Welcome to your Favourite Choice of Foods
```

Sample Program-5

Write a program to get the length of the sides of a triangle and determine whether it is equilateral triangle or isosceles triangle or scalene triangle.

```
# Sample program to classify a triangle based on its sides

# Get user input for the lengths of the sides
side1 = float(input("Enter the length of side 1: "))
side2 = float(input("Enter the length of side 2: "))
side3 = float(input("Enter the length of side 3: "))

# Check the type of triangle based on the lengths of its sides
if side1 == side2 == side3:
    print("It is an equilateral triangle.")
elif side1 == side2 or side1 == side3 or side2 == side3:
    print("It is an isosceles triangle.")
else:
    print("It is a scalene triangle.")
```

```
Enter the length of side 1: 7
Enter the length of side 2: 7
Enter the length of side 3: 11
It is an isosceles triangle.
```

Looping Statements

Looping statements in programming languages allow you to execute a block of code repeatedly. In Python, there are mainly two types of looping statements: for loop and while loop.

For loop

For loop iterates through a portion of a program based on a sequence, which is an ordered collection of items.

The ***"for"*** keyword is used to start the loop. The loop variable takes on each value in the specified sequence (e.g., list, string, range). The colon (:) at the end of the for statement indicates the start of the loop body. The statements within the loop body are executed for each iteration. Indentation is used to define the scope of the loop body. All statements indented under the for statement are

considered part of the loop. It is advisable to utilize a for loop when the exact number of iterations is known in advance.

Syntax

```
for <control-variable> in <sequence/items in range>:
        <statements inside body of the loop>
```

Example -1

```
for i in range(5):
    print("Python")
```

```
Python
Python
Python
Python
Python
```

Example-2

```
for i in range(5):
    print(i)
```

```
0
1
2
3
4
```

Point to be Noted

In the above program

- range (5) returns the values 0,1,2,3,4
- For each iteration of the loop variable i receives these values.
- First iteration of the loop i=0 (one time print("Python") executes, similarly with i=1,2,3,4 also print statement works.
- Whatever is given inside the loop executes repeatedly. In the first example 5 times Python was printed, but in example-2, as i is to be printed it displayed 0 1 2 3 4

Example -3

```
for i in range(10):
    if i>3:
        break
    print(i)
```

```
0
1
2
3
```

The **for** loop iterates over each item in the sequence until it reaches the end of the sequence or until the loop is terminated using a **break** statement. Here when the value of i is greater than 3, for loop is terminated.

For loop is a powerful construct for iterating over collections of data and performing operations on each item.

Sample Program-6

Write a program to display even numbers and their squares between 100 and 110.

```
for num in range(100,110,2):
    square=num*num
    print(num, "squared is", square)
```

```
100 squared is 10000
102 squared is 10404
104 squared is 10816
106 squared is 11236
108 squared is 11664
```

Sample Program-7

Write a program to read a list, display each element and its type. ***(****use type( ) to display the data type.****)***

```
lst = [ 25, "fruit", 17.7, ('a', 'b'), 100]
for word in lst:
    print (word, type(word))
```

```
25 <class 'int'>
fruit <class 'str'>
17.7 <class 'float'>
('a', 'b') <class 'tuple'>
100 <class 'int'>
```

Point to be Noted

In the above program

- the control variable word gets each element of the list. Hence in print statement each element and its type is displayed
- Same program can be written using the following code also

  for i in range (len (lst)):

  print (lst[i] , type (lst[i]))

  Here we take i as index number, lst[0]= 25 & lst[-1] = 100

  len(lst) gives the length of the list (total number of elements in the list)

Sample Program-8

Write a program to read a string. Split the string into list of words and display each word.

```
Str="Iam studying in Jyothis Central School"
wordlist=Str.split()
print ("Words in list format", wordlist)
for word in wordlist:
    print(word)
```

```
Words in list format ['Iam', 'studying', 'in', 'Jyothis', 'Central', 'School']
Iam
studying
in
Jyothis
Central
School
```

Sample Program-9

Write a simple program to display the values stored in dictionary

```
dict = {'S1':"Bio-Math", 'S2':"Math-Comp", 'S3': "Bio-Psy", 'S4': "Math-AI"}
for key in dict:
    print(dict[key])
```

```
Bio-Math
Math-Comp
Bio-Psy
Math-AI
```

UNDERSTANDING CSV file (Comma Separated Values)

CSV files are delimited files that store tabular data (data stored in rows and columns). It looks similar to spread sheets, but internally it is stored in different format. In csv file, values are separated by comma. Data Sets used in AI programming are easily saved in csv format. Each line in a csv file is a data record. Each record consists of more than one fields(columns). The csv module of Python provides functionality to read and write tabular data in CSV format.

Let us see an example of opening, reading and writing formats for a file student.csv with file object file. student.csv contains the columns rollno, name and mark.

importing library	import csv
Opening in reading mode	file= open(“student.csv”, “r”)
Opening in writing mode	file= open(“student.csv”, “w”)
closing a file	file.close( )
writing rows	wr=csv.writer(file) wr.writerow([12, “Kalesh”, 480])
Reading rows	details = csv.reader(file) for rec in details: print(rec)

(Note: -csv files can be created easily using spreadsheets saved with extension .csv)

Sample Program-10

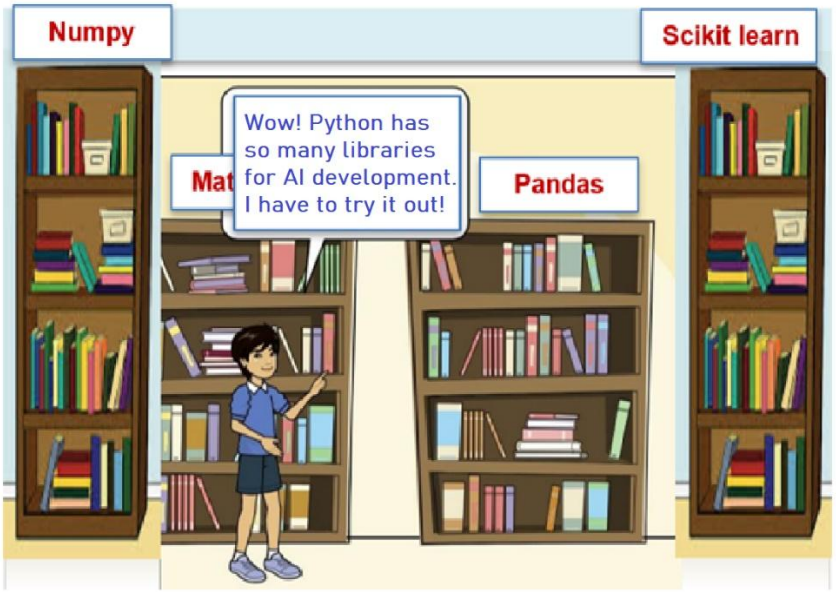
Write a Program to open a csv file students.csv and display its details.

```
import csv
file = open("D:\JPB\Python\students.csv", "r")
details=csv.reader(file)
for rec in details:
    print(rec)
```

```
['RollNo', 'Name', 'class', 'TrName']
['11', 'Akshith', 'II', 'Sruthy ']
['12', 'Ashmitha', 'VII', 'Ruby']
['13', 'M J Anakha', 'X', 'Jayasankar']
```

INTRODUCING LIBRARIES

A **library** in Python typically refers to a collection of reusable modules or functions that provide specific functionality. Libraries are designed to be used in various projects to simplify development by providing pre-written code for common tasks. Concept of libraries are very easy to understand.



In Python, functions are organized within libraries similar to how library books are arranged by subjects such as physics, computer science, and economics. For example, the "math" library contains numerous functions like sqrt(), pow(), abs(), and sin(), which facilitate mathematical operations and calculations. To utilize a library in a program, it must be imported. For example, if we wish to use the sqrt() function in our program, we include the statement "**import math**". This allows us to access and utilize the functionalities provided by the math library.

Python offers a vast array of libraries for various purposes, making it a versatile language for different domains such as web development, data analysis, machine learning, scientific computing, and more. Now, let us explore some libraries that are incredibly valuable in the realm of Artificial Intelligence.

NUMPY

NumPy, which stands for Numerical Python, is a powerful library in Python used for numerical computing. It is a general-purpose array-processing package. NumPy provides the **ndarray** (N-dimensional array) data structure, which represents arrays of any dimension. These arrays are homogeneous (all elements are of the same data type) and can contain elements of various numerical types (integers, floats, etc.)

Where and why do we use the Numpy library in Artificial Intelligence?
Suppose you have a dataset containing exam scores of students in various subjects, and you want

to perform some basic analysis on this data. You can utilize NumPy arrays to store exam scores for different subjects efficiently. With NumPy's array operations, you can perform various calculations such as calculating average scores for each subject, finding total scores for each student, calculating the overall average score across all subjects., identifying the highest and lowest scores. NumPy's array operations streamline these computations, making them both efficient and convenient. This makes NumPy an indispensable tool for data manipulation and analysis in data science applications.

NumPy can be installed using Python's package manager, pip.

```
pip install numpy
```

Creating a Numpy Array - Arrays in Numpy can be created in multiple ways. Some of the ways are programmed here:

➡ **Using List of Tuples**

```
import numpy as np
ar = np.array( [ (99, 88, 77), (44, 55, 66)])
print ("Numpy Array:\n", ar)
```

```
Numpy Array:
 [[99 88 77]
 [44 55 66]]
```

➡ **Using values from the user (using empty( )--** *The empty() function in Python is used to return a new array of a given size)*

```
import numpy as np
n=int(input("Enter the size of an array"))
ar=np.empty(n)
for i in range(n):
    ar[i]=int(input("Enter a number"))
print("Array\n", ar)
```

```
Enter the size of an array4
Enter a number34
Enter a number67
Enter a number85
Enter a number92
Array
 [34. 67. 85. 92.]
```

PANDAS

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" . Pandas is a powerful and versatile library that simplifies tasks of data manipulation in Python . Pandas is built on top of the NumPy library which means that a lot of structures of

NumPy are used or replicated in Pandas and Pandas is particularly well-suited for working with tabular data, such as spreadsheets or SQL tables. Its versatility and ease of use make it an essential tool for data analysts, scientists, and engineers working with structured data in Python.

Where and why do we use the Pandas library in Artificial Intelligence?
Suppose you have a dataset containing information about various marketing campaigns conducted by the company, such as campaign type, budget, duration, reach, engagement metrics, and sales performance. We use Pandas to load the dataset, display summary statistics, and perform group-wise analysis to understand the performance of different marketing campaigns. We then visualize the sales performance and average engagement metrics for each campaign type using Matplotlib, a popular plotting library in Python. Pandas provides powerful data manipulation and aggregation functionalities, making it easy to perform complex analysis and generate insightful visualizations. This capability is invaluable in AI and data-driven decision-making processes, allowing businesses to gain actionable insights from their data.

Pandas can be installed using:

```
pip install pandas
```

Pandas generally provide two data structures for manipulating data, They are: Series and Data Frame.

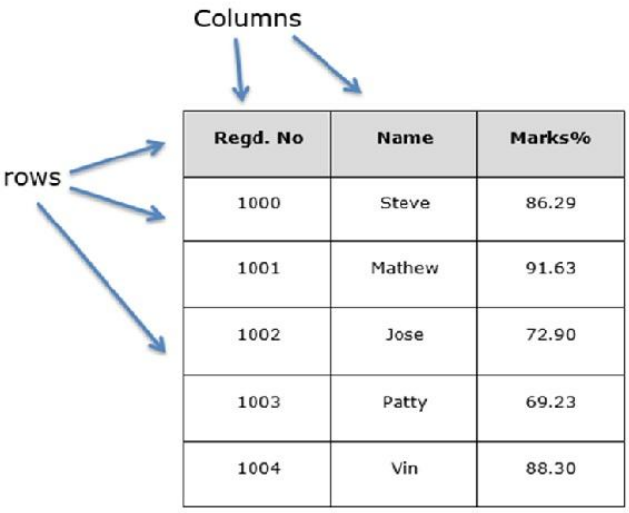
Series

A Series is a one-dimensional array containing a sequence of values of any data type (int, float, list, string, etc.) which by default have numeric data labels starting from zero. The data label associated with a particular value is called its **index**. We can also assign values of other data types as index. We can imagine a Pandas Series as a column in a spreadsheet as given here.

In data science, we often encounter datasets with two-dimensional structures. This is where Pandas Data Frames come into play.

A Data Frame is used when we need to work on multiple columns at a time, i.e., we need to process the tabular data.

Index	Data
0	Mark
1	Justin
2	John
3	Vicky



For example, the result of a class, items in a restaurant's menu, reservation chart of a train, etc.

A DataFrame is a two-dimensional labeled data structure like a table of MySQL. It contains rows and columns, and therefore has both a row and column index. Each column can have a different type of value such as numeric, string, boolean, etc., as in tables of a database.

Creation of DataFrame

There are several methods to create a DataFrame in Pandas, but here we will discuss two common approaches:

Using NumPy ndarrays-

```
import numpy as np
import pandas as pd
array1 = np.array([10,20,30])
array2 = np.array([100,200,300])
array3 = np.array([-10,-20,-30])
dFrame = pd.DataFrame( [array1, array2, array3] , columns = ['col1', 'col2', 'col3'])
print(dFrame)
```

```
   col1  col2  col3
0    10    20    30
1   100   200   300
2   -10   -20   -30
```

Using List of Dictionaries

```
listDict = [{'Dance':10, 'Music':20}, {'Dance':15,'Music':10,'Painting':20}, {'Painting': 12}]
a= pd.DataFrame(listDict, index=['X', 'XI', 'XII'])
print(a)
```

```
     Dance  Music  Painting
X     10.0   20.0       NaN
XI    15.0   10.0      20.0
XII    NaN    NaN      12.0
```

Point to be Noted

- ➔ Dictionary keys become column labels by default in a DataFrame, and the list elements become the row values.
- ➔ NaN (Not a Number) is inserted if a corresponding value for a column is missing.
- ➔ Pandas uses isnull() function to identify NaN values in a DataFrame.

Dealing with Rows and Columns

Based on the DataFrame **'Result'** provided below, we can observe various operations related to rows and columns. Each operation statement is accompanied by its corresponding output from the Result DataFrame.

DataFrame: **Result**

Result

	Rajat	Amrita	Meenakshi	Rose	Karthika
Maths	90	92	89	81	94
Science	91	81	91	71	95
Hindi	97	96	88	67	99

Adding a New Column to a DataFrame:

We can add a new column 'Fathima', by mentioning column name as given below

```
Result['Fathima']=[89,78,76]
print(Result)
```

	Rajat	Amrita	Meenakshi	Rose	Karthika	Fathima
Maths	90	92	89	81	94	89
Science	91	81	91	71	95	78
Hindi	97	96	88	67	99	76

Adding a New Row to a DataFrame:

We can add a new row to a DataFrame using the DataFrame.loc[ ] method. Let us add marks for English subject in Result ➔

```
Result.loc['English'] = [90, 92, 89, 80, 90, 88]
print(Result)
```

	Rajat	Amrita	Meenakshi	Rose	Karthika	Fathima
Maths	90	92	89	81	94	89
Science	91	81	91	71	95	78
Hindi	97	96	88	67	99	76
English	90	92	89	80	90	88

Deleting Rows and Columns from a DataFrame:

We need to specify the names of the labels to be dropped and the axis from which they need to be dropped. To delete a row, the parameter axis is assigned the value 0 and for deleting a column, the parameter axis is assigned the value 1.

Deleting a row "Hindi"

```
Result = Result.drop('Hindi', axis=0)
print(Result)
```

	Rajat	Amrita	Meenakshi	Rose	Karthika	Fathima
Maths	90	92	89	81	94	89
Science	92	84	90	72	96	88
English	90	92	89	80	90	88

Delete the columns having labels 'Rajat', 'Meenakshi' and 'Karthika'':

```
Result = Result.drop(['Rajat','Meenakshi','Karthika'], axis=1)
print(Result)
```

```
         Amrita  Rose  Fathima
Maths        92    81       89
Science      84    72       88
English      92    80       88
```

Point to be Noted

During Data Analysis, DataFrame.drop() method is used to remove the rows and columns.

Accessing DataFrame Elements

Data elements in a DataFrame can be accessed using different ways. Two common ways of accessing are using loc and iloc. DataFrame.loc[ ] uses label names for accessing and DataFrame.iloc[ ] uses the index position for accessing the elements of a DataFrame. Let us check an example

```
Result.loc['Science']
```

/

```
Result.iloc[1]
```

```
Rajat          91
Amrita         81
Meenakshi      91
Rose           71
Karthika       95
Fathima        78
Name: Science, dtype: int64
```

Understanding Missing Values

Missing Data or Not Available data can occur when no information is provided for one or more items or for a whole unit. During Data Analysis, it is common for an object to have some missing attributes. If data is not collected properly it results in missing data.

In DataFrame it is stored as NaN (Not a Number). For example, while collecting data, some people may not fill all the fields while taking the survey. Sometimes, some attributes are not relevant to all.

Pandas provide a function isnull() to check whether any value is missing or not in the DataFrame. This function checks all attributes and returns True in case that attribute has missing values, otherwise returns False.Now, we can explore different operations related to missing values based on the DataFrame 'listDict' provided below.

```
     Dance  Music  Painting
X     10.0   20.0       NaN
XI    15.0   10.0      20.0
XII    NaN    NaN      12.0
```

listDict . isnull( ) →

	Dance	Music	Painting
X	True	True	False
XI	True	Tre	True
XII	False	False	True

Point to be Noted

Finding any missing value in a column ➔ listDict['Music'] . isnull() . any( ) →True
Finding total number of NaN ➔ listDict . isnull() . sum() → 3
Deleting entire row with NaN values ➔ listDict . dropna( )
Replacing NaN values (here by 1) ➔ listDict . fillna (1)

Attributes of DataFrames

Attributes are the properties of a DataFrame that can be used to fetch data or any information related to a particular dataframe.

The syntax of writing an attribute is:

DataFrame_name . attribute

Let us understand the attributes of Dataframes with the help of DataFrame Teacher

DataFrame:Teacher

	IX	X	XI	XII
Physics	Jayasankar	Shanthini	Sruthy	Anand Raj
Maths	Snitha	Haripriya	Praseetha	Sobhana Beegum
Artificial Intelligence	Noufiya	Rejila	Lekshmi	Jyoti

Displaying Row Indexes - Teacher.index

```
In [2]: Teacher.index
Out[2]: Index(['Physics', 'Maths', 'Artificial Intelligence'], dtype='object')
```

Displaying column Indexes - Teacher.columns

```
In [3]: Teacher.columns
Out[3]: Index(['IX', 'X', 'XI', 'XII'], dtype='object')
```

Displaying datatype of each - Teacher.dtypes

```
In [4]: Teacher.dtypes
Out[4]: IX      object
        X       object
        XI      object
        XII     object
        dtype: object
```

Displaying data in Numpy Array form - Teacher.values

```
In [5]: Teacher.values

Out[5]: array([['Jayasankar', 'Shanthini', 'Sruthy', 'Anand Raj'],
               ['Snitha', 'Haripriya', 'Praseetha', 'Sobhana Beegum'],
               ['Noufiya', 'Rejila', 'Lekshmi', 'Jyoti']], dtype=object)
```

Displaying total number of rows and columns (row, column) - Teacher.shape

```
In [6]: Teacher.shape

Out[6]: (3, 4)
```

Displaying first n rows (here n = 2) - Teacher. head (2)

```
In [7]: Teacher.head(2)
```

Out[7]:

	IX	X	XI	XII
Physics	Jayasankar	Shanthini	Sruthy	Anand Raj
Maths	Snitha	Haripriya	Praseetha	Sobhana Beegum

Displaying last n rows (here n = 2) - Teacher. tail (2)

```
In [8]: Teacher.tail (2)
```

Out[8]:

	IX	X	XI	XII
Maths	Snitha	Haripriya	Praseetha	Sobhana Beegum
Artificial Intelligence	Noufiya	Rejila	Lekshmi	Jyoti

Importing and Exporting Data between CSV Files and DataFrames

We can create a DataFrame by importing data from CSV files. Similarly, we can also store or export data in a DataFrame as a .csv file.

Importing a CSV file to a DataFrame

Using the read_csv() function, you can import tabular data from CSV files into pandas dataframe by specifying a parameter value for the file name

Syntax: pd.read_csv("filename.csv")

Example: Reading file students.csv

```
import pandas as pd
import csv

df = pd.read_csv("D:\JPB\Python\students.csv",sep =",", header=0)
print(df)
```

```
   RollNo         Name class      TrName
0      11      Akshith    II     Sruthy
1      12     Ashmitha   VII        Ruby
2      13  M J Anakha     X   Jayasankar
```

Point to be Noted

- read_csv() is used to read the csv file with its correct path
- sep specifies whether the values are separated by comma, semicolon, tab, or any other character. The default value for sep is a space.
- The parameter header marks the start of the data to be fetched. header=0 implies that column names are inferred from the first line of the file. By default, header=0.

Exporting a DataFrame to a CSV file

We can use the to_csv() function to save a DataFrame to a text or csv file. For example, to save the DataFrame Teacher into csv file resultout, we should write

Teacher.to_csv(path_or_buf='C:/PANDAS/resultout.csv', sep=',')

When we open this file in any text editor or a spreadsheet, we will find the above data along with the row labels and the column headers, separated by comma.

Scikit-learn (Learn)

Note to Teacher: This topic can be taught after teaching the Machine Learning Unit.

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling via a consistent interface in Python. Sklearn is built on (relies heavily on) NumPy, SciPy and Matplotlib. .

Key Features:

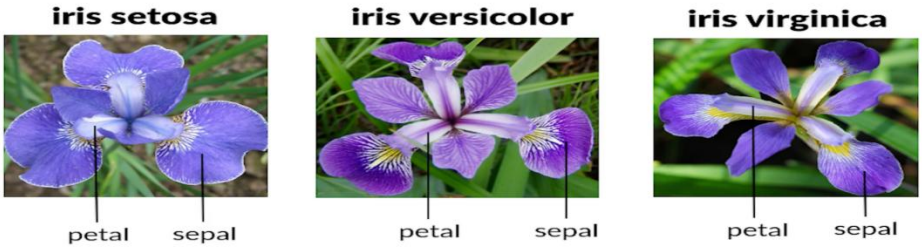
- Offers a wide range of supervised and unsupervised learning algorithms.
- Provides tools for model selection, evaluation, and validation.
- Supports various tasks such as classification, regression, clustering, dimensionality reduction, and more.
- Integrates seamlessly with other Python libraries like NumPy, SciPy, and Pandas.

Install scikit-learn using the statement

```
pip install scikit-learn
```

load_iris (In sklearn.datasets)

The Iris dataset is a classic and widely used dataset in machine learning, particularly for classification tasks. It comprises measurements of various characteristics of iris flowers, such as sepal length, sepal width, petal length, and petal width, along with the corresponding species of iris to which they belong. The dataset typically includes three species: setosa, versicolor, and virginica.



from sklearn.datasets import load_iris	importing iris dataset
iris = load_iris()	calls the “load_iris()” function to load the Iris dataset
X = iris.data	X is a variable and assigned as feature vector.The feature vectors contain the input data for the machine learning model
y= iris.target	Y is a variable and assigned as target variable.The target variable contains the output or the variable we want to predict with the model.

Sample output – First 10 rows of X

```
Feature names: ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)']
Target names: ['setosa' 'versicolor' 'virginica']

First 10 rows of X:
 [[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]
 [5.4 3.9 1.7 0.4]
 [4.6 3.4 1.4 0.3]
 [5.  3.4 1.5 0.2]
 [4.4 2.9 1.4 0.2]
 [4.9 3.1 1.5 0.1]]
```

Here, each row represents a sample (i.e., an iris flower), and each column represents a feature (i.e., a measurement of the flower).

For example, the first row [5.1 3.5 1.4 0.2] corresponds to an iris flower with the following measurements:

- Sepal length: 5.1 cm
- Sepal width: 3.5 cm
- Petal length: 1.4 cm
- Petal width: 0.2 cm

train_test_split (In sklearn.model_selection)

Datasets are usually split into training set and testing set. The training set is used to train the model and testing set is used to test the model.

Most common splitting ratio is 80 : 20 . (Training -80%, Testing-20%)

from sklearn.model_selection import train_test_split	importing train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 1)	
X_train, y_train	the feature vectors and target variables of the training set respectively.
X_test, y_test	the feature vectors and target variables of the testing set respectively.
test_size = 0.2	specifies that 20% of the data will be used for testing, and the remaining 80% will be used for training.

random_state = 1	Ensures reproducibility by fixing the random seed. This means that every time you run the code, the same split will be generated.

KNeighborsClassifier (In sklearn.neighbors)

Scikit-learn has wide range of Machine Learning (ML) algorithms which have a consistent interface for fitting, predicting accuracy, recall etc. Here we are going to use KNN (K nearest neighbors) classifier.

from sklearn.neighbors import KNeighboursClassifier	importing KneighboursClassifier (type of supervised learning algorithm used for classification tasks.)
knn = KNeighborsClassifier(n_neighbors =3)	we create an instance of the KNeighborsClassifier class . n_neighbors = 3 indicates that the classifier will consider the 3 nearest neighbors when making predictions. This is a hyperparameter that can be tuned to improve the performance of the classifier.
knn.fit(X_train, y_train)	trains the KNeighborsClassifier model using the fit method. it constructs a representation of the training data that allows it to make predictions based on the input features.
y_pred = knn.predict(X_test)	The knn object contains the trained model, make predictions on new, unseen data.

metrics

```
from sklearn import metrics
Accuracy = metrics.accuracy_score(y_test, y_pred))
```

This calculates the accuracy of the model by comparing the predicted target values (y_pred) with the actual target values (y_test). The accuracy_score represents the proportion of correctly predicted instances out of all instances in the testing set.

Scikit-learn offers a variety of modules that simplify the process of building, training, and evaluating machine learning models, making it a popular choice for various tasks in this domain. In our session, we utilized the 'load_iris()' function to load the Iris dataset. Upon loading, we split the dataset into training and test sets using the 'train_test_split' function. Subsequently, we trained our model using the K-Nearest Neighbors Classifier ('KNeighborsClassifier') and evaluated its performance using appropriate metrics. This workflow represents a typical data analysis pipeline in AI project development.

Now, to validate the model's predictive accuracy, we can use some sample data.

```
sample = [[5, 5, 3, 2], [2, 4, 3, 5]]
preds = knn.predict(sample)
pred_species=[]
for p in preds:
   pred_species.append(iris.target_names[p])
print("Predictions:", pred_species)
```

The provided code snippet demonstrates how to use the trained classifier to make predictions on sample data. After initializing the sample data as [5, 5, 3, 2], the classifier predicts the species of iris flowers based on these measurements. Finally, the predicted species are printed to the console.

This is a program that combines different parts of our project to make it complete and understandable.

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics

# Load the Iris dataset
iris = load_iris()
X = iris.data
y = iris.target

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)

# Train KNN classifier
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(X_train, y_train)

# Predict on test set
y_pred = knn.predict(X_test)

# Calculate accuracy
accuracy = metrics.accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

# Sample predictions
sample = [[5, 5, 3, 2], [2, 4, 3, 5]]
preds = knn.predict(sample)

# Convert numeric predictions to class names
pred_species = []
for p in preds:
    pred_species.append(iris.target_names[p])

print("Predictions:", pred_species)
```

Output-

```
Accuracy: 1.0
Predictions: ['versicolor', 'virginica']
```

Using this model, we can identify the type of flower in the iris dataset. By analyzing the length and width of the sepals and petals, we can compare them with the features of the setosa, versicolor, and virginica species to determine the flower's species.

Point to be Noted

Links to explore python more

Tutorials

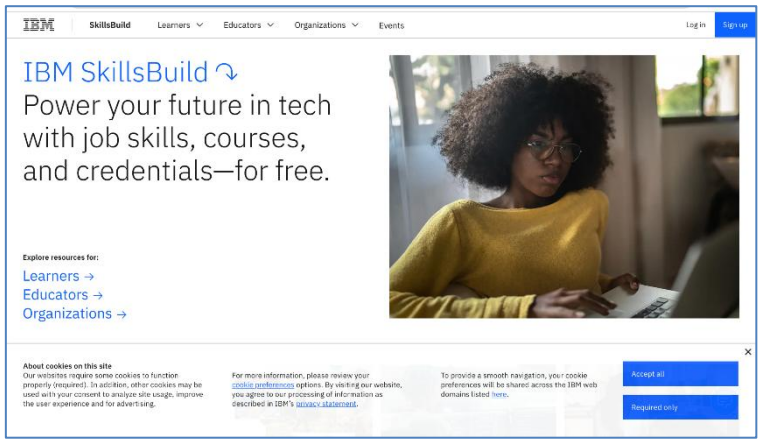
1. https://www.programiz.com/python-programming
2. *https://www.analyticsvidhya.com/blog/2021/05/data-types-in-python/*
3. *https://www.w3schools.com/python/default.asp*
4. *https://www.geeksforgeeks.org/pandas-tutorial/*
5. *https://www.learnpython.org/en/Pandas_Basics*
6. *https://www.geeksforgeeks.org/python-programming-language/*
7. *https://scikit-learn.org/stable/tutorial/basic/tutorial.html*
8. *https://pandas.pydata.org/docs/user_guide/10min.html*

Courses

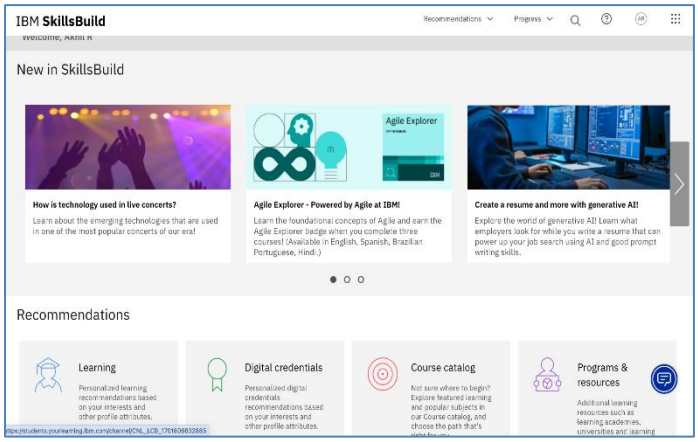
1. *https://aistudent.community/single_course/2021*
2. *https://www.kaggle.com/learn/pandas*
3. *https://www.udemy.com/course/pandas-with-python/*

Step-by-Step guide for students to use the IBM Skills Build website to learn Python:

Step 1: Visit the IBM SkillsBuild website using the link - https://skillsbuild.org/ and sign up for an account.



Step 2: Locate and click on the "High School Student" option, then proceed to click on the "Sign Up" button.



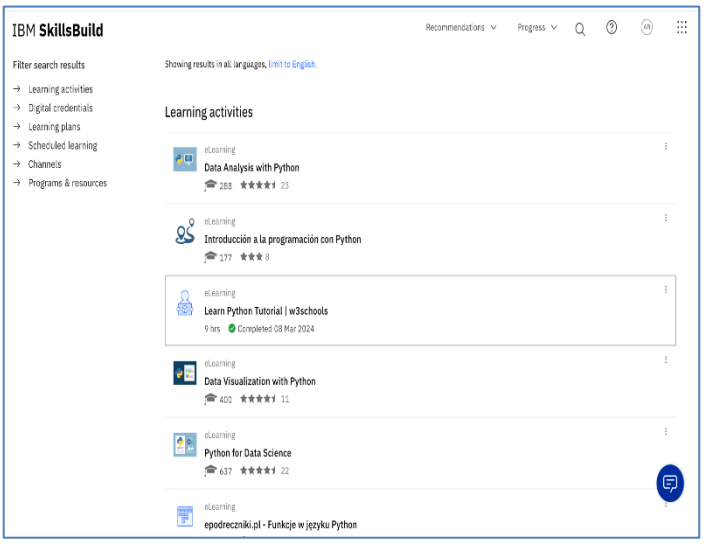
Step 3: Fill in the required information to create an account. You can sign up using your email address, LinkedIn ID, or IBM ID.

Step 4: Upon successfully completing this, you will be redirected to your dashboard. This is where you can explore a variety of courses.

Step 5: To start learning Python, use the search option at the top of the page and type in "Python" to find relevant courses.

Step 6: Browse, select a course, complete the tutorial and exercise.

Step 7: Monitor your progress on the IBM Skills Build platform and feel free to explore additional courses or resources to further enhance your understanding of Python and other related topics.



EXERCISES

A. Multiple choice questions

1. Identify the datatype L = “45”

   a. String b. int c. float d. tuple

2. Which of the following function converts a string to an integer in python?

   a. int(x) b. long(x) c. float(x) d.str(x)

3. Which special symbol is used to add comments in python?

   a. $ b.// c. /*.... */ d.#

4. Which of the following variable is valid?

   a. Str name b.1str c._str d.#Str

5. Elements in the list are enclosed in _____ brackets

   a. ( ) b. { } c. [ ] d. /* */

6. Index value of last element in list is _____

   a. 0 b.-10 c. -1 d.10

7. What will be the output of the following code?

```
a = [10,20,30,40,50]
print(a[0])
```

   a.20 b.50 c. 10 d.40

8. Name the function that displays the data type of the variable.

   a. data( ) b. type( ) c. datatype( ) d. int( )

9. Which library helps in manipulating csv files?

   a. files b.csv c. math d. print

10. Which keyword can be used to stop a loop?

    a. stop b.break c. brake d. close

11. What is the primary data structure used in NumPy to represent arrays of any dimension?

    a) Series b) DataFrame c) ndarray d) Panel

12. Which of the following is not a valid method to access elements of a Pandas DataFrame?
    a) Using column names as attributes.
    b) Using row and column labels with the .loc[] accessor.
    c) Using integer-based indexing with the .iloc[] accessor.
    d) Using the .get() method.
13. What is the purpose of the head() method in Pandas?
    a) To display the first few rows of a DataFrame.
    b) To display the last few rows of a DataFrame.
    c) To count the number of rows in a DataFrame.
    d) To perform aggregation operations on a DataFrame.
14. Which method is used to drop rows with missing values from a DataFrame in Pandas?
    a) drop_rows() b) remove_missing() c) dropna() d) drop_missing_values
15. Which is not a module of Sklearn?
    a) load_iris b)train_test_split c)metrics d)Scikit

ANSWERS
1. a. String
2. a. int(x)
3. d. #
4. c. _str
5. c. [ ]
6. c. -1
7. c. 10
8. b. type( )
9. b. .csv
10. b. break
11. c. ndarray
12. d. Using the .get( ) method
13. a) To display the first few rows of a DataFrame
14. c. dropna( )
15. d. Scikit

B. Answer the following questions

1. input() function accepts the value as string only. How can you convert string to int?
   Using int( ) function together with input( ), we can convert string to int.

2. What are variables? What are the rules of declaring variables in Python?
   Named labels whose value can be used and processed during program run. Generally, keywords (list given above) are not used as variables. Variable names cannot start with digit and also it can't contain any special characters except underscore.

3. What do you mean by type casting?
   A variable of particular datatype can be converted into another datatype using some functions. The explicit conversion of an operand to a specific type is called type casting.

4. “Python supports dynamic typing”, True or False. Justify your answer.
   True.
   . Python automatically determines the variable's data type based on its value. It infers the data type from the value, and this is why Python is often called a "dynamically-typed" language.

5. Name any four features of python language.
   - ❖ High Level language
   - ❖ Interpreted Language
   - ❖ Free and Open Source
   - ❖ Platform Independent (Cross-Platform)

6. Give examples for keywords.
   and, as, continue, if, not try, del, pass

7. Expand CSV.
   Comma Separated Values

8. How do you read data from a CSV file into a Pandas DataFrame?
   import pandas as pd
   pd.read_csv("filename.csv")

C. Long Answer Questions

1. Describe the data types supported by Python, providing relevant examples.

Integer	Stores whole number	a=10
Floaing Point	Stores numbers with fractional part	x=5.5
Complex	Stores a number having real and imaginary part	num=a+bj
String	immutable sequences Stores text enclosed in single or double quote or triple quotes	name= “Ria”
List	mutable sequences Stores list of comma separated values of any data type between square [ ]	lst=[25, 15.6, “car”, “XY”]

Tuple	Immutable sequence Stores list of comma separated values of any data type between parentheses ( )	tup=(11, 12.3, “abc”)
Dictionary	Unordered set of comma-separated key:value pairs within braces {}	dict= { 1 : “One”, 2: “Two”, 3: “Three”}

2. Define an operator and provide examples of different operators along with their functions.

Operators are symbols or keywords that perform operations on operands to produce a result. Python supports a wide range of operators:

- Arithmetic operators (+, -, *, /, %)
- Relational operators (==, !=, <, >, <=, >=)
- Assignment operators (=, +=, -=)
- Logical operators (and, or, not)
- Bitwise operators (&, |, ^, <<, >>)
- Identity operators (is, is not)
- Membership operators (in, not in)

D. Practice Programs

1. Write a Tipper program where the user inputs the total restaurant bill. The program should then display two amounts: 15 percent tip and 20 percent tip.

```
bill_total = float(input("Enter the restaurant bill total: "))

# Calculate the 15% tip
tip_15_percent = bill_total * 0.15

# Calculate the 20% tip
tip_20_percent = bill_total * 0.20

# Display the tip amounts
print("15% Tip: Rs", format(tip_15_percent, '.2f'))
print("20% Tip: Rs", format(tip_20_percent, '.2f'))
```

```
Enter the restaurant bill total: 1235
15% Tip: Rs 185.25
20% Tip: Rs 247.00
```

2. Write a program to check whether the user is eligible for driving license or not.

```
age = int(input("Enter your age: "))

# Check if the user is eligible for a driving license
if age >= 18:
    print("Congratulations! You are eligible for a driving license.")
else:
    print("Sorry, you are not eligible for a driving license yet. You need to be at least 18 years
```

```
Enter your age: 18
Congratulations! You are eligible for a driving license.
```

3. Your father always gives his car for service after 15000 km. Check whether his car needs service or not. Read the kilometer reading from the user and give the output.

```
kilometer_reading = float(input("Enter the kilometer reading of the car: "))

# Define the service threshold (in kilometers)
service_threshold = 15000

# Check if the kilometer reading exceeds the service threshold
if kilometer_reading >= service_threshold:
    print('''Your father's car needs service. It has exceeded the service
          threshold of''', service_threshold, "kilometers.")
else:
    print('''Your father's car does not need service yet. It has not exceeded the service
           threshold of", service_threshold, "kilometers.")
```

```
Enter the kilometer reading of the car: 16543
Your father's car needs service. It has exceeded the service threshold of 15000 kilometers.
```

4. Write a program to display the first ten even natural numbers (use for loop).

```
count = 0

# Iterate through natural numbers starting from 1
for num in range(1, 101):  # We iterate up to 100 to ensure we get the first ten even natural numbe
    # Check if the number is even
    if num % 2 == 0:
        # Print the even number
        print(num, end=" ")
        # Increment the count of even numbers
        count += 1
        # Check if we have printed the first ten even numbers
        if count == 10:
            break
```

```
2 4 6 8 10 12 14 16 18 20
```

5. Write a program to accept the Basic salary from the user and calculate the Net Salary.

Net Salary= Basic Salary + HRA + DA -PF

HRA=30% of Basic DA=20% of Basic PF=12% of Basic

```
basic_salary = float(input("Enter the Basic Salary: "))
# Calculate HRA (30% of Basic)
hra = 0.30 * basic_salary
# Calculate DA (20% of Basic)
da = 0.20 * basic_salary
# Calculate PF (12% of Basic)
pf = 0.12 * basic_salary
# Calculate Net Salary
net_salary = basic_salary + hra + da - pf

# Display the Net Salary
print("Basic: ", basic_salary)
print("House Rent Allowance: ", hra)
print("Dearness Allowance: ", da)
print("Provident Fund: ", pf)
print("Net Salary:", net_salary)
```

```
Enter the Basic Salary: 13425
Basic:  13425.0
House Rent Allowance:  4027.5
Dearness Allowance:  2685.0
Provident Fund:  1611.0
Net Salary: 18526.5
```

6. Write a program to create series from an array in Python.

```
import numpy as np
import pandas as pd

# Define a NumPy ndarray
data_array = np.array([10, 20, 30, 40, 50])

# Create a series from the ndarray
series = pd.Series(data_array)

# Display the series
print("Series from ndarray:")
print(series)
```

```
Series from ndarray:
0    10
1    20
2    30
3    40
4    50
dtype: int32
```

7. Consider the following admission.csv and answer the following questions:

Name	CLASS	Gender	Marks
Amit	10	M	75
Ashu	9	F	95
Abhinav	9	M	86
Ravi	10	M	57
Rashmi	11	F	78
Ramesh	10	M	72
Mohit	9	M	53
Manavi	10	F	47
Dhruv	9	M	76

a. Create a dataframe from the admission.csv

```
import pandas as pd
import csv
df=pd.read_csv("admission.csv")
print(df)
```

b. Display first 3 rows of the dataframe

```
print(df.head(3))
```

c. Display the details of Ravi

```
print(df . loc['Ravi'])
```

d. Display the total number of rows and columns in the data frame

```
print(df.shape)
```

e. Display the column "Gender"

```
print(df['Gender'])
```

E.Competency Based Questions

1. Help Priya to differentiate the given information into various datatypes of Python. Name of the student, email id, student id, marks in 5 subjects which can be changed at any point, 3 extra subjects chosen which cannot be changed later.

   Ans: The different datatypes of Python for the given information are:

   a. Name of the student-String type
   b. Email id-String type
   c. Student id-Integer type
   d. Marks in 5 subjects which can be changed at any point-List type
   e. Extra 3 subjects which cannot be changed later-Tuple type.

2. For any analysis to be done, a huge amount of data needs to be collected and stored in a proper format. Rohan has stored the information in a delimited file that stores tabular data which is separated by comma. Which type of file is Rohan using?

   Ans: CSV file (Comma Separated Values)

3. Athrav is confused about the different libraries of Python. Help him choose the correct library for the following tasks.
   a. Data manipulation and aggregation functionalities
   b. Numerical computing
   c. Mathematical operations like square root, cosine values
   d. Machine learning and statistical modeling

   Ans: a. pandas b. numpy c. math d. scikit-learn

4. Rohit wants to input the data about the runs scored by his 50 classmates. Which is the most appropriate loop to be used in this case?
   Ans: 'for' loop

5. Samhita needs guidance to identify these concepts related to Python. As a Python language expert, help her.
   a. ___ is inserted if a corresponding value for a column is missing. (NaN).
   b. ____ is used when we need to work on multiple columns at a time. (Data Frame).
   c. Unordered set of comma-separated key:value pairs within braces {} – dictionary.
   d. _____ are symbols or keywords that perform operations on operands to produce a result. (Operators)
   e. _____ is a dataset which is classic and widely used in ML, particularly for classification tasks. (Iris dataset)
   f. It is a type of supervised learning algorithm used for classification tasks. (KNN – K nearest neighbors).

UNIT 4: Introduction to Capstone Project

Title: Introduction to Capstone Project	**Approach**: Hands on, Team Discussion, Web search, Case studies

Summary: In this unit, students will be able to understand the concept of the Capstone project and its importance as a final project of an academic program. They will understand that all the knowledge gathered will be integrated in the form of a project known as Capstone Project. The chapter will help the students understand how real-life simple day-to-day problems or situations can be handled and solved using a methodology known as Design Thinking and how to leverage AI in addressing these issues. The students will also learn to associate the issues with Sustainable Development Goals. They will be able to exchange their points of view based on experiences and discuss potential solutions to the problem.

Learning Objectives:

- Understand the meaning of the Capstone Project and its goals.
- Understand how problems can be identified, decomposed and solved using Design Thinking Methodology.
- Learn the steps of Design Thinking and apply for solving simple issues.
- Learn to create Empathy maps.
- Understand the importance of 5W1H in Design Thinking and Capstone Project development.
- Relate the importance of Sustainable Development Goals and how these issues can be aligned with Capstone Project.

Key Concepts: Design Thinking, Empathy Map, SDG's, Capstone Project

Learning Outcomes:

By the end of this unit, students will be able to

- Decompose any problem using the 5W1H method
- Apply Design thinking methodology
- Create empathy maps
- Align problems to SDGS
- Apply all the learnings in solving real world problems
- Comfortably express their solution to a problem in non-technical words

Pre-requisites: Basic knowledge about problem solving, issues around them, good communication skills to express their views about any issue, basic knowledge about AI, scientific temper.

Launching Your Journey: A Teacher's Guide to the Capstone Project Introduction

This lesson equips you to introduce the exciting Capstone Project, a cornerstone of the course and their final assessment.

1. Setting the Stage:

- **Capstone Overview:** Begin by explaining the Capstone Project – a student-driven project they will develop throughout the course and submit for their final examination.
- **Long-term Focus:** Emphasize the project's importance and how it allows them to integrate and apply skills learned throughout the course.

2. Early Start, Strong Finish:

- **Spreading the Workload:** Highlight the benefit of starting the project early in Class XI. This allows for a more manageable workload and avoids last-minute pressure in Class XII.

3. Considering Different Viewpoints:

- **Multiple Perspectives:** Encourage students to approach the project from various viewpoints:
    - User Perspective: Who will benefit from the project?
    - Business Perspective: How does the project address a business need or opportunity?
    - Project Developer Perspective: What are the technical considerations and limitations?

This multi-faceted approach fosters well-rounded project development.

4. Building the Foundation (Class XII):

- **Detailed Methodology:** Explain that Class XII Unit 1 (Data Science Methodology) will delve deeper into the steps involved in developing a Capstone Project. This includes:
    - Data Collection Methods: Exploring various ways to gather data for the project.
    - Data Storage Options: Discussing different methods to store the collected data.
    - Data Analysis Techniques: Examining various tools and approaches for analyzing data.

5. The Power of the Project Logbook:

- **Introducing the Logbook:** Introduce the concept of a project logbook – a dedicated notebook or digital document where students will track their progress.
- **Starting Early:** Encourage students to begin completing portions of the logbook in Class XI itself. This might include:
    - Brainstorming project ideas.
    - Researching the chosen topic.
    - Defining initial project goals and objectives.

Additional Tips:

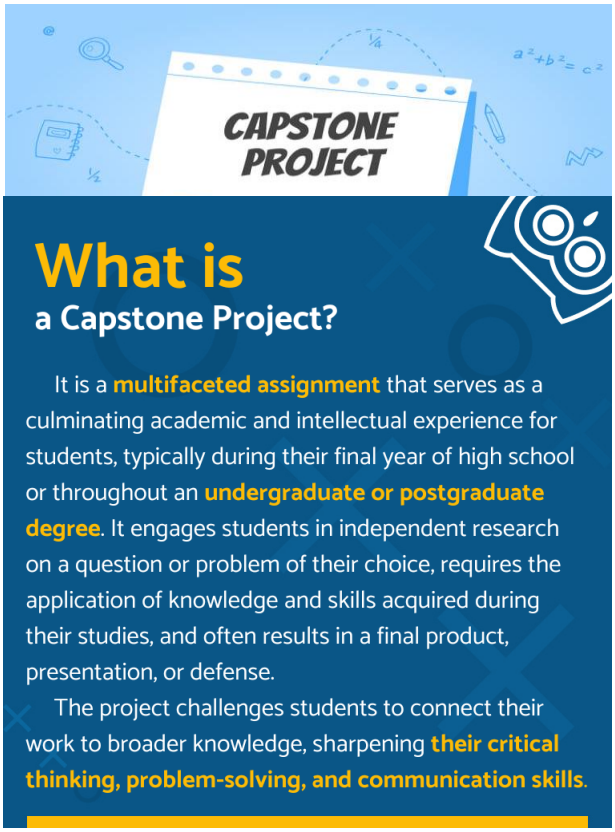
- Facilitate brainstorming sessions to spark project ideas aligned with student interests.
- Invite guest speakers from relevant fields to share their project development experiences.
- Provide opportunities for students to present their initial project ideas and receive feedback from peers and instructors.

By implementing these strategies, you can ignite student enthusiasm for the Capstone Project, set them up for success, and empower them to create insightful and impactful projects.

Introduction to Capstone Project

Teachers can ask the following questions:

1. **Can you think of a problem or challenge in your everyday life, schoolwork, or hobbies that you might be able to solve using technology? Is there a pattern involved in this problem?** This gets them thinking about how AI can be applied to real-world problems and introduces the concept of patterns being crucial for AI solutions.
2. **Imagine you are working on a team for an AI Capstone project. What skills would be important for each team member to have in order to be successful?** This gets them thinking about the collaborative nature of AI projects and the different skills needed beyond just technical knowledge.



A capstone project is a project where students must research a topic, deeply understand the subject matter and integrate all their knowledge gathered about a particular subject to develop a solution to a problem.

The AI Capstone project will give learners an opportunity to implement the AI skills as a final step in the learning path to showcase their expertise in the subject matter.

While doing a Capstone project, students will learn more about problems which they can solve to improve their lives and make the world a better place. Also, they learn many important skills including:

- How to work as a team member?
- How to clearly identify an issue? And how is it affecting (the user)?
- How to brainstorm solutions and select the best one?
- How to decide which type of AI may be useful for the proposed solution?
- How to ethically gather and use data to train a computer to help solve the issue?
- How to test the prototype with users and use their feedback to improve the solution?
- How to pitch their solution to people who will be able to help them take action?

Some Examples for Capstone Project are given below:

1. Stock Prices Predictor
2. Develop A Sentiment Analyzer
3. Movie Ticket Price Predictor
4. Students Results Predictor
5. Human Activity Recognition using Smartphone Data set
6. Classifying humans and animals in a photo

The list of examples is huge as every year students come up with new ideas and innovations for new projects.

Asking the right question before starting the AI Capstone Project

Before Starting a Project, one should analyze whether a problem has a pattern associated with it. If there is no pattern, then the problem cannot be solved with AI technology. It is fundamental that this question, ***"Is there a Pattern?"*** is asked before deciding to embark on an AI development journey.

There are various AI techniques which can be applied to check patterns and solve the problems. Predictive Analysis helps in identifying the patterns by asking different questions which fall into different approaches. Some of the questions are given below:

1) Which category? (Classification)
2) How much or how many? (Regression)
3) Which group? (Clustering)
4) Is this unusual? (Anomaly Detection)
5) Which option should be taken? (Recommendation)

Now, once these questions are answered, we can decide that the problem can be solved using AI. Complex problems cannot be dealt with as such. Then we try to simplify the complex problem to analyze it. This is known as Problem Decomposition.

Problem decomposition steps

1. Understand the problem and then restate the problem in your own words
    - Know what the desired inputs and outputs are
    - Ask questions for clarification (in class these questions might be to your instructor, but most of the time they will be asking either yourself or your collaborators)
2. Break the problem down into a few large pieces.
    - Write these down, either on paper or as comments in a file.
3. Break complicated pieces down into smaller pieces.
    - Keep breaking down until all the pieces are small.
4. Code one small piece at a time.
    - Think about how to implement it and write the code/query
    - Test it on its own and Fix problems, if any

Activity

1. Many people are facing serious issues with traffic jams during office hours which may be caused due to multiple reasons. Considering this as a major issue to resolve, write down the problem decomposition steps to make it simple to solve.

First Step → Understand the Problem and Restate it.

Public is facing difficulty due to the problem of traffic jams especially during the busy office hours.

Second Step → Break the problem down into a few large pieces.

Split the traffic issue separately for each reason

a. Bad Roads
b. Accidents
c. Office Hours
d. Rash Driving
e. Inappropriate Signals
f. Over Crowded Area

Third Step → Break complicated pieces down into smaller pieces.

From the above sub steps (reasons), if any reason is of complicated type, we can split it into more simpler units. For example, Rash Driving can be further split into

a. Over Speed
b. New Drivers with less experience
c. Careless Driving

Fourth Step –>Take each issue at a time and try to find the solution

2. Critical and Creative Thinking

Definition: Critical thinking is the ability to analyze a situation and make a judgment based on facts and data. Creative thinking on the other hand refers to the ability to come up with new ideas or rather solutions.

So, these two skills are very important to raise questions against problems, gather and assess relevant information, brainstorm ideas and formulate well-reasoned conclusions and solutions. Creative thinking and critical thinking are necessary skills for equipping individuals to be the social change makers, leaders and innovators. Complex problems can be easily solved using critical and creative thinking which has a systematic approach to address any problem which must be solved. This approach is known as Design Thinking.

Introduction to Design Thinking

Definition: Design thinking is a non-linear, iterative process that expert teams use to understand users, challenge assumptions, redefine problems and create innovative solutions to prototype and test.

Design Thinking framework

Design Thinking is a design methodology that provides a solution-based approach to solving problems. It's extremely useful in tackling complex problems that are ill-defined or unknown.

The five stages of Design Thinking are as follows: Empathize, Define, Ideate, Prototype, and Test.

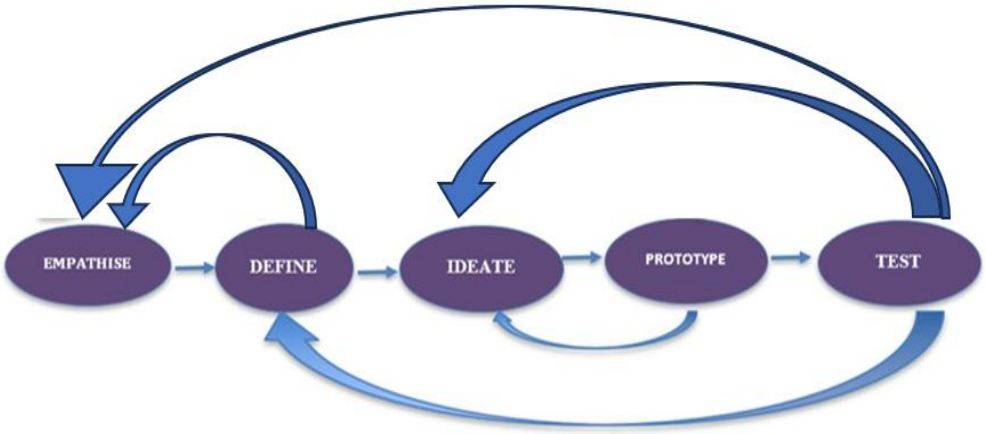


Figure 1: Stages of Design Thinking Framework

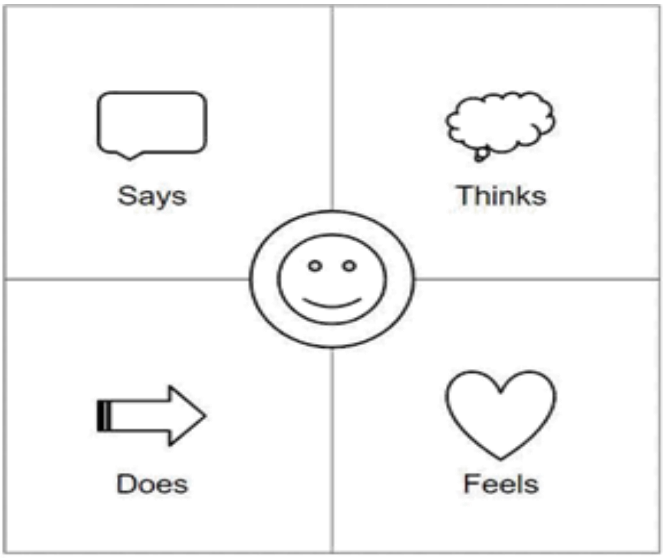
i) Empathize

Design thinking begins with empathy. This requires doing away with any preconceived notions and immersing oneself in the context of the problem for better understanding. In simple words, through empathy, one can put oneself in other people's shoes and connect with how they might be feeling about their problem, circumstance, or situation.

There is a challenge one needs to solve. How does one approach it? Empathy starts from here. As a designer of the solution to a challenge, one should always understand the problem from an end-user perspective. This is done by observation, interaction or by imagination.

Designers are expected to interact with customers / users very frequently to gather detailed facts about the problems and user's expectations. A detailed analysis of these facts leads to solving the problem in the best possible way.

Empathy Map An extremely useful tool for understanding the users' needs and gaining a deeper insight into the problem at hand is the empathy map. It also helps in deepening that understanding, gaining insight into the user's behavior.



To create a “persona” or profile for the user, you can use the empathy map activity to create a realistic general representation of the user or users. Personas can include details about a user’s education, lifestyle, interests, values, goals, needs, thoughts, desires, attitudes, and actions.

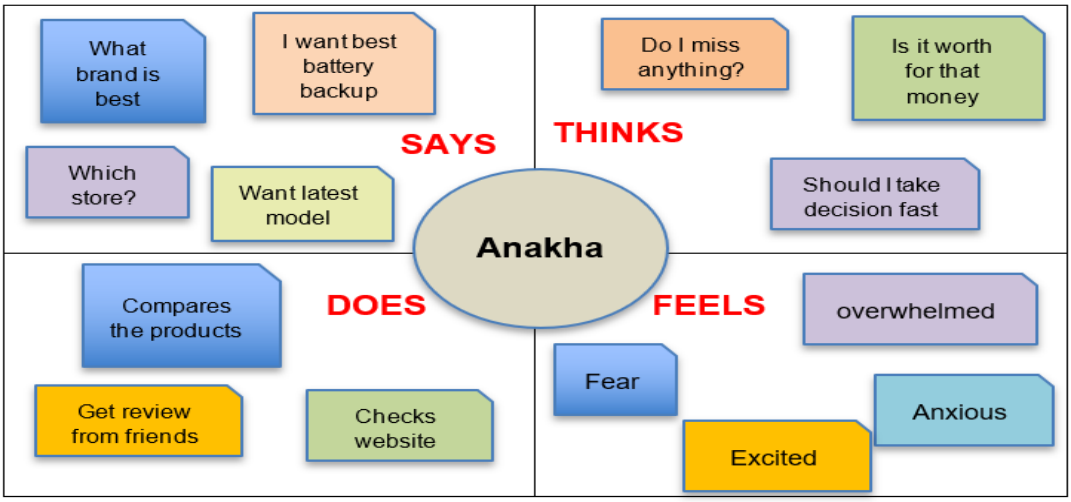
An Empathy Map is divided into 4 quadrants. —Says, Thinks, Does, Feels

- Says – This quadrant contains whatever the user says aloud
- Thinks – This quadrant depicts the thoughts which user have about the problem
- Does –This quadrant depicts the actions of user
- Feels – This quadrant displays the emotional status of the user.

Creating Empathy Map

Draw the Empathy map grid as shown above. Write down the respective observations in corresponding quadrants. These observations can be written directly or can paste a sticky note with the observations in the quadrants. Let us understand through an example. Anakha has a desktop computer. She is planning to buy one new Laptop for her educational purpose. Let us create an Empathy map for Anakha.

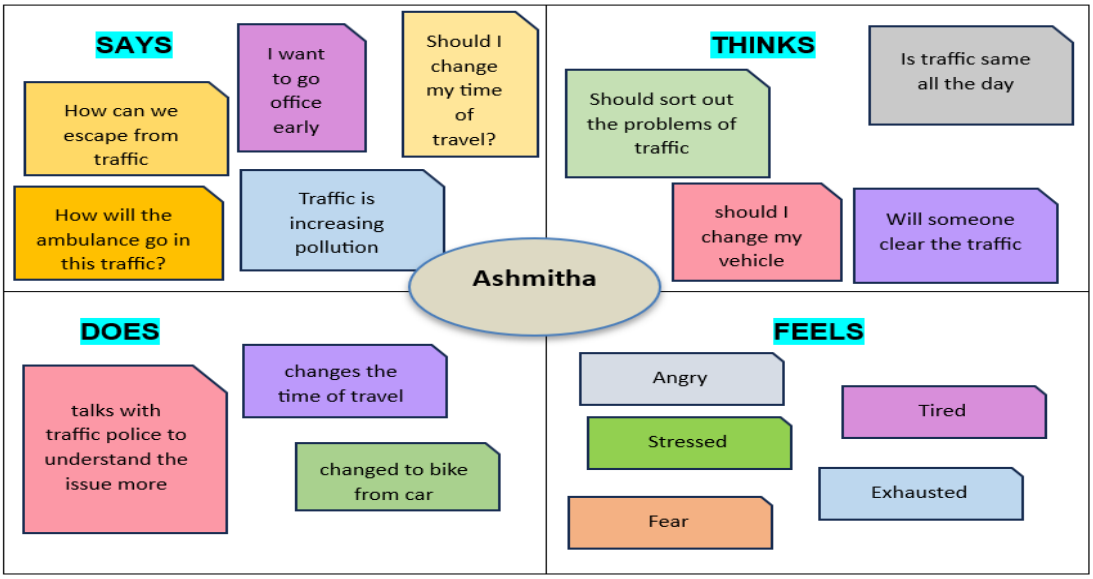
Empathy Map for Anakha



Online tool to create Empathy Map
https://online.visual-paradigm.com/diagrams/features/empathy-map-template/
Reference:https://www.ibm.com/design/thinking/page/toolkit/activity/empathy-map

Activity

1. Ashmitha drives to her office and back daily. The Office is hardly a 30 minutes' drive from her home. However, due to traffic jams, it takes more than 1 hour. Ashmitha is hoping for a solution to this traffic issue. Prepare an Empathy map related to Ashmitha.



ii) Define

In the Define stage, information collected during Empathize is used to draw insights and is instrumental in stating the problem that needs to be solved. It's an opportunity for the design thinker to define the challenge or to write the problem statement in a human-centered manner with a focus on the unmet needs of the users.

To extract / gather relevant facts and information from users/customers, it is recommended to use this simple and reliable method of questioning: the **5W1H method**–i.e. asking questions starting with Who? What? When? Where? Why? and How?

For Example,

For the traffic issue in a city, we could define the problem in detail after getting the answers of 5W1H questions. Sample questions about the problem are given below:

Who ?	What ?	Where ?
• Who are the sufferers? • Who is responsible? • Who can control it? • Who all are involved in it? • Who needs immediate help? • Who questions it? • Who identifies it?	• What can be done? • What should be done? • What is the issue? • What causes the issue? • What steps are necessary? • What to be avoided? • What may be the reason?	• Where is the problem seen? • Where else it can occur? • Where to start the solution? • Where to concentrate more? • Where are the issues severe? • Where can we ignore issue? • Where is solution impossible
When ?	**Why ?**	**How ?**
• When does traffic occurs? • When can we start clearing the traffic? • When is it more problematic • When to think about a quick solution? • When is traffic ignorable?	• Why traffic occurs? • Why it becomes a problem? • Why should we find a solution to it? • Why to clear traffic in that city? • Why can't we ignore traffic?	• How will we find a solution? • How to deal with the traffic? • How to manage the users affected by traffic? • How to solve without creating more issues? • How to identify the traffic?

Activity

1. Considering Ashmitha's traffic issue answer the following questions given below to help define the problem.

a. Who is the sufferer of the Problem?

All the people who travel through the traffic area.

b. What is the Problem?

Due to heavy traffic, people are not able to reach the office on time, all are wasting a lot of time.

c. Where does it happen?

In busy streets, places with bad roads, Accident prone areas.

d. When can it happen usually?

Especially during the morning and evening hours when schools/offices start and ends.

e. Why does it happen?

It may happen due to many reasons such as:
a. Bad Roads, b. Accidents, c. Office Hours, d. Rash Driving, e. Inappropriate Signals, f. Over Crowded Area

f. How can it be solved?

It can be solved by developing a system which could prevent intense traffic by giving some suggestions

iii) Ideate

By now the problem is obvious and it is time to brainstorm ways and methods to solve it. At this stage, numerous ideas are generated as a part of the problem-solving exercise. In short, ideation is all about idea generation. During brainstorming, one should not be concerned if the ideas generated are possible, feasible, or even viable. The only task of the thinkers is to think of as many ideas as possible for them. It requires "going wide" mentally in terms of concepts and outcomes. There are many brainstorming tools that can be used during this stage. Here is an overview of the most essential ideation techniques employed to generate numerous ideas.

Brainstorm

During a Brainstorming session, students leverage the synergy of the group to generate new innovative ideas by building on others' ideas. Participants should be able to discuss their ideas freely without fear of criticism. Many ideas are collected so that different options are available for solving the challenge.

Brain dump

Brain dump is very similar to Brainstorm; however, it's done individually. It allows the concerned person to open the mind and let the thoughts be released and captured onto a piece of paper. The participants write down their ideas onto paper or post-it notes and share their ideas later with the larger group.

Brain writing

Brain writing is also very similar to a Brainstorm session and is known as 'individual brainstorming'. At times only the most confident of team members share their ideas while the introverts keep the ideas to themselves. Brainwriting gives introverted people time to write them down instead of sharing their thoughts out loud with the group. The participants write down their ideas on paper and, after a few minutes, pass on their own piece of paper to another participant who then elaborates on the first person's ideas and so forth. In this way all participants pass their papers on to someone else and the process continues. After about 15 minutes, the papers are collected and posted for instant discussion.

During brainstorming, one should not be concerned if the ideas generated are possible, feasible, or even viable. The only task of the thinkers is to think of as many ideas as possible for them. It requires "going wide" mentally in terms of concepts and outcomes. It is all about creativity and imagination; all types of ideas are encouraged, whether stupid or wise – it hardly matters as long as the solution is imagined.

After everyone shares their ideas, specific measures are applied to evaluate the ideas without being judgmental or critical to narrow the list. It may so happen that the solution comes from the unlikeliest of ideas. So, at this point focus is on quantity over quality of ideas. The most feasible ideas are chosen for further exploration. Storyboarding, or making a visual mock-up of an idea, can also be useful during ideation.

Activity

1. For the traffic problems in a city, brainstorm and generate different ideas as a team.

Some of the ideas generated by brainstorming a team is given below

1. Give alternate route suggestions through Digital boards kept at important junctions identifying the traffic so that it could control traffic getting worsen
2. Make some system so that the percentage of road issues and the severity of traffic issue can be predicted. It may be seriously taken by officials and they will consider repairing of roads.
3. Keep track of rash driving persons. If more than one time the rash driving person causes traffic issues, his/her license can be cancelled.
4. Make some systems to clear the area immediately when accidents occur, so that traffic doesn't cause issue
5. Signaling system should be rearranged if found inappropriate
6. Announce the traffic areas live, so other people won't drive into traffic area
7. Introduce Intelligent Traffic Signal.

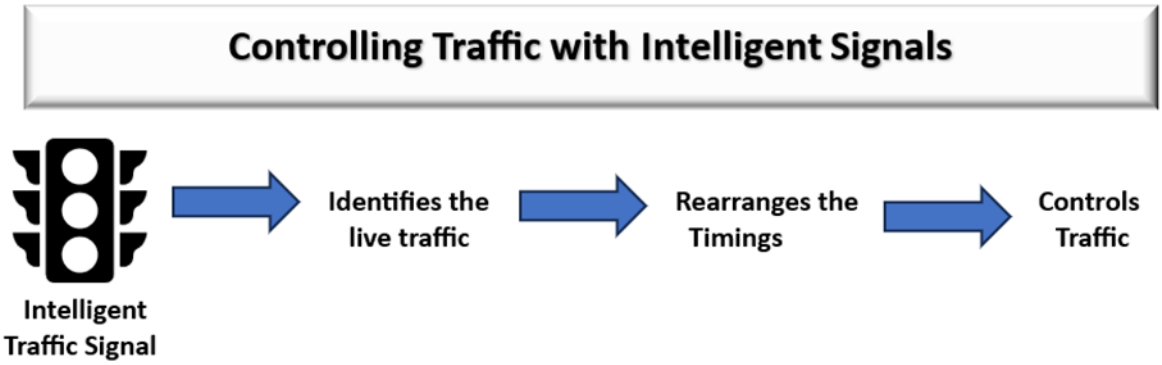
[All these are ideas and does not require to be implemented]

iv) Prototype

The prototype stage involves creating a model designed to solve consumers' problems which is tested in the next stage of the process. Creating a prototype is not a detailed process. It may include a developing simple drawing, poster, group role-playing, homemade "gadget, or a 3d printed product." The prototypes must be quick and easy to develop and cheap. Therefore, prototypes are visualized as rudimentary forms of what a final product is expected to look like. Prototyping is intended to answer questions that get you closer to your final solution. Prototypes, though quick and simple to make, bring out useful feedback from users. Prototypes can be made with everyday materials also.

Activity

1. Through a poster or simple drawing depict a prototype for any one solution for the traffic issue.



v) Test

One of the most important parts of the design thinking process is to test the prototypes with the end users. This step is often seen going parallel to prototyping. During testing, the designers receive feedback about the prototype(s) and get another opportunity to interact and empathize with the people they are finding solutions for. Testing focuses on what can be learned about the user and the problem, as well as the potential solution.

Activity

1. Show the prototype made by you to your parents/teachers/relatives who usually get stuck in traffic issues. Collect the feedback from them and understand whether your prototype will work or not. Write down the views your users shared and the conclusion you arrived.

Response after Testing the Prototype:

- Very good if it controls the traffic
- What happens if from all sides equal traffic block comes in a road
- Will be good before traffic jam.

3. Sustainable Development Goals

Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs. The global community is witnessing a transition towards a sustainable future. To achieve a sustainable future, 17 goals have been announced by the United Nations which are termed as the Sustainable Development Goals.

Projects undertaken all over the world play a crucial role in fulfilling this task by aligning their strategies and vision with the sustainable development goals and targets.

Students can select any problem based on SDG to find an AI solution for it.

For example:

SDG 2 was adopted to achieve “Zero Hunger”. That means it is essential to recognize that the successful transformation of our current food systems (shift to a sustainable food system) should tackle the main global challenges in a sustainable, healthy, and inclusive manner. Healthy food habits, food Supply to all areas, consumption of food as per the production, Seasonal Foods etc. are the areas where students can concentrate for identifying problems and try to derive simple AI based solution.

Some other use cases of the SDGs are given below

- Good health and well being
- Gender equality
- Renewable energy
- Combating hunger and poverty
- Protecting resources
- Promoting responsible consumption and production
- Providing access to education
- Sustainable and safe cities and communities

Activity

1. Write down a few problems which you notice from any five SDGs which can be taken up as a Capstone project.

a. Quality Education:

i. Providing education remotely, leveraging hi-tech, low-tech and no-tech approaches;

ii. Ensure coordinated responses and avoid overlapping efforts;

iii. Ensuring return of students to school when they reopen to avoid an upsurge in dropout rates.

b. Reduced inequalities:

i. Reduction of relative economic inequalities inequality in some countries having poorest and most vulnerable communities.

ii. Improving the situations in countries with weaker health systems.

c. Life on Land:

i. Prevention of Deforestation caused by humans and restoration of land

ii. Preventions and cure of diseases that are transmissible between animals and humans

d. No Poverty

i. Creation of Strong social protection systems to prevent people from falling into poverty

ii. Reduction of social exclusion, and high vulnerability of certain populations to disasters and diseases.

iii. Responsible distribution of resources.

e. Clean Water and Sanitation

i. To increase access to clean drinking water and sanitation mostly in rural areas

ii. Managing our water sustainably to manage our production of food and energy

4. Project Abstract Creation Using Design Thinking Framework

We can now choose a problem (capstone project) and check how it can be detailed using Design Thinking Framework.

Format for Project Abstract Creation

1. Project Name:
2. Team Members Name:
3. Problem Selection –Important issue which you think can be solved using AI by doing survey or interview and which is aligned to any of the SDGs.
4. Users Affected by the Problem
5. Empathise – create Empathy Map
6. Define the Problem – Write down the 5W1H questions and write the problem statement
7. Ideate – Brainstorm and write down the ideas
8. Prototype- create prototype by simple drawing, poster, group role-playing, homemade "gadget, or a 3d printed product."
9. Test (optional at this stage)

Now let us see an example

1. Project Name: {Suitable name related to your problem and your creativity}

2. Team Members Name: { Students can fill by themselves}

3. Problem Selection: After doing a survey amongst the school students who are in Class X, it was observed that most of the students are confused about the subjects that they have to choose in future.

Details about the issue:

While taking admission to plus-two courses, students are confused with subject selection. To provide quality education students should be provided with the subjects in which they have aptitude. A system to analyze their interests, choose subjects from the available options and choose the institutions which provide those subjects are a big problem.

After seeing the issue, it was aligned to SDG-4, Quality Education.

4. Users:

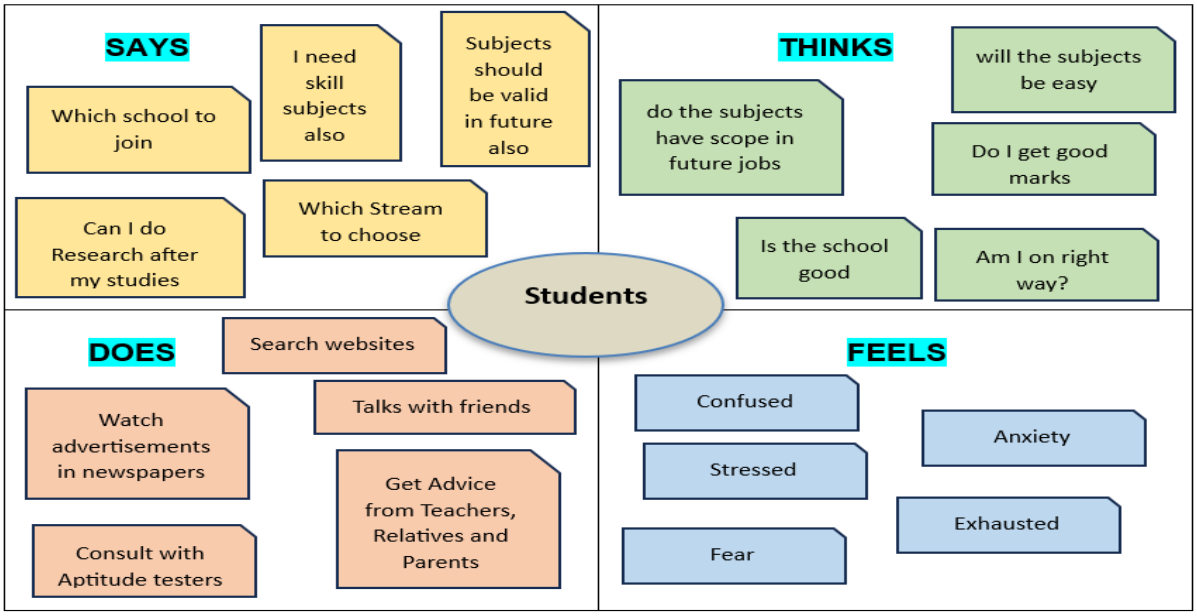
All the students who pass class X face the issue of selecting an appropriate subject for their higher studies as it is their turning point to move on according to their interest.

5. Empathise:

Two categories of students can be met and interviewed.

- First—Those students who are studying in plus-two (to know how difficult it was to choose a subject and a school)
- Second—Those students who are going to take admission (can know about their anxieties and expectations)

After the discussion with students, Team members can sit together and finalize the points they understood from the users(students). They can write each in a chit of paper, classify it into the categories SAYS, THINKS, DOES & FEELS. Now they can create an Empathy Map. A sample empathy map is given below.



6. Define

After Empathetically understanding the students, the problem may be defined.

Get the Answers of 5W1H Questions. Write down the Problem Statement.

5W1H Questions & Answers

Who?

1. Who are the users?

Students seeking admission in plus-two course

2. Who else can be the users?

Parents/well-wishers who seek admission for their wards

3. What do you know about the users?

Users are students who are interested in taking plus-two admissions. Students have well defined future plans. Now they must select the appropriate subject as per their plan.

What?

1. What is the Problem?

Not able to select the subject of their choice

2. What else are the Problems?

They can't choose the school which provides these subjects.

Where?

1.Where is the problem/ In which situation does the problem arises.?

All the places where admission takes place & all the place where students wish to take admission the problem persists.

When?

1. When does the problem occur?

Problem occurs during the time of admission

Why?

1. Why does the problem occur?

Problem occurs due to the lack of knowledge of subjects available as part of the curriculum

2. Why should it be solved?

Only if students pursue the course, they have aptitude, a well-developed society can be built in future.

How

1. How can we solve it?

Through a system which will analyze a student's aptitude and suggests the subject to take. After checking the database, the system can suggest the schools giving the subject combinations also.

2. How will the solution improve the situation?

The solution could help students in choosing the appropriate subjects and could give effective education.

PROBLEM STATEMENT

Our students have the problem of not being able to choose the subjects of their interest while taking admission to their plus-two course and it can be solved by developing a system which will analyze students' aptitude and suggest the subject to take.

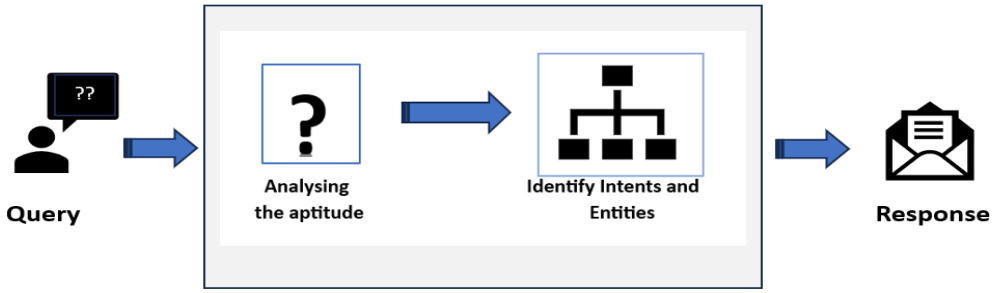
7. Ideate

As a team students can discuss and put forward different ideas to implement the solution

1. Create an application to input the interest and get suggestions on subjects
2. Create a chatbot to converse with
3. Make a robot to discuss with and give advice
4. Create a mobile application where through different sources queries can be given and the application will give the result.

8. Prototype

Select any one idea and prepare the prototype of the solution - Here we select the idea of creating a chatbot to give the solution



Prototype of Working of a chatbot

Details of intents and entities can be drawn separately to display the prototype in detail. (Leaving it to the creativity of students)

Note for Teachers

- Students can elaborate any part of the Prototype if they wish.
- Students should Write some intents and its entities and the expected outcomes. They can compare these expected outcomes with others opinion who checks these. Thus, testing phase can be conducted. (if needed)

EXERCISES

A. OBJECTIVE TYPE QUESTIONS

1. Which is NOT correct about the Capstone project?
    a. It is a project where students must research independently
    b. Students can integrate all their knowledge to develop a solution to the problem
    c. It is the final project of an academic program.
    d. Students are made to look at real world situations, exchange their point of view and discuss potential solutions to the problem.

2. Which is the fundamental question which is asked before deciding to embark on an AI development journey?
    a. Should we do the project?
    b. Is there a solution?
    c. Can we leave the problem?
    d. Is there a pattern?
3. The question, 'which option should be taken?' comes under which category?
    a. Classification
    b. Regression
    c. Anomaly detection
    d. Recommendation
4. Which stage in Design Thinking is missing (Prototype, Ideate, Test, Define)?
    a. Evaluation
    b. Empathise
    c. Evolution
    d. Enrichment
5. Real computational tasks are complicated. To accomplish them, you need to _____ before coding.
    a. Test the problem causing variables
    b. Ask for finance help from stakeholders
    c. Break down the problem into smaller units
    d. Write it in a paper and read it to your partners
6. Which one is not a part of the 5W1H method?
    a. Who
    b. While
    c. Where
    d. When
7. _____ is the process of generating ideas and solutions through sessions such as sketching, brainstorming etc.
    a. Define
    b. Empathise
    c. Ideate
    d. Prototype
8. Which one among the following is known as individual brainstorming?
    a. Brainwriting
    b. Brain dump
    c. Brainstorm
    d. Brain explore

9. _____ is a useful process during ideation.
    a. Storyboarding
    b. Empathy map
    c. Prototyping
    d. Data collection
10. The _____ stage involves creating a model designed to solve consumer's problems in Design Thinking.
    a. Modeling
    b. Prototype
    c. Training
    d. Deployment
11. _____ Analysis helps in identifying the patterns by asking different questions which fall into different approaches.
    a. Predictive
    b. Prescriptive
    c. Probability
    d. Prototype

ANSWERS

1. a. It is a project where students must research independently
2. d. Is there a pattern?
3. d. Recommendation
4. b.Empathise
5. c. Break down the problem into smaller units
6. b.While
7. c.Ideate
8. a.Brain Writing
9. a.Storyboarding
10. b. Prototype
11. a.Predictive

B. Short Answer Questions

1. What do you mean by a capstone project

The final project of an academic program, typically integrating all of the learning from the program is called the Capstone Project. A capstone project is a project where students must research a topic, deeply understand the subject matter and integrate all their knowledge to develop a solution to a problem.

2. Name some of the skills which students will acquire by doing capstone project

While doing a Capstone project, students will learn many important skills including:

- how to work as a team member
- how to clearly identify an issue and who it is affecting (the user)
- how to brainstorm solutions and select the best one
- how to decide which type of AI may be useful for the proposed solution
- how to ethically gather and use data to train a computer to help solve the issue
- how to test the prototype with users and use their feedback to improve the solution
- how to pitch their solution to people who will be able to help them take action

3. Give some examples of the use case of SDG's

Some use cases of the SDGs are given below

- Renewable energy
- Reporting and communicating
- Combating hunger and poverty
- Protecting resources
- Promoting responsible consumption
- Providing access to education

4. Students are promoted to develop critical and creative thinking. Why?

Students must be taught to think critically and creatively so that they could communicate and work together effectively. They could raise vital questions against problems, gather and assess relevant information, brainstorm ideas and formulate a well-reasoned conclusions and solutions. Creative thinking and critical thinking are necessary skills for equipping individuals to be the social change makers, leaders and innovators.

5. Define Design Thinking

Design Thinking is a design methodology that provides a solution-based approach to solving problems. It's extremely useful in tackling complex problems that are ill-defined or unknown.

C. Long Answer Questions

1. Write down the steps of Problem decomposition

Problem decomposition steps

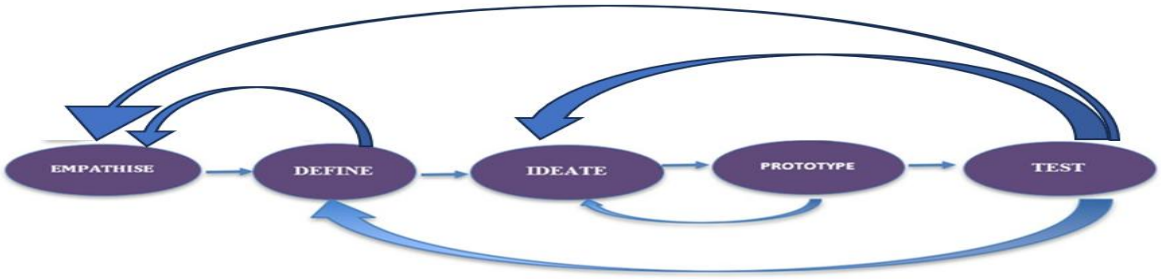
1. Understand the problem and then restate the problem in your own words
   - Know what the desired inputs and outputs are
   - Ask questions for clarification (in class these questions might be to your instructor, but most of the time they will be asking either yourself or your collaborators)
2. Break the problem down into a few large pieces.
   - Write these down, either on paper or as comments in a file.
3. Break complicated pieces down into smaller pieces.

- Keep breaking down until all of the pieces are small.

4. Code one small piece at a time.
   - Think about how to implement it and write the code/query
   - Test it on its own and Fix problems, if any

2. Explain briefly the different stages of Design thinking

There are five stages of Design Thinking. They are: Empathize, Define, Ideate, Prototype, and Test.



i)Empathize:
Design thinking begins with empathy. It allows designers to understand, empathize and share the feelings of the users. As a designer of the solution to a challenge, one should always understand the problem from the end-user perspective. This is done by observation, interaction or by imagination.
Designers interact with customers / users very frequently to gather detailed facts about the problems and user's expectations. A detailed analysis of these facts leads to approaching the problem in the best possible way.

ii) Define
In the Define stage, information collected during Empathize is used to draw insights and is instrumental in stating the problem that needs to be solved. It's an opportunity for the design thinker to define the challenge or to write the problem statement in a human-cantered manner with a focus on the unmet needs of the users. 5W1H method of asking questions can be implemented here.

iii) Ideate
Ideation is the process of generating ideas and solutions through sessions such as sketching, brainstorming etc. At this stage, numerous ideas are generated as a part of the problem-solving exercise. In short, ideation is all about idea generation. Brainstorm, Brain write and Brain dump can be used for ideate stage.

iv) Prototype
The prototype stage involves creating a model designed to solve consumers' problems which is tested in the next stage of the process. Creating a prototype may include a developing simple drawing, poster, group role-playing, homemade "gadget, or a 3d printed product." The prototypes must be quick and easy to develop and cheap. Therefore, prototypes are visualized as rudimentary forms of what a final product is expected to look like.

v) Test

During testing, the designers receive feedback about the prototype(s), and get another opportunity to interact and empathize with the people they are finding solutions for. Testing focuses on what can be learned about the user and the problem, as well as the potential solution.

3. Explain the different types of Ideation techniques

Different ideation techniques are brainstorm, brain dump and brain writing

Brainstorm

During a Brainstorming session, students leverage the synergy of the group to generate new innovative ideas by building on others' ideas. Participants should be able to discuss their ideas freely without fear of criticism. A large number of ideas are collected so that different options are available for solving the challenge.

Brain dump

Brain dump is very similar to Brainstorm; however, it's done individually. It allows the concerned person to open the mind and let the thoughts be released and captured onto a piece of paper. The participants write down their ideas onto paper or post-it notes and share their ideas later with the larger group.

Brain writing

Brain writing is also very similar to a Brainstorm session and is known as 'individual brainstorming'. At times only the most confident of team members share their ideas while the introverts keep the ideas to themselves. Brainwriting gives introverted people time to write them down instead of sharing their thoughts out loud with the group. The participants write down their ideas on paper and, after a few minutes, pass on their own piece of paper to another participant who then elaborates on the first person's ideas and so forth. In this way all participants pass their papers on to someone else and the process continues. After about 15 minutes, the papers are collected and posted for instant discussion.

D. Case Study

1. Based on SDG-7, Affordable and clean energy, formulate a Problem and write down the Design thinking steps associated with it

SDG-7, Affordable and clean energy

Problem – Inefficient usage of Electric/Electronic Appliances and wastage of energy

Design Thinking Steps

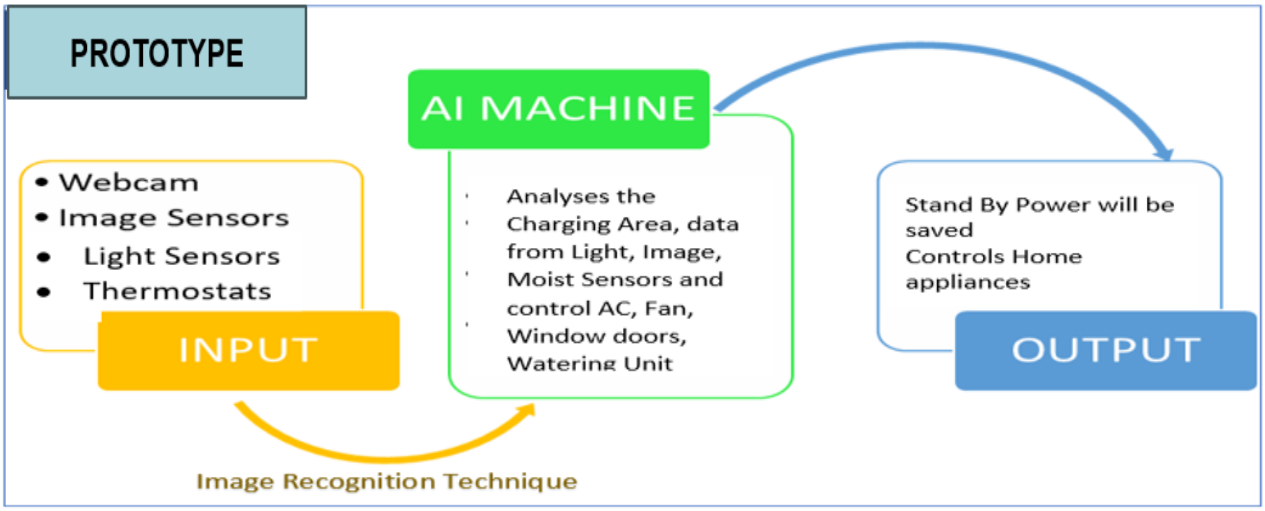
Empathize: Talk with people and understand how they are using the electric/electronic devices. Are they using it whole day or in a limited. Do they know how to use. Whether they are using it smoothly or roughly without knowing the usage. Understand whether there is wastage of energy and people are aware of it.

Define: To particularly understand the problem, ask 5W1H questions and with the answers of those questions understand that the problem to be solved is "Not knowing the efficient usage of Appliances"

Ideate: Brainstorm and list down the ideas generated. From these ideas one of the ideas can be selected as solution for implementation

Solution Selected: An AI machine can be made which will monitor all the appliances and controls the usage of electricity.

Prototype: The prototype of the solution can be drawn like this



Test: This Prototype can be tested by analysing it with Parents and Teachers. Feedback collected could improve the previous stages.

E. Competency Based Questions

1. As you near the completion of your Artificial Intelligence program, you are eager to put your newfound skills to good use. You have always been fascinated by the potential of AI to improve healthcare. Recently, you have learned about the challenges of early detection for certain diseases, leading to poorer prognoses. What do you call the project in which you must research a topic, deeply understand the subject matter, and integrate all their knowledge gathered about a particular subject to develop a solution to a problem?

Ans: Capstone Project

2. Shilpa Balagangadharan, a psychology teacher, is concerned about the negative impact of low exam scores on student mental health. To address this, she is implementing a multi-step process to develop and pilot a program that promotes resilience and reduces feelings of depression. Here, the tasks are NOT arranged in a logical order:
   a. Conduct interviews and focus groups with students, parents, teachers, and mental health professionals.
   b. Develop mockups of potential solutions like support groups, scripts for intervention training, or pilot programs for mindfulness exercises.

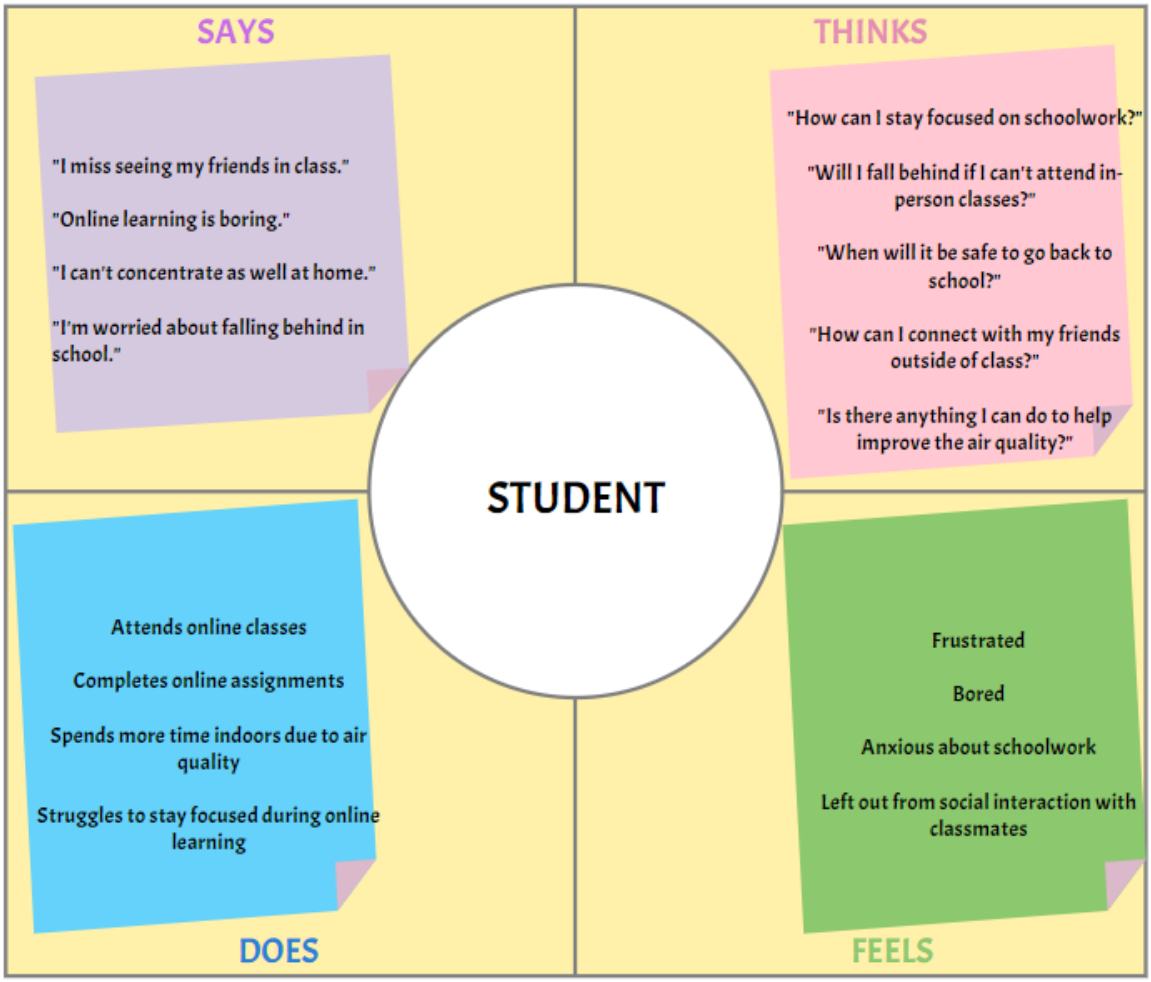
c. Arrange peer support programs based on the feedback and best practices identified. Create a "help-seeking is cool" campaign to encourage students to utilize available resources. Continuously monitor and adapt the program based on its effectiveness.
d. Analyze the empathy map to define the core problem statement.
e. Implement the training materials (scripts, mockups) in small groups and gather feedback from participants.

Arrange the steps in the correct order.

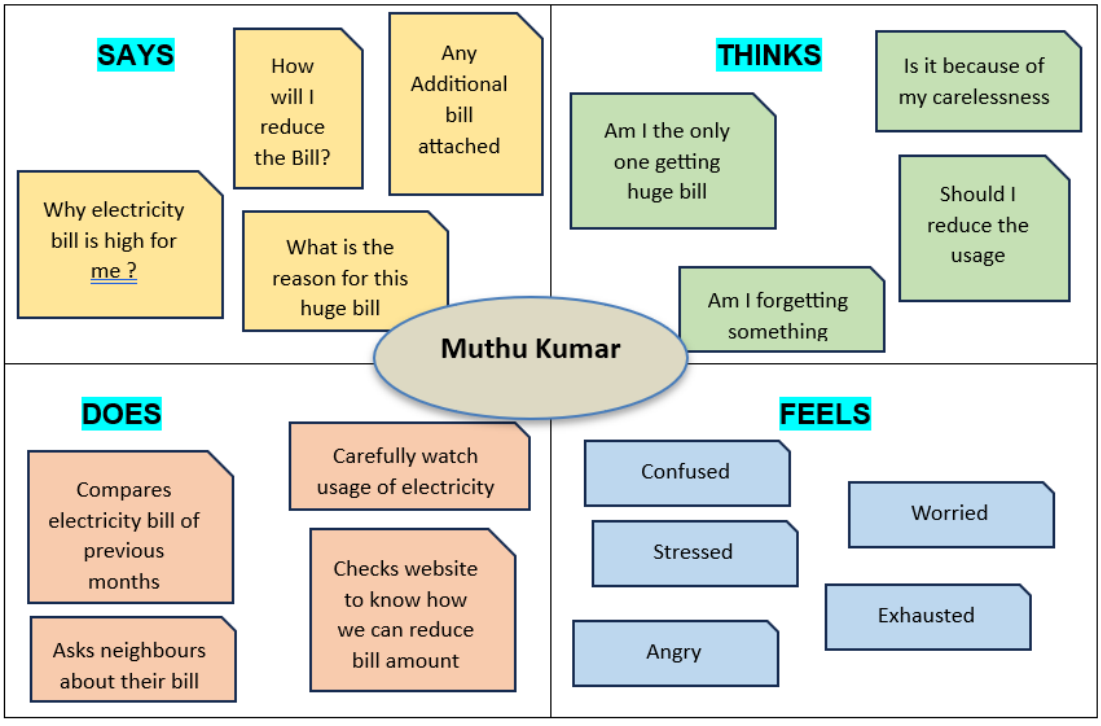
Ans- The correct order for the steps in Shilpa's student wellbeing initiative, aligned with the design thinking process, is:

a. Conduct interviews and focus groups with students, parents, teachers, and mental health professionals. (**Empathize** Stage)
d. Analyze the empathy map to define the core problem statement. (**Define** Stage)
b. Develop mockups of potential solutions like support groups, scripts for intervention training, or pilot programs for mindfulness exercises. (**Ideate** Stage)
e. Implement the training materials (scripts, mockups) in small groups and gather feedback from participants. (**Prototype** Stage)
c. Arrange peer support programs based on the feedback and best practices identified. Create a "help-seeking is cool" campaign to encourage students to utilize available resources. Continuously monitor and adapt the program based on its effectiveness. (**Test** Stage)

3. You live in Darjeeling, a beautiful hill station. Due to heavy rains, landslides have damaged the roads, leaving the town temporarily isolated. Your school is closed, and online learning is the only option available. How do you feel about not being able to attend school in person and facing challenges with remote learning? Draw an empathy map to illustrate your thoughts, feelings, and experiences during this unexpected situation.



4. Mr. Muthu Kumar is facing a serious issue of high electricity bill every month. He is trying a lot to reduce it. Considering his issue is worth solving, draw an empathy map related to Muthu Kumar.



5. *A small island nation in the Pacific Ocean, known for its beautiful beaches and colorful coral reefs, is in trouble. The sea level is rising, which is causing the shores to erode and making the land salty where plants used to grow. Most of the buildings and roads are near the coast, so they are at risk from the rising sea and strong storms. Because of these problems, it is harder for people to grow food like they used to, and they might not have enough to eat.*
   Which Sustainable Development Goals (SDGs) can you align the above scenario with?

Ans:

- Zero Hunger (SDG 2)
- Sustainable Cities and Communities (SDG 11)
- Climate Action (SDG 13).

UNIT 5: Data Literacy – Data Collection to Data Analysis

Title: Data Literacy - Data Collection to Data Analysis	**Approach:** Team Discussion, Web search**,** Case studies
Summary: This unit will introduce students to the basics of data literacy, focusing on data collection and its sources, Level of Measurements, Statistical analysis of data, Matrices and Data preprocessing. Students will learn how to collect different types of data, how to store data effectively and visualise it.	
Learning Objectives: 1. To understand the importance of data literacy in AI. 2. To explore various data collection methods and their applications. 3. To analyse data using basic Statistical analysis techniques. 4. To identify matrices and their role in representing data like images. 5. To understand the preparation of data to suit the models.	
Key Concepts: 1. What is Data Literacy? 2. Data Collection 3. Exploring Data 4. Statistical Analysis of data 5. Representation of data, Python Programs for Statistical Analysis and Data Visualization 6. Introduction to matrices 7. Data Preprocessing 8. Data in Modelling and Evaluation	
Learning Outcomes: Students will be able to - 1. Explain the importance of data literacy in AI. 2. Identify different data collection methods and their applications. 3. Apply basic data analysis techniques to analyse data. 4. Visualize the data using different techniques.	
Pre-requisites: Basic computer skills and basic maths skills	

Unveiling the Many Faces of Data: A Teacher's Guide

This lesson equips you to introduce students to the diverse world of data and its analysis in Python.

Demystifying Data Types:

Before diving into projects, establish a strong foundation by exploring data types:

- **Nominal, Ordinal, Interval, Ratio:** Unveil the different scales of measurement for data points. Nominal (category labels), Ordinal (ranked order), Interval (consistent units but no true zero), and Ratio (true zero allows meaningful comparisons).
- **Quantitative vs. Qualitative:** Distinguish between data that is numerical (quantitative) and data that describes characteristics (qualitative).
- **Discrete vs. Continuous:** Help students differentiate between data with distinct values (discrete) and data that can take any value within a range (continuous).
- **Structured, Semi-structured, Unstructured:** Introduce the various data organization formats. Structured data follows a predefined schema, semi-structured allows some flexibility, and unstructured data has no fixed format.

The Cyclical Nature of Data Projects:

Emphasize the iterative nature of data projects. Data is used at every stage of the cycle:

- **Problem Definition:** Data informs the problem you're trying to solve.
- **Data Collection:** Gathering the right data is crucial for analysis.
- **Data Cleaning and Preparation:** Data is processed to ensure accuracy and usability.
- **Data Analysis and Exploration:** Data is analysed to uncover patterns and insights.
- **Visualization and Communication:** Findings are communicated through visualizations.

Coding for Data Analysis:

- **Python Environments:** Introduce Python programming environments like IDLE or Jupyter Notebook. Jupyter Notebook provides a user-friendly interface for interactive coding and visualization.

Essential Libraries:

- **Statistics Library:** Introduce the 'statistics' library for performing common statistical calculations on data (e.g., mean, median, standard deviation).
- **Matplotlib Library:** Showcase the 'matplotlib' library for creating various data visualizations (e.g., histograms, scatter plots) to represent data insights effectively.

Critical Thinking in Data Analysis:

Nurturing critical thinking is vital for responsible data analysis. Encourage students to:

- **Question Assumptions:** Challenge underlying assumptions in data collection and analysis.
- **Evaluate Sources:** Critically assess the credibility and potential biases of data sources.
- **Consider Alternative Explanations:** Explore other interpretations of data to avoid oversimplification.

The Power of Scepticism and Curiosity:

Emphasize the importance of scepticism and intellectual curiosity throughout the data analysis process. These qualities pave the way for robust and unbiased insights.

1. WHAT IS DATA LITERACY?

Teachers can ask the following questions:

- **Think about all the information you encounter daily – online, in books, from friends. Can you categorize this information in any way? Is it all the same?** (This question primes students to think about data as information and how it can be presented in different forms.)
- **Imagine you're trying to solve a mystery. What kind of clues would you need to gather and analyse to figure things out?** (This question connects data analysis to real-world problem-solving, making it relatable to students.)

Data can be defined as a representation of facts or instructions about some entity (students, school, sports, business, animals etc.) that can be processed or communicated by human or machines. It is a widely known fact that Artificial Intelligence (AI) is essentially data-driven. AI involves converting large amounts of raw data into actionable information that carry practical value and is usable.

Data literacy means being able to find and use data effectively. This includes skills like collecting data, organizing it, checking its quality, analysing it, understanding the results and using it ethically.

Data may be **structured, semi structured or unstructured**. It should be collected, organized and analysed properly to know whether the input for AI models is valid and appropriate or not. AI data analysis involves using AI techniques and data science to improve the processes of cleaning, inspecting, and modelling both structured and unstructured data. The primary objective is to extract valuable information that can support decision-making and drawing conclusions.

2. DATA COLLECTION

Teachers can ask the following questions:

- **Think about your favourite movie recommendation platform. How do you think they use data to suggest movies you might like?** (This question primes students to think about data collection for predictions in a familiar context)
- **Imagine you wanted to build a robot that could sort recycling. What kind of data would you need to collect to train it to recognize different materials?** (This question introduces the concept of data collection for a specific purpose and the importance of diverse data.)

Data collection allows you to capture a record of past events so that we can use data analysis to find recurring patterns. From those patterns, you build predictive models using machine learning algorithms that look for trends and predict future changes.

Data collection means pooling data by scraping, capturing, and loading it from multiple **sources**, including offline and online sources. High volumes of data collection or data creation can be the hardest part of a machine learning project, especially at scale.

How much data you need depends on how many features there are in the data set. It is recommended to collect as much data as possible for good predictions. You can begin with small batches of data and see the result of the model. The most important thing to consider while data collection is diversity. Diverse data will help your model cover more scenarios. So, when focusing on how much data you need, you should cover all the scenarios in which the model will be used.

The **quantity of data** also depends on the complexity of your model. If it is as simple as license plate detection then you can expect predictions with small batches of data. But if you are working on higher levels of Artificial intelligence like medical AI, you need to consider huge volumes of data.

Before collecting the data, data scientists may understand the problem, its preferable solution and the data requirements. Based on these data requirements, sources of data will be identified and data will be collected. Data is the main ingredient of any Project. throughout the development of the project, data is required. Hence the process of identifying the data requirements, its collection and analysis will be done iteratively.

There are mainly two sources of data collection: Primary and Secondary.

Primary Sources are sources which are created to collect the data for analysis. Some of the examples are given below

Method	Description	Example
Survey	Gathering data from a population through interviews, questionnaires, or online forms. Useful for measuring opinions, behaviors, and demographics.	A researcher uses a questionnaire to understand consumer preferences for a new product.
Interview	Direct communication with individuals or groups to gather information. It can be structured, semi-structured, or unstructured.	An organization conducts an online survey to collect employee feedback about job satisfaction.
Observation	Watching and recording behaviors or events as they occur. Often used in ethnographic research or when direct interaction is not possible.	Observing children's play patterns in a schoolyard to understand social dynamics.
Experiment	Manipulating variables to observe their effects on outcomes. Used to establish cause-and-effect relationships.	Testing the effectiveness of different advertising campaigns on a group of people.

Marketing Campaign (using data)	Utilizing customer data to predict behavior and optimize campaign performance.	A company personalizes email marketing campaigns based on past customer purchases.
Questionnaire	A specific tool used within surveys - a list of questions designed to gather data from respondents. You can collect quantitative (numerical) or qualitative (descriptive) information.	A questionnaire might ask respondents to rate their satisfaction on a scale of 1 to 5 and also provide open-ended feedback.

Secondary data sources are where the data is already stored and ready for use. Data given in Books, journals, News Papers, Websites, Internal transactional databases, etc can be reused for data analysis. Some methods of collecting secondary data are:

Method	Description	Example
Social Media Data Tracking	Collecting data from social media platforms like user posts, comments, and interactions.	Analyzing social media sentiment to understand audience reception towards a new product launch.
Web Scraping	Using automated tools to extract specific content and data from websites.	Scraping product information and prices from e-commerce websites for price comparison.
Satellite Data Tracking	Gathering information about the Earth's surface and atmosphere using satellites.	Monitoring weather patterns and environmental changes using satellite imagery.
Online Data Platforms	Websites offering pre-compiled datasets for various purposes.	Kaggle, GitHub etc.

3. EXPLORING DATA

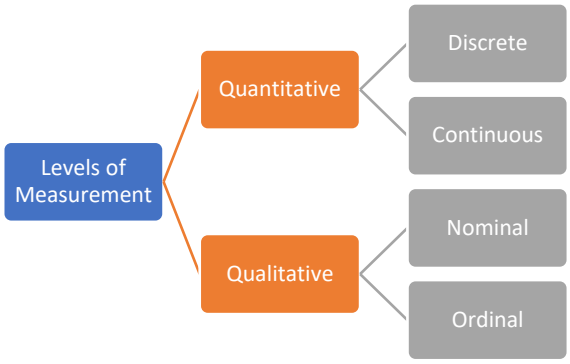
Teachers can ask the following questions:

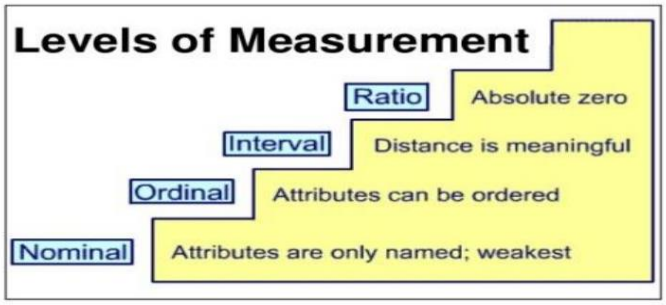
- Imagine you're collecting data on student's favourite movie genres. Could you rank the genres from most to least popular (ordinal)? Or would you just be able to say which genre is the favourite (nominal)?
- When measuring temperature, we can say it's 20 degrees Celsius today, which is 10 degrees warmer than yesterday. Can we say it's twice as hot today (interval)? Why or why not?

Exploring data is about "getting to know" the data: and its values - whether they are typical, unusual, spread out, or whether they are extremes. More importantly, during the process of exploration one gets an opportunity to identify and correct any problems in the data that would affect the conclusions drawn in any way during analysis.

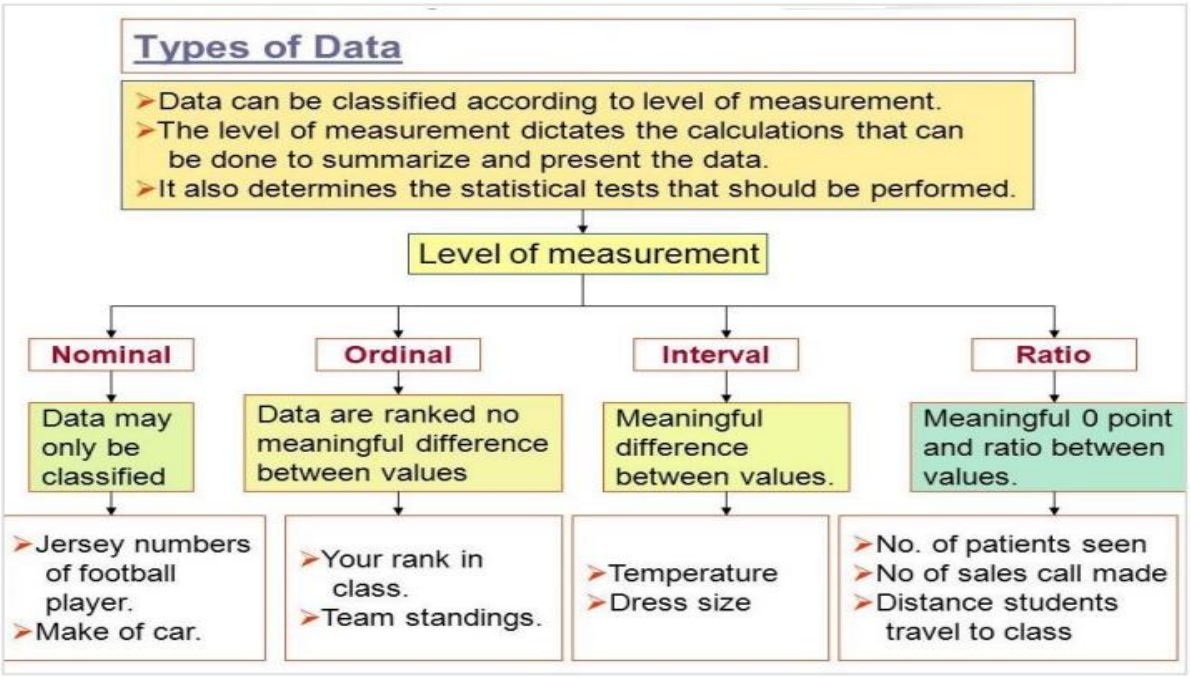
Levels of Measurement

The way a set of data is measured is called the level of measurement. Not all data can be treated equally. It makes sense to classify data sets based on different criteria. Some are quantitative, and some qualitative. Some data sets are continuous and some are discrete. Qualitative data can be nominal or ordinal. And quantitative data can be split into two groups: interval and ratio.





https://slideplayer.com/slide/8137745/



1. Nominal Level

Nominal variables are like categories such as Mercedes, BMW or Audi, or like the four seasons – winter, spring, summer and autumn. They aren't numbers, and cannot be used in calculations and neither in any order or rank. The nominal level of measurement is the simplest or lowest of the four ways to characterize data. Nominal means "in name only".

Colours of eyes, yes or no responses to a survey, gender, smartphone companies, etc all deal with the nominal level of measurement. Even some things with numbers associated with them, such as a number on the back of a cricketer's T-shirt are nominal since they are used as "names" for individual players on the field and not for any calculation purpose.

Example 1:
Please indicate your marital status.
___ Married ___ Single ___ Separated ___ Divorced ___ Widowed

Example 2:
Do you like or dislike chocolate ice cream?
___ Like ___ Dislike

Example 3:
Which of the following supermarkets have you shopped at in the last 30 days? Please check all that apply.
___ Albertson's ___ Winn-Dixie ___ Publix ___ Safeway ___ Walmart

Example 4:
Please indicate your gender:
___ Female ___ Male ___ Transgender

2. Ordinal Level

Ordinal data, is made up of groups and categories which follow a strict order. For e.g. if you have been asked to rate a meal at a restaurant and the options are: unpalatable, unappetizing, just okay, tasty, and delicious. Although the restaurant has used words not numbers to rate its food, it is clear that these preferences are ordered from negative to positive or low to high, thus the data is qualitative, ordinal. However, the difference between the data cannot be measured. Like the nominal scale data, ordinal scale data cannot be used in calculations. A Hotel industry survey where the responses to questions about the hotels are accepted as, "excellent," "good," "satisfactory," and "unsatisfactory." These responses are ordered or ranked from the excellent service to satisfactory response to the least desired or unsatisfactory. But the differences between the two pieces of data as seen in the previous case cannot be measured. Another common example of this is the grading system where letters are used to grade a service or good. You can order things so that A is higher than a B, but without any other information, there is no way of knowing how much better an A is from a B.

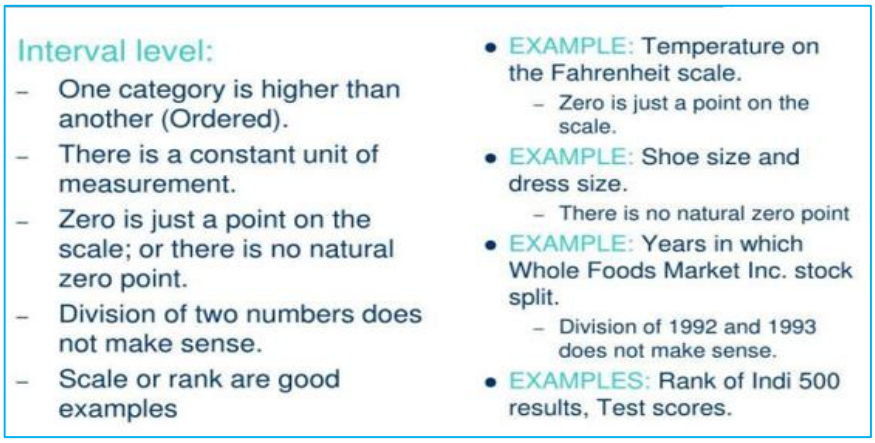
Example 1:
How likely are you to recommend the Santa Fe Grill to a friend?
Definitely Will Not Recommend 1 2 3 4 5 6 7 Definitely Will Recommend

Example 2:
Using a scale of 0–10, with "10" being Highly Satisfied and "0" being Not Satisfied At All, how satisfied are you with the banking services you currently receive from (read name of primary bank)?
Answer: _____

Example 3:
Please indicate how frequently you use different banking methods. For each of the banking methods listed below, circle the number that best describes the frequency you typically use each method.

Banking Methods	Never Use										Use Very Often
Inside the bank	0	1	2	3	4	5	6	7	8	9	10
Drive-up window	0	1	2	3	4	5	6	7	8	9	10
24-hour ATM	0	1	2	3	4	5	6	7	8	9	10
Debit card	0	1	2	3	4	5	6	7	8	9	10
Bank by mail	0	1	2	3	4	5	6	7	8	9	10
Bank by phone	0	1	2	3	4	5	6	7	8	9	10
Bank by Internet	0	1	2	3	4	5	6	7	8	9	10

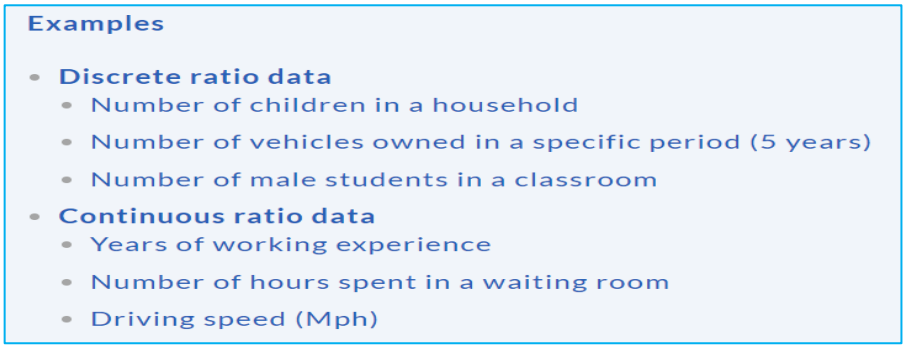
3. Interval Level



Data that is measured using the interval scale is similar to ordinal level data because it has a definite ordering but there is a difference between the two data. The differences between interval scale data can be measured though the data does not have a starting point i.e. zero value. Temperature scales like Celsius (° C) and Fahrenheit (° F) are measured by using the interval scale.

In both temperature measurements, 40° is equal to 100° minus 60°. Differences make sense. But 0 degrees does not because, in both scales, 0 is not the absolute lowest temperature. Temperatures like -20° F and -30° C exist and are colder than 0. Interval level data can be used in calculations, but the comparison cannot be done. 80° C is not four times as hot as 20° C (nor is 80° F four times as hot as 20° F). There is no meaning to the ratio of 80 to 20 (or four to one)

4. Ratio Scale Level



Ratio scale data is like interval scale data, but it has a 0 point and ratios can be calculated. For example, the scores of four multiple choice statistics final exam questions were recorded as 80, 68, 20 and 92 (out of a maximum of 100 marks). The grades are computer generated. The data can be put in order from lowest to highest: 20, 68, 80, 92 or vice versa. The differences between the data have meaning. The score 92 is more than the score 68 by 24 points. Ratios can be calculated. The smallest score is 0. So, 80 is four times 20. The score of 80 is four times better than the score of 20. So, we can add, subtract, divide and multiply the two ratio level variables. Egg: Weight of a person. It has a real zero point, i.e. zero weight means that the person has no weight. Also, we can add, subtract, multiply and divide weights at the real scale for comparisons

Activity-1

Student Health Survey – Fill in the response and mention appropriate Level of Measurement

Query	Response	Level of Measurement
Sex (Male/ Female)	Female	Nominal
Height (in metres)	158	Ratio
Weight (in kilograms)	60	Ratio
Rate overall health (Excellent; Good; Average; Below Average)	Good	Ordinal
Pulse rate (in BPM)	80	Ratio
Body temperature (in Fahrenheit)	36	Interval
Country of residence	India	Nominal

Activity-2.

Indicate whether the variable is ordinal or not. Write the variable type, if it is not ordinal.

- Opinion about a new law (favour or oppose) Not Ordinal -- Nominal
- Letter grade in an English class (A, B, C, etc.) Ordinal
- Student rating of teacher on a scale of 1 – 10. Not Ordinal -- Interval

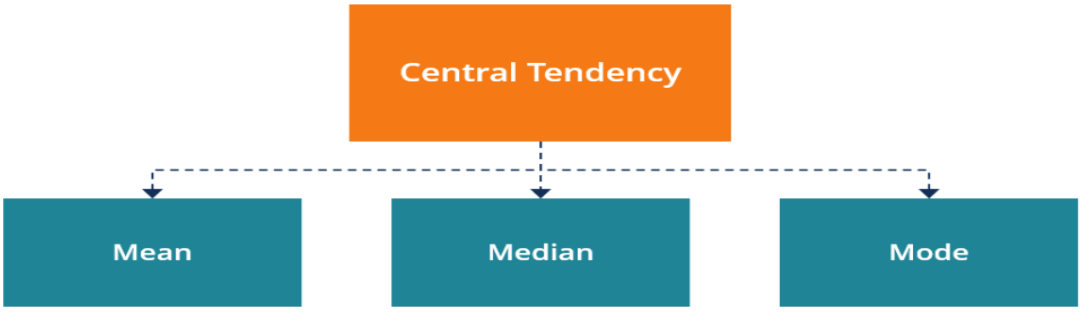
4. STATISTICAL ANALYSIS OF DATA

Teachers can ask the following questions:

- Imagine you have a dataset of the heights of all students in your class. How would you find the "average" height? Is there just one way, or could there be different ways to measure it depending on the data?
- We often hear about data being "spread out" or "clumped together." How can we describe how spread out the data in a set is, besides just knowing the average value?

Measure of Central Tendency

Statistics is the science of data, which is in fact a collection of mathematical techniques that helps to extract information from data. For the AI perspective, statistics transforms observations into information that you can understand and share. Usually,



Statistics deals with large dataset and Central tendency is used for the understanding and analysis purpose of data. **“Central tendency” is stated as the summary of a data set in a single value that represents the entire distribution of data domain (or data set).**

We can perform Statistical Analysis using Python programming language**.** For that we have to import the library statistics into the Program. Some important functions which we will use in future programs in this module are

mean ( ) →returns the mean of the data
median ( ) →returns the median of the data
mode ( ) →returns the mode of the data
variance ( ) →returns the variance of the data
stdev ( ) →returns the standard deviation of the data

Mean

In statistics, the mean (more technically the arithmetic mean or sample mean) can be estimated from a sample of examples drawn from the domain. It is a quotient obtained by dividing the total of the values of a variable by the total number of their observations or items.

$$M = \sum fx / n$$

where M = Mean

∑ = Sum total of the scores
f = Frequency of the distribution
x = Scores
n = Total number of cases

Example -1

The set S = {5,10,15,20,30}
Mean of set S = 5+10+15+20+30/5 = 80/5 = 16

Example- 2 Calculate the mean of the following grouped data.

Class	Frequency
2 - 4	3
4 - 6	4
6 – 8	2
8 – 10	1

Solution

Class	Frequency (f)	Mid value (x)	$f \cdot x$
2 -4	3	3	9
4 - 6	4	5	20
6 – 8	2	7	14
8 – 10	1	9	9
	n=10		$\sum f \cdot x$=52

Mean (M) = $\sum fxn \; / \; n$

=52 / 10
= 5.2

USING PYTHON

Program-1

There are 25 students in a class. Their heights are given below. Write a Python Program to find the mean.

heights → 145, 151, 152, 149, 147, 152, 151,149, 152, 151, 147, 148, 155, 147,152,151, 149,145, 147, 152,146, 148, 150, 152, 151

```
import statistics
height = [145,151, 152, 149, 147, 152, 151,149,
          152, 151, 147, 148, 155, 147,152,151,
          149,145, 147, 152,146, 148, 150, 152, 151]
print ("Mean height of students", statistics.mean(height))
```

OUTPUT

```
Mean height of students 149.56
```

Median

The median is another measure of central tendency. It is positional value of the variables which divides the group into two equal parts, one part comprising all values greater than median and other part smaller than median.

Example-3

Following series shows marks in mathematics of students learning AI

17	32	35	15	21	41	32	11	10	20	27	28	30

We arrange this data in an ascending or descending order.

10, 11, 15, 17, 20, 21, 27, 28, 30, 32, 32, 35, 40

As 27 is in the middle of this data position wise, therefore Median = 27

Program-2

There are 25 students in a class. Their heights are given below. Write a Python Program to find the median.

heights → 145, 151, 152, 149, 147, 152, 151,149, 152, 151, 147, 148, 155, 147,152,151, 149,145, 147, 152,146, 148, 150, 152, 151

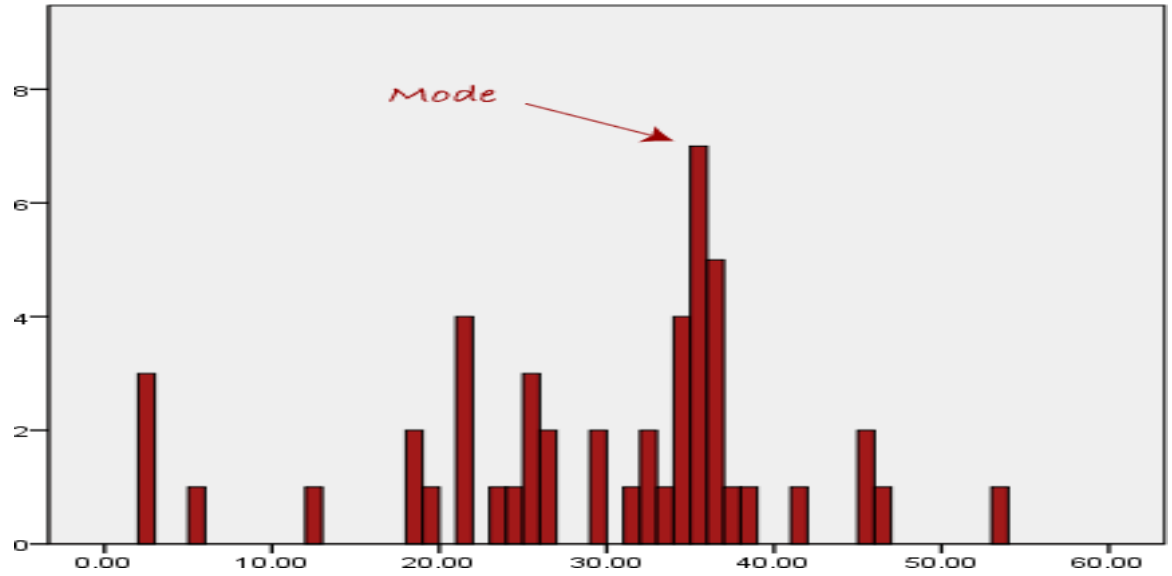
```
import statistics
height = [145,151, 152, 149, 147, 152, 151,149,
          152, 151, 147, 148, 155, 147,152,151,
          149,145, 147, 152,146, 148, 150, 152, 151]
print ("Median of height of students", statistics.median(height))
```

OUTPUT

```
Median of height of students 150
```

Mode

Mode is another important measure of central tendency of statistical series. It is the value which occurs most frequently in the data series. It represents the highest bar in a bar chart or histogram. An example of a mode is presented below:



Example-4 Age of 15 students of a class

Age (years) 22, 24, 17, 18, 17, 19, 18, 21, 20, 21, 20, 23, 22, 22, 22,22,21,24

- We arrange this series in ascending order as 17,17,18,18,19,20,20,21,21,22,22,22,
- An inspection of the series shows that 22 occurs most frequently, hence

Mode=22

Program – 3

Write a program to find the mode

(heights → 145,151, 152, 149, 147, 152, 151,149, 152, 151, 147, 148, 155, 147,152,151, 149, 145, 147, 152,146, 148, 150, 152, 151)

```
import statistics
height = [145,151, 152, 149, 147, 152, 151,149,
          152, 151, 147, 148, 155, 147,152,151,
          149,145, 147, 152,146, 148, 150, 152, 151]
print ("Mode of height of students", statistics.mode(height))
```

OUTPUT

```
Mode of height of students 152
```

In summary, when do we use mean, median and mode:

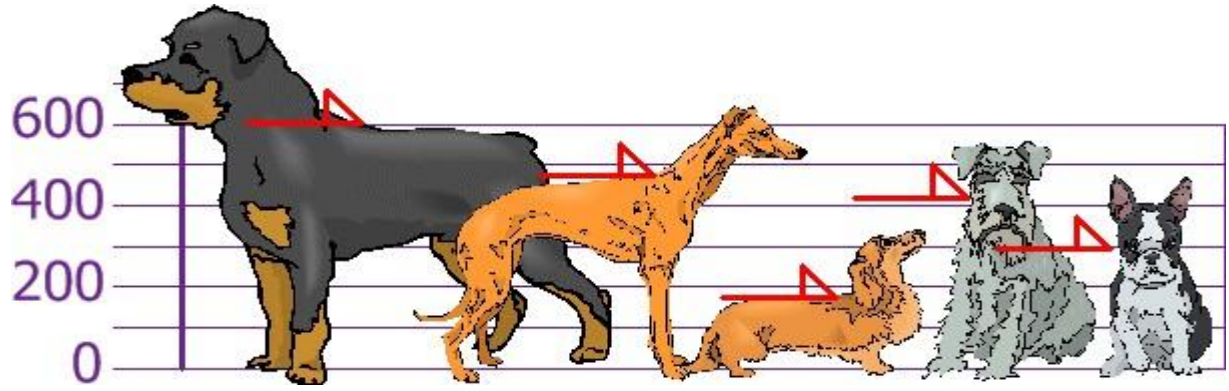
Mean	Median	Mode
• The mean is a good measure of the central tendency when a data set contains values that are relatively evenly spread with no exceptionally high or low values.	• The median is a good measure of the central value when the data include exceptionally high or low values. The median is the most suitable measure of average for data classified on an ordinal scale.	• Mode is used when you need to find the distribution peak and peak may be many. For example, it is important to print more of the most popular books; because printing different books in equal numbers would cause a shortage of some books and an oversupply of others.

Variance and Standard Deviation

Measures of central tendency (mean, median and mode) provide the central value of the data set. Variance and standard deviation are the measures of dispersion (quartiles, percentiles, ranges), they provide information on the spread of the data around the centre.

Let us understand these two using a diagram:

Measure the height (at the shoulder) of 5 dogs (in millimetres)



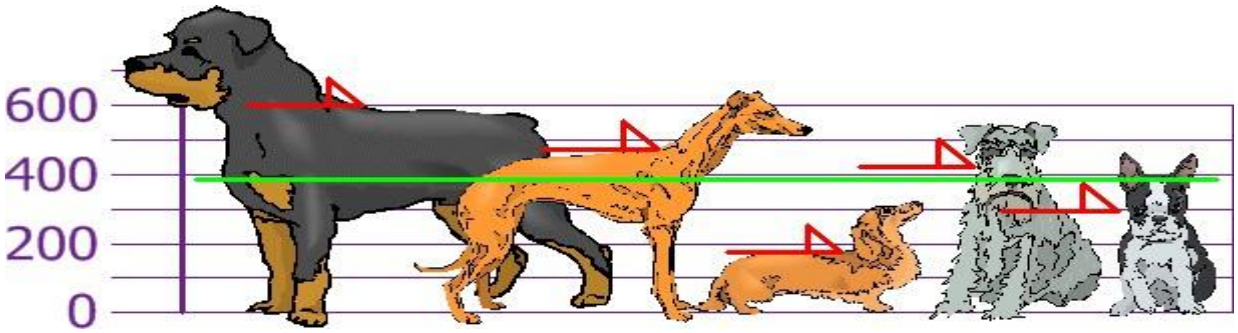
As you can see, their heights are:

600mm,
470mm,
170mm,
430mm and
300mm.

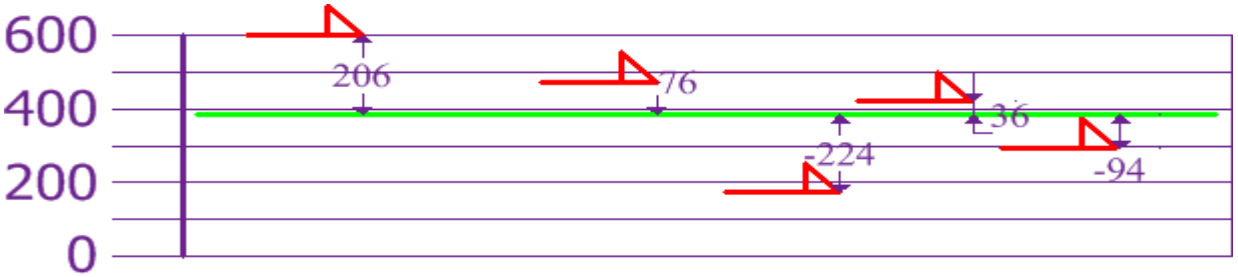
Let us calculate their mean,

Mean = (600 + 470 + 170 + 430 + 300) / 5 = 1970 / 5 = 394 mm

Now let us plot again after taking mean height (The green Line)



Now, let us find the deviation of dogs' height from the mean height



Calculate the difference (from mean height), square them, and find the average. This average is the value of the variance.

Variance = [$(206)^2 + (76)^2 + (-224)^2 + (36)^2 + (-94)^2$] / 5 = 108520 / 5 = 21704

And standard deviation is the square root of the variance.

Standard deviation = $\sqrt{21704}$ = 147.32

FORMULA

Variance	$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$

Some important facts about variance and standard deviation

- A small variance indicates that the data points tend to be very close to the mean, and to each other.
- A high variance indicates that the data points are very spread out from the mean, and from one another.
- A low standard deviation indicates that the data points tend to be very close to the mean.
- A high standard deviation indicates that the data points are spread out over a large range of values.

USING PYTHON

Program -4

Write a program to find the variance and standard deviation. (heights → 145,151, 152, 149, 147, 152, 151,149, 152, 151, 147, 148, 155, 147,152,151, 149,145, 147, 152,146, 148, 150, 152, 151)

```
import statistics
height = [145,151, 152, 149, 147, 152, 151,149,
          152, 151, 147, 148, 155, 147,152,151,
          149,145, 147, 152,146, 148, 150, 152, 151]
print ("Variance in the height of students", statistics.variance(height))
print ("Standard Deviation", statistics.stdev(height))
```

OUTPUT

```
Variance in the height of students 6.756666666666667
Standard Deviation 2.5993588953175872
```

5. REPRESENTATION OF DATA

Teachers can ask the following questions:

- **Imagine you have a big bag of colourful candies. How would you easily describe the different colours and how many candies there are of each colour without counting them all one by one?** (This question relates data to something familiar and prompts students to think about summarizing information)
- **Have you ever seen a weather map or a graph in a book? Why do you think people use pictures and charts instead of just writing out numbers?** (This question introduces the concept of data visualization and its benefits over raw data)

According to Wikipedia, “Statistics is the discipline that concerns the collection, organization, analysis, interpretation and presentation of data. To achieve this task, statisticians summarize a large amount of data in a format that is compact and produces meaningful information. Without displaying values for each observation (from populations), it is possible to represent the data in brief while keeping its meaning intact using certain techniques called 'data representation'. It can also be defined as a technique for presenting large volumes of data in a manner that enables the user to interpret the important data with minimum effort and time.

Data representation techniques are broadly classified in two ways:

Non-Graphical technique:

Tabular form and case form: This is the old format of data representation not suitable for large datasets. Non-graphical techniques are not so suitable when our objective is to make some decisions after analysing a set of data.

Graphical Technique:

The visual display of statistical data in the form of points, lines, dots and other geometrical forms is most common. For a complex and large quantity of data, human brain is more comfortable in dealing if represented through visual format means Graphical or pictorial representation of the data using graph, chart, etc. is known as **Data visualization**.

It would not be possible to discuss the methods of construction of all types of diagrams and maps primarily due to time constraint. We will, therefore, describe the most commonly used graphs and the way they are drawn.

- Line graphs
- Bar diagrams
- Pie diagram
- Scatter Plots
- Histogram

Data Visualization is possible in python using the library **Matplotlib**. It is a comprehensive library that can be used to create a wide variety of plots, including line plots, bar charts, histograms, scatter plots, and more. Matplotlib is also highly customizable, allowing users to control the appearance of their plots in great detail. **pyplot** is a submodule of Matplotlib that provides a MATLAB-like interface to the library. pyplot also provides a number of convenience functions that make it easy to create simple plots.

Installing Matplotlib

```
pip install matplotlib
or
python – m pip install – U matplotlib
```

In the program we have to import the library.

import matplotlib.pyplot

Some of the common functions and its description is given below

Function Name	Description
title ( )	Adds title to the chart/graph
xlabel ( )	Sets label for X-axis
ylabel ( )	Sets label for Y-axis
xlim ( )	Sets the value limit for X-axis
ylim( )	Sets the value limit for Y-axis
xticks ( )	Sets the tick marks in X-axis
yticks( )	Sets the tick marks in Y-axis
show ( )	Displays the graph in the screen
savefig("*adrress*")	Saves the graph in the address specified as argument.
figure (figsize = *value in tuple format*)	Determines the size of the plot in which the graph is drawn. Values should be supplied in tuple format to the attribute figsize which is passed as argument.

List of Markers and its descriptions:

marker	symbol	description	marker	symbol	description
"."	•	point	"P"	✚	plus (filled)
","	·	pixel	"*"	★	star
"o"	●	circle	"h"	⬢	hexagon1
"v"	▼	triangle_down	"H"	⬢	hexagon2
"^"	▲	triangle_up	"+"	+	plus
"s"	■	square	"x"	×	x
"p"	⬟	pentagon	"D"	◆	diamond

List of Graph Colour Codes:

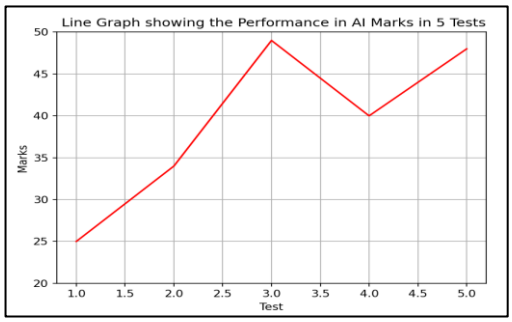
1. Line Graph

A line graph is a powerful tool used to represent **continuous data** along a numbered axis. It allows us to visualize **trends and changes** in data points over time. Line graphs are suitable for data that can take on any value within a specific range. The line can slope upwards, indicating an **increase**, or downwards, signifying a **decrease**, reflecting the changes in the data over time.

Example-5:

Kavya's AI marks for 5 consecutive tests is given below. Draw a line graph to Analyse her performance.

Test-1	Test-2	Test-3	Test-4	Test-5
25	34	49	40	48



USING PYTHON

Activity -3: Construct a simple line graph to represent the rainfall data of Kerala as shown in the table below

Month	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Rainfall (cm)	7.5	6.3	3.5	1.8	1.2	25.8	19.7	20.3	15.9	22.4	18.6	11.2

Line chart is plotted in python using the function plot ( ). Colour of the line can be mentioned by giving the colour codes inside the plot function.

Attributes of plot function which are used inside plot ( ) function are:

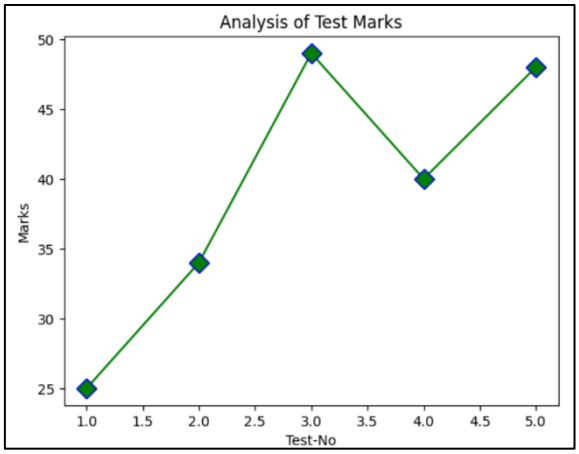
line width	Sets the width of the line
line style	determines the style of line (solid, dashed, dot, dashdot)
marker, markersize, markeredgecolor	determines the marker's shape, size and marker edge colour respectively

Program-5

Write a program to draw a line chart, we use plot function (use Example 1)

```
import matplotlib.pyplot as pl
Test=[1,2,3,4,5]
Marks=[25, 34, 49, 40, 48]

pl.title ("Analysis of Test Marks")
pl.xlabel("Test-No")
pl.ylabel("Marks")

pl.plot(Test, Marks,'g', marker='D', markersize=10, markeredgecolor='blue', linestyle='solid')
pl.show()
```

OUTPUT

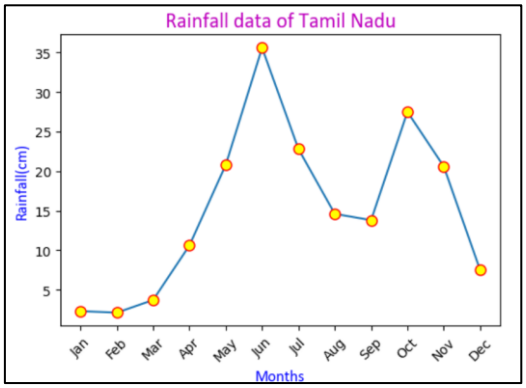


Program -6

Write a program to draw a line chart to visualize the comparative rainfall data for 12 months in Tamil Nadu using the CSV file "rainfall.csv".

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv("rainfall.csv")
x=df['Months']
y=df['Rainfall(cm)']
plt.figure(figsize=(6,4))
plt.plot(x,y,marker='o', markersize=8, markeredgecolor='red', markerfacecolor='yellow')
plt.xticks(rotation = 45)
plt.xlabel("Months",fontname='Calibri',color='b',fontsize=12)
plt.ylabel("Rainfall(cm)",fontname='Calibri',color='b',fontsize=12)
plt.title("Rainfall data of Tamil Nadu",fontname='Calibri',color='m',fontsize=16)
plt.show()
```

OUTPUT



2. Bar Graph

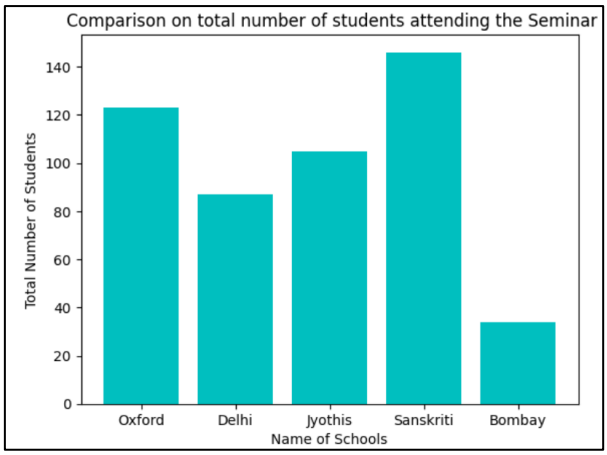
A bar chart or bar graph is a graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. It is a good way to show relative sizes, i.e., to show comparison between different categories. The relative sizes of the bars allow for easy comparison between different categories.

Example-6

Create a bar graph to illustrate the distribution of students from various schools who attended a seminar on "Deep Learning". The total number of students from each school is provided below.

Oxford Public School	Delhi Public School	Jyothis Central School	Sanskriti School	Bombay Public School
123	87	105	146	34

```
import matplotlib.pyplot as pl
a=["Oxford", "Delhi", "Jyothis", "Sanskriti", "Bombay"]
b=[123, 87, 105, 146, 34]

pl.xlabel("Name of Schools")
pl.ylabel("Total Number of Students")

pl.title( " Comparison on total number of students attending the Seminar")
pl.bar(a,b,color='c')

pl.show()
```



Bar chart is plotted in python using the function bar ( ).

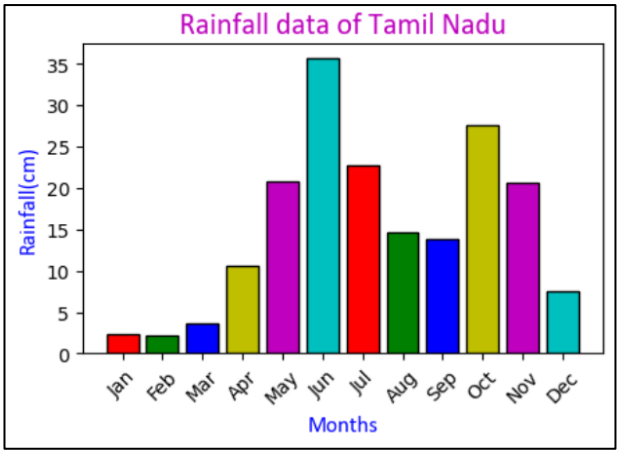
Attributes of bar function which are used inside bar ( ) functions are:

color	determines the color of the bars
edgecolor	determines the colour of the bar edges
width	determines the width of the bars

Program – 7

Write a program to draw a bar chart to visualize the comparative rainfall data for 12 months in Tamil Nadu using the CSV file "rainfall.csv".

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv("rainfall.csv")
x=df['Months']
y=df['Rainfall(cm)']
plt.figure(figsize=(6,4))
c=['r','g','b','y','m','c']
plt.bar(x,y,color=c,edgecolor='k')
plt.xticks(rotation = 45)
plt.xlabel("Months",fontname='Calibri',color='b',fontsize=12)
plt.ylabel("Rainfall(cm)",fontname='Calibri',color='b',fontsize=12)
plt.title("Rainfall data of Tamil Nadu",fontname='Calibri',color='m',fontsize=16)
plt.show()
```



3. Histogram

Histograms are graphical representations of data distribution, with vertical rectangles depicting the frequencies of different value ranges. They are drawn on a natural scale, making it easy to interpret the central tendency, such as the mode, of the data. Despite their simplicity and ease of understanding, histograms have a limitation: they can only represent one data distribution per axis.

Example -7

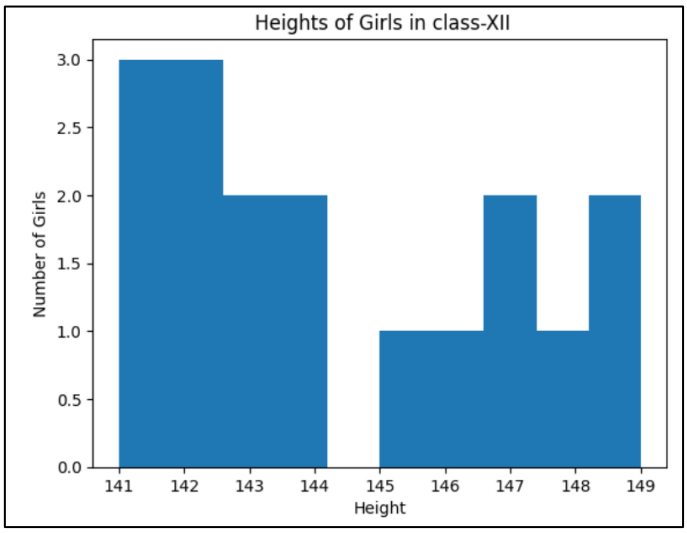
Given a dataset containing the heights of girls in class XII, construct a histogram to visualize the distribution of heights.

141,145,142,147,144,148,141,142,149,144,143,149,146,141, 147, 142, 143

Solution:

To draw a histogram from this, we first need to organize the data into intervals. These intervals are also called logical ranges or **bins**. After computing the number of girls in each interval, draw the graph. Histogram is plotted in python using the function hist ( ).

```
import matplotlib.pyplot as pl
a=[141, 145, 142, 147, 144, 148, 141, 142, 149, 144, 143,
   149, 146, 141, 147, 142, 143]
pl.ylabel("Number of Girls")
pl.xlabel("Height")
pl.title( " Heights of Girls in class-XII")
pl.hist(a)
pl.show()
```



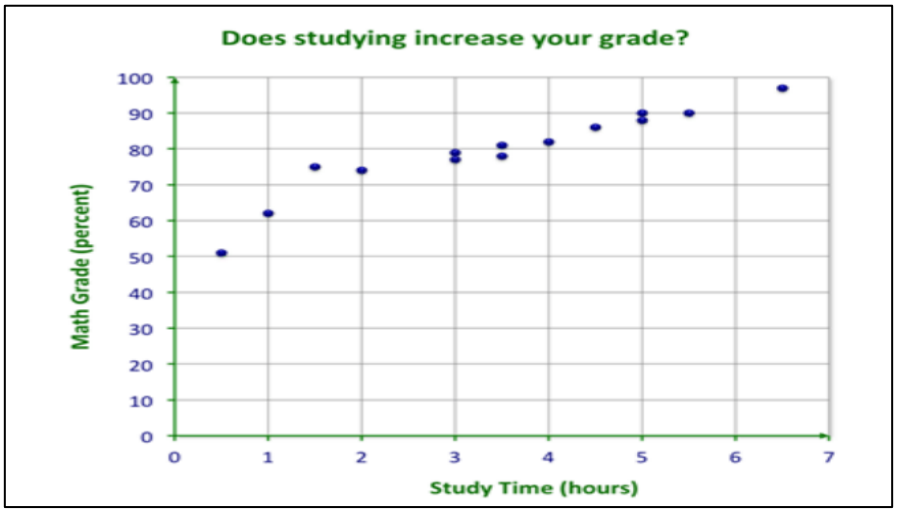
4. Scatter Graph

Scatter plots visually represent relationships between two variables by plotting data points along both the x and y axes. They reveal correlations, whether positive or negative, within paired data, showcasing trends and patterns. Essentially, scatter plots illustrate connections between variables through ordered pairs, making them useful for analyzing paired numerical data and situations where the dependent variable varies across different values of the independent variable. Their strength lies in their ability to clearly depict trends, clusters, and relationships within datasets.

Example-8 A student had a hypothesis for a science project. He believed that the more the students studied Math, the better their math scores would be. He took a poll in which he

Study Time (Hours)	4	3.5	5	2	3	6.5	0.5	3.5	4.5	5
Maths Grade (%)	82	81	90	74	77	97	51	78	86	88

asked students the average number of hours that they studied per week during a given semester. He then found out the overall percentage that they received in their Math classes. His data is shown in the table below:
To understand this data, he decided to make a scatter plot. The independent variable, or input data, is the study time because the hypothesis is that the Math grade depends on the study time. That means that the Math grade is the dependent variable, or the output data. The input data is plotted on the x-axis and the output data is plotted on the y-axis.



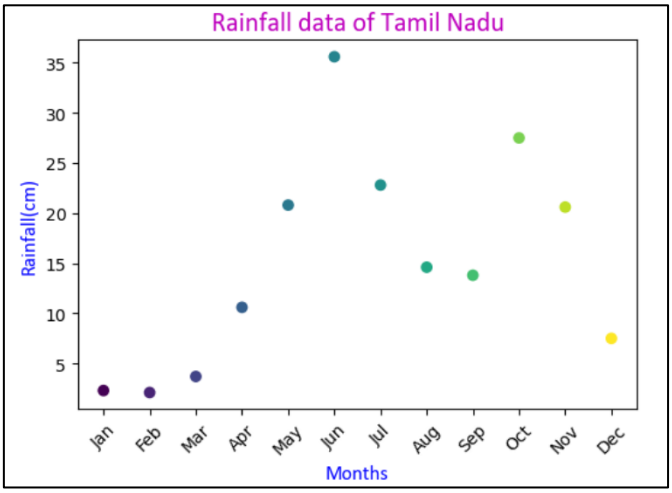
Scatterplot is plotted using the function scatter ( )

Program-8

Write a program to draw a scatter chart to visualize the comparative rainfall data for 12 months in Tamil Nadu using the CSV file "rainfall.csv".

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv("rainfall.csv")
x=df['Months']
y=df['Rainfall(cm)']
plt.figure(figsize=(5,3))
plt.figure(figsize=(6,4))
colors = np.array([0, 10, 20, 30, 40, 45, 50,  60, 70, 80, 90, 100])
plt.scatter(x,y,c=colors,cmap='viridis')
plt.title("Rainfall data of Tamil Nadu",fontname='Calibri',color='m',fontsize=16)
plt.xticks(rotation = 45)
plt.xlabel("Months",fontname='Calibri',color='b',fontsize=12)
plt.ylabel("Rainfall(cm)",fontname='Calibri',color='b',fontsize=12)
plt.show()
```

OUTPUT



5. Pie Chart

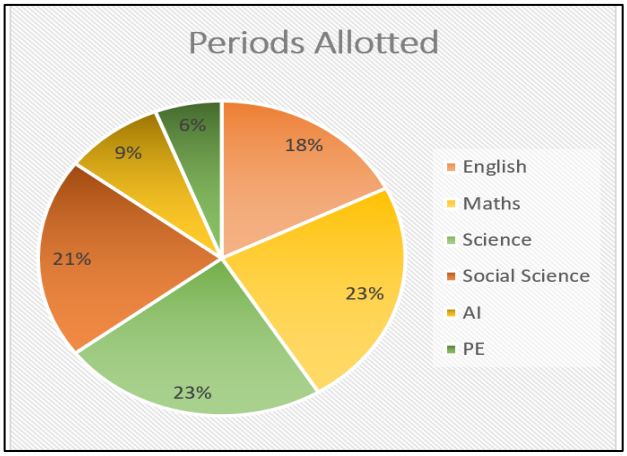
A pie chart is a circular graph divided into segments or sections, each representing a relative proportion or percentage of the total. Each segment resembles a slice of pie, hence the name. Pie charts are commonly used to visualize data from a small table, but it is recommended to limit the number of categories to seven to maintain clarity. However, zero values cannot be depicted in pie charts.

While useful for illustrating compositions or comparing parts of a whole, pie charts can be challenging to interpret and compare with data from other charts. They are not suitable for showing changes over time. Pie charts find applications in various domains such as business, education, and personal finance. In business, they can indicate the success or failure of products or services. In education, they can depict time allocations for different subjects. At home, pie charts can help visualize monthly expenses relative to income.

Example-9

Below given is a Pie chart drawn with the periods allotted for each subject in a week.

Subject	Periods Allotted
English	6
Maths	8
Science	8
Social Science	7
AI	3
PE	2

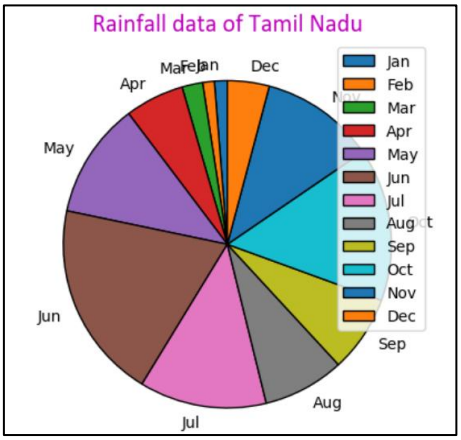


Pie Chart is plotted using the function pie ( )

Program-9

Write a program to draw a pie chart to visualize the comparative rainfall data for 12 months in Tamil Nadu using the CSV file "rainfall.csv".

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv("rainfall.csv")
x=df['Months']
y=df['Rainfall(cm)']
wp = { 'linewidth' : 1, 'edgecolor' : "black" }
plt.pie(y,labels=x, startangle=90,wedgeprops=wp)
plt. legend(loc='upper right')
plt.title("Rainfall data of Tamil Nadu",fontname='Calibri',color='m',fontsize=16)
plt.show()
```



6. INTRODUCTION TO MATRICES

The knowledge of matrices is necessary in all branches of mathematics. Matrix is one of the most powerful tools in Mathematics. In mathematics, matrix (plural matrices) is a rectangular arrangement of numbers. The numbers are arranged in tabular form as rows and columns. Matrices play a huge role in computer vision domain of AI. On the computer, the image is represented as a combination of pixels. This is represented mathematically as matrices!

Let us understand with the help of an example: Consider

Aditi bought 25 pencils 5 erasers

Adit bought 10 pencils 2 erasers

Manu bought 5 pencils 1 eraser

The above information can be arranged in tabular form as follows

	Pencils	**Erasers**
Aditi	25	5
Adit	10	2
Manu	5	1

And this can be represented as

Row1
Row2
Row3

$$\begin{bmatrix} 25 & 5 \\ 10 & 2 \\ 5 & 1 \end{bmatrix}$$

The entries in the rows represent number of pencils and erasers bought by Aditi, Adit and Manu respectively. Or in another form as

Col1	Col2	Col3
Row1		
Row2		

$$\begin{bmatrix} 25 & 10 & 5 \\ 5 & 2 & 1 \end{bmatrix}$$

Here, the entries in the columns represent number of pencils and erasers bought by Aditi, Adit and Manu respectively. We denote matrices by capital letters, for example

$$A = \begin{bmatrix} 5 & 15 \\ -7 & \sqrt{2} \\ 12 & 0 \end{bmatrix}$$

Order of a matrix

A matrix has *m* rows and *n* columns. It is called a matrix of *order m* × *n* or simply *m* × *n* matrix (read as an *m* by *n* matrix). So, the matrix A in the above example is a 3 × 2 matrix. The number of elements are m x n => 3 x 2 = 6 elements. Each individual element is represented as a_{ij} where i represents row and j represents column. In general a_{ij}, is an element lying in the i^{th} row and j^{th} column. We can also call it as the $(i, j)^{th}$ element of the matrix.

$$P = \begin{bmatrix} a11 & a12 \\ a21 & a22 \\ a31 & a32 \end{bmatrix}$$

Operations on Matrices

1. ***Addition of matrices*** **-** the sum of two matrices is obtained by adding the corresponding elements of the given matrices. Also, the two matrices have to be of the same order. Example:

$$A = \begin{bmatrix} 3 & 2 \\ 4 & -1 \\ 2 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 6 & 3 \\ 5 & 9 \\ 3 & 2 \end{bmatrix}$$

$$A+B = \begin{bmatrix} 3+6 & 2+3 \\ 4+5 & -1+9 \\ 2+3 & 0+2 \end{bmatrix} = \begin{bmatrix} 9 & 5 \\ 9 & 8 \\ 5 & 2 \end{bmatrix}$$

2. **Difference of matrices** - The difference A – B is defined as a matrix where each element is obtained by subtracting the corresponding elements $(a_{ij} – b_{ij})$. Matrices A and B must be of the same order. Example

$$A = \begin{bmatrix} -2 & 1 \\ 6 & 10 \\ 5 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} -1 & 3 \\ 2 & 9 \\ 3 & 1 \end{bmatrix}$$

$$A\text{-}B = \begin{bmatrix} -2+1 & 1-3 \\ 6-2 & 10-9 \\ 5-3 & 3-1 \end{bmatrix} = \begin{bmatrix} -1 & -2 \\ 4 & 1 \\ 2 & 2 \end{bmatrix}$$

3. **Transpose of a matrix** – a matrix obtained by interchanging the rows and columns. Transpose of a matrix A is denoted by A' or A^T. Example

$$A = \begin{bmatrix} 8 & 7 \\ 2 & 5 \\ 4 & 6 \end{bmatrix} \qquad A^T = \begin{bmatrix} 8 & 2 & 4 \\ 7 & 5 & 6 \end{bmatrix}$$

Order = 3x2 Order = 2x3

Applications of matrices in AI

Matrices are used throughout the field of machine learning for computing:

- Image Processing – Digital images can be represented using matrices. Each pixel on the image has a numerical value. These values represent the intensity of the pixels. A grayscale or black-and-white image has pixel values ranging from 0 to 255. Smaller values closer to zero represent darker shades, whereas bigger ones closer to 255 represent lighter or white shades. So, in a computer, every image is kept as a matrix of integers called a Channel.
- Recommender systems use matrices to relate between users and the purchased or viewed product(s)
- In Natural Language processing, vectors depict the distribution of a particular word in a document. Vectors are one-dimensional matrices.

7. DATA PREPROCESSING

Teachers can ask the following:

- **Imagine you have a giant bag of mixed candies. How can we sort through the candies to make it easier to find the Flavors we want?** (This relates to data cleaning and organization)
- **Sometimes data can be like a bag of candy with a few weird pieces mixed in. How can we make sure all the data makes sense of what we're trying to learn from it?** (This relates to identifying and handling outliers and inconsistencies)

Data preprocessing is a crucial step in the machine learning process aimed at making datasets more machine learning-friendly. It involves several processes to clean, transform, reduce, integrate, and normalize data:

1. Data Cleaning

1. **Missing Data:**
   Missing data occurs when values are absent from the dataset, which can happen due to various reasons. Strategies for handling missing data include deleting rows or

columns with missing values, inputting missing values with estimates, or using algorithms that can handle missing data.

2. **Outliers:**
   Outliers are data points that significantly differ from the rest of the data, often due to errors or rare events. Dealing with outliers involves identifying and removing them, transforming the data, or using robust statistical methods to reduce their impact.
3. **Inconsistent Data**
   Data with typographical errors, different data types etc are corrected and consistency among the data is observed.
4. **Duplicate Data**
   Duplicate data will be identified and removed to ensure data integrity.

2. Data Transformation

Categorical variables are converted to Numerical variable. New features are identified and existing features are modified if needed

3. Data Reduction
Dimensionality reduction, i.e. reducing the number of features of data set is done. If data set is too large to handle sampling techniques are applied

4. Data Integration and Normalization
If data is stored in multiple sources or formats, they are merged or aggregated together. Then the data is normalized to ensure that all features have a similar scale and distribution which can improve machine learning models.

5. Feature Selection
The most relevant features that contribute the most to the target variable are selected and irrelevant data are removed.

8. DATA IN MODELLING & EVALUATION
Teachers can ask the following:

- **Imagine you're lost in a new city. You have a map, but it might be helpful to have a separate practice map to try figuring things out before using it on the actual streets. Why do you think splitting data into training and testing sets might be similar to this idea?** (This question connects the concept of data splitting to a familiar scenario and highlights the purpose of each set)
- **When making a decision, do you consider all the information available or just some of it? Why do you think using techniques like cross-validation might be important when evaluating how well a machine learning model performs?** (This question relates the importance of evaluating models to make informed decisions and introduces the concept of cross-validation)

After the data is pre-processed, it is splitted into two --Training data set and Testing data set. The training set is used to train the machine learning models, while the testing set is used to evaluate the performance of the trained models. While modelling, appropriate machine learning algorithms are chosen based on the nature of the problem (e.g., classification, regression, clustering) and the characteristics of the dataset.

Techniques such as train-test split, cross-validation, and error analysis are employed to estimate the model's generalization ability and identify areas for improvement. Train-Test Split trains the model with its training set and evaluates using the test set. Cross Validation ensures that the model's performance is consistent across different subsets of the data.
Different types of evaluation techniques are applied on the model depending on the data. For classification problems, metrics like accuracy, precision, recall, F1-score, and ROC curve are commonly used. For regression problems, metrics like mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared are often used.

In today's world, knowing how to work with data is important. As artificial intelligence becomes more and more common, understanding data helps us use information better. It is like having a map to find your way through a big city. Being good with data helps us make smart decisions and use technology wisely.

EXERCISES:

I. Multiple-choice questions

1. Which of the following best defines data literacy?
   A) The ability to read and write data
   B) The ability to find and use data effectively
   C) The ability to analyse data using AI
   D) The ability to collect and store data securely

2. What is the purpose of data preprocessing?
   A) To make data more complex
   B) To make data less accessible
   C) To clean and prepare data for analysis
   D) To increase the size of the dataset

3. How can missing data be handled in a dataset?
   A) By ignoring it
   B) By replacing missing values with estimates
   C) By deleting rows or columns with missing values
   D) By converting missing values to zero

4. Which of the following statements about the quantity of data needed for machine learning projects is true?
   A) More data is always better for good predictions.
   B) Small batches of data are sufficient for complex models.
   C) Data quantity depends solely on the number of features.
   D) Data diversity is not essential for model performance.

5. Which of the following is an example of a primary source of data collection?
A) Web scraping B) Social media data tracking
C) Surveys D) Kaggle datasets

6. What method of data collection involves direct communication with individuals or groups to gather information?
A) Observations B) Experiments C) Interviews D) Marketing campaigns

7. Which of the following is an example of ratio scale data?
A) Grading students' exam papers as "A," "B," "C," "D," and "F"
B) Measuring the temperature in Celsius
C) Rating a meal at a restaurant as "unpalatable," "unappetizing," "just okay," "tasty," and "delicious"
D) Recording the weight of a person in kilograms

8. What is the distinguishing feature of ratio scale data?
A) It involves categories without a specific order
B) It has a zero point and allows for ratios to be calculated
C) It involves categories with a strict order but no measurable differences between categories
D) It has a definite order, but the differences between categories cannot be measured

9. Which statistical measure is most suitable for data sets with evenly spread values and no exceptionally high or low values?
A) Mean B) Median C) Mode D) Variance

10. What is the term used to describe the graphical or pictorial representation of data?
A) Statistical summary B) Data organization
C) Data visualization D) Data interpretation

ANSWERS
1. B) The ability to find and use data effectively
2. C) To clean and prepare data for analysis
3. B) By replacing missing values with estimates
4. A) More data is always better for good predictions.
5. C) Surveys
6. C) Interviews
7. C) Recording the weight of a person in kilograms
8. B) It has a zero point and allows for ratios to be calculated
9. A) Mean
10. C) Data visualization

II. Short answer questions:

1. Explain the concept of data literacy and its importance in today's digital age.

Data literacy refers to the ability to find, understand, and use data effectively. In today's digital age, where data is abundant, data literacy is crucial for making informed decisions, understanding trends, and solving complex problems. It helps individuals and organizations extract valuable insights from data to drive innovation and growth.

2. What is data preprocessing?

Data preprocessing is the process of cleaning and preparing raw data before it is used for analysis or modelling. It involves handling missing data and outliers to ensure the data is accurate and reliable.

3.What is data visualization and why is it important?

Data visualization is the graphical representation of data to help people understand the significance of data by summarizing and presenting it in a visual form such as charts, graphs, or maps.

Data visualization is important because it allows for the exploration and understanding of data patterns, trends, and outliers that may not be apparent in raw data. It helps in making data-driven decisions and communicating information clearly and efficiently.

4. How does a line graph differ from a bar graph?

A line graph is used to show trends over time with continuous data, while a bar graph is used to compare different categories of data with discrete values.

5. When would you use a scatter plot?

A scatter plot is used to show the relationship between two variables in a set of paired data, helping to identify correlations or trends between the variables.

6. What is data?

Data can be defined as a representation of facts or instructions about some entity (students, school, sports, business, animals etc.) that can be processed or communicated by human or machines.

7. What do you mean by web scraping?

Web scraping is the process of using bots to extract content and data from a website. Web scraping extracts underlying HTML code and, with it, data stored in a database. The scraper can then replicate entire website content elsewhere.

8. If a matrix has 6 elements, what are the possible orders it can have?

Answer: 4 orders – (1x6), (6x1), (2x3) and (3x2)

9.Construct a 3x2 matrix where each element is given by $a_{ij} = i * j$

Answer: 3x2 matrix means 6 elements

a11 = 1x1 a12 = 1x2

a21=2x1 a22=2x2

a31=3x1 a32=3x2

Putting all elements in matrix form we get:

$$A=\begin{bmatrix}1 & 2\\ 2 & 4\\ 3 & 6\end{bmatrix}$$

9. Find the transpose of the matrix B = $\begin{bmatrix}5 & -1 & 4\\ 2 & 3 & 6\end{bmatrix}$

Answer: B = $\begin{bmatrix}5 & 2\\ -1 & 3\\ 4 & 6\end{bmatrix}$

III. Long answer questions:

1. Discuss the advantages and limitations of using a pie chart in data visualization. Provide examples to illustrate your points.

One advantage of using a pie chart is that it can effectively show the proportion of each category in a dataset. For example, a pie chart can be used to visualize the market share of different companies in a specific industry. However, pie charts have limitations, such as difficulty in comparing multiple datasets or showing trends over time. For example, a pie chart would not be suitable for visualizing changes in sales over the course of a year, as it cannot effectively convey this type of information.

2. Explain the terms mean, median and mode.

Mean	Median	Mode
The mean is a good measure of the central tendency when a data set contains values that are relatively evenly spread with no exceptionally high or low values.	The median is a good measure of the central value when the data include exceptionally high or low values. The median is the most suitable measure of average for data classified on an ordinal scale.	Mode is used when you need to find the distribution peak and peak may be many. For example, it is important to print more of the most popular books; because printing different books in equal numbers would cause a shortage of some books and an oversupply of others.

3. Explain the four levels of measurement.

Four levels of measurements are Nominal, Ordinal, Interval and Ratio

The nominal level of measurement is the simplest or lowest of the four ways to characterize data. Nominal means "in name only". Colours of eyes, yes or no responses to a survey, gender, smartphone companies, etc all deal with the nominal level of measurement.

Ordinal data, is made up of groups and categories which follow a strict order. For e.g. if you have been asked to rate a meal at a restaurant and the options are: unpalatable, unappetizing, just okay, tasty, and delicious. The options is Ordinal and the data is qualitative,. Like the nominal scale data, ordinal scale data cannot be used in calculations.

Data that is measured using the interval scale is similar to ordinal level data because it has a definite ordering but in interval scale data does not have a starting point i.e. zero value. Temperature scales like Celsius (oC) and Fahrenheit (F) are measured by using the interval scale. Ratio scale data is like interval scale data, but it has a 0 point and ratios can be calculated. We can add, subtract, divide and multiply the two ratio level variables. Eg: Weight of a person. It has a real zero point. Hence it can be considered as ratio value.

4. Given the matrices A and B. Calculate A+ B and B – A.

$$A = \begin{bmatrix} 2 & 7 \\ 4 & 12 \\ 15 & -3 \end{bmatrix} \qquad B = \begin{bmatrix} -2 & 4 \\ 12 & 6 \\ 7 & 0 \end{bmatrix}$$

Answer:

$$A+B = \begin{bmatrix} 0 & 11 \\ 16 & 18 \\ 22 & -3 \end{bmatrix} \qquad B - A = \begin{bmatrix} -4 & -3 \\ 8 & -6 \\ -8 & 3 \end{bmatrix}$$

IV. Python Programs

1. The ages of a group of people in a community are: 25, 28, 30, 35, 40, 45, 50, 55, 60, 65. Write a program to calculate the mean, median, and mode of the ages.

```
import statistics
age=[25, 28, 30, 35, 40, 45, 50, 55, 60, 65]
mean=statistics.mean(age)
median=statistics.median(age)
mode = statistics.mode(age)
print("Mean of age", mean)
print("Median of age", median)
print("Mode of age", mode)
```

```
Mean of age 43.3
Median of age 42.5
Mode of age 25
```

2. A company recorded the daily temperatures (in degrees Celsius) for five consecutive days: 20°C, 22°C, 25°C, 18°C, and 23°C. Determine the variance and standard deviation of the temperatures.

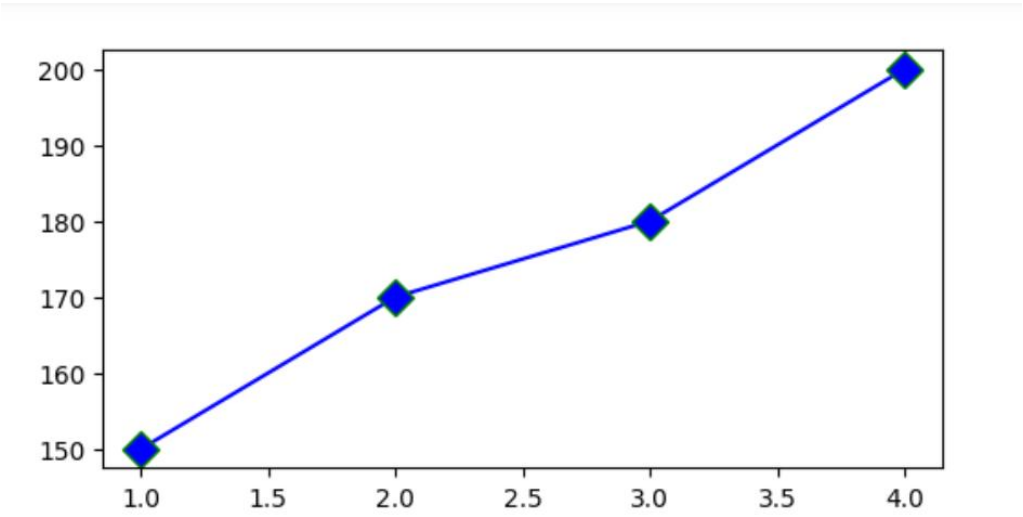
```
import statistics
temp=[20, 22, 25, 18, 23]
var=statistics.variance(age)
std=statistics.stdev(age)

print("Variance: ", var)
print("Standard Deviation", std)
```

```
Variance:  195.56666666666666
Standard Deviation 13.984515246037907
```

3. Plot a line chart representing the weekly number of customer inquiries received by a customer service center:

- Week 1: 150 inquiries
- Week 2: 170 inquiries
- Week 3: 180 inquiries
- Week 4: 200 inquiries

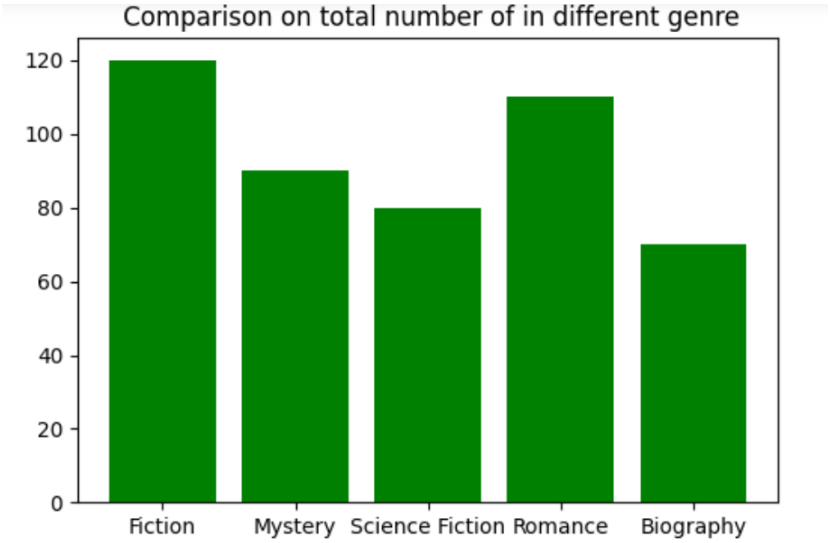
```
import matplotlib.pyplot as pl
Week=[1,2,3,4]
Inquiries=[150, 170, 180, 200]

pl.title ("Customer inquires in service center")
pl.xlabel("Week")
pl.ylabel("Inquiries")

pl.plot(Week, Inquiries,'b', marker='D', markersize=10, markeredgecolor='green', linestyle='solid')
pl.show()
```



4. Plot a bar chart representing the number of books sold by different genres in a bookstore:

- Fiction: 120 books
- Mystery: 90 books
- Science Fiction: 80 books
- Romance: 110 books
- Biography: 70 books

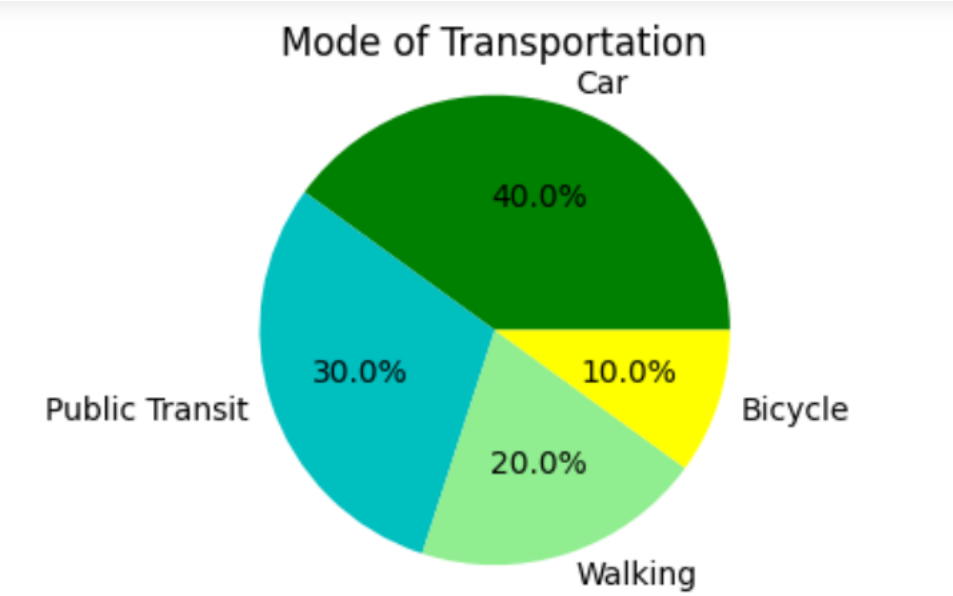
```
import matplotlib.pyplot as pl
genre=["Fiction", "Mystery", "Science Fiction", "Romance", "Biography"]
books=[120, 90, 80, 110, 70]

pl.xlabel("Genre of Books")
pl.ylabel("Total Number of Books")
pl.figure(figsize=(5,4))

pl.title( " Comparison on total number of in different genre")
pl.bar(genre,books,color='g')

pl.show()
```



5. Visualize the distribution of different types of transportation used by commuters in a city using a pie chart:

- Car: 40%
- Public Transit: 30%
- Walking: 20%
- Bicycle: 10%

```
import matplotlib.pyplot as plt
labels = ['Car', 'Public Transit', 'Walking', 'Bicycle']
sizes = [40, 30, 20, 10]
colors = ['green', 'c','lightgreen','yellow']
plt.figure(figsize=(3, 3))
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%')
plt.axis('equal')
plt.title('Mode of Transportation')
plt.show()
```



V. Competency Based Questions

1. Anakha is working on a project involving mood analysis of individuals experiencing depression. She plans to visit hospitals, yoga instructors, mental health care providers, and healthcare institutes. How can Anakha collect data from these organizations and individuals for her project?
   Ans: Surveys, Interviews, Questionnaire

2. Kalaimathi and her friend are planning to build an admission chatbot for her school. They have collected data from different sources with the help of their friends. Now, they need to arrange it in proper order to analyse it. They plan to classify the data based on the levels of measurement. The data they have collected is as follows:
   Student Name, Age, Gender, GPA (Grade Point Average), ECA (Extra-Curricular activities), Place, Parent Name, Parent Education level, Distance from school, Fees, Interview rating, No. of years in last school, Admission Test score
   Ans:

NOMINAL	ORDINAL	INTERVAL	RATIO
Student Name	Parent Educational level	No. of years in last school	Age
Gender	Interview rating	Admission test score	GPA
Place		Distance from school to home	ECA
Parent Name			Fees

3. During a sales analysis, metrics such as sales revenue per month, average sales revenue, and the most popular products sold are examined to comprehend overall performance and aid in decision-making regarding marketing strategies, inventory management, and resource allocation. Which measurements in statistics will facilitate the sales analysis process?
   Ans: Mean, Median & Mode

4. Selvan's Textiles operates multiple showrooms in a city. Fahad is working on a project to predict the annual sales percentage for the upcoming year. He intends to analyse the trend of sales percentage over time and also the sales percentages in different regions. To facilitate data analysis, he plans to visualize the data using graphs.
    a. Which type of graph would be most appropriate for visualizing the trend of sales percentage over time?
    b. Which type of graph would be most suitable for comparing sales percentages across different regions?

    Ans:

    a. Line Chart
    b. Pie Chart
5. Akshith wrote a program to visualize the data analysis of five test and marks got.

```
import matplotlib.pyplot as pl
Test=[1,2,3,4,5]
Marks=[25, 34, 49, 40, 48]

pl.title ("Analysis of Test Marks")
pl.xlabel("Test-No")
pl.ylabel("Marks")

pl.plot(Test, Marks,'g', marker='D')
```

The program did not have any errors. But the line graph was not showing up. Could you find the reason why the graph is not shown even though the program has no errors?

Ans:

In order to view the graph we have to use show( ) function in the program. If we write pl.show( ), the graph window will be displayed.

UNIT 6: Machine Learning Algorithms

Title: Machine Learning Algorithms	**Approach:** Interactive/ Collaborative / Hands on Activity

Summary: In this unit, the students will be able to understand the concept of Machine Learning and its connection with AI. They will understand the different ways in which machines can learn or can be taught using different methodologies. The main types of machine learning methods- Supervised, Unsupervised and Reinforcement methods will be introduced to the students.

Learning Objectives:

1. Understand Machine Learning and the various machine learning algorithms
2. Understand regression as a type of supervised learning.
3. Understand classification as a type of supervised learning.
4. Understand clustering as a type of unsupervised learning.
5. List of algorithms for regression, classification and clustering
6. Differentiate between regression problem, classification problem and clustering problem.

Key Concepts:

1. Machine Learning in a nutshell
2. Types of Machine Learning
3. Supervised Learning
   Regression – Understanding Correlation, Regression, Finding the line, Linear Regression algorithm
   Classification – How it works, Types, k – Nearest Neighbour algorithm
4. Unsupervised Learning
   Clustering – How it works, Types, k -means Clustering algorithm

Learning Outcomes:

By the end of this unit, students will be able to:

1. Differentiate the different types of machine learning methods.
2. They will be able to understand the concept behind each machine learning methods.
3. Apply these methods to develop simple solutions for some day-to-day situations.
4. Build up this knowledge to the next level to apply during Capstone Project development.

Pre-requisites: Basic Knowledge about Al, domains of AI, applications of AI.

Unveiling the Magic of Machine Learning: A Teacher's Guide

This lesson equips you to introduce students to the fascinating world of Machine Learning (ML) and its applications in our daily lives.

Hooking Students with Everyday ML:

- Start with a bang! Discuss familiar experiences where AI and ML play a role:
    - Shopping apps recommending socks after a shoe purchase.
    - Entertainment platforms suggesting similar movies based on viewing history.
    - Social media platforms using facial recognition for tagging friends in photos.

Machine Learning: The Core Concepts:

- Explain the basic principles of Machine Learning: how machines learn from data to make predictions or decisions without explicit programming.

Ethics in AI and ML: A Debate Worth Having:

- Spark critical thinking with class debates or discussions on the ethical considerations surrounding AI and ML. Explore topics like:
    - Algorithmic bias and fairness.
    - Data privacy and security concerns.
    - The potential impact of AI on jobs and society.

Demystifying Machine Learning Approaches:

- Introduce the three main types of Machine Learning:
    - **Supervised Learning:** Machines learn from labeled data (data with predefined outputs) to make predictions for new, unseen data.
    - **Unsupervised Learning:** Machines identify patterns and relationships in unlabeled data (data without predefined outputs) for tasks like grouping or anomaly detection.
    - **Reinforcement Learning:** Machines learn through trial and error, receiving rewards for desired actions to optimize their behavior in a given environment.

Exploring Real-World Applications:

- Illustrate how each type of Machine Learning tackles real-world problems:
    - Supervised Learning: Spam filtering, medical diagnosis, stock price prediction.
    - Unsupervised Learning: Customer segmentation, image categorization, fraud detection.
    - Reinforcement Learning: Self-driving cars, game playing AI, robot control.

Diving Deeper: Supervised Learning

- Define Supervised Learning and its key components: data, algorithms, and models.

- Explain how labeled data is used to train the model, allowing it to learn the relationship between inputs and outputs.

Supervised Learning Tasks:

- Discuss common supervised learning tasks:
    - **Classification:** Predicting a category for new data points (e.g., email spam or not spam).
    - **Regression:** Predicting a continuous value for new data points (e.g., house price prediction).

Supervised Learning Algorithms:

- Introduce popular supervised learning algorithms:
    - **Linear Regression:** A simple yet powerful algorithm for predicting continuous outputs based on a linear relationship with the input data. (e.g., predicting sales based on marketing spend)

K-Nearest Neighbors (KNN): A Closer Look

- Explain the KNN algorithm and its core concept of similarity.
- Discuss how KNN classifies new data points based on the class of their nearest neighbors (k most similar data points in the training data).
- Emphasize the importance of choosing the optimal value for K (number of neighbors) to achieve accurate classification.

Unsupervised Learning and Clustering:

- Move on to Unsupervised Learning and the concept of clustering.
- Provide real-life examples like customer segmentation (grouping customers with similar characteristics) or image categorization (categorizing images based on content).

K-Means Clustering: Grouping Similar Data

- Introduce K-Means clustering, an unsupervised learning technique that aims to group similar data points together.
- Explain the steps involved in the K-Means algorithm:
    - Initialization: Choosing an initial set of cluster centers (centroids).
    - Assignment: Assigning each data point to the nearest cluster center.
    - Update: Re-computing the cluster centers based on the assigned data points.
- Highlight the importance of choosing the optimal value for K (number of clusters) to achieve meaningful groupings.

This lesson provides a foundation for students to explore the exciting world of Machine Learning further.

Teachers can ask the following questions:

- **Think about the examples mentioned earlier (online shopping recommendations, facial recognition). Can you guess how these technologies might work? What kind of information do you think they might use?** This gets them thinking about the underlying concepts behind the AI they encounter daily and activates their prior knowledge.
- **Imagine you have a huge pile of data, like all the movies ever made and how many people watched them. Can you think of any way a computer program could learn from this data and make predictions?** This question introduces the concept of data and how machine learning can use it to learn and make decisions.

Students, you have likely experienced the following scenarios in your daily lives:

- When browsing through online shopping apps, you receive recommendations to purchase socks immediately after looking at shoes.
- Online entertainment platforms suggest movies similar to ones you have previously watched.
- Facial recognition technology identifies you in photos on social media platforms.
- Chatbots engage in conversations with you in a natural manner.

MACHINE LEARNING IN A NUTSHELL

Machine Learning (ML) is a part of artificial intelligence (AI) that focuses on teaching computers to learn from data and make decisions without being explicitly programmed. Unlike traditional programming where developers provide precise instructions, ML algorithms learn from patterns and relationships in data. This allows them to generalize and make decisions on new, unseen data.

- ML algorithms learn from various **types of data**, including images, text, sensor readings, and historical records. Instead of hardcoding rules, ML models identify patterns and relationships within the data to make predictions or decisions.
- Some common ML **algorithms** include decision trees, neural networks, and support vector machines. Trained models serve as representations of the learned data, such as recognizing handwritten digits using a neural network.
- The **applications** of ML are vast and diverse. It powers recommendation systems like those used by Netflix, speech recognition, medical diagnosis, and autonomous vehicles. ML is also behind chatbots, personalized ads, and fraud detection systems.
- However, ML also presents **challenges**. Overfitting, where models become too specialized on training data, can lead to poor performance on new data. Bias in training data can result in biased predictions, and some models are difficult to interpret, acting as black boxes. Despite these challenges, ML transforms data into knowledge, enabling computers to learn, adapt, and make decisions autonomously.
- Artificial intelligence (AI) and machine learning (ML) have significantly **impacted** various aspects of our lives. From transportation and finance to healthcare and

entertainment, AI algorithms are pervasive. They power self-driving cars, fraud detection systems, personalized shopping experiences, and virtual assistants like Siri and Alexa. As technology continues to evolve, the influence of AI and ML is only expected to grow, shaping the future of our society and culture.

Activity 1: Autodraw - Experience the power of machine learning with Autodraw! Autodraw combines machine learning with the creativity of talented artists, allowing you to draw things quickly and effortlessly. Visit the following link to play the game:

TYPES OF MACHINE LEARNING

Teachers can ask the following questions:

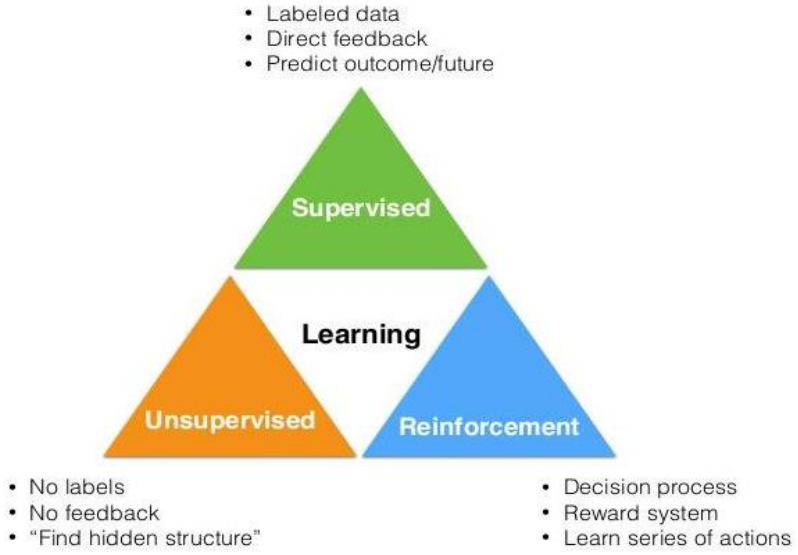
1. **Think back to the different AI applications we discussed earlier (e.g., recommendation systems, facial recognition). Can you imagine how they might learn from data? Do they all learn in the same way?** This prompts them to make connections between real-world applications and the concept of machine learning. It also primes them for the idea of different learning approaches.
2. **Imagine you have a dataset with pictures of different animals. How could a computer program learn to group these pictures into categories like cats, dogs, and birds? (supervised vs unsupervised learning)** This is a specific scenario that introduces the concept of labelled data (supervised learning) vs. unlabelled data (unsupervised learning).
3. **Can you think of a situation where learning by trial and error might be a good approach for a machine learning program?** This taps into their understanding of learning through experience and lays the groundwork for reinforcement learning.

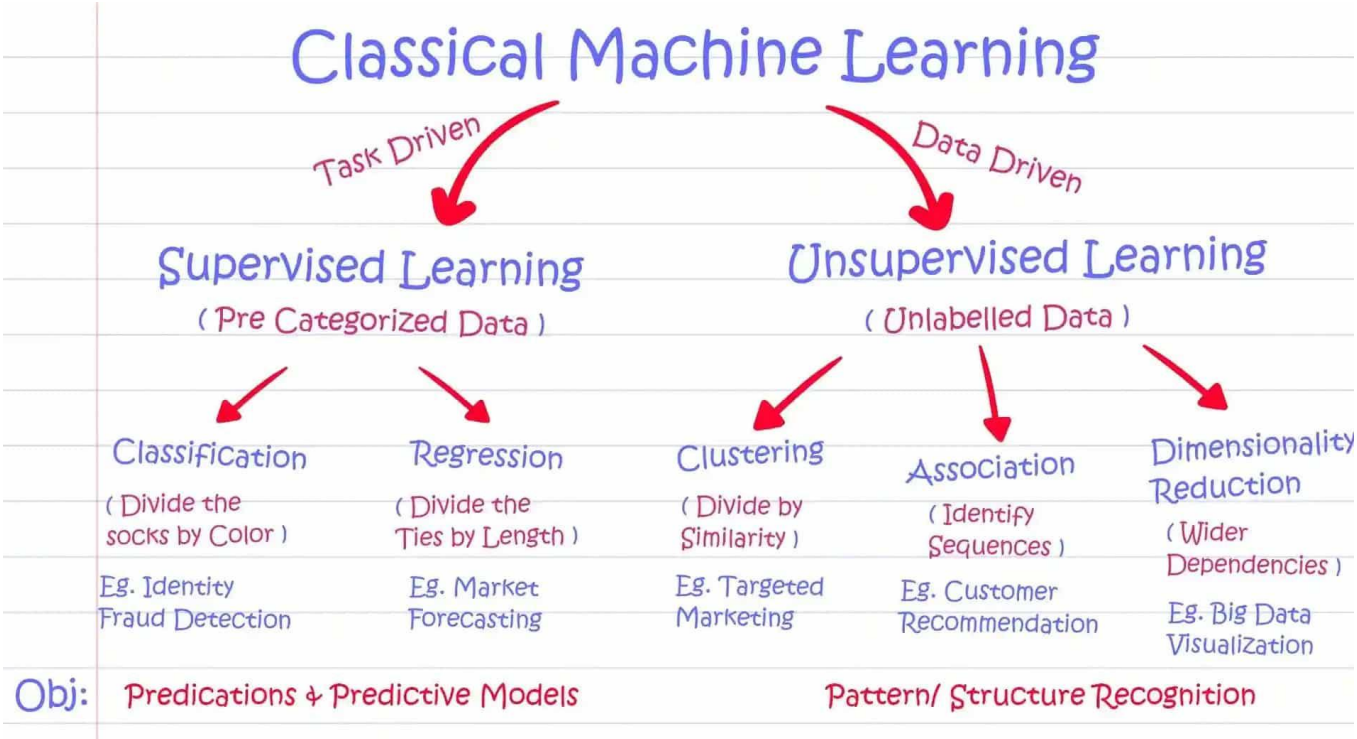
In Chapter 1, we introduced you to the fascinating world of artificial intelligence (AI) and its various learning mechanisms. We discussed three main types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. These terms represent the algorithms that drive AI systems, serving as the building blocks for programming intelligent behavior and decision-making processes. Now, let us delve deeper into how these algorithms shape the landscape of AI applications.

- **Supervised learning** involves the model learning from **labeled data**, where the input data is accompanied by the correct output. The algorithm learns to **map input data to output labels** based on example input-output pairs provided during training. The goal is to learn a mapping function so that the model can make predictions on unseen data. Examples include *linear regression, logistic regression, decision trees, support vector machines, and neural networks*.
- **Unsupervised learning**, on the other hand, deals with **unlabelled data**, where the algorithm tries to **find hidden patterns or structure** without explicit guidance. The goal of this is to explore and discover inherent structures or relationships within the

data, such as clusters or associations. Examples include *k-means clustering, hierarchical clustering, principal component analysis, and autoencoders*.

- Finally, **reinforcement learning** involves an agent learning to make decisions by interacting with an environment to maximize cumulative rewards. Through **trial and error**, the agent learns a policy or strategy to take actions that lead to the highest cumulative reward over time. These **rewards** serve as feedback, guiding the agent towards favorable actions, while **penalties** discourage undesirable behavior. This iterative learning approach is particularly effective in scenarios requiring a sequence of decisions, such as game-playing or robot control, where the agent adapts behavior based on past experiences and outcomes. Examples include *Q-learning, deep Q-networks, policy gradients, and actor-critic methods.*





https://miro.medium.com/v2/resize:fit:1100/format:webp/0*Uzqy-gqZg77Wun0e.jpg

A. SUPERVISED LEARNING

Supervised learning stands out as one of the foundational types of Machine Learning. It is a powerful approach that allows machines to learn from labeled data, making predictions or decisions based on that learning. Within supervised learning, two primary types of algorithms emerge:

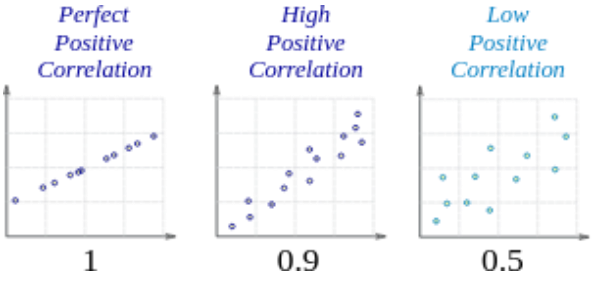
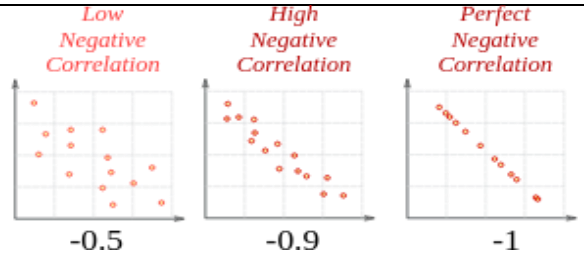
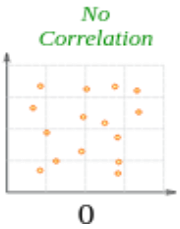
1. Regression – works with continuous data
2. Classification – works with discrete data

1. REGRESSION

Understanding Correlation: The Foundation of Regression Analysis

In data analysis, correlation is a fundamental concept that helps us grasp the relationship between variables, laying the groundwork for predictive modeling and insightful analysis. Correlation is a measure of the strength of a linear relationship between two quantitative variables (e.g. price, sales). If the change in one variable appears to be accompanied by a change in the other variable the two variables are said to be **correlated** and this inter dependence is called **correlation**.

Types of Correlation:

1. Positive Correlation: In a positive correlation, both variables move in the same direction. As one variable increases, the other also tends to increase, and vice versa.	
2. Negative Correlation: Conversely, in a negative correlation, variables move in opposite directions. An increase in one variable is associated with a decrease in the other, and vice versa.	
3. Zero Correlation: When there is no apparent relationship between two variables, they are said to have zero correlation. Changes in one variable do not predict changes in the other	

Correlation can have a value:

- 1 is a perfect positive correlation
- 0 is no correlation (the values don't seem linked at all)
- -1 is a perfect negative correlation

PEARSON'S R

Pearson's correlation coefficient (often denoted as Pearson's **r**) is one of the crucial factors to consider when assessing the appropriateness of regression analysis. **Pearson's r measures the strength and direction of the linear relationship between two continuous variables.** In the context of regression analysis, a high degree of correlation between the independent and dependent variables suggests that there may be a meaningful relationship to explore using regression techniques.

The **requirements** when considering the use of Pearson's correlation coefficient are:

1. Scale of measurement should be interval or ratio.
2. Variables should be approximately normally distributed.
3. The association should be linear.
4. There should be no outliers in the data.

Pearson's r is calculated using the formula:

$$r = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2} \sqrt{\sum_i (y_i - \overline{y})^2}}$$

***r* can take a range of values from +1 to -1**

- A value of 0 indicates that there is no association between the two variables.
- A value greater than 0 indicates a positive association; that is, as the value of one variable increases, so does the value of the other variable.
- A value less than 0 indicates a negative association; that is, as the value of one variable increases, the value of the other variable decreases.

Example 1

In the example below of 6 people with different ages and different weight, let us try calculating the value of the Pearson r.

Sr. No	Age (x)	Weight (y)
1	40	78
2	21	70
3	25	60
4	31	55
5	38	80
6	47	66

Solution:

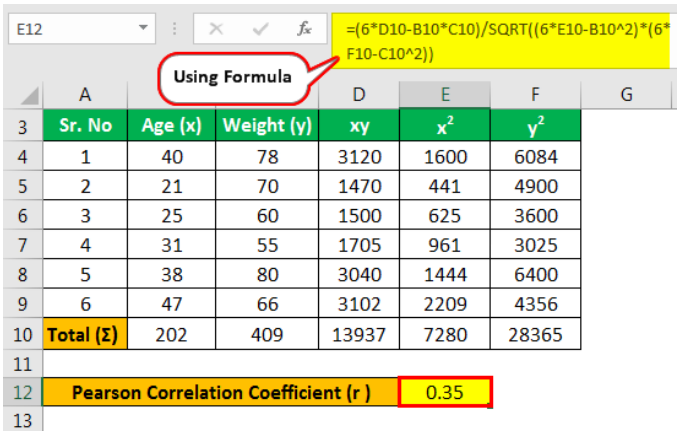
For the Calculation of the Pearson Correlation Coefficient, we will first calculate the following values:

Sr. No	Age (x)	Weight (y)	xy	x^2	y^2
1	40	78	3120	1600	6084
2	21	70	1470	441	4900
3	25	60	1500	625	3600
4	31	55	1705	961	3025
5	38	80	3040	1444	6400
6	47	66	3102	2209	4356
Total (Σ)	202	409	13937	7280	28365

Here the total number of people is 6 so, **n=6**

(6*D10-B10*C10)/SQRT((6*E10-B10^2)*(6*F10-C10^2))

Now the calculation of the Pearson R is as follows:



	A	D	E	F	G
3	Sr. No	Age (x)	Weight (y)	xy	x^2	y^2
4	1	40	78	3120	1600	6084
5	2	21	70	1470	441	4900
6	3	25	60	1500	625	3600
7	4	31	55	1705	961	3025
8	5	38	80	3040	1444	6400
9	6	47	66	3102	2209	4356
10	Total (Σ)	202	409	13937	7280	28365
11						
12	Pearson Correlation Coefficient (r)			0.35		
13						

r = (n (∑xy)- (∑x)(∑y))/(√ [n ∑x2-(∑x)2][n ∑y2– (∑y)2)

r = (6 * (13937)- (202)(409)) / (√ [6 *7280 -(202)2] * [6 * 28365- (409)2)

r = (6 * (13937)- (202) * (409))/(√ [6 *7280 -(202)2] * [6 * 28365- (409)2)

r = (83622- 82618)/(√ [43680 -40804] * [170190- 167281)

r = 1004/(√ [2876] * [2909)

r = 1004 / (√ 8366284)

r = 1004 / 2892.452938

r = 0.35

The value of the Pearson correlation coefficient is **0.35**

It is important to note that, regression analysis may not be suitable in certain situations:

1. **No Correlation**: If there is no correlation between the variables, meaning they change independently of each other, regression analysis will not provide meaningful insights or predictions.
2. **Non-linear Relationships**: While regression can model linear relationships well, it may not capture more complex, non-linear relationships effectively. In such cases, alternative techniques like polynomial regression or non-linear regression may be more appropriate.

3. **Outliers**: Outliers, or extreme data points, can disproportionately influence the regression model and lead to inaccurate predictions. In the presence of outliers, it is essential to assess their impact and consider alternative modeling approaches.
4. **Violation of Assumptions**: Regression analysis relies on certain assumptions, such as the linearity of relationships and the absence of multicollinearity (high correlation between predictor variables). If these assumptions are violated, the results of the regression analysis may be unreliable.

REGRESSION

Regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. Its primary objective is to understand and predict the value of the dependent variable based on the values of the independent variables. In simpler terms, regression helps us understand how changes in one or more variables are associated with changes in another variable.

Regression analysis is particularly useful when dealing with **continuous data,** where variables can take on any value within a certain range. For example, variables such as height, temperature, salary, and time are all continuous, meaning they can be measured along a continuous scale. In regression, these continuous variables are used to predict or explain the variability in another continuous variable, known as the dependent variable. By analyzing the relationship between the independent and dependent variables, regression allows us to make predictions and understand how changes in one variable may impact the other. This makes regression a powerful tool for forecasting, prediction, and understanding complex relationships in various fields such as economics, social sciences, and healthcare

When we make a distribution in which there is an involvement of more than one variable, then such an analysis is called Regression Analysis. It generally focuses on finding or rather predicting the value of the variable that is dependent on the other.

Let there be two variables *x and y*. If *y* depends on *x,* then the result comes in the form of a simple regression. Furthermore, we name the variables *x* and *y* as:

y – Regression / Dependent / Explained Variable. It is the variable we want to predict or understand.

x – Independent /Predictor / Explanator Variable It is used to predict or explain changes in the dependent variable.

Therefore, if we use a simple linear regression model where *y* depends on *x,* then the regression line of *y* on *x* is:

$$y = a + bx + e$$

In this equation,

- **a** represents the **intercept** of the regression line with the y-axis.
- **b** represents the **slope** of the regression line, indicating the rate of change in y for a unit change in x.
- **e** represents the **error or residual**, which accounts for the difference between the observed values of y and the values predicted by the regression equation.

FINDING THE LINE

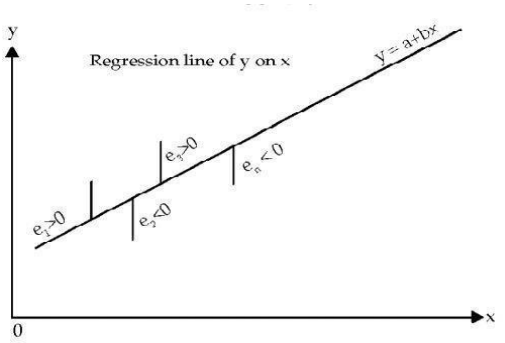
Regression analysis relies on the concept of the regression line or curve, which represents the best-fit relationship between the variables involved. This line or curve is determined by minimizing the differences between the observed values of the dependent variable and the values predicted by the regression model.

The **least squares method** is commonly employed to find this best-fit line or curve. This method minimizes the squared differences between observed and predicted values, ensuring that the regression line captures the overall trend or pattern in the data as accurately as possible. By systematically estimating the parameters of the regression model, such as the intercept and slope, the least squares method provides a precise description of the relationship between the variables.

Through the least squares method, regression analysis yields estimate of the regression coefficients that define the best-fit relationship between the variables. These coefficients allow for making predictions about the dependent variable based on the values of the independent variable(s) with greater accuracy and reliability. As a result, this is widely used in regression analysis.

Properties of the Regression line:

- The line minimizes the sum of squared difference between the observed values (actual y-value) and the predicted value (ŷ value)
- The line passes through the mean of independent and dependent features.



Example 1

In the example of 6 people with different ages and different weight, let us draw the line of best fit in Excel.

Sr. No	Age (x)	Weight (y)
1	40	78
2	21	70
3	25	60
4	31	55
5	38	80
6	47	66

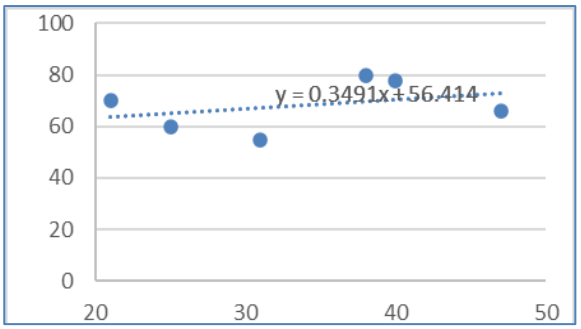
Solution:

Step 1: Select the Age and Weight.

Step 2: Insert a scatter chart and make changes to the following:

Trendline Name: **Linear**, check **Display Equation on Chart**

X axis **minimum: 20**



Step 3: Let us verify the values of slope and intercept using slope() and intercept() function in excel.

Step 4: Click on any cell and type =slope (Now, select the values of Weight, and then type comma. Now, select the values of Age and press enter.

Step 5: Click on any cell and type =intercept (Now, select the values of Weight, and then type comma. Now, select the values of Age and press enter.

Some of the regression algorithms include Linear Regression, Logistic Regression, Decision Tree Regression, Random Forest Regression. Let us learn about Linear Regression.

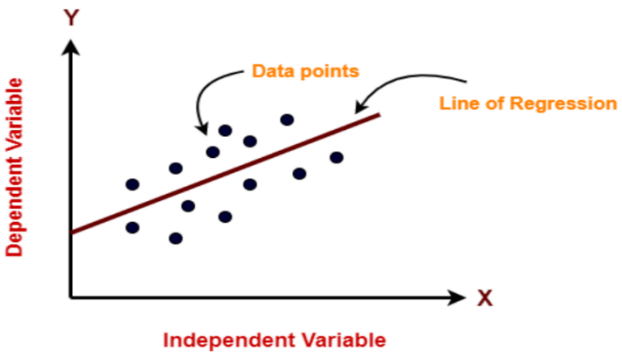
Linear Regression

Linear regression is one of the most basic types of regression in machine learning. The linear regression model consists of a predictor variable and a dependent variable related linearly to each other. In case the data involves more than one independent variable, then linear regression is called multiple linear regression models.

Linear regression is further divided into two types:

a) **Simple Linear Regression:** The dependent variable's value is predicted using a single independent variable in simple linear regression.

b) Multiple Linear Regression: In multiple linear regression, more than one independent variable is used to predict the value of the dependent variable.



Applications of Linear Regression:

- **Market Analysis**: Linear regression helps understand how different factors like pricing, sales quantity, advertising, and social media engagement relate to each other in the market.
- **Sales Forecasting:** It predicts future sales by analyzing past sales data along with factors like marketing spending, seasonal trends, and consumer behavior**.**
- **Predicting Salary Based on Experience:** Linear regression estimates a person's salary based on their years of experience, education, and job role, aiding in recruitment and compensation planning.

- **Sports Analysis:** Linear regression analyzes player and team performance by considering statistics, game conditions, and opponent strength, assisting coaches and team management in decision-making.
- **Medical Research:** Linear regression examines relationships between factors like age, weight, and health outcomes, helping researchers identify risk factors and evaluate interventions.

Advantages of Linear regression

- Simple technique and easy to implement
- Efficient to train the machine on this model

Disadvantages of Linear regression

1. Sensitivity to outliers, which can significantly impact the analysis.
2. Limited to linear relationships between variables.

https://www.javatpoint.com/linear-regression-in-machine-learning

Python program for Linear regression- (For Advanced Learners)**

Import **scipy** and draw the line of Linear Regression:

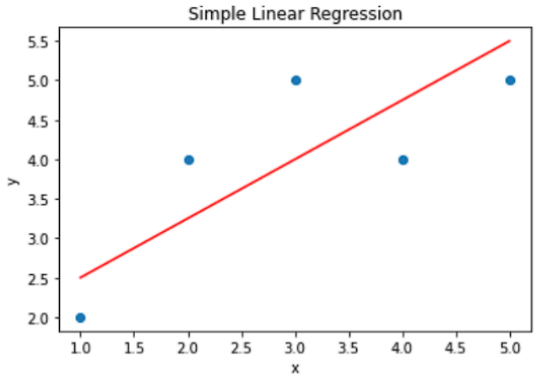
```
import numpy as np
import matplotlib.pyplot as plt
# Sample data
x = np.array([1, 2, 3, 4, 5])
y = np.array([2, 4, 5, 4, 5])
# Calculate mean and standard deviation
x_mean = np.mean(x)
y_mean = np.mean(y)
x_std = np.std(x)
y_std = np.std(y)
# Calculate covariance and slope
covariance = np.sum((x - x_mean) * (y - y_mean)) / (len(x) - 1)
slope = covariance / (x_std**2)
# Calculate y-intercept (b)
intercept = y_mean - slope * x_mean
# Predicted values
y_pred = slope * x + intercept
# Plot data and regression line
plt.scatter(x, y)
plt.plot(x, y_pred, color='red')
# Add labels and title
plt.xlabel('x')
plt.ylabel('y')
plt.title('Simple Linear Regression')
```

```
# Show the plot
plt.show()
# Print slope and intercept
print(f"Slope: {slope:.2f}")
print(f"Intercept: {intercept:.2f}")
```

This program:

- Imports numpy for numerical calculations and matplotlib.pyplot for plotting.
- Defines sample data for x and y. You can replace this with your own data.
- Calculates mean, standard deviation, covariance, and slope.
- Calculates y-intercept based on slope and mean.
- Predicts y values for given x using the linear equation.
- Plots the data points and the regression line.
- Prints the estimated slope and intercept values.

The expected output of the above program would be




```
Slope: 0.75
Intercept: 1.75
```

REFERENCES

Video links:

- https://www.youtube.com/watch?v=rHeaoaiBM6Y
- https://www.youtube.com/watch?v=VWCRDH1_rv0
- https://www.youtube.com/watch?v=YUPagM-OB_M
- https://www.youtube.com/watch?v=CtsRRUddV2s&t=70s

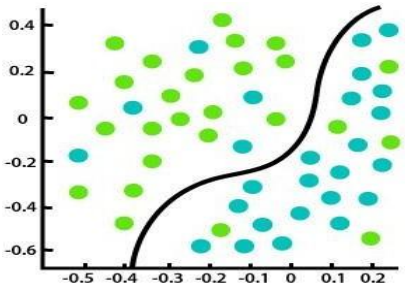
2. CLASSIFICATION

Classification is a fundamental concept in artificial intelligence and machine learning that involves categorizing data into predefined classes or categories. The main objective of classification is to assign labels to data instances based on their features or attributes. In classification, the data is typically labeled with class labels or categories, and the goal is to build a model that can accurately assign these labels to new, unseen data instances. This process is supervised learning, where the model learns from labeled training data to make predictions on unseen data.

For example, let us say, you live in a gated housing society and your society has separate dustbins for different types of waste: paper waste, plastic waste, food waste and so on. What you are basically doing over here is classifying the waste into different categories and then labeling each category. In the picture given below, we are assigning the labels 'paper', 'metal', 'plastic', and so on to different types of waste.

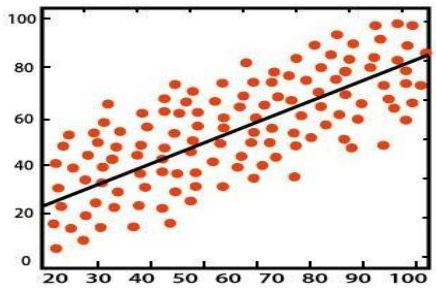
Look at the two graphs below and suggest which graph represents the classification problem.

Graph 1



_____

Graph 2



_____

How Classification Works

In classification tasks within machine learning, the process revolves around categorizing data into distinct groups or classes based on their features. Here is how it typically works:

- **Classes or Categories**: Data is divided into different classes or categories, each representing a specific outcome or group. For example, in a binary classification scenario, there are two classes: positive and negative.
- **Features or Attributes**: Each data instance is described by its features or attributes, which provide information about the instance. These features are crucial for the classification model to differentiate between different classes. For instance, in email classification, features might include words in the email text, sender information, and email subject.
- **Training Data**: The classification model is trained using a dataset known as training data. This dataset consists of labelled examples, where each data instance is

associated with a class label. The model learns from this data to understand the relationship between the features and the corresponding class labels.

- **Classification Model**: An algorithm or technique is used to build the classification model. This model learns from the training data to predict the class labels of new, unseen data instances. It aims to generalize from the patterns and relationships in the training data to make accurate predictions.
- **Prediction or Inference**: Once trained, the classification model is used to predict the class labels of new data instances. This process, known as prediction or inference, relies on the learned patterns and relationships from the training data.

Types of classification

The four main types of classification are:

1) Binary Classification
2) Multi-Class Classification
3) Multi-Label Classification
4) Imbalanced Classification

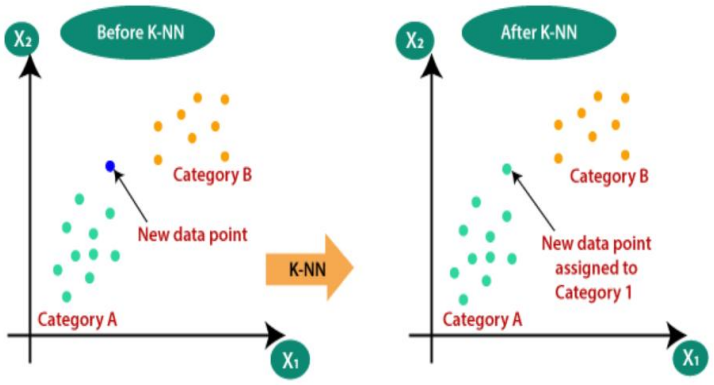
Classification Type	Binary Classification	Multi-Class Classification	Multi-Label Classification	Imbalanced Classification
Description	Classification tasks with two class labels.	Classification tasks with more than two class labels.	Classification tasks where each example may belong to multiple class labels.	Classification tasks with unequally distributed class labels, typically with a majority and minority class.
Example	• Email spam detection - spam or not • Conversion prediction - buy or not • Medical test - Cancer detected or not • Exam results - pass/fail	• Face classification • Plant species classification • Optical character recognition • Image classification into thousands of classes	• Photo classification - objects present in the photo (bicycle, apple, person, etc.)	• Fraud detection • Outlier detection • Medical diagnostic tests

K- Nearest Neighbour algorithm (KNN)

The K-Nearest Neighbors algorithm, commonly known as KNN or k-NN, is a versatile non-parametric supervised learning technique used for both classification and regression tasks. It operates based on the principle of proximity, making predictions or classifications by considering the similarity between data points.

Why KNN Algorithm is Needed:

KNN is particularly useful when dealing with classification problems where the decision boundaries are not clearly defined or when the dataset does not have a well-defined structure. It provides a simple yet effective method for identifying the category or class of a new data point based on its similarity to existing data points.



Steps involved in k-NN

- Select the number K of the neighbors
- Calculate the Euclidean distance of K number of neighbors
- Take the K nearest neighbors as per the calculated Euclidean distance.
- Among these k neighbors, count the number of the data points in each category.
- Assign the new data points to that category for which the number of the neighbor is maximum.
- Our model is ready.

Applications of KNN:

- Image recognition and classification
- Recommendation systems
- Healthcare diagnostics
- Text mining and sentiment analysis
- Anomaly detection

Advantages of KNN:

- Easy to implement and understand.
- No explicit training phase; the model learns directly from the training data.
- Suitable for both classification and regression tasks.
- Robust to outliers and noisy data.

Limitations of KNN:

- Computationally expensive, especially for large datasets.
- Sensitivity to the choice of distance metric and the number of neighbors (K).
- Requires careful preprocessing and feature scaling.
- Not suitable for high-dimensional data due to the curse of dimensionality.

Python Program for K Nearest Neighbour Algorithm- (For Advanced Learners)**

```
# importing libraries
import numpy as nm
import matplotlib.pyplot as mtp
import pandas as pd
#importing datasets
data_set= pd.read_csv('user_data.csv')

#Extracting Independent and dependent Variable
x= data_set.iloc[:, [2,3]].values
y= data_set.iloc[:, 4].values
# Splitting the dataset into training and test set.
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test= train_test_split(x, y, test_size= 0.25, random_state=0)
#feature Scaling
from sklearn.preprocessing import StandardScaler
st_x= StandardScaler()
x_train= st_x.fit_transform(x_train)
x_test= st_x.transform(x_test)
```

Reference: https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning
Dataset link: rainfall.csv

Note to the teacher: Can introduce Teachable Machine to develop any image classifier problem.
Website: https://teachablemachine.withgoogle.com/

REFERENCES

Video Session:

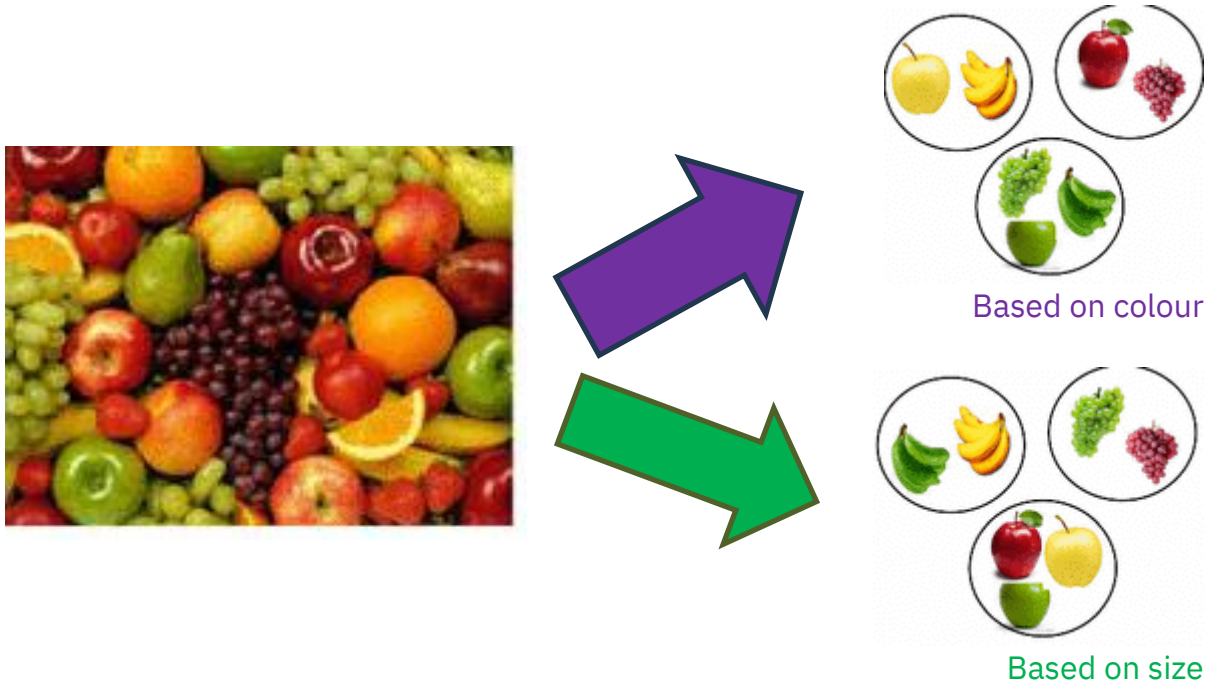
Classification: https://www.youtube.com/watch?v=hBKI7XvD8R8
KNN Algorithm: https://www.youtube.com/watch?v=0p0o5cmgLdE

B. UNSUPERVISED LEARNING

3. CLUSTERING

Clustering, or cluster analysis, is a machine learning technique used to group **unlabeled dataset** into clusters or groups based on similarity. Clustering aims to organize data points into groups where points within the same group are more similar to each other than to those in other groups. It involves finding patterns or structures in the data without the need for predefined class labels. It does it by finding some similar patterns in the unlabelled dataset such as shape, size, color, behavior, etc., and divides them as per the presence and absence of those similar patterns. It is an **unsupervised learning** method, hence no supervision is provided to the algorithm, and it deals with the unlabeled dataset.
The clustering technique is commonly used for **statistical data analysis.**

Example: Let us consider the clustering technique using a real-world example. Imagine you are visiting a shopping center where items are grouped together based on their similarities. For instance, in the fruits section, you will find apples, bananas, and grapes neatly arranged together. This organization makes it convenient for shoppers to locate specific items they are looking for.



In a similar way, clustering algorithms group similar data points together based on common characteristics or features. This approach helps in organizing and making sense of large datasets in various tasks, such as market segmentation, image recognition, and customer segmentation.

How Clustering Works:

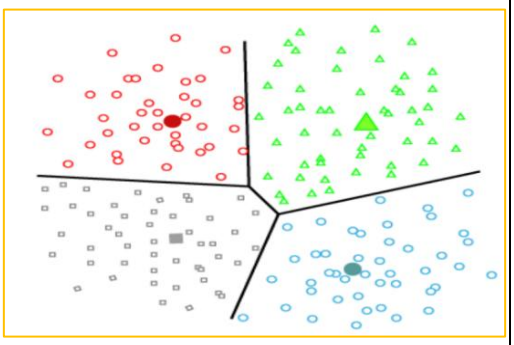
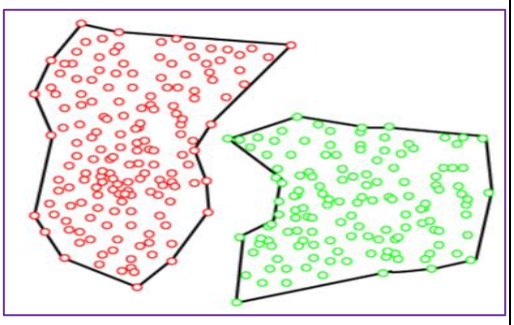
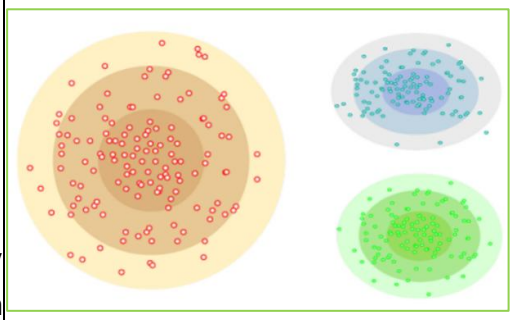
To cluster data effectively, follow these key steps:

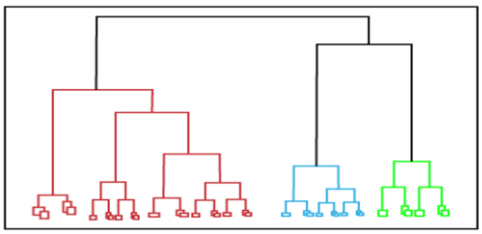
1) **Prepare the Data:** Select the right features for clustering and make sure the data is ready by scaling or transforming it as needed.
2) **Create Similarity Metrics:** Define how similar data points are by comparing their features. This similarity measure is crucial for clustering.
3) **Run the Clustering Algorithm:** Apply a clustering algorithm to group the data. Choose one that works well with your dataset size and characteristics.
4) **Interpret the Results:** Analyze the clusters to understand what they represent. Since clustering is unsupervised, interpretation is essential for assessing the quality of the clusters.

Types of Clustering Methods

Some of the common clustering methods used in Machine learning are:

1) Partitioning Clustering
2) Density-Based Clustering
3) Distribution Model-Based Clustering
4) Hierarchical Clustering

1. Partitioning Clustering It is a type of clustering that divides the data into non-hierarchical groups. It is also known as the **centroid-based method**. The most common example of partitioning clustering is the K-Means Clustering algorithm. In this type, the dataset is divided into a set of k groups, where k is used to define the number of pre-defined groups. The cluster center is created in such a way that the distance between the data points of one cluster is minimum as compared to another cluster centroid.	
2. Density-Based Clustering The density-based clustering method connects the highly-dense areas into clusters, and the arbitrarily shaped distributions are formed as long as the dense region can be connected. This algorithm does it by identifying different clusters in the dataset and connects the areas of high densities into clusters. The dense areas in data space are divided from each other by sparser areas. These algorithms can face difficulty in clustering the data points if the dataset has varying densities and high dimensions.	
3. Distribution Model-Based Clustering Distribution model-based clustering assumes data within each cluster follows a specific probability distribution, often a **Gaussian** (bell-shaped) distribution, and uses this to calculate the likelihood of data points belonging to each cluster. For example, Gaussian Mixture Models (GMM) employ this approach with the **Expectation-Maximization algorithm**.	

4. Hierarchical Clustering Hierarchical clustering can be used as an alternative for the partitioned clustering as there is no requirement of pre-specifying the number of clusters to be created. In this technique, the dataset is divided into clusters to create a tree-like structure, which is also called a **dendrogram**. The observations or any number of clusters can be selected by cutting the tree at the correct level. The most common example of this method is the **Agglomerative Hierarchical algorithm.**	

K- Means clustering

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. The k-means algorithm is one of the most popular clustering algorithms. It classifies the dataset by dividing the samples into different clusters of equal variances. The number of clusters must be specified in this algorithm.

Steps involved K-Means Clustering:

The working of the K-Means algorithm is explained in the below steps:

- Select the number K to decide the number of clusters.
- Select random K points or centroids. (It can be other from the input dataset).
- Assign each data point to their closest centroid, which will form the predefined K clusters.
- Calculate the variance and place a new centroid of each cluster.
- Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.
- If any reassignment occurs, then go to step-4 else go to FINISH.
- The model is ready.

Activity: Visual AI: This tool allows you to visualize K-means clustering in real-time. Upload your own data or use provided examples, adjust parameters, and see how clusters change visually using the link Visualise k-means

Applications of K-Means Clustering:

- **Market Segmentation:** group customers based on similar purchasing behaviours or demographics for tailored marketing strategies.
- **Image Segmentation:** partition images into regions of similar colours to aid in tasks like object detection and compression.
- **Document Clustering:** categorize documents based on content similarity, aiding in organization and information retrieval.

- **Anomaly Detection:** identify outliers by clustering normal data points and detecting deviations.
- **Customer Segmentation:** segment customers for targeted marketing and personalized experiences.

Advantages of K-Means Clustering:

- Easy to implement, making it suitable for users of all levels.
- Handles large datasets with low computational resources.
- Works well with numerous features and data points.
- Are easy to understand, aiding in decision-making.
- Applicable across various domains and data types.

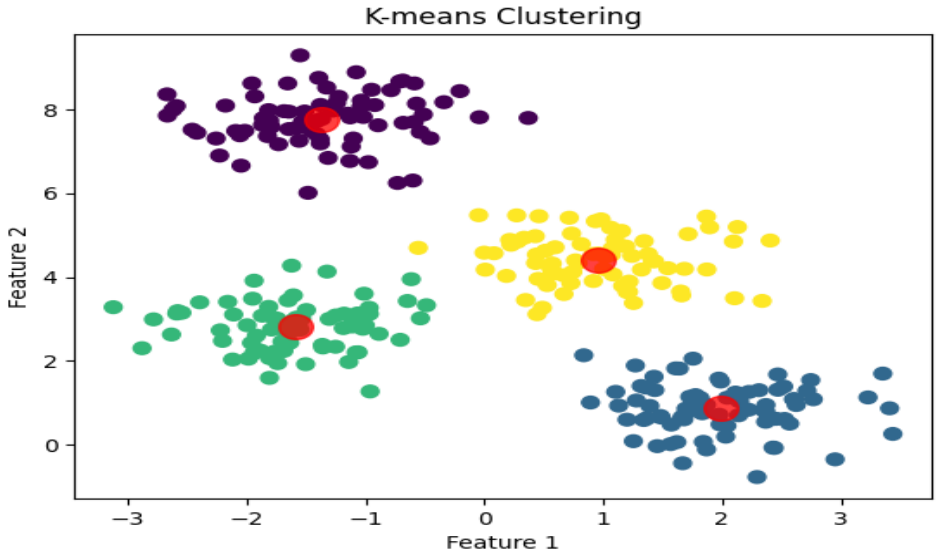
Limitations of K-Means Clustering:

- Results can vary based on initial centroid placement.
- Assumes clusters are spherical, which is not always true.
- Number of clusters must be known beforehand.
- Outliers can distort clusters due to their influence on centroids.
- May converge to suboptimal solutions instead of the global optimum.

Python Program for K Means Clustering- (For Advanced Learners)**

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans

# Generate synthetic data
X, _ = make_blobs(n_samples=300, centers=4, cluster_std=0.60, random_state=0)

# Apply K-means clustering
kmeans = KMeans(n_clusters=4)
kmeans.fit(X)
y_kmeans = kmeans.predict(X)

# Plot the data points and centroids
plt.scatter(X[:, 0], X[:, 1], c=y_kmeans, s=50, cmap='viridis')
centers = kmeans.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='red', s=200, alpha=0.75)
plt.title('K-means Clustering')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.show()
```

Output



This program does the following:

1. Generates synthetic data using make_blobs from sklearn.datasets.
2. Applies K-means clustering with n_clusters=4.
3. Plots the data points colored by their cluster assignments and shows the centroids as red circles.

You can adjust the parameters like the number of clusters, standard deviation, and number of samples in make_blobs to observe different clustering scenarios.

REFERENCES

Video Session

Videos: Clustering: https://www.youtube.com/watch?v=wk2ylI1qgU0
K-means clustering: https://www.youtube.com/watch?v=4b5d3muPQmA&t=119s

EXERCISES

A. Multiple Choice Questions

1. Which of the following are the types of correlation?
    a. Positive correlation
    b. Negative Correlation
    c. No correlation
    d. All of the above
2. Which of the following techniques is an analysis of the relationship between two variables to provide the prediction mechanism?
    a. Standard error
    b. Correlation
    c. Regression
    d. None of the above
3. Which of the given plots is suitable for testing the linear relationship between a dependent and independent variable?
    a. Bar chart
    b. Scatter plot
    c. Histograms
    d. All of the above

4. Which of the following scatter plots represents a positive correlation?
    a. points scattered randomly with no apparent trend
    b. points forming a diagonal line and bottom left to top right
    c. points forming a diagonal line from top left to bottom right
    d. points clustered around a central point
5. Which regression technique is used when there is only one independent variable?
    a. logistic regression
    b. multiple linear regression
    c. simple linear regression
    d. polynomial regression
6. What is one advantage of linear regression analysis?
    a. it is robust to outliers
    b. it can capture nonlinear relationships between variables
    c. it is simple and easy to interpret
    d. it is suitable for classification tasks
7. What is supervised learning in Artificial Intelligence?
    a. training a computer algorithm on input data that is not labelled.
    b. training a computer algorithm on input data that has been labelled for a specific output.
    c. training a computer algorithm without any input data
    d. training a computer algorithm to perform unsupervised tasks.
8. Which type of classification involves categorizing data into two distinct classes?
    a. multi-class classification
    b. binary classification
    c. unsupervised classification
    d. regression classification
9. What is logistic regression commonly used for in binary classification?
    a. categorizing observations into multiple classes
    b. predicting continuous values for input data
    c. categorizing observations into two distinct classes
    d. identifying unstructured data patterns
10. What is the primary goal of classification in AI?
    a. categorizing data into random groups
    b. locating and classifying things or concepts into predefined groups
    c. predicting continuous values for input data
    d. identifying unstructured data patterns
11. Which algorithm is commonly used for binary classification?
    a. Decision trees
    b. Support Vector Machine
    c. Logistic Regression
    d. k-Nearest Neighbors
12. The K-Nearest Neighbors (KNN) algorithm assigns a class to new data point by considering:
    a. Distance from the data point to a predefined decision boundary
    b. Majority vote of its K nearest neighbors in the training data
    c. Similarity of the data point to a cluster centroid

d. probability of each class given the data point's features.

13. What does a classification model in AI ultimately want to achieve?
    a. to identify patterns and associations in data
    b. to predict continuous numerical values
    c. to categorize input data into predefined classes or labels
    d. to optimize decision-making processes
14. What are some challenges in applying classification models to real-world problems?
    a. Data bias and fairness
    b. Interpretability and explainability
    c. overfitting and underfitting
    d. All of the above
15. What is clustering?
    a. Grouping labeled dataset
    b. Dividing data into different clusters
    c. Finding linear association between variables
    d. Predicting future behaviors of a dependent variable
16. Which type of learning does clustering belong to?
    a. Supervised learning
    b. Unsupervised learning
    c. Semi-supervised learning
    d. Reinforcement learning
17. Which method is used to group highly dense areas into clusters?
    a. Partitioning clustering
    b. Density-based clustering
    c. Distribution model-based clustering
    d. Hierarchical clustering
18. Which algorithm is an example of partitioning clustering?
    a. Mean-shift algorithm
    b. DBSCAN algorithm
    c. K-Means algorithm
    d. Fuzzy clustering algorithm
19. Which clustering method allows data objects to belong to more than one group or cluster?
    a. Partitioning clustering
    b. Density-based clustering
    c. Distribution model-based clustering
    d. Fuzzy clustering
20. Which clustering algorithm is sensitive to outliers?
    a. K-Means algorithm
    b. Mean-shift algorithm
    c. DBSCAN algorithm
    d. Hierarchical clustering

B. Fill in the blanks

1. In _____ type of ML, the models are not trained in labeled data sets.
2. The _____ measures the linear relationship between the independent and dependent variables.
3. _____predicts continuous numerical values, while Logistic regression predicts discrete categories.
4. _____ are data points on the scatterplot that do not follow the pattern of the dataset.
5. _____ algorithm operates based on the principle of proximity, making predictions by considering the similarity between data points.
6. Clustering is a machine learning technique used to group _____ dataset.
7. Partitioning clustering divides the data into non-hierarchical groups, also known as _____ method.
8. Density-based clustering connects highly dense areas into clusters, separated by areas of _____.
9. The primary requirement for the number of clusters in K-Means algorithm is _____ beforehand.
10. Clustering is widely used in applications such as market segmentation and _____.

C. True or False:

1. Clustering is a supervised learning technique.
2. Hierarchical Clustering requires pre-specifying the number of clusters.
3. Fuzzy clustering is a hard clustering method.
4. Classification is an unsupervised learning technique.
5. In k-NN algorithm, k is the number of nearest data points.
6. K-Means algorithm requires specifying the number of clusters.

ANSWERS

A. Multiple Choice Questions

1. a. All of the above
2. c. Regression
3. b. Scatter plot
4. b. points forming a diagonal line and bottom left to top right
5. c. simple linear regression
6. c. it is simple and easy to interpret
7. b. training a computer algorithm on input data that has been labelled for a specific output.
8. b. binary classification

9. c. categorizing observations into two distinct classes
10. b. locating and classifying things or concepts into predefined groups
11. c. Logistic Regression
12. b. Majority vote of its K nearest neighbors in the training data
13. c. to categorize input data into predefined classes or label
14. d. All of the above
15. b. Dividing data into different clusters
16. b. Unsupervised learning
17. c. Distribution model-based clustering
18. c. K-Means algorithm
19. d. Fuzzy clustering algorithm
20. a. K-Means algorithm

B. Fill in the blanks

1. Unsupervised Learning
2. Correlation coefficient
3. Linear Regression
4. Outlier
5. K-nearest neighbors (KNN) algorithm
6. unlabelled dataset
7. centroid-based method
8. low point density
9. Specified
10. Data Analysis

C. True or False:

1. False 2. False 3. False 4. False 5. True 6. True

D. Short answer type questions:

1. What is Machine learning? Name the three methods of machine learning.

Ans. Machine learning (ML) is a type of artificial intelligence (AI) focused on building computer systems that learn from data. It uses algorithms that learn from data to make predictions. The predictions can be generated through three methods known as supervised learning, unsupervised learning and reinforcement learning. In supervised learning algorithms learn patterns from existing data, in unsupervised learning, they discover general patterns in data and reinforcement learning where they learn through reward and punishment methods.

2. How are correlation measures used in AI applications?

Ans. Feature selection: Identify features highly correlated with the target variable, potentially indicating relevance for prediction. Exploratory data analysis: Understand relationships between variables and identify potential or anomalies. Recommender systems: Recommend items based on past user behaviour and correlations between items purchased together.

3. Name some examples of regression algorithms?

Ans. Examples of regression algorithms include Linear Regression, Polynomial Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Support Vector Regression (SVR), Decision Tree Regression, Random Forest Regression, and Gradient Boosting Regression. These algorithms are used to predict continuous numerical values are widely applied in various fields such as finance, economics and engineering.

4. What are regression algorithms used for?

Ans. Regression algorithms are used for predicting continuous numerical values based on input features. They are widely applied in various fields such as finance for stock price forecasting, economics for predicting economic indicators, healthcare for disease progression estimation, and engineering for predicting product performance. Regression analysis helps uncover relationships between variables and make informed predictions for future data points.

5. What is Linear regression? Give two applications of regression in machine learning?

Ans. Linear Regression is a supervised learning algorithm. It makes use of one independent variable X to predict the outcome of a second dependent variable Y. In machine learning, regression is used to predict outputs and forecast trends.

6. How can outliers impact regression analysis?

Ans. An outlier is a data point that differs significantly from other observations. An outlier may be due to a variability in the measurement, an indication of data which may be collected or it may be the result of experimental error. They can significantly skew the results of regression analysis by distorting the regression line and affecting the accuracy of predictions.

7. What is the primary difference between classification and regression?

Ans. Classification predicts discrete values, while regression predicts continuous values.

8. Provide examples of classification problems in real-life scenarios.

Ans. Examples of classification problems include email spam detection, handwritten character recognition, and sentiment analysis in social media posts.

9. What are some common applications of clustering techniques?

Ans. Common applications of clustering techniques include market segmentation, statistical data analysis, social network analysis, image segmentation, and anomaly detection.

10. List the types of clustering methods.

Ans. Types of clustering methods include partitioning clustering, density-based clustering, distribution model-based clustering, hierarchical clustering, and fuzzy clustering.

E. Long Answer Questions:

1. How does classification model work?

Ans.

- **Classes or Categories:** Data is organized into different groups, such as "positive" and "negative," representing distinct outcomes.
- **Features or Attributes:** Each data instance is described by specific characteristics or attributes, providing information about the instance.
- **Training Data:** The classification model learns from a dataset containing labeled examples, associating each instance with a class label.
- **Classification Model:** An algorithm or technique is applied to the training data to build a model that can predict the class labels of new instances.
- **Prediction or Inference:** Once trained, the model is used to classify new data instances based on the patterns learned during training.

2. Explain the types of clustering.

Ans.

- **Partitioning Clustering**: This method divides data into non-hierarchical groups using a centroid-based approach, where data points are grouped into k clusters based on the proximity to cluster centroids. Eg: K-Means Clustering algorithm.
- **Density-Based Clustering**: This technique identifies clusters by connecting highly dense areas in the dataset, allowing for arbitrarily shaped clusters to form as long as dense regions are connected. Eg: DBSCAN algorithm.
- **Distribution Model-Based Clustering**: Here, data is clustered based on the probability of belonging to a particular distribution, often assuming Gaussian distributions. Eg: Expectation-Maximization Clustering algorithm, using Gaussian Mixture Models (GMM).
- **Hierarchical Clustering**: This approach creates a tree-like structure, or dendrogram, to cluster data without requiring the pre-specification of the number of clusters. Eg: Agglomerative Hierarchical algorithm.

3. Write any two advantages and disadvantages of linear regression.

Ans:

Advantages of Linear regression

- Simple technique and easy to implement
- Efficient to train the machine on this model

Disadvantages of Linear regression

- Sensitivity to outliers, which can significantly impact the analysis.
- Limited to linear relationships between variables.

4. What are the steps involved in k-NN algorithm?

Ans:

- Select the number K of the neighbors
- Calculate the Euclidean distance of K number of neighbors
- Take the K nearest neighbors as per the calculated Euclidean distance.
- Among these k neighbors, count the number of the data points in each category.
- Assign the new data points to that category for which the number of the neighbor is maximum.
- Our model is ready.

5. What are the steps involved in k-means clustering?

Ans:

- Select the number K to decide the number of clusters.
- Select random K points or centroids. (It can be other from the input dataset).
- Assign each data point to their closest centroid, which will form the predefined K clusters.
- Calculate the variance and place a new centroid of each cluster.
- Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.
- If any reassignment occurs, then go to step-4 else go to FINISH.
- The model is ready.

F. Competency Based Questions

1. Asmita is developing an AI-driven recommendation system for a retail e-commerce platform. What type of machine learning method might she have used to:
   a. Train the model with details of past purchases, user interactions, and product ratings?
   b. Identify groups of similar users or products based on their browsing behavior?

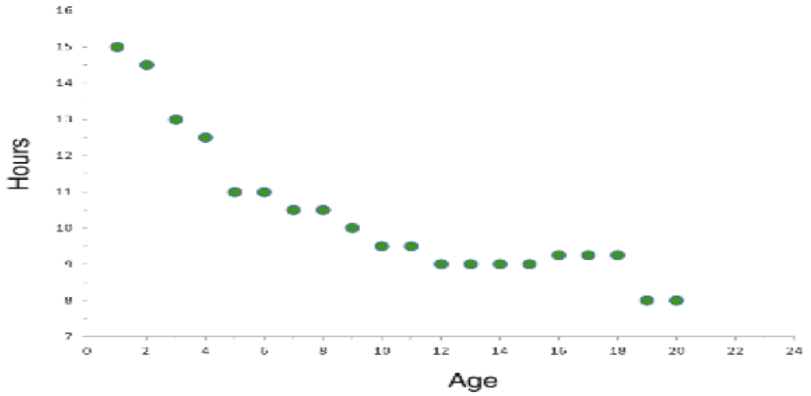
Ans:

a. Supervised Learning
b. Unsupervised Learning

2. Suppose you are a sales manager tasked with forecasting sales for the upcoming quarter. Describe how you would use linear regression in this scenario, including the data you would collect and the steps involved in the analysis.

Ans. In sales forecasting, linear regression can be used to **predict future sales** based on historical sales data, marketing spends, seasonality, and other factors. The sales manager would collect historical sales data along with relevant variables such as advertising expenditures, promotional activities, and economic indicators. By analysing this data using linear regression, the sales manager can forecast future sales trends and adjust strategies accordingly.

3. Observe the scatter plot showing the amount of sleep needed per day by age.



What type of correlation is shown here?

Ans. As age increases (moving along the x-axis toward greater numbers), the amount of sleep needed decreases (y-values decreasing). This is a **negative correlation**. This indicates that as individuals grow older, they generally require less sleep.

4. Ramesh is working on an assignment where he needs to categorize real-world applications of Artificial Intelligence (AI) into two groups: Classification and Clustering. While his initial attempt seems partially correct, his teacher identified a mistake.

Classification	Clustering
Medical Diagnosis	E-mail Spam Detection
Sentiment Analysis	Identifying high risk patient groups
Fraud Detection	Anomaly detection in network traffic

Identify the mistake

Ans: E-mail Spam Detection is categorized under Clustering, but it should be under Classification.

5. Researchers are developing a new blood test to detect cancer early. The test analyzes various biomarkers (indicators) in a patient's blood sample. The test results need to be categorized accurately. A positive result should indicate the presence of cancer cells, while a negative result should indicate no cancer. Which type of classification algorithm would be most suitable for this new cancer detection blood test?

Ans: Binary Classification

UNIT 7: Leveraging Linguistics and Computer Science

Title: Leveraging linguistics and computer science	**Approach:** Interactive/ Collaborative / Activity

Summary:

Combining the methods and ideas from computer science with linguistics allows for the improvement of many applications pertaining to language understanding and processing. Significant progress is made in natural language processing, machine translation, sentiment analysis, information retrieval, speech recognition, syntax parsing, computational semantics, and dialogue systems by fusing linguistic theories with computer approaches. Computers can now comprehend, analyze, and produce human language more effectively because to this multidisciplinary approach, which also improves their performance in tasks like sentiment analysis, speech recognition, search engine optimization, translation, and conversation interaction.

Learning Objectives:

1. Understand the challenges of natural language processing (NLP) and its importance in modern technology.
2. Explore the components and processes involved in NLP, including lexical analysis, syntactical analysis, semantic analysis, discourse integration, and pragmatic analysis.
3. Learn about the applications of NLP in various fields such as sentiment analysis, smart assistants, email filtering, predictive text, document analysis, and automatic summarization.

Key Concepts:

1. Understanding Human Language Complexity
2. Introduction to Natural Language Processing (NLP) - Emotion Detection and Sentiment Analysis, Classification Problems, Chatbot
3. Phases of NLP
4. Applications of NLP

Learning Outcomes:

Develop a better understanding of the complexities of language and the challenges involved in NLP tasks. Learn new techniques and algorithms for NLP tasks.

Pre-requisites: Knowledge in language, NLP, Data, AI, ML, NLP

Unlocking the Secrets of Language: A Teacher's Guide to Natural Language Processing (NLP)

This lesson equips you to introduce students to the fascinating world of Natural Language Processing (NLP) – how computers understand and process human language.

Hooking Students with the Challenges of Communication:

- **Break the Ice:** Start with a fun activity! Play a classic game of telephone or try deciphering a coded message. Discuss how even simple communication can be easily misunderstood.

Lost in Translation? Computers and Human Language:

- Spark curiosity with a discussion on how computers, unlike humans, struggle to understand the complexities of human language. Explore challenges like:
    - **Ambiguity:** Words can have multiple meanings depending on context.
    - **Sarcasm and Humor:** Computers struggle to interpret these subtleties.
    - **Informal Language:** Slang and colloquialisms pose difficulties for machines.

Introducing NLP: Bridging the Gap:

- Enter NLP! Introduce Natural Language Processing – the field that bridges the gap between human language and computers.

Demystifying NLP Concepts:

- Explain key NLP concepts with real-world applications:
    - **Tokenization:** Breaking down text into individual units like words or phrases. (e.g., "This is a sentence" becomes ["This", "is", "a", "sentence"])
    - **Sentiment Analysis:** Extracting opinions and emotions from text. (e.g., Analysing product reviews to understand customer sentiment)

Looking Beyond Words: Recognizing Emotions in Text:

- Explore how NLP goes beyond words to analyze attitudes and emotions in text. Discuss:
    - **Applications:** Social media monitoring, customer service chatbots, market research.
    - **Challenges:** Understanding context, sarcasm, and cultural nuances.

Chatbots: Conversational AI in Action:

- Introduce chatbots – computer programs that simulate conversation with humans. Explore examples like virtual assistants or customer service bots.

Applying NLP in the Real World:

- Guide students in exploring how NLP tackles real-world tasks:
    - **Sentiment Analysis:** Analyzing online reviews or social media posts to understand public opinion.
    - **Text Summarization:** Automatically creating concise summaries of lengthy documents.
    - **Language Translation:** Breaking down language barriers with machine translation.
    - **Chatbot Development:** Building interactive chatbots for various applications.

Designing Your Own Chatbot:

- Take it to the next level! Guide students through the design process for their own chatbot:
    - **User Personas:** Define who the target user of the chatbot is.
    - **Goals:** Determine what the chatbot should accomplish.
    - **Conversational Flows:** Map out the conversation flow and responses.
    - **Implementing NLP:** Integrate NLP techniques for understanding user input.
    - **Iteration and Feedback:** Test, refine, and improve the chatbot based on user feedback.

This lesson provides a captivating introduction to NLP and empowers students to explore its potential applications in our ever-evolving world.

Understanding Human Language Complexity

Teachers can ask the following questions:

1. **What are some challenges people face when communicating with each other, even if they speak the same language?** (This question primes students to think about the complexities of human language that the passage will address.)
2. **Have you ever interacted with a technology that seems to understand what you say? What is it?** (This question activates students' prior knowledge about virtual assistants and NLP, which will be discussed later in the passage.)

Linguistics is a field of study that focuses on the strategic application of linguistic principles and practices to meet specific objectives or goals. The field of linguistics is often used in marketing, advertising, communications, education, natural language processing, etc. Understanding how language works (including its structure, semantics and pragmatics, as well as sociolinguistics) allows individuals and organizations to tailor their messages, improve communication and influence behavior.

Understanding human language is difficult, even for people who have grown up with it. Human language is incredibly complex, full of strange expressions that seem to contradict each other, metaphors that require cultural knowledge to understand, and grammatical structures that sometimes turn simple ideas into **tongue-twisters**. Machines require systems that research scientists call **Natural Language Processing**, or NLP, to understand human language.

NLP is a branch of Artificial Intelligence (AI) that allows computers to understand, create, and manipulate human speech. NLP has the capability to query the data with natural-language text or voice. It is also known as “language in”. Most consumers have interacted with NLP. For example, NLP is at the core of the technology behind virtual assistants like ODA, Siri, Cortana, Alexa, etc. NLP can be applied to both written text and speech. Some examples of tools that are powered by NLP are: Web search, Email spam filtering, Auto-translate text or speech Document summarization Sentiment analysis Grammar/spell checking.

Activity 1: Find five interesting facts about “IBM Project Debater” and write in the space provided below.

Answer: Real-Time Argument Generation, Massive Data Processing, Natural Language Understanding, Debate with Humans, Speech Synthesis and Delivery

Introduction to Natural Language Processing

Teachers can ask the following questions:

1. **What are some ways that computers typically organize information?** (This question primes students to think about structured data, which will be contrasted with unstructured human language in the passage.)
2. **Can you think of any examples from your own experience where language can be confusing or misleading?** (This question activates students' prior knowledge about ambiguity in language, which will be discussed later in the passage.)

Computers are best at working with structured data, in which everything is neatly grouped and labelled. Unfortunately for machines, human language is anything but structured. You have been using language for most of your life. Your brain accomplishes this through some of the most complicated neural circuitry on Earth. But it is very difficult to create machines that can work with human language.

In NLP, machines segment sentences and extract meaning from "tokens" of human language. In everyday language, a "token" is defined as an individual unit of text—typically a word—that is used for analyzing and processing information. Although it is loosely held together by rules of grammar, human language is largely unstructured. Unlike structured information, which can be arranged in tables or matrices with neatly labeled rows and columns, unstructured information is messy and difficult to understand. To see why, consider this famous joke by Groucho Marx.

"One morning I shot an elephant in my pajamas. How he got in my pajamas, I don't know."

-Adapted from Groucho Marx, 20th century comedian and movie star

To deal with the "messiness" of unstructured information, computers begin with one sentence at a time. This is called **sentence segmentation**. Computers then break the information into small chunks of information, called **tokens**, that can be individually classified. Once the tokens in text have been sorted into a structure based on what they mean, NLP can work with them.

The following activities show you how Groucho Marx's joke can be tokenized into useful categories called **entities** and **relationships.**

- An **entity** is a noun representing a person, place, or thing. It is not an adjective, verb, or other article of speech.

Activity 2:

Now keeping this in mind, identify the entities in the sentence "**I shot an elephant in my pajamas**."

Answer: I, elephant, pajamas

- A **relationship** is a group of two or more entities that have a strong connection to one another.

Activity 3:

Recall Groucho Marx's quote:

"One morning, I shot an elephant in my pajamas. How he got my pajamas, I don't know."

Now keeping this in mind, identify the relationship between the entities in Groucho Marx's statement:

I + elephant, I + pajamas, in + pajamas, elephant + pajamas, I + shot

Relationship between two entities	Not a relationship between two entities
I + elephant, I + pajamas	**I + shot**
elephant + pajamas	**in + pajamas**

Once an AI has classified entities and relationships in text or speech, the AI can begin structuring the information as a step toward understanding it. Your brain, by the way, does the same thing, which might have helped you find entities and relationships in the previous activities. For example, consider the following two sentences: "Manu broke the toy. He always breaks toys." Notice that there is a relationship between the two sentences: the word **he** is related to the word **Manu**. The machine uses NLP to identify this relationship.

- A **concept** is something implied in a sentence but not actually stated. This is trickier because it involves matching ideas rather than the specific words present in the sentence.

Activity 4:

Recall, once more, Groucho Marx's statement:

"One morning, I shot an elephant in my pajamas. How he got my pajamas, I don't know."

Based on this statement, select the words that are concepts implied, but not stated.

- ☐ Safari
- ☐ Rifle
- ☐ Photographed
- ☐ Pajamas

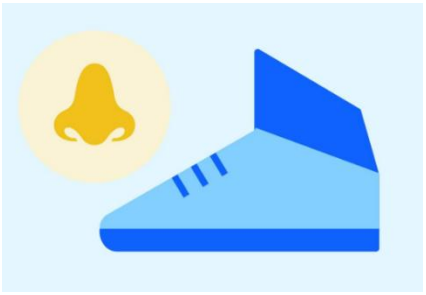
Answers: Rifle, Photographed, Safari

Emotion Detection and Sentiment Analysis

Aspect	Emotion Detection	Sentiment Analysis
Definition	Identifies distinct human emotion types.	Measures the strength of an emotion.
Examples	Determining if an expression is anger, happiness, etc.	Assessing if data is positive, negative, or neutral.
Use Cases	Analyzing user ratings, comments in surveys, etc.	Reading social media posts, customer service chats, etc.

Aspect	Emotion Detection	Sentiment Analysis
AI Training	Can be trained to classify emotions.	Utilizes a sliding scale between positive and negative.
Purpose	Identifying emotional tokens to understand context.	Assessing the overall tone or sentiment of text data.

Classification Problem



Here is an old-fashioned riddle:

Why does your nose run and your feet smell?

Human language is full of terms that are vague or have double meanings. This is called a **classification problem**.

In this riddle, the phrases "your nose run" and "your feet smell" are used in a humorous way to highlight the ambiguity of language, which poses a classification problem. For instance, "a runny nose" typically refers to having a cold and needing tissues to wipe one's nose, while "a smelly foot" indicates that one's foot has an unpleasant odour. The challenge arises because the words "run" and "smell" have multiple meanings.

Similarly, in everyday language, phrases like "***shipping a box by train***" or "***filling in a form by filling it out***" may seem contradictory or confusing due to the double meanings of the words used. While humans can quickly grasp the intended meaning based on context, AI systems may struggle to accurately classify such phrases without a comprehensive understanding of language nuances and context.

To address this problem,

- An AI system utilizes machine learning techniques such as supervised learning.
- By feeding the system with a large dataset containing examples of language usage and their corresponding classifications, the AI system learns patterns and relationships between words, phrases, and their meanings.
- Over time and with exposure to more data, the AI system improves its classification accuracy by adjusting its internal parameters based on the observed patterns.
- AI systems may not achieve perfect classification accuracy. There will always be some degree of uncertainty or error associated with the system's classifications. To address this, well-designed AI systems not only provide a response but also a confidence value, indicating the system's level of certainty in its classification.

Chatbots

Teachers can ask the following questions:

1. **Imagine you are playing a game with a friend where you can only give pre-programmed responses. What are some challenges you might face in communicating with your friend?** (This question primes students to think about the limitations of rule-based systems, which will be introduced in the context of chatbots later in the passage.)
2. **Have you ever interacted with a voice assistant or messaging bot? What did you find interesting or challenging about your experience?** (This question activates students' prior knowledge about chatbots and conversational interfaces, which will be discussed in detail in the passage.)

Chatbots are software applications or computer programs designed to simulate conversation with human users, typically through text-based or voice-based interactions. They use artificial intelligence (AI), natural language processing (NLP), and machine learning techniques to understand user queries and provide appropriate responses. Chatbots can be integrated into various platforms such as websites, messaging apps, and voice assistants, and they serve a wide range of purposes including customer support, information retrieval, task automation, and entertainment.

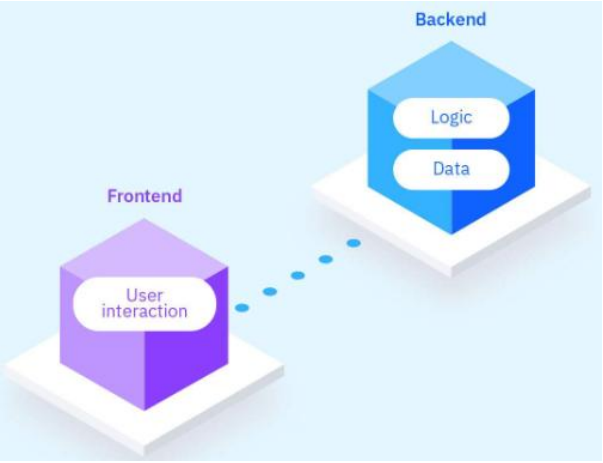
Chatbots, utilized in various fields from retail to healthcare, offer constant availability for online engagement. While not always capable of providing answers, they serve to listen and respond to repetitive inquiries, relieving businesses from the need to allocate human resources for such tasks. *Chatbots operate effectively with small-scale data*, catering to specific needs like movie queries for a cinema chain or broader inquiries for AI systems scanning social media.

Chatbots	Rule-based Chatbots	AI-powered Chatbots
Description	Operate on predefined rules and decision trees. Follow programmed rules to respond to user input.	Utilize natural language processing (NLP) and machine learning algorithms. Also known as chat agents or virtual assistants.
Advantages	- Easy to develop and maintain. - Provide consistent and accurate answers to specific questions.	- 24/7 availability for immediate and consistent support. - Offer personalized interactions based on user preferences and history. - Improve efficiency and cost savings by automating tasks and reducing service costs.
Limitations	- Struggle with understanding complex language. - Unable to adapt to situations beyond programmed rules.	- High development costs and resource requirements. - Prone to biases from training data and lack of transparency in decision-making. - Ethical considerations regarding privacy, manipulation, and responsible use.
Use Cases	- Customer service tasks like answering common questions and	- Entertainment and Gaming: Engage users with interactive stories and personalized gaming experiences.

Chatbots	Rule-based Chatbots	AI-powered Chatbots
	providing order updates. - Guiding users through specific processes.	- Finance and Banking: Answer queries about accounts, transactions, and financial products, and process simple requests.

Structure of a chatbot

A chatbot has a “frontend” and a “backend”.



- The **frontend** of a chatbot serves as the messaging channel through which users interact, providing a user-friendly interface. However, one limitation of the frontend is that it may lack contextual understanding, meaning it might struggle to grasp the full meaning or context of user messages beyond the immediate input.
- The **backend** of a chatbot is where the hard work takes place. The backend operates application logic and has enough memory to remember earlier parts of a conversation as dialog continues.

Chatbots **understand** a question by breaking it into parts and relating those parts to things in its memory. A chatbot’s goal is to identify **entities** and **intents**, then use what it is found to trigger a **dialog**. Let us understand what intent means.

Intent

An intent is a purpose: the reason why a user is contacting the chatbot. Think of it as something like a ***verb***: a kind of action. Users may have various intents when interacting with a chatbot, such as filing a complaint, asking for directions, or speaking to a salesperson. Institutions often have multiple intents that they want their chatbots to address.

For instance, imagine you are tasked with creating a chatbot for a restaurant chain. One potential intent could be to inquire about the restaurant's operating hours. To ensure the chatbot can effectively handle this intent, you might interview a staff member who frequently fields such queries over the phone. Then, your goal would be to compile a comprehensive list of the diverse ways in which customers might inquire about the restaurant's opening hours. The following table provides many examples of possible user inputs that map to this kind of intent.

Intent	Possible user inputs
Open	When do you open?
	What are your hours?
	You open now?
	How late are you open?
	Can I walk in at 7 pm?

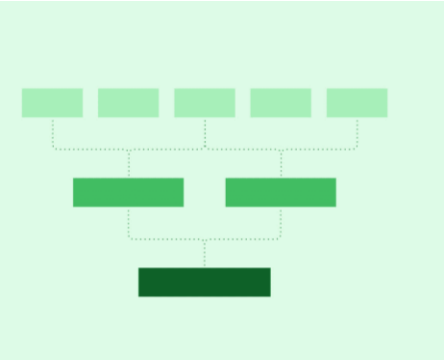
Entity

An entity is a noun: a person, place, or thing. Once you have a list of the intents you want your chatbot to fulfill, you are ready to continue. If a user asks, "What are the hours for the Bangalore office?", then providing business hours is the **intent** and Bangalore is the **entity**. A chatbot needs a full list of entities to be helpful. The following table lists examples of entities that map to the intent and possible user inputs of the previous restaurant chain example.

Intent	Possible user inputs	Entities
Open	When do you open?	Bangalore
	What are your hours?	Schedule
	You open now?	Time
	How late are you open??	Time
	And so forth; there are many **inputs** mapping to this intent.	And so forth; there are many **entities** implied by this intent.

Dialog

A **dialog** is a flowchart—an IF / THEN tree structure that illustrates how a machine will respond to user intents. A dialog is what the machine replies after a human asks a question. Even if a human uses run-on sentences, poor grammar, chat messaging expressions, and so on, artificial intelligence allows the NLP to understand well enough to provide a response.

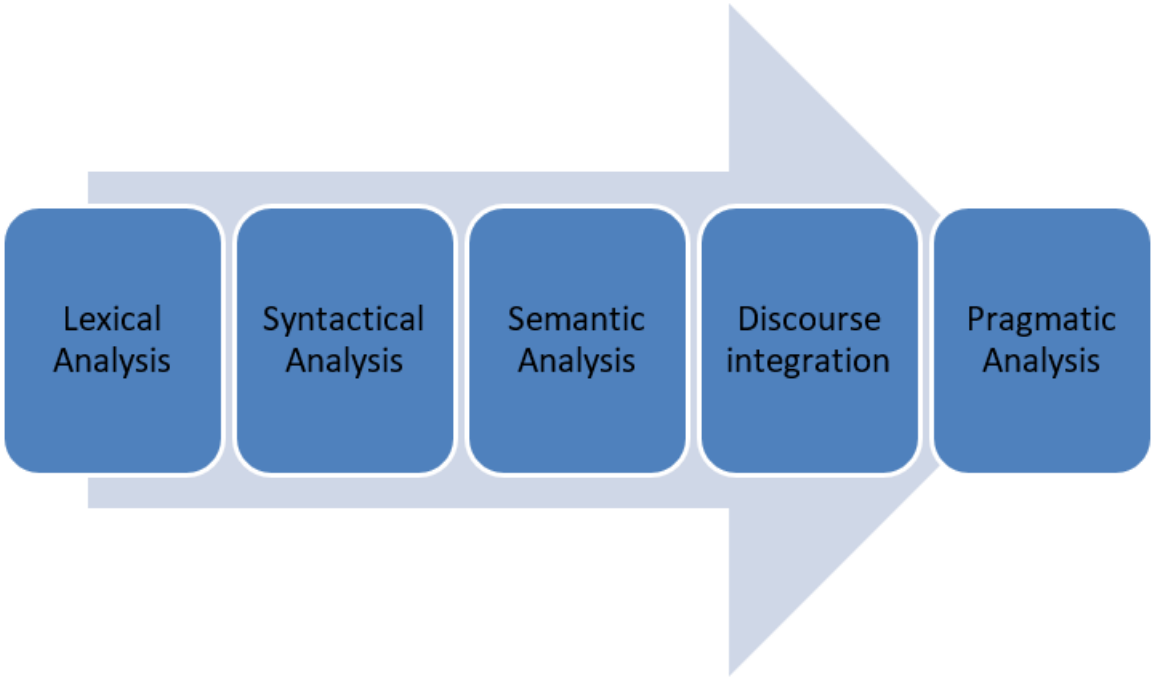


The dialog represents each possible word or phrase a user might enter, the matched response for the chatbot, and the many possible subsequent replies a user might make next. That is too much for an ordinary flowchart to show (you might need three or four dimensions!), so chatbot software condenses each moment of the conversation into a node. A node contains a statement by the chatbot and a long, expandable list of possible replies.

Creating the flowchart for a chatbot's responses entails a comprehensive process. Each potential user input following the chatbot's initial greeting must be assigned a corresponding reply. For instance, in the case of inquiries about a restaurant's operating hours, all conceivable questions would trigger a single response. This mapping of a vast array of possible questions to a limited number of responses persists throughout the conversation until its conclusion.

Natural Language Processing – Converting Speech to Text & analysing its intent

Natural language processing (NLP) involves a series of five phases that enable machines to analyse, categorize, and understand both spoken and written language.



These steps utilize deep neural network-style machine learning techniques to mimic the brain's ability to process data accurately. Businesses leverage tools and algorithms aligned with these NLP stages to extract insights from vast amounts of data and facilitate informed decision-making. Effective comprehension of these NLP stages is essential for optimizing the use of NLP in text and voice applications. Let's explore each step-in detail:

1. **Lexical analysis**

   This step involves understanding and examining the structure of words in a language. It breaks down the text into paragraphs, phrases, and words. Lexical normalization techniques like stemming and lemmatization are commonly used to reduce words to their base forms.

   - Stemming reduces words to their root form, such as removing suffixes like "ing", "ly", "es", and "s".
   - Lemmatization reduces words to their dictionary form, considering factors like parts of speech (POS) to determine their meaning in context.

2. **Syntactical Analysis**

   Syntactic Analysis is used to check grammar, word layouts, and word relationships.

   Example: Mumbai travels to the Anuj.

   The line "Mumbai travels to Anuj" makes no sense, hence it is rejected by the Syntactic Analyzer. **Syntactical parsing** is the analysis of words in a sentence for grammar. Dependency Grammar and Part of Speech (POS) tags are significant syntactic elements.

3. **Semantic Analysis**

   Semantic analysis aims to understand the various meanings conveyed by a sentence in

a clear and contextually appropriate manner. It extracts relevant insights from the text to comprehend its intended message.

4. **Discourse Integration**
   This involves understanding the context of a statement or word based on preceding sentences or words. It helps interpret references like pronouns and proper nouns by identifying their connections with earlier parts of the conversation.
   Example- Arti wants it.
   We can observe from the following sentence that the "it" keyword makes no sense. In reality, it applies to anything we don't know. That is all this "it" word depends on the prior sentence, which is not provided. So, if we know what "it" is, we can simply find the reference.
5. **Pragmatic Analysis**
   It denotes the study of meanings in a particular language. Process of extracting insights from a text. It involves verbal repetition, such as "who said what to whom?" It recognizes how individuals communicate with one another, the context in which they are speaking, and many other factors.

Let us see how each stage works with the help of an example sentence.

"The cat sat on the mat."

Now, let's go through the **five stages of NLP** using this sentence:

Analysis Step	Description	Example/Outcome
1. Lexical Analysis (Tokenization & Morphological Analysis)	Breaks down the sentence into individual words (tokens) and each word is analysed for its root form using lemmatization or stemming.	Tokens: ["The", "cat", "sat", "on", "the", "mat"]
2. Syntactic Analysis (Parsing & Grammar check)	Checks the grammatical structure of the sentence according to formal syntax rules. (Subject-Verb-Object (SVO) structure)	Subject: "The cat" Verb: "sat" Phrase: "on the mat"
3. Semantic Analysis (Meaning Extraction)	Analyses the meaning of individual words and ensures the sentence makes logical sense.	"cat" identified as an animal. "sat" as an action. "mat" as an object.
4. Discourse Integration (Context Understanding)	Considers the context of the sentence within a conversation or a paragraph to extract additional meaning.	If the previous sentence was "It was raining outside". The model might infer that the **cat** sat inside to avoid the rain.
5. Pragmatic Analysis (Real-World Interpretation)	Looks at real-world meaning and intent, adapting the interpretation based on the context of who said it and why.	In everyday conversation: A person might be casually describing a cat's action. In a chatbot response to a pet-related query: The bot might suggest a comfortable pet mat.

Video Session: Five Phases of NLP (Natural Language Processing) (with Tools and Applications) -
https://www.youtube.com/watch?v=8SSAliFwCy8

Applications of NLP

Businesses can use natural language processing tools to analyse data, discover insights, automate time-consuming operations, and gain a competitive advantage.

- **Sentiment Analysis:** Natural language processing (NLP) may evaluate consumer comments, social media posts, product reviews, and other text data to determine the sentiment (positive, negative, or neutral) associated with a specific brand, product or service. This data aids firms in understanding client impressions and sentiments.
- **Voice Assistants:** Voice assistants are becoming increasingly popular! Whether it's Siri, Alexa, or Google Assistant, practically everyone uses one to make calls, create reminders, plan meetings, set alarms, browse the internet, and so on. How do they work? They use a complicated blend of voice recognition, natural language interpretation, and natural language processing to understand and respond to what humans say. The long-term purpose of voice assistants is to serve as a bridge between humans and the internet, providing a variety of services through simple speech interaction. However, they are still a long way from achieving that aim, as Siri occasionally fails to grasp what you are saying!
- **Email Filtering:** Email is a part of our daily lives. We find ourselves bombarded with emails about job, study, and a variety of other topics. We receive emails from a variety of sources; some are work-related or from our dream school or institution, while others are spam or promotional in nature. Here, Natural Language Processing comes into play. It classifies incoming emails as "important" or "spam" and assigns them accordingly.
- **Document Analysis**: Document analysis is another use of natural language processing. Companies, institutions, and schools, among other places, are constantly inundated with data that must be properly organized, stored, and searched. All of this may be accomplished with NLP. It not only searches a keyword but also categorizes it according to the instructions, saving us from the tedious and time-consuming task of searching for a single person's information from a large number of files. It is not only restricted to this, but it also assists users in making informed decisions about claims and risk management.
- **Automatic Summarization:** Data has grown in line with technological advancements. This rise of data has broadened the scope of data processing. Still, manual data processing is time-consuming and error-prone. NLP provides a solution for this as well; it can not only summarize the meaning of information but also identify the emotional meaning hiding within it. As a result, the summary process becomes more efficient and precise.

Activity: Check out https://sites.research.google/versebyverse/

This is an experimental AI-powered muse that helps you write poetry inspired by classic American poets!

Activity: Creating a Chatbot

Create a chatbot on ordering ice-creams using any of the following platforms:

- Google Dialogflow
- Botsify.com
- Botpress.com

Video session (for Google Dialogflow) : https://www.youtube.com/watch?v=bIXkqDZMgaI

Activity: Program to print the POS tags of a statement- (For Advanced Learners)**

```
pip install nltk

import nltk
from nltk.tokenize import word_tokenize
from nltk import pos_tag

nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')

# Sample sentence
sentence = "The quick brown fox jumps over the lazy dog."

# Tokenize the sentence into words
words = word_tokenize(sentence)

# Perform part-of-speech tagging
pos_tags = pos_tag(words)

# Print the tokens with their respective parts of speech
for word, pos in pos_tags:
    print(f"Word: {word}, POS: {pos}")

Output:
Word: The, POS: DT
Word: quick, POS: JJ
Word: brown, POS: NN
Word: fox, POS: NN
Word: jumps, POS: VBZ
Word: over, POS: IN
Word: the, POS: DT
Word: lazy, POS: JJ
Word: dog, POS: NN
Word: ., POS: .
```

Activity: Creating a simple rule based chatbot using Python- (For Advanced Learners)**

```
import random

def get_response(user_input):
    user_input = user_input.lower()
    if "hello" in user_input:
        return "Hi there! How can I assist you?"
    elif "how are you" in user_input:
```

```
        return "I'm just a bot, but thanks for asking!"
    elif "bye" in user_input:
        return "Goodbye! Have a great day!"
    else:
        return "I'm sorry, I didn't understand that."

def main():
    print("Welcome to the Simple Chatbot!")
    print("Type 'bye' to exit.")

    while True:
        user_input = input("You: ")
        if user_input.lower() == 'bye':
            print("Chatbot: Goodbye! Have a great day!")
            break
        else:
            response = get_response(user_input)
            print("Chatbot:", response)

if __name__ == "__main__":
    main()
```

Output:

```
Welcome to the Simple Chatbot!
Type 'bye' to exit.
You: hello good afternoon
Chatbot: Hi there! How can I assist you?
You: how are you
Chatbot: I'm just a bot, but thanks for asking!
You: bye
Chatbot: Goodbye! Have a great day!
```

EXERCISES

A. Multiple Choice Questions:

1. Which of the following is NOT a common task in NLP?
   a) Machine translation
   b) Text summarization
   c) Speech recognition
   d) Image recognition
2. What is the main challenge/s of NLP?
   a) Handling Ambiguity of Sentences
   b) Handling Tokenization
   c) Handling POS-Tagging
   d) All of the mentioned
3. What is a chatbot?
   a) A physical robot used for chatting purposes.
   b) A computer program designed to simulate conversation with human users, especially over the internet.
   c) An advanced form of search engine.
   d) A tool used for sending automated emails.

4. Which of the following is an application of Natural Language Processing (NLP)?
   a. Autonomous vehicles
   b. Predicting stock prices
   c. Sentiment analysis
   d. Virtual reality gaming
5. Which of the following statements about Voice Recognition Interfaces is true?
   a. They solely rely on text-based inputs.
   b. They are incapable of understanding multiple languages.
   c. They convert spoken language into text or commands.
   d. They require a physical keyboard for interaction.

ANSWERS

1. d. Image recognition
2. a. Handling Ambiguity of Sentences
3. b. A computer program designed to simulate conversation with human users, especially over the internet.
4. c. Sentiment analysis
5. c. They convert spoken language into text or commands.

B. Short answer questions:

1. How does NLP help in email filtering? Give a real-life example.

Ans. NLP plays a crucial role in classifying incoming emails as "important" or "spam," improving email sorting and enhancing user experience. It utilizes language processing techniques to analyze email content, identify patterns, and make accurate filtering decisions.

Real-life Example:

Scenario: Imagine a professional's inbox flooded with various emails, including work-related messages, promotions, and potential spam.

NLP algorithms can scan through the email content, looking for indicators of importance like keywords, sender reputation, and email structure. By leveraging NLP for email filtering, important work emails are prioritized and separated from less critical or spam messages.

2. List the steps of NLP Processing.

Ans. The steps of Natural Language Processing (NLP) processing include:

Lexical Analysis, Syntactic Analysis, Semantic Analysis, Discourse Integration, Pragmatic Analysis

3. Briefly explain the two types of chatbots.

Ans.

Rule-based Chatbots:	AI-powered Chatbots
They work on predefined rules and decision trees to provide consistent and accurate answers to specific questions. Advantages: Easy to develop and maintain, Provide 24/7 availability for immediate and consistent support. Limitations: Struggle with understanding complex language High development costs and resource requirements.	Utilize natural language processing (NLP) and machine learning algorithms to respond based on learned patterns and interactions, providing personalized responses. Advantages: Offer personalized interactions based on user preferences and history. Improve efficiency and cost savings by automating tasks and reducing service costs. Limitations: Prone to biases from training data and lack of transparency in decision-making Ethical considerations regarding privacy, manipulation, and responsible use.

4. Briefly explain the classification problem. Give at least two examples.

Ans. Classification is a type of machine learning problem where the goal is to predict the categorical class labels of new observations based on past data. In classification, the output variable is a category or class label.

Two examples of classification problems are:

Email Spam Detection: Given a set of emails labeled as spam or not spam, the task is to classify new email messages as either spam or not spam.

Handwritten Digit Recognition: Classifying images of handwritten digits (0-9) into their corresponding numeric values.

5. Define the following:

(a) Intent (b)Entity (c) Dialog

Ans. (a) In the context of the provided document summaries, the term "intent" primarily refers to the purpose or objective behind a user's contact with chatbots or the desired outcome of a particular communication.

(b) An entity is a noun: a person, place, or thing. Eg: . If a user asks, “What are the hours for the Bangalore office?”, then providing business hours is the intent and Bangalore is the entity.

(c) A dialog is a flowchart—an IF / THEN tree structure that illustrates how a machine will respond to user intents. A dialog is what the machine replies after a human asks a question.

C. Long Answer Questions:

1. Explain the structure of a chatbot.

Ans. A chatbot has a "frontend" and a "backend".

- The frontend of a chatbot serves as the messaging channel through which users interact, providing a user-friendly interface. The one limitation of the frontend is that it may lack contextual understanding, meaning it might struggle to grasp the full meaning or context of user messages beyond the immediate input.
- The backend of a chatbot is where the hard work takes place. The backend operates application logic and has enough memory to remember earlier parts of a conversation as dialog continues.
- Chatbots understand a question by breaking it into parts and relating those parts to things in its memory. A chatbot's goal is to identify entities and intents, then use what it is found to trigger a dialog.

2. "A syntax tree is created as part of the procedure to visually represent semantic links." Identify the phase of NLP processing?

Ans. The phase of NLP processing that involves the creation of a syntax tree to visually represent semantic links is the Syntactical Analysis phase.

In syntactical analysis, the aim is to check the grammar, word layouts, and word relationships in a given text. One of the key tasks in this phase is to create a syntax tree, also known as a parse tree, which represents the grammatical structure of the sentence and visually displays the relationships between words. This helps in understanding the syntactical constructs and semantic relationships within the text, thereby aiding in the overall comprehension and analysis of the language.

D. Case Study questions:

1. Imagine you are a customer service manager at a global e-commerce company facing increasing customer inquiries across multiple channels. To alleviate the strain on your support team and enhance customer satisfaction, you decide to implement a chatbot solution. Develop a case study outlining the challenges you faced, the criteria you used to select a chatbot platform.

   Ans.

   - High Volume of Customer Inquiries: Managing the increasing number of customer inquiries across various channels, including email, live chat, and social media, was overwhelming for the support team.
   - Response Time and Availability: Ensuring timely responses and 24/7 availability for customer queries without round-the-clock human support.
   - Scalability: The need for a solution that could scale with the company's growth and handle a diverse range of customer queries effectively.

- Consistency in Responses: Maintaining consistency in responses across different channels and ensuring accurate information delivery to customers.

➢ Criteria for Selecting a Chatbot Platform:

- Natural Language Processing (NLP) Capabilities: A chatbot platform with advanced NLP capabilities to understand and respond effectively to natural language queries.
- Integration with Multiple Channels: Ability to seamlessly integrate with various communication channels like website chat, social media etc.
- Customization and Personalization: Platform that allows customization to reflect the brand tone and offers personalized experiences to customers.
- Analytics and Reporting: Provision for detailed analytics on chatbot interactions, customer satisfaction metrics etc.
- Scalability and Flexibility: Platform that can scale with the company's growth and adapt to evolving customer needs.
- Training and Support: Adequate training resources and customer support from the chatbot platform provider to ensure smooth implementation and ongoing maintenance.

2. Imagine you are a customer experience lead at a telecommunications company, currently relying on a rule-based chatbot to handle customer queries. However, due to limitations in scalability and adaptability, you are considering a transition to an AI-based chatbot solution. Detail the challenges faced during the transition, and the observed impact on customer service efficiency and satisfaction.

Ans. Challenges Faced During Transition:

➢ Knowledge Base Transition
➢ Training Data Quality
➢ Technical Integration
➢ Training and Skill Development
➢ Regulatory Compliance

Observed Impact on Customer Service Efficiency and Satisfaction:

➢ Enhanced Scalability
➢ Adaptability to Varied Queries
➢ Personalized Customer Interactions
➢ Reduced Response Time
➢ Analytics and Insights

E.Competency Based Questions:

1.Reena’s e-commerce website struggles with a high volume of customer inquiries. They're considering a chatbot to help. Should it be rule-based or AI-powered?
Here's how Reena can decide:
Rule-based chatbot:

- Pros: Faster development, lower cost, perfect for answering simple, repetitive questions (e.g., order tracking, delivery status).
- Cons: Limited understanding of complex questions, can't adapt to new inquiries, frustrating for users if it can't answer their specific needs.

AI-powered chatbot:

- Pros: Handles complex questions and learns over time, improves customer experience with more natural conversations.
- Cons: Higher development cost, requires training data and ongoing maintenance, may not be perfect initially and could misunderstand complex inquiries.

What factors Reena must evaluate to decide which chatbot best suits their needs and budget.
Ans- Factors to consider:

- Types of customer inquiries: If most questions are repetitive (tracking, returns), a rule-based chatbot might suffice. Complex inquiries (product recommendations, troubleshooting) would benefit from AI.
- Budget: AI chatbots require more investment.
- Technical expertise: Implementing and maintaining AI chatbots might require additional resources.

2.You're a compliance officer for a bank launching a new chatbot for customer service. The chatbot can answer basic questions about accounts, transfers, and suspicious activity. However, some customers might ask for specific account details like balances or transaction history. How can you ensure the chatbot complies with data protection regulations and protects sensitive financial information, while still being helpful to customers?

Ans-To address this, you can implement a two-pronged approach:

1. Security: Require multi-factor authentication (PIN, password, etc.) before revealing sensitive information. The chatbot can also deflect requests for specific details by offering alternative solutions (e.g., suggesting the mobile app for checking balances).
2. Transparency: Inform users upfront that the chatbot can't share sensitive data without authentication. Additionally, provide an easy option to connect with a live customer service representative for complex issues requiring such information disclosure.

3.Rahul's marketing team developed an app that understands text input. He wants to use the app to analyse customer reviews and improve the clarity of their promotional messages. How can Rahul leverage discourse integration techniques based on customer conversations to craft more effective marketing messages?

Ans- Rahul can analyse references and connections within reviews to understand what features resonate with customers (e.g., long battery life) and how they talk about their needs (e.g., "perfect for travel"). By using techniques like reference ("We heard you need a long-lasting battery...") and connectives ("This phone boasts a powerful battery, therefore keeping you connected all day"), Rahul can create clear and relevant messages that address customer priorities.

4. Jyoti is tasked with improving a healthcare chatbot for a hospital. Patients often ask the chatbot repetitive questions about symptoms, appointment scheduling, and basic medical information. How can Jyoti leverage NLP techniques to make the chatbot more helpful and engaging for patients?

Ans- Jyoti can use techniques like intent recognition and entity recognition to understand the user's goal (e.g., scheduling an appointment) and specific details (e.g., type of doctor). By implementing information retrieval from a medical knowledge base, the chatbot can answer basic medical questions accurately.

5. Ashmita's cinema chain is launching a new mobile app with a chatbot for ticket purchases and movie recommendations. Moviegoers often ask repetitive questions about showtimes and struggle to discover new movies they might enjoy. How can Ashmita leverage the chatbot functionalities to address these issues and improve customer experience?

Ans- Ashmita can focus on functionalities that answer FAQs with a clear interface (e.g., showtimes, ticket prices) and integrate a recommendation engine. The chatbot can analyze user data (past purchases) to suggest similar movies or upcoming releases that align with their preferences. This personalized approach can help users discover new favorites while addressing their most common questions, leading to a more engaging and satisfying moviegoing experience.

UNIT 8: AI Ethics and Values

Title: AI Ethics and Values	**Approach**: Team discussion, Web search, Case studies

Summary: In the present age of extensive use of AI tools by people of different walks of life, this chapter will guide the students to keep the guidelines in mind during the development and usage of AI. Also, students will be able to understand the ethical implications of different AI tools which have created a lot of commotion in the community because of different types of bias. They will understand the present-day challenges related to AI ethics.

Learning Objectives:

- Understand the fundamental concepts of ethics and its relevance in the context of artificial intelligence (AI).
- Identify bias arising from various sources present in AI systems and understand their societal implications.
- Understand the importance of mitigating bias in AI systems and be able to identify strategies for reducing bias in AI technologies.
- Understand the importance of developing AI policies.

Key Concepts:

1. Ethics in Artificial Intelligence
2. The five pillars of AI Ethics
3. Bias, Bias Awareness, Sources of Bias
4. Mitigating Bias in AI Systems
5. Developing AI Policies
6. Moral Machine Game
7. Survival of the Best Fit Game

Learning Outcomes:

Students will be able to -

- Demonstrate an understanding of the fundamental principles of ethics and gain insight into ethical considerations related to AI technologies.
- Develop an understanding of AI bias, its sources, and its real-world implications, as well as the ethical considerations.
- Identify and apply strategies for mitigating bias in AI systems to promote fairness and transparency in technology.
- Recognize the significance of AI policies in promoting responsible, safe, and ethical use of AI technologies.

Pre-requisites: Understanding the concept of data and reasonable fluency in English language. Ability to understand visual data.

The Moral Compass of AI: A Teacher's Guide to AI Ethics

This lesson equips you to introduce students to the critical topic of AI ethics and how it shapes responsible AI development.

1. The Importance of AI Ethics:

- **Introduction:** Begin by defining AI ethics and its crucial role in ensuring responsible AI development.
- **Key Considerations:** Discuss core ethical challenges surrounding AI:
    - Bias: How can AI systems perpetuate or amplify biases present in data or algorithms?
    - Fairness: How can we ensure AI decisions are fair and don't discriminate against specific groups?
    - Privacy: How can we protect user privacy when AI systems collect and analyse data?
    - Transparency: How can we understand how AI systems reach their decisions?

2. AI Bias: Real-World Consequences:

- **Exploring Examples:** Present real-world scenarios where AI bias caused issues:
    - Facial recognition systems misidentifying people of colour.
    - Algorithmic discrimination in the credit domain- loan decisions unfairly impacting certain demographics.

3. The Pillars of AI Ethics:

- **Defining the Framework:** Introduce the key pillars of AI ethics and their practical application:
    - **Accountability:** Who is responsible for the actions and decisions of AI systems? (e.g., Developers, users, regulators)
    - **Transparency:** How can we understand how AI systems arrive at their conclusions? (e.g., Explainable AI techniques)
    - **Fairness:** How can we ensure AI systems are just and unbiased? (e.g., Auditing data for bias, diversifying data sets)
    - **Privacy:** How can we protect user data privacy when AI interacts with it? (e.g., Strong data security practices)
    - **Safety:** How can we ensure AI systems operate safely and securely? (e.g., Rigorous testing and risk assessments)

4. Fostering Critical Thinking in Ethics:

- **Moral Dilemmas:** Encourage students to analyse ethical dilemmas related to AI.
- **Decision-Making:** Guide discussions on how to approach difficult ethical decisions in the context of AI development and use.

5. Identifying Bias in AI:

- **Unveiling Bias Sources:** Educate students on potential sources of bias in AI:

    - Data Collection: How data selection and representation can lead to biased outcomes.
    - Algorithm Design: How algorithms might inherit biases from their creators.
    - Decision-Making Processes: How human intervention could introduce bias.
- **Mitigating Bias:** Discuss strategies to identify and mitigate bias in AI systems.

6. Understanding AI Policies:

- **The Purpose of AI Policies:** Explain the role of AI policies in promoting responsible development.
- **Key Areas Addressed:** Explore the main areas addressed by AI policies:
    - Data Privacy: Protecting user data rights and ensuring responsible data collection practices.
    - Security: Mitigating risks associated with AI vulnerabilities and cyberattacks.
    - Bias: Implementing strategies to address bias in AI systems.
    - Accountability: Defining who is responsible for the actions and decisions made by AI.

7. Moral Game Design for AI:

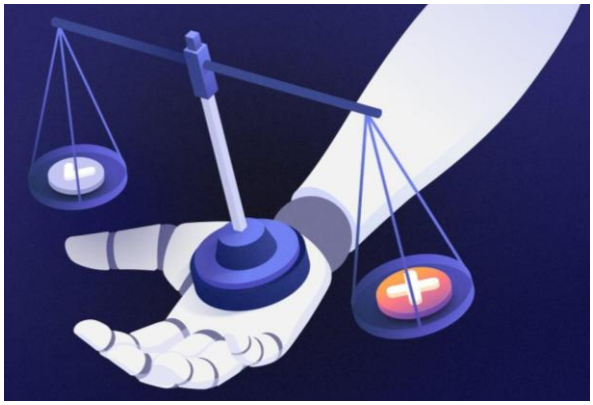
- **Ethical Games:** Introduce the concept of moral game design.
- **Emphasize Consequences:** Discuss how games can be used to explore ethical dilemmas in AI and highlight the consequences of AI decisions.
- **Critical Thinking and Moral Reasoning:** Utilize games to promote critical thinking and moral reasoning skills in the context of AI.

By incorporating these elements, your lesson will spark insightful discussions and equip students to become responsible participants in the evolving world of AI.

ETHICS IN ARTIFICIAL INTELLIGENCE

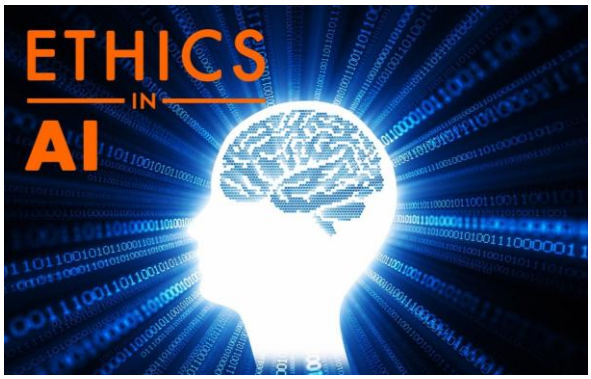
Teachers can ask the following questions:

1. **Imagine you create a new game with your classmates. What rules would you establish to make sure the game is fair and everyone has a chance to win?** (This question primes students to think about the concept of fairness, which is a key ethical principle in AI.)
2. **Can you think of a situation where a seemingly objective decision might have unintended consequences?** (This question activates students' prior knowledge about bias and its potential impact, which will be connected to the ethical considerations of AI later in the passage.)



In today's rapidly evolving technological landscape, artificial intelligence (AI) has emerged as a transformative force with the potential to revolutionize various aspects of human society. However, with this technological advancement comes a myriad of ethical considerations and challenges. This chapter serves as a foundational exploration into the ethical dimensions of AI, aiming to equip students with the necessary knowledge and tools to navigate the complex ethical landscape of AI technologies.

Ethics refers to the moral principles that govern human behavior and decision-making. It encompasses concepts such as right and wrong, fairness, justice, and accountability. Ethical considerations guide individuals and organizations in making responsible choices that align with societal values and norms. In the context of AI, ethics play a crucial role in ensuring that AI technologies are developed and deployed in a manner that upholds ethical principles and respects human rights. Ethical considerations are essential in addressing issues such as bias, transparency, accountability, privacy, and the societal impact of AI.



AI Ethics refers to the ethical principles and guidelines that govern the design, development, and deployment of AI technologies. AI ethics aims to ensure that AI systems are developed and used in ways that are fair, transparent, accountable, and aligned with human values.

Example 1:

Suppose a CCTV camera was to spot your face in a crowd outside a sports stadium. In the police data center somewhere in the city/ country, an artificial neural network analyzes images from the CCTV footage frame by-frame. A floating cloud in the sky causes a shadow on your face and the neural network (by mistake) finds your face similar to the face of a wanted criminal.

- If the police were to call you aside for questioning and tell you they had reason to detain you, how would you defend yourself?

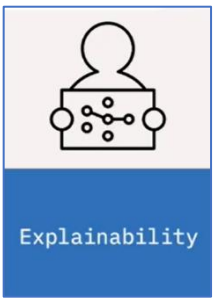
- Was it your fault that your shadowed face has resemblance by few degrees with a person in the police record?

Example 2:

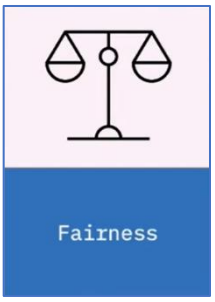
This happened in the USA in 2018. An AI system was being used to allocate care to nearly 200 million patients in the US. It was discovered later that the AI system was offering a lower standard of care to the black patients. Across the board, black people were assigned lower risk scores than white people. This in turn meant that black patients were less likely to be able to access the necessary standard of care.

The problem stemmed from the fact that the AI algorithm was allocating risk values using the predicted cost of healthcare. Because black patients were often less able to pay or were perceived as less able to pay for the higher standard of care, the AI essentially learned that they were not entitled to such a standard of treatment. Though the system was fixed / improved after being discovered, the big question is – whose problem was this? The AI system developers or the US black people data (which was true to an extent)?

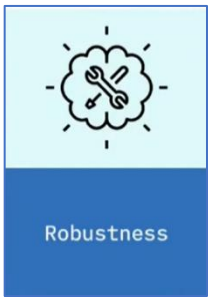
THE FIVE PILLARS OF AI ETHICS



- **Explainability** refers to the transparency and interpretability of AI systems, allowing users to understand how algorithms make decisions and predictions. Explainable AI enables stakeholders to comprehend the underlying logic, factors, and considerations driving algorithmic outcomes, fostering trust, accountability, and ethical use of AI technologies. Explainability is essential for ensuring that AI systems are transparent, accountable, and aligned with ethical principles.



- ***Fairness*** in AI seeks to remove bias and discrimination from algorithms and decision-making models. Machine learning fairness addresses and eliminates algorithmic bias from machine learning models based on sensitive attributes like race and ethnicity, gender, sexual orientation, disability, and socioeconomic class.



- ***Robustness*** in AI systems indeed refers to their ability to consistently provide accurate and reliable results regardless of the conditions they encounter and for extended periods. It is all about making sure that AI algorithms and systems operate as expected without any unexpected errors or deviations from their intended behavior. This involves ensuring stability in the algorithms, being able to reproduce results, and maintaining consistent performance across different datasets and environments. Achieving reliability in AI systems requires thorough testing, validation, and quality assurance at every stage of development.

- ***Transparency*** involves openness and disclosure about the design, operation, and implications of AI systems. Transparent AI frameworks provide clear documentation, disclosure, and communication about the data, algorithms, and decision-making processes used in AI applications. Transparency promotes accountability, scrutiny, and informed decision-making, enabling users and stakeholders to assess the ethical implications and societal impacts of AI technologies.

- ***Privacy*** refers to the right of individuals to control their personal information and to be free from unwanted intrusion into their lives. It encompasses the ability to keep certain aspects of one's life private, such as personal communications, activities, and personal data. Privacy is essential for safeguarding personal autonomy, dignity, and freedom from unwarranted interference.

Activity:

Organize the class into groups. Reflect on the following points based on the video links given

- Video: AI for Good
- Reflect on the video "The Ethical Robot" and identify two ethical dilemmas that stood out to you. Document these questions.
- Using "How to build a moral robot" as a reference point, list the moral and ethical principles you wish to embed in your robot. Consider the video as inspiration, but do not feel constrained by its content. Feel free to expand your thoughts with creativity and innovation.
- Assemble a team of five students and collectively watch the video "Humans need not apply." It is recommended to view the video multiple times. Following your discussions, compile a group paper summarizing your insights and interpretations from the video.

BIAS, BIAS AWARENESS, SOURCES OF BIAS

Bias, in simple terms, means having a preference or tendency towards something or someone over others, often without considering all the relevant information fairly. It can lead to unfair treatment or decisions based on factors like personal beliefs, past experiences, or stereotypes. In everyday life, bias can affect how we perceive and interact with people, situations, or ideas. In the context of artificial intelligence, bias refers to when AI systems make unfair or inaccurate decisions due to flawed data or built-in assumptions, which can result in unfair outcomes for certain groups of people.

In today's interconnected world, artificial intelligence (AI) technologies play an increasingly prominent role in various aspects of our lives, from healthcare to finance to criminal justice. However, as AI systems become more pervasive, it is essential to recognize and address the presence of bias in these technologies. **Bias awareness** means understanding that AI systems might have unfair preferences because of different things like the information they were taught with, the rules they follow, or the ideas they were built upon. So, being aware of bias in AI is like knowing that sometimes AI might make unfair choices or judgments because of how it was trained or made.

Activity:

Question. 1: Why are most images that show up when you do an image search for “vacation” seen as beaches?

Question 2: Why are most images that show up when you do an image search for “nurse” seen as females?

Question 3: Organize students into groups and ask them to find answers for the questions given below after going through the link Amazon Recruitment Tool.

Discussion Questions:

- How do algorithmic hiring systems function, and what criteria are typically used to evaluate job applicants?
- What are the ethical implications of using biased algorithms in hiring processes, particularly regarding fairness, equal opportunity, and diversity?
- How might biased hiring algorithms perpetuate systemic inequalities in employment and hinder efforts to promote inclusivity in the workforce?

Reference: https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/

AI bias, also referred to as machine learning bias or algorithm bias, refers to AI systems that produce biased results that reflect and perpetuate human biases within a society, including historical and current social inequality. Bias can be found in the initial training data, the algorithm, or the predictions the algorithm produces. When bias goes unaddressed, it hinders people's ability to participate in the economy and society. It also reduces AI's potential.

The source of bias in AI

Teachers can ask the following questions:

1. **Imagine you are training a new dog trick. How can you make sure your training data is fair and clear so your dog understands what you want it to do?** (This question primes students to think about the importance of fair and unbiased training data, which is crucial for AI systems as well.)
2. **Have you ever played a game where the rules seemed unfair to one player or another? Why is it important to have fair rules in games and activities?** (This question activates students' prior knowledge about fairness and its connection to bias. This will be helpful when they learn about how bias can affect AI systems.)

Eliminating AI bias requires drilling down into datasets, machine learning algorithms and other elements of AI systems to identify sources of potential bias.

1. Training data bias

AI systems learn to make decisions based on training data, so it is essential to assess datasets for the presence of bias. One method is to review data sampling for over- or underrepresented groups within the training data. For example, training data for a facial recognition algorithm that over-represents white people may create errors when attempting facial recognition for people of color. Similarly, security data that includes information gathered in geographic areas that are predominantly black could create racial bias in AI tools used by police. Bias can also result from how the training data is labeled. For example, AI recruiting tools that use inconsistent labeling or exclude or over-represent certain characteristics could eliminate qualified job applicants from consideration.

2. Algorithmic bias

Using flawed training data can result in algorithms that repeatedly produce errors, unfair outcomes, or even amplify the bias inherent in the flawed data. Algorithmic bias can also be caused by programming errors, such as a developer unfairly weighting factors in algorithm decision-making based on their own conscious or unconscious biases. For example, indicators

like income or vocabulary might be used by the algorithm to unintentionally discriminate against people of a certain race or gender.

3. **Cognitive bias**

When people process information and make judgments, we are inevitably influenced by our experiences and our preferences. As a result, people may build these biases into AI systems through the selection of data or how the data is weighted. For example, cognitive bias could lead to favoring datasets gathered from Americans rather than sampling from a range of populations around the globe.

Examples of AI bias in real life

- **Healthcare**—Underrepresented data of women or minority groups can skew predictive AI algorithms. For example, computer-aided diagnosis (CAD) systems have been found to return lower accuracy results for black patients than white patients.
- **Online advertising**—Biases in search engine ad algorithms can reinforce job role gender bias. Independent research at Carnegie Mellon University revealed that Google's online advertising system displayed high-paying positions to males more often than to women.
- **Image generation**—Academic research found bias in the generative AI art generation application Midjourney. When asked to create images of people in specialized professions, it showed both younger and older people, but the older people were always men, reinforcing gender bias of the role of women in the workplace.

Activity: Role Play

Share the following examples of biased AI systems and their potential consequences and ask students to do a role play to present each scenario:

- **Facial Recognition Technology:**
  - **Example**: Facial recognition systems have been shown to exhibit bias against certain demographic groups, particularly people with darker skin tones and women.
  - **Consequences**: Biased facial recognition algorithms can lead to misidentification and wrongful arrests, disproportionately affecting marginalized communities and eroding trust in law enforcement.
- **Predictive Policing Algorithms:**
  - **Example**: Predictive policing algorithms use historical crime data to forecast future criminal activity and allocate law enforcement resources. However, studies have found that these algorithms can perpetuate racial and socioeconomic biases, leading to over-policing of minority neighborhoods.
  - **Consequences**: Biased predictive policing algorithms may exacerbate racial profiling and discrimination, fueling tensions between law enforcement agencies and communities of color and undermining public trust in the criminal justice system.
- **Algorithmic Hiring Systems:**
  - **Example**: AI-powered hiring systems are used by companies to screen job applications and identify potential candidates. However, research has shown that these systems can perpetuate gender and racial biases, favoring certain demographic groups over others.

- **Consequences**: Biased hiring algorithms may reinforce existing disparities in employment opportunities, leading to discrimination against underrepresented groups and hindering efforts to promote diversity and inclusion in the workforce.

- **Healthcare Algorithms:**
  - **Example**: AI algorithms are increasingly used in healthcare for tasks such as diagnosing diseases and predicting patient outcomes. However, studies have identified biases in healthcare algorithms that can result in differential treatment recommendations based on factors such as race or socioeconomic status.
  - **Consequences**: Biased healthcare algorithms may lead to disparities in patient care, with certain demographic groups receiving suboptimal or inequitable treatment. This can contribute to worsened health outcomes and perpetuate healthcare inequalities.
- **Credit Scoring Systems:**
  - **Example**: AI-powered credit scoring systems are used by financial institutions to assess individuals' creditworthiness and determine their eligibility for loans and other financial products. However, these systems have been found to exhibit biases that disproportionately disadvantage certain demographic groups, such as low-income individuals and people of color.
  - **Consequences**: Biased credit scoring algorithms may limit access to financial opportunities for marginalized communities, perpetuating socioeconomic inequalities and hindering economic mobility.

MITIGATING BIAS IN AI SYSTEMS

Mitigating bias in AI systems is essential for several reasons. Firstly, when AI systems have bias, they can make existing problems like unfairness and discrimination even worse. For example, biased algorithms used in hiring processes may unfairly disadvantage certain groups, leading to systemic discrimination. Secondly, biased AI makes people trust technology less. If people don't trust AI to make fair decisions, they might not want to use it, which can cause problems for everyone. Lastly, addressing bias is essential for upholding ethical principles and ensuring that AI technologies are developed and used responsibly.

Strategies for Mitigating Bias

There are several strategies and techniques for mitigating bias in AI systems:

- **Using Diverse Data**: To reduce bias, we should use lots of different kinds of information to teach AI. This way, the AI can learn from many different examples and viewpoints, making it less likely to be biased.
- **Detecting Bias**: We need ways to find and measure bias in AI systems before they are used. This could mean looking at how the AI makes decisions for different groups of people or using special tools to see if the AI is being fair.
- **Fair Algorithms**: We can make AI systems fairer by using special algorithms that are designed to be fair. These algorithms make sure to consider fairness when making decisions, helping to reduce bias.

- **Being Transparent**: It is important for AI systems to be clear and explain how they make decisions. When people understand how AI works, they can see if there is any bias and fix it.
- **Inclusive Teams**: When creating AI, it is helpful to have a team of people from different backgrounds and experiences. This way, they can spot biases that others might miss and make sure the AI is fair for everyone.

Activity:

Allow students to examine various forms of media, such as news articles, advertisements, or social media posts, and identify instances of bias based on factors like race, gender, or socio-economic status. Encourage them to discuss how bias can influence perceptions and stereotypes.

Do you know?

IBM's AI Fairness 360 is an open-source and comprehensive toolkit that offers a wide range of tools and resources for addressing bias in machine learning models. The toolkit includes over 70 fairness metrics which help users detect bias in their models. This suggests that AI Fairness 360 provides robust mechanisms for identifying potential sources of bias. The toolkit offers over 10 algorithms for mitigating bias once it has been detected. These bias mitigating algorithms include optimizing the preprocessing stage, prejudice remover, and regular algorithms, among others. With its diverse set of features, educational resources, and validation mechanisms, it aims to support users in promoting fairness and equity in AI applications.

DEVELOPING AI POLICIES

Developing AI policies is essential for ensuring that AI technologies are used responsibly, safely, and ethically, while also promoting innovation and public trust.

- Rules for AI should start with being good to people and respecting their rights. This means treating everyone fairly, being honest about how AI works, making sure it is safe, and being accountable if something goes wrong.
- We need clear rules and standards for how AI is used. These rules should cover important things like protecting people's information, making sure AI does not have unfair biases, keeping it safe, and making sure people can ask questions about how AI works.
- When making these rules, it is important to talk to lots of different people. This includes government people, business leaders, scientists, community groups, and regular people. Everyone's opinion matters because AI affects everyone.
- Before using AI, we should check to see if there are any problems or risks. This means thinking about what could go wrong and making plans to fix it.

Understanding the components of AI policies involves examining guidelines and principles established by various organizations and regulatory bodies.

1. **IBM AI Ethics Board:**

   Focus: Ethical development and deployment of AI technologies across various industries.

   Components:

   - Development of ethical principles and guidelines for AI research and development.
   - Recommendations for addressing ethical considerations such as fairness, transparency, accountability, and bias mitigation in AI systems.

- Engagement with stakeholders, including researchers, policymakers, and industry partners, to promote dialogue and collaboration on ethical AI practices.
- Support for educational initiatives and resources to raise awareness and understanding of AI ethics among developers, users, and the public.

2. **Microsoft's Responsible AI Page:**
   Focus: Corporate responsibility and ethics in AI
   Components:
   - Principles for responsible AI development and deployment, including fairness, reliability, privacy, and inclusivity.
   - Tools and resources for integrating ethical considerations into AI projects, such as fairness assessments and bias detection algorithms.
   - Case studies and best practices for implementing responsible AI practices across various industries and domains.
3. **Artificial Intelligence at Google:**
   Focus: Corporate AI ethics and governance
   Components:
   - Google's principles for ethical AI development, encompassing areas such as fairness, safety, privacy, and accountability.
   - Guidelines for designing AI systems that prioritize human values and societal well-being.
   - Commitments to transparency, collaboration, and continuous improvement in AI governance and decision-making.
4. **European Union's Ethics Guidelines for Trustworthy AI—Press Release:**
   Focus: Ethical guidelines for AI development and deployment in the EU
   Components:
   - Principles for trustworthy AI, including respect for human autonomy, prevention of harm, fairness, and accountability.
   - Requirements for transparency, explainability, and auditability in AI systems.
   - Recommendations for ensuring human oversight and accountability mechanisms in AI applications with high societal impact.

Further reading:

- IBM 's AI Ethics Board
- Microsoft's Responsible AI page
- Artificial Intelligence at Google: Our Principles
- European Union's Ethics Guidelines for Trustworthy AI—Press Release

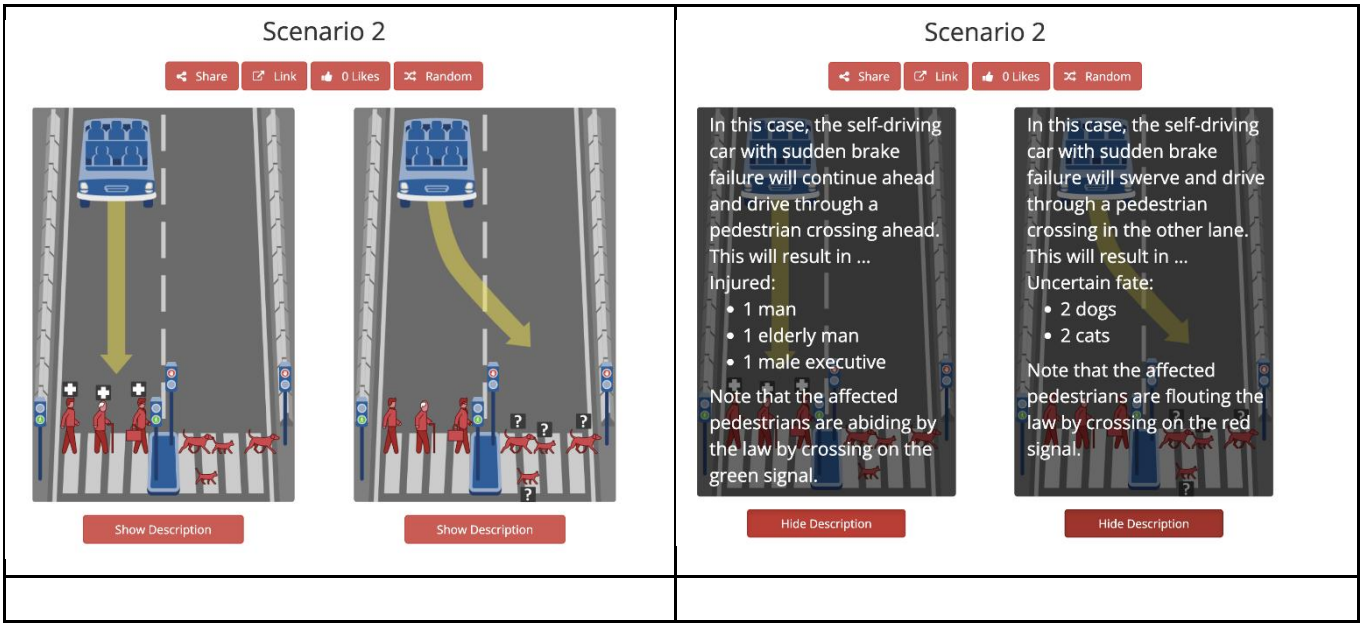
MORAL MACHINE GAME

An **ethical dilemma** is a situation in which a person or group is faced with conflicting moral principles or choices. In other words, it is a scenario where there is no clear "right" or "wrong" decision, and any action taken may have both positive and negative consequences. Ethical dilemmas often involve difficult choices between two or more options, each of which may involve values or principles that are important to the decision-maker.

An **ethical dilemma in the context of artificial intelligence (AI)** arises when there is a conflict between moral principles or values in the design, development, deployment, or use of AI technologies. These dilemmas often emerge due to the complex and multifaceted nature of AI systems, which can have far-reaching implications for individuals, society, and the environment.

The **Moral Machine**, developed by researchers at the Massachusetts Institute of Technology (MIT), is an online platform designed to explore ethical dilemmas in AI through interactive decision-making scenarios. Users are presented with various hypothetical situations where autonomous vehicles must make split-second decisions that could result in harm or even loss of life. The scenarios often involve difficult choices, such as deciding whether to prioritize the safety of passengers or pedestrians, obey traffic laws, or consider factors like age, gender, and social status.

The Moral Machine presents users with a series of thought-provoking scenarios, each with its own set of moral considerations and trade-offs. For example, imagine you are the operator of a self-driving car that encounters a situation where it must choose between swerving to avoid hitting a group of pedestrians, potentially endangering its passengers, or staying the course and risking harm to those on the road. What decision would you make? And more importantly, why?



While the scenarios presented on the Moral Machine are hypothetical, they reflect real-world ethical dilemmas that AI developers, policymakers, and society at large must grapple with. As AI continues to advance and become more pervasive, it is essential that we engage in meaningful dialogue and reflection on the ethical implications of these technologies. The Moral Machine serves as a powerful tool for sparking conversation, raising awareness, and promoting ethical awareness in the age of AI.

Activity:

1. **Begin with: The Ethical Challenges of Autonomous Vehicles**

   Discussion Points:

   - What ethical issues are highlighted by the presenter concerning the application of AI in military contexts, and how do these resonate with your viewpoints?

- What steps can legislators take to tackle the ethical dilemmas presented by autonomous weaponry, particularly regarding accountability and the preservation of human dignity?

2. Present the Moral Machine game as a tool for delving into ethical decision-making challenges related to AI - accessible at https://www.moralmachine.net/

SURVIVAL OF THE BEST FIT GAME

Activity:

Survival of the Best Fit is an educational game about hiring bias in AI. We aim to explain how the misuse of AI can make machines inherit human biases and further inequality.

Students can do this activity in the practical class to understand the concept of bias while people are hired.

Present this tool for deeper understanding of bias during hiring of people in any company or organization- https://www.survivalofthebestfit.com/

EXERCISES

A. Multiple Choice Questions

1. What is the primary focus of AI ethics?
   a. Ensuring AI systems are efficient and error-free
   b. Guiding the development and use of AI technologies according to ethical principles
   c. Maximizing the profitability of AI technologies
   d. Focusing solely on the technological advancements in AI
2. Which ethical principle is concerned with fairness and justice in AI systems?
   a. Transparency
   b. Accountability
   c. Bias and fairness
   d. Privacy
3. What role does transparency play in AI ethics?
   a. Ensuring AI systems are invisible to users
   b. Making the AI decision-making process clear and understandable
   c. Keeping the AI technologies a secret
   d. None of the above
4. What is a major ethical concern related to AI and privacy?
   a. AI systems improving data security automatically
   b. AI technologies accessing and using personal data without consent
   c. AI enhancing user privacy by default
   d. None of the above
5. How can bias in AI systems impact society?
   a. By making AI systems more efficient
   b. By perpetuating and exacerbating existing inequalities
   c. By improving fairness and justice automatically
   d. None of the above

6. Which of the following strategies is suggested for mitigating bias in AI systems?
    a. Ignoring the bias
    b. Increasing the complexity of AI algorithms
    c. Regular audits and inclusive data practices
    d. Focusing solely on the technological aspects
7. What is the purpose of ethical frameworks and guidelines in AI?
    a. To restrict the development of AI technologies
    b. To ensure AI development aligns with human values and ethical principles
    c. To make AI systems less transparent
    d. None of the above
8. Who is responsible for ensuring the ethical use of AI systems?
    a. AI developers only
    b. Government bodies only
    c. Users only
    d. All stakeholders, including developers, governments, and users
9. What is an example of AI assisting humanity as mentioned in the unit?
    a. Reducing energy efficiency in buildings
    b. Enhancing mental health care through AI-powered chatbots
    c. Increasing bias in hiring processes
    d. Decreasing crop yield predictions accuracy
10. What is the role of AI ethics in the context of autonomous vehicles?
    a. To ensure vehicles are aesthetically pleasing
    b. To navigate ethical dilemmas such as decision-making in critical situations
    c. To make autonomous vehicles less safe
    d. None of the above

B. True/False

1. Ethical considerations in AI are only relevant for AI researchers and developers.
2. Bias in AI systems can lead to unfair and discriminatory outcomes.
3. Transparency in AI systems means making the algorithm's code publicly available.
4. Bias in AI systems can lead to unfair outcomes and reinforce existing societal inequalities.
5. Transparency is not important in AI decision-making processes as long as the outcomes are beneficial.
6. Understanding the fundamental concepts of ethics is irrelevant in the context of artificial intelligence.
7. Analyzing real-world examples of AI bias provides insights into the ethical challenges in AI development.
8. Identifying bias in AI systems has no impact on society.
9. Evaluating strategies to mitigate bias in AI systems is unnecessary.
10. The ability to critically analyze the ethical implications of AI technologies does not require considering their impact on society.

C. Fill in the Blanks

1. Understanding the fundamental principles of ethics is crucial to applying ethical considerations in the field of _____.
2. The ability to critically analyze the ethical implications of AI decision-making processes requires a deep understanding of their _____ on individuals and society.
3. Investigating various types of bias in AI systems enables students to understand their _____ implications.
4. _____ in AI systems can lead to unfair and discriminatory outcomes, making it essential to address issues of bias, fairness, and equity.
5. In the context of AI, _____ is important for making the decision-making processes of AI systems clear and understandable to users.

ANSWERS

A. Multiple Choice Questions

1. b. Guiding the development and use of AI technologies according to ethical principles
2. c. Bias and fairness
3. b. Making the AI decision-making process clear and understandable
4. b. AI technologies accessing and using personal data without consent
5. b. By perpetuating and exacerbating existing inequalities
6. c. Regular audits and inclusive data practices
7. b. To ensure AI development aligns with human values and ethical principles
8. d. All stakeholders, including developers, governments, and users
9. b. Enhancing mental health care through AI-powered chatbots
10. b. To navigate ethical dilemmas such as decision-making in critical situations

B. True/False

1. False	2. True	3. False	4. True	5. False
6. False	7. True	8. False	9. False	10. False

C. Fill in the Blanks

1. Artificial Intelligence	2. Individuals	3. Societal
4. Bias	5. Clear	

D. Short Answer Type Questions

1. What is algorithmic bias, and can you give an example?

Ans. Algorithmic bias occurs when AI systems display prejudice in their output, such as facial recognition software misidentifying individuals from certain ethnic groups more frequently than others.

2. How can bias in AI systems be reduced?

Ans. Implementing regular audits of AI algorithms and training data can identify and correct biases, ensuring the AI's decisions are fair and equitable across all user groups.

3. Define AI ethics.

Ans. AI ethics involves the principles and values guiding the development, deployment, and use of artificial intelligence technologies to ensure they benefit society while minimizing harm and respecting human rights.

4. Why is transparency important in AI systems?

Ans. Transparency in AI systems is crucial for users to understand and trust AI decision-making processes, ensuring accountability and enabling scrutiny for potential biases or errors.

5. How can AI impact privacy and data protection?

Ans. AI technologies can infringe on privacy by collecting, analyzing, and sharing personal data without adequate consent, highlighting the need for robust data protection measures.

6. What role does fairness play in AI systems?

Ans. Fairness in AI systems aims to ensure equitable treatment and outcomes for all individuals, preventing discrimination based on race, gender, or other characteristics.

7. How can ethical dilemmas in AI be addressed?

Ans. Ethical dilemmas in AI can be addressed through the application of ethical frameworks, stakeholder engagement, and the development of policies that prioritize human welfare.

8. What strategies can mitigate bias in AI?

Ans. Mitigating bias in AI involves diverse data collection, algorithmic transparency, fairness audits, and continuous monitoring to identify and correct biases.

9. What challenges do policymakers face in regulating AI?

Ans. Policymakers face challenges in regulating AI, including keeping up with rapid technological advancements, balancing innovation with ethical considerations, and addressing global disparities in AI governance.

E. Case Study Analysis

Scenario:

A technology company has been contracted to develop a sophisticated facial recognition system intended for law enforcement applications. This system is designed to assist in identifying suspects by matching their facial features against a database. Initially celebrated for its high level of accuracy and efficiency in processing and identifying faces, the system, however, has come under scrutiny after reports emerged that it disproportionately misidentifies people of color. Investigations revealed that the misidentification rate for individuals of color was significantly higher than for white individuals, leading to a higher incidence of wrongful detentions and arrests among these communities. Further analysis indicated that the training data used to develop the algorithm lacked sufficient diversity, predominantly featuring faces of white individuals. This oversight has not only raised questions about racial bias inherent in AI technologies but also about the ethical implications of deploying such systems in sensitive areas like law enforcement without thorough vetting for fairness and impartiality.

1. What ethical problems are evident in this scenario?

Ans. The scenario highlights ethical concerns including racial bias in technological applications, the risk of wrongful arrests due to biased AI, and the potential violation of rights for people of color.

2. What is one ethical approach to mitigate the bias found in the facial recognition system without compromising its accuracy?

Ans. An ethical solution involves enhancing the diversity of the dataset used to train the facial recognition system, ensuring it accurately represents all racial groups. This would reduce bias while preserving the system's overall accuracy.

3. How does the lack of diversity in training data contribute to the problem of algorithmic bias in this scenario?

Ans. The lack of diversity in the training data has directly contributed to the algorithmic bias by failing to accurately represent the variety of human facial features across different races, leading to the system's inability to correctly identify individuals of color.

4. What measures can be taken to ensure that the deployment of AI systems in law enforcement is both ethical and effective?

Ans. To ensure the ethical and effective deployment of AI systems in law enforcement, comprehensive measures such as implementing strict regulatory standards, conducting regular bias audits, and involving diverse groups in the system's development and evaluation process are crucial.

5. Discuss the potential long-term impacts on public trust and societal fairness if biases in AI systems like facial recognition are not addressed.

Ans. If biases in AI systems remain unaddressed, it can lead to a significant erosion of public trust, particularly among marginalized communities disproportionately affected by such biases. Over time, this may exacerbate social inequalities and injustice, undermining the legitimacy of law enforcement agencies and the broader justice system.

F. Ethical Dilemma

Read the following ethical dilemma and provide your response:

In a complex urban environment, an autonomous vehicle faces an imminent collision dilemma. The vehicle's path is obstructed by a pedestrian who has unexpectedly entered the roadway, violating traffic signals. Simultaneously, to the vehicle's side, a group of cyclists legally occupies the bike lane. The vehicle's AI must make an instantaneous decision: continue on its trajectory, risking harm to the pedestrian, or alter its course, endangering the cyclists. This decision-making process involves critical ethical considerations, including the valuation of human life, the assessment of potential harm, and the prioritization of legal and moral obligations in split-second scenarios.

Discussion Question: Consider the ethical dilemmas presented by AI technologies and explore how lawmakers, business leaders, and the public can work together to tackle these issues and encourage the ethical creation and implementation of AI.

Ans. Reflecting on the ethical challenges posed by AI technologies, especially in autonomous vehicle dilemmas, requires a multifaceted approach from policymakers, industry stakeholders, and the public. To navigate these challenges and foster responsible AI development and deployment, collaboration across these groups is essential. Policymakers should establish clear ethical guidelines and regulatory frameworks that mandate safety, transparency, and accountability in AI systems. Industry stakeholders must commit to ethical AI development practices, prioritizing inclusivity in testing phases to ensure diverse scenarios and outcomes are considered. Moreover, public engagement through education and dialogue can inform more nuanced AI ethics policies and encourage societal acceptance of AI technologies. Together, these efforts can lead to the development of AI systems that are not only technologically advanced but also ethically responsible and socially beneficial.

G. Competency Based Question

1. Imagine you are a data scientist working on training an AI model for facial recognition. During the development phase, you discover that the AI system shows a significant bias against individuals with darker skin tones, leading to higher error rates in identifying them accurately. As an ethical data scientist, how would you approach this issue to ensure fairness and minimize discrimination in the facial recognition AI model?"
   Ans -Train the AI model with a diverse dataset including samples from various skin tones to address bias.

2. You're developing an AI for self-driving cars. In an emergency, the car must decide between hitting pedestrians or swerving and risking passenger safety. What factors would you consider when programming this ethical dilemma?
   Ans- When programming an AI system for autonomous vehicles to make split-second ethical decisions, several factors need consideration to prioritize between the safety of passengers and pedestrians:
   Safety Metrics and Risk Assessment, Ethical Principles and Guidelines, Contextual Awareness, Legal and Regulatory Compliance

3. A company develops an AI algorithm for automated hiring processes. Concerns are raised about the lack of transparency in how the algorithm makes hiring decisions, leading to potential biases and discrimination. As an AI ethics expert, how would you advocate for greater transparency in AI algorithms used for hiring? What measures would you recommend to ensure fairness and accountability in automated hiring systems?

   Ans- Here's how we can push for fairer AI hiring:

   - Show, don't tell: Let candidates see the factors the AI considers during selection. Explain why they were chosen or not (when possible).
   - Measure for fairness: Track how the AI treats different groups of applicants. Regularly check for hidden biases.
   - Many minds are better: Include HR professionals, ethics experts, and even job seekers in developing and testing the AI. This ensures it's fair and accountable.

4. As a product manager for a gaming company, you are tasked with designing an AI-powered chatbot to engage users with interactive stories and personalized gaming experiences. How would you approach the design process to ensure the chatbot provides engaging and personalized experiences? Additionally, what strategies would you implement to address potential biases in the chatbot's decision-making algorithms, particularly concerning fairness and inclusivity?"

   Ans - Designing an AI-powered chatbot for a gaming company would involve creating interactive stories, personalized gaming experiences, and engaging user interactions. This could include features such as character customization, in-game assistance and tips, progress tracking, and social integration for multiplayer experiences. To address potential biases in decision-making algorithms, the chatbot's training data would be curated carefully to avoid stereotypes or discriminatory patterns. Regular testing and user feedback would also be essential to refine the chatbot's performance and ensure an immersive gaming experience for users.

5. Scenario: You are presented with a scenario on the Moral Machine where an autonomous vehicle must choose between swerving to avoid hitting a group of elderly pedestrians or staying the course and potentially harming younger pedestrians.
   Question: How would you approach making such a decision in the context of programming AI systems for autonomous vehicles? What ethical principles would you prioritize?

   Ans- Consider factors like age, vulnerability, and potential impact of injuries to make a balanced decision

Class XI| **Artificial Intelligence** |Teacher Handbook