KVS RO DELHI

ARTIFICIAL INTELLIGENCE

CLASS XII

STUDENT'S SUPPORT MATERIAL (2025-26)

KVS RO

KENDRIYA VIDYALAYA SANGATHAN, DELHI REGION

STUDENT SUPPORT MATERIAL

ARTIFICIAL INTELLIGENCE

CLASS XII

CONTENT CREATORS:

NAME OF MEMBER	DESIGNATION	KV NAME
Created By:		
Ms Shilpa Dahiya	PGT-CS	PM SHRI KV No1 Delhi Cantt
Ms Renu Singh	PGT-CS	PM SHRI KV No1 Delhi Cantt
Ms Sarwari Begum	PGT-CS	PM SHRI KV No1 Delhi Cantt
Mr Pradeep Kumar Sharma	PGT-CS	PM SHRI KV No2 Delhi Cantt
Ms Geeta Singh	PGT-CS	PMSHRI KV No2 Delhi Cantt
Mr Pawanjeet Singh	PGT-CS	PM SHRI KV No2 Delhi Cantt
Checked By:		
Ms Shilpa Dahiya	PGT-CS	PM SHRI KV No1 Delhi Cantt

TABLE OF CONTENTS

SYLLABUS

ARTIFICIAL INTELLIGENCE (SUBJECT CODE-843)

CLASS–XII (SESSION 2025-2026)

	UNITS	NO.OF HOURS		MAX MARKS
PART-A	**EMPLOYABILITYSKILLS**			
	Unit1:Communication Skills-IV	15		2
	Unit2:Self-Management Skills-IV	10		2
	Unit3:ICT Skills-IV	15		2
	Unit4:Entrepreneurial Skills-IV	10		2
	Unit5:Green Skills-IV	10		2
	TOTAL	**60**		**10**
PART-B	**SUBJECTSPECIFICSKILLS**	**Th.**	**Prac.**	
	Unit1:Python Programming–II*	6	18	(*tobeevaluatedinpracticals only)
	Unit2:Data Science Methodology: An Analytic Approach to Capstone Project	8	12	8
	Unit3:Making Machines See	6	12	6
	Unit4:AI with Orange Data Mining Tool*	4	18	(*tobeevaluatedinpracticals only)
	Unit5:Introduction to Big Data and Data Analytics	7	12	6
	Unit6:Understanding Neural Networks	8	12	8
	Unit7:Generative AI	6	12	7
	Unit8:Data Storytelling	5	4	5

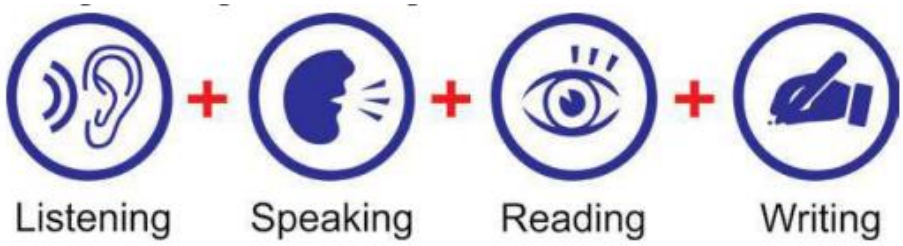
Part A: EMPLOYABILITY SKILLS

UNIT 1: Communication Skills

1. Introduction to Communication

Definition: Communication is a two-way process through which information or message is exchanged between individuals using language, symbols, signs or behavior.

Parts of Communication:
- Speaking
- Listening
- Reading
- Writing



Key Communication Skills: Listening, Speaking, Reading and Writing.

Importance:
- Speaking more than one language can help a person communicate with people around the world.
- Learning English can help in communicating with people besides the mother tongue.
- Effective communication is essential in today's information age to deal with people and customers.

Elements of Communication:
- Sender: Encodes and sends the message.
- Message: The information to be communicated.
- Channel: The medium used to send the message.
- Receiver: Decodes the message.
- Feedback: Response to check understanding.

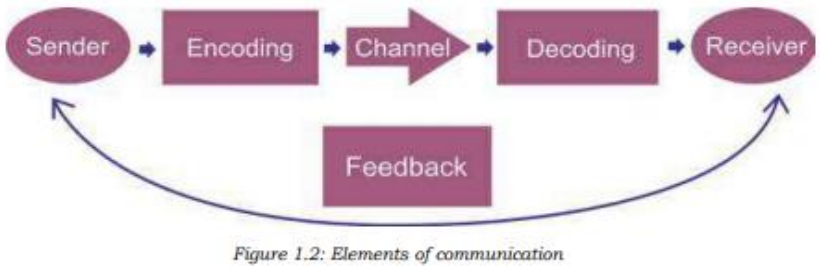


Figure 1.2: Elements of communication

2. Active Listening

Effective communication involves skills that can be utilized to send messages that are clear, concise and accurate.

A clear statement conveys the exact message.
A concise statement is brief and to the point.
An accurate statement is factual and verifiable.

Listening Skill:
Every effective conversation starts with listening. Listening skill is one of the most important skills in communication.
Reasons why listening attentively is important:
- We listen to obtain information.
- We listen to understand.
- We listen to enjoy.
- We listen to learn.
- We listen to build and maintain relationships.
- We listen to resolve conflicts.

Active Listening:
Active listening is an art, which comprises both a desire to comprehend and offer support and empathy to the speaker.
Factors affecting active listening:
- Eye contact
- Gestures
- Avoiding distractions
- Giving feedback

Stages of Active Listening:
1. Receiving
2. Understanding
3. Remembering
4. Evaluating
5. Responding

RESPECT Acronym:
R – Remove distractions
E – Eye contact
S – Show attentiveness
P – Pay attention
E – Empathies
C – Clarify doubts
T – Tune yourself to the speaker's timing

Barriers to Active Listening:

- Being pre-occupied
- Noise and visual distractions
- Past experiences or mindset
- Personal factors3. Parts of Speech

In any language, parts of speech are categories of words based on their function in a sentence. They are the building blocks of a language.

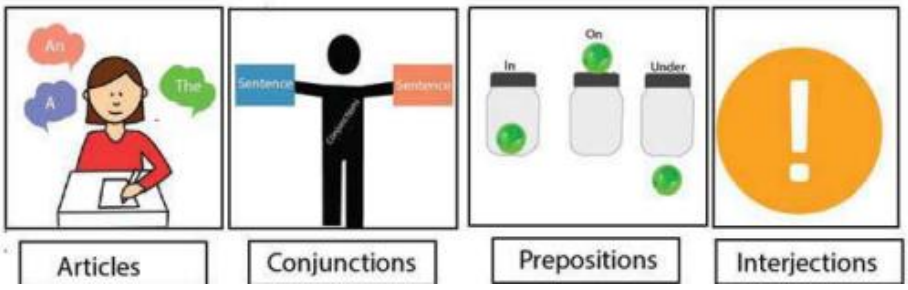
Sentence: A group of words that communicates a complete meaning, thought or action.
Phrase: A group of words which does not make complete sense.

Capitalisation Rules (MINTS):
M – Months
I – The letter 'I' when used alone
N – Names (people, places, rivers, mountains, etc.)
T – Titles (Dr., Mr., etc.)
S – Starting letter of sentences

Punctuation: There are 15 basic punctuation marks including full stop, comma, question mark, exclamation mark, apostrophe, colon, semi-colon, dash, hyphen, parenthesis, quotation mark, bracket, brace, ellipsis and bullet point.

Basic Parts of Speech:
- Noun: Names of person/place/thing/idea.
- Pronoun: Word used in place of a noun.
- Adjective: Describes a noun.
- Verb: Shows action.
- Adverb: Adds meaning to verb/adjective/adverb.

Supporting Parts of Speech:



- Articles: a, an, the
- Conjunctions: and, or, but
- Prepositions: in, on, at, under
- Interjections: Wow! , Oh!, Help!

4. Writing Sentences

Sentence: A group of words that together expresses a complete idea.

A sentence must be understood by others and generally contains subject, verb, and object.

Parts of a Sentence:

- Subject: Person/thing performing the action.
- Verb: Action word.
- Object: Person/thing receiving the action.

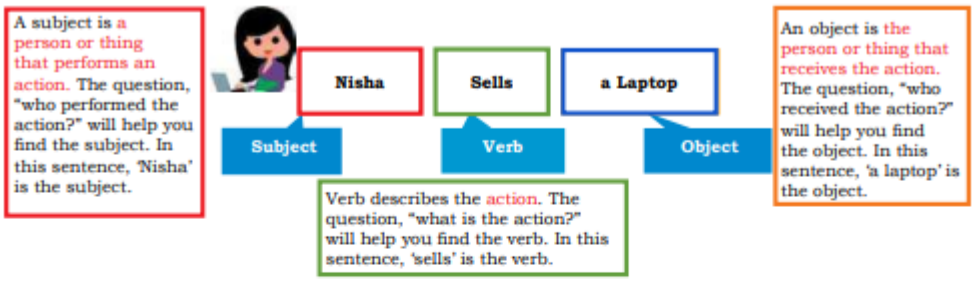


Figure 1.5: Parts of a sentence

Table 1.7: Parts of a sentence

Sentence	Subject	Verb	Object
I wrote a letter.	I	wrote	a letter
He called the customer.	He	called	the customer
She packed the product.	She	packed	the product
Dia and Sanjay booked a cab.	Dia and Sanjay	booked	a cab

Types of Objects:

- Direct Object: Answers 'what?'
- Indirect Object: Answers 'to whom/for whom?'

Voice:

- Active Voice: Subject performs the action.

Active voice	Passive voice
She wrote an email.	An email was written by her.
He opened the door for the customer.	The door was opened by him for the customer.
Mohan sold a bike.	The bike was sold by Mohan.

Active	Passive
I did not beat him.	He was not beaten by me.
Kapil made a profit yesterday.	A profit was made by Kapil yesterday.
The tiger was chasing the deer.	The deer was being chased by the tiger.

- Passive Voice: Subject receives the action.

Types of Sentences:

- Declarative (statement)
- Interrogative (question)
- Exclamatory (emotion)
- Imperative (command)

Sentence	Type
Where is my I-card?	Interrogative
My arms ache from planting those saplings!	Exclamatory
Reading mythology will make you more aware.	Declarative
Come with us right now.	Imperative
No way! I don't want a tattoo!	Exclamatory
Get out of the bed immediately!	Imperative

Paragraph: A group of sentences that focus on one idea. A new idea requires a new paragraph.

MCQ:

1. Communication is a _____process in which persons communicate information or messages using language, symbols, signs, or action.

a. Two-Way process b. Three-Way process

c. Four-Way process d. One-Way process

Answer

a. Two-Way process

2. _____ are the parts of communication in communication skills.

a. Speaking b. Listening c. Reading & Writing d. All of the above

Answer

d. All of the above

3. _____involves a sender, who encodes and sends a message through a channel, and a receiver, who decodes the message and gives feedback.

a. Listening b. Communication c. Speaking d. Reading & Writing

Answer

b. Communication

4**. _____is important in communication as it helps in knowing whether the receiver has understood the message or not.**

a. Sender b. Receiver c. Feedback d. All of the above

Answer

c. Feedback

5**. _____ involves skills that can be utilized to send messages that are clear, concise and accurate.**

a. Clear communication b. Effective communication

c. Consist communication d. None of the above

Answer

b. Effective communication

6. A _____ statement is one which conveys the exact message that your are trying to convey to the other person.

a. Consist b. Clear c. Both a) and b) d. None of the above

Answer

b. Clear

7. Effective communication skills help us to communicate the message _____.

a. Correctly b. Precisely c. Completely d. All of the above

Answer

d. All of the above

8. Why is listening attentively important?

a. We listen to obtain information b. We listen to understand

c. We listen to learn d. All of the above

Answer

d. All of the above

9. What are the various factors that affect active listening?

a. Eye contact b. Gestures c. Avoid distractions d. All of the above

Online learning platform

Answer

d. All of the above

10. What is the stage of active listening?

a. Receiving & Understanding b. Remembering

c. Evaluating & Responding d. All of the above

Answer

d. All of the above

11. What are the barriers of active listening?

a. Noise and Visual distractions b. Being pre – occupied & Personal factor

c. Past experiences or mindset d. All of the above

Answer

d. All of the above

12. You work at the front desk of a telecom company. A customer approaches you while you are working. The customer has a query regarding a bill. What would you do?

a. Not pay attention to the customer

b. Keep the work aside and help the customer

c. Continue doing your work while talking inattentively to the customer

d. Ask the customer to talk to someone else

Answer

b. Keep the work aside and help the customer

13. Which of the following can be a barrier to active listening?

a. Noisy environment

b. Not maintaining an eye contact with the speaker

c. Not being attentive

d. All of the above

Answer

d. All of the above

14. Which of the following is not a stage of active listening?

a. Receiving b. Understanding c. Non-responding d. Evaluating

Answer

c. Non-responding

15. What are the characteristics of an ideal message?

a. Clear b. Concise c. Accurate d. All of the above

Answer

d. All of the above

16. _____ known as building blocks of a language.

a. Sentence b. Words c. Parts of Speech d. None of the above

Answer

c. Parts of Speech

17. A _____ is a group of words that communicates a complete meaning, thought or action

a. Sentence b. Phrase c. Communication d. None of the above

Communication skills course

Answer

a. Sentence

18. A group of words, which does not make complete sense, is known as a _____.

a. Sentence b. Phrase c. Communication d. None of the above

Answer

b. Phrase

19. All the sentences begin with a _____ letter.

a. Small b. Capital c. Both a) and b) d. None of the above

Answer

b. Capital

20. The different types of words we use in sentences are called _____.

a. Sentence b. Parts of Speech c. Communication d. None of the above

Answer

b. Parts of Speech

21. The full stop and the comma are known as _____. They help us to separate the sections of a phrase and explain its meaning.

a. Punctuation b. Sentence c. Exclamatory marks d. None of the above

Answer

a. Punctuation

22. In the English language, there are _____ basic parts of speech.

a. 6 b. 7 c. 8 d. 9

Answer

c. 8

23. Word that related to a person, place, thing, or idea are called _____ These are referred to as 'naming words.'

a. Noun b. Pronoun c. Adjective d. Verb

Answer

a. Noun

24. A _____ is a word substituted in the place of a noun.

a. Noun b. Pronoun c. Adjective d. Verb

Answer

b. Pronoun

25. Words that describe other words are called _____ .

a. Noun b. Pronoun c. Adjective d. Verb

Answer

c. Adjective

26. _____ is words that show action.

a. Noun b. Pronoun c. Adjective d. Verb

Answer

d. Verb

27. Words that add meaning to verbs are known as _____ .

a. Noun b. Pronoun c. Adverb d. Verb

Answer

c. Adverb

28. The terms 'a', 'an' and 'the' are referred to as _____.

a. Article b. Conjunctions c. Prepositions d. Interjections

Answer

a. Article

29. Words that connect two nouns, phrases, or sentence are called _____ .

a. Article b. Conjunctions c. Prepositions d. Interjections

Answer

b. Conjunctions

30. _____ is a term that is used to express the relationship between two words.

a. Article b. Conjunctions c. Prepositions d. Interjections

Answer

c. Prepositions

31. Strong emotions, such as delight, surprise, wrath, or pain, are expressed using _____ words. They are punctuated by an exclamation mark at the end.

a. Article b. Conjunctions c. Prepositions d. Interjections

Answer

d. Interjections

32. What is a sentence?

a. A group of ideas that form a complete paragraph

b. A group of words that communicates a complete thought

c. A set of rules that we must follow to write correctly

d. A set of words that contains basic punctuation marks

Answer

b. A group of words that communicates a complete thought

33. Which of these sentences is capitalized correctly?

a. I am Hungry. b. Divya and sunil are reading.

c. The bucket is Full of water. d. She lives in Delhi.

Answer

d. She lives in Delhi.

34. A _____ is one that has only one subject and one predicate or has only one finite verb.

a. Simple Sentence b. Complex Sentence

c. Direct Sentence d. Indirect Sentence

Answer

a. Simple Sentence

35. A _____ is one, which consists of two or more coordinate clauses, joined by a coordinating conjunction.

a. Simple Sentence b. Complex Sentence

c. Direct Sentence d. Indirect Sentence

Answer

b. Complex Sentence

36. The person or thing that does an action is known as _____.

a. Subject b. Object c. Both a) and b) d. None of the above

Answer

a. Subject

37. In a sentence, the object can be either _____ or _____.

a. Direct or Indirect b. Complete or Not Complete

c. Both a) and b) d. None of the above

Answer

a. Direct or Indirect

38. Sentences where the subject performs an action are called as _____.

a. Active voice b. Passive Voice c. Indirect Voice d. None of the above

Answer

a. Active voice

39. Subject receives an action in sentence are known as _____.

a. Active voice b. Passive Voice c. Indirect Voice d. None of the above

Answer

b. Passive Voice

40. A group/ collection of sentences forms a _____.

a. Sentence b. Paragraph c. Both a) and b) d. None of the above

Answer

b. Paragraph

41. In the line "The children played football." identify the subject.

a. The children b. Children played c. Played d. Football

Answer

a. The children

42. In the line "The children played football." identify the object.

a. The children b. Children played c. Played d. Football

Answer

d. Football

43. Which sentence contains both indirect and direct object?

a. I am watching TV. b. She bought a blue pen.

c. The girls played cricket. d. He wrote his sister a letter.

Answer

d. He wrote his sister a letter.

44. The passive voice is used in which of the following statements?

a. They are watching a movie. b. The clock was repaired by Ramesh.

c. In the room, he is asleep. d. My pet horse bit the policeman.

Answer

b. The clock was repaired by Ramesh.

45. Type of sentence?

a. Declarative b. Interrogative c. Exclamatory & Imperative d. All of the above

Answer

d. All of the above

46. _____ sentence provides information or states a fact. It always ends with a full stop.

a. Declarative b. Interrogative c. Exclamatory d. Imperative

Answer

a. Declarative

47. _____ sentence asks a question. It always ends with question mark.

a. Declarative b. Interrogative c. Exclamatory d. Imperative

Answer

b. Interrogative

48. An _____ sentence expresses a strong emotion, such as joy, sadness, fear or wonder. It always ends with an exclamation mark.

a. Declarative b. Interrogative c. Exclamatory d. Imperative

Answer

c. Exclamatory

49. _____ sentences show an order, command, request or advice. It can end with a full stop or an exclamation mark.

a. Declarative b. Interrogative c. Exclamatory d. Imperative

Answer

d. Imperative

50. Identify the subject in the sentence, "The children played football".

a. The children b. Children played c. Played d. Football

Answer

b. Children played

51. Identify the object in the sentence, "The children played football."

a. The children b. Children played c. Played d. Football

Answer

d. Football

52. Which of these sentences has/have both indirect and direct objects?

a. I am working on a presentation. b. She bought a blue pen.

c. The girls played cricket. d. He wrote a letter to his sister.

Answer

d. He wrote a letter to his sister.

53. Which of these sentences is/are in passive voice?

a. They are watching a movie. b. The clock was repaired by Raju.

c. He is sleeping in the room. d. My pet dog bit the postman.

Answer

b. The clock was repaired by Raju.

LONG ANSWER QUESTIONS:

1. To effective communication, What skills are required?

Answer – To learn language you need to develop four key skills –

a. Listening

b. Speaking

c. Reading

d. Writing

2. What are various elements of a communication cycle?

Answer – The various elements of a communication cycle are –

a. Sender: the person beginning the communication.

b. Message: the information that the sender wants to convey.

c. Channel: the means by which the information is sent.

d. Receiver: the person to whom the message is sent.

e. Feedback: the receiver's acknowledgement and response to the message.

3. Why is Listening Important?

Shop for bestsellers

Soft skills training

Answer – Listening is important because –

a. We listen to obtain information

b. We listen to understand

c. We listen to enjoy

d. We listen to learn

e. We listen to build and maintain relationships

f. We listen to resolve conflicts

4. What are the different factors that affect active listening?

Answer – The various factors that affect active listening are as follows –

a. Eye contact – It is a form of body language. It's one of the most important components of the entire communication process. Maintaining eye contact with the person you're speaking to sends a signal to the speaker that "Yes, I am talking to you or listening to you". Avoiding eye contact could indicate that you don't want to hear what the other person has to say.

Soft skills training

b. Gestures –These indicate to the speaker if you are listening or not. Keep your hand and feet still while talking to someone.

Avoiding Distractions – You need to identify the things that distract you. You must physically remove the distractions in order to listen attentively.

c. Giving feedback – Positive or negative feedback is possible. However, in both circumstances, one must be polite so that the person to whom the feedback is being given is not hurt or offended.

5. What are the different stages of active listening?

Soft skills training

Skill development programs

Answer – The five stages of active listening are as follows –

a. Receiving

b. Understanding

c. Remembering

d. Evaluating

e. Responding

6. How to ensure active listening?

Answer – To ensure active listening are –

a. Remove any distractions that may interfere with your ability to listen. For example, When talking to someone, turn down the volume of the television, radio, or mobile phone.

High school curriculum

b. Looking at the speaker while listening is referred to as eye contact.

c. Use gestures to show that you're paying attention to the speaker.

d. Pay close attention to what the speaker is saying.
e. Feel the speaker's feelings and empathize with them. Empathy is the ability to understand and share another person's thoughts or experiences by thinking what it would be like to be in their position.
f. Clarify doubts. Ask questions to clarify doubts.
g. Tune in to the speaker's timing, i.e., wait for the speaker to finish before responding.

7. What are the overcoming barriers to active listening?

Answer – The overcoming barriers to active listening are –
a. Being pre – occupied -When you're preoccupied, you could not pay attention to what someone is saying.
b. Noise and Visual distractions – In a noisy location, you might not be able to hear the other person clearly.
c. Past experiences or mindset – You may have developed biases or prejudices based on past experiences and interactions.
d. Personal factors – Your personal feelings may affect your listening, for example, your preconceptions about the other person.

8. What are the Capitalization rules in the English Language?

Answer – Capitalization rules of English Language are –
a. Capitalize the first letter in the names of all months.
b. Capitalize the letter 'I' when used as a word.
c. Capitalize the first letter in the names of people, places, rivers, seas and oceans, mountains, islands and days.
High school curriculum
d. Capitalize the first letter in the titles used before people's names.
e. Capitalize the first letter in every sentence.

9. What are the basic parts of speech in the English Language?

Answer – There are eight basic parts of speech in the English language. These are none, pronoun, verb, adjective, adverb, preposition, conjunction and interjection.
a. Noun – Nouns is the name of a person, place, animal or thing. This is also called 'naming words.
b. Pronoun – A pronoun is used in place of a noun.
c. Adjectives – Adjectives are words that describe other words.
d. Verbs – Verbs are words that show action.
e. Adverbs – Adverbs are words that add meaning to verbs, adjectives, or other adverbs.
f. Preposition – A preposition is a word palace before a noun or pronoun.
g. Conjunction – A conjunction is a word that connects two words, phrases, or clauses.
h. Interjection – A word used to communicate emotion is called an interjection.

10. What do you mean by supporting parts of Speech?

Answer – We are using many words, phrases, clauses or sentences to connect words, these words are known as supporting parts of speech.
Supporting parts of speech examples are –
a. Articles
b. Conjunctions
c. Prepositions
d. Interjections

11. What do you mean by Simple sentence?
Answer – A simple sentence is one in which there is only one subject and one predicate, or one finite verb.

12. What do you mean by Complex sentences?
Answer – A complex sentence is one, which consists of two or more coordinate clauses, joined by a coordinating conjunction.

13. How many types of objects are there in the English language?
Answer – Object normally comes after the verb phrase. The presence or absence of an object is determined by the verb's meaning. There are two types of objects in English language
a. Direct – Direct objects are the results of action, A direct object answers the questions 'what' or 'whom'. Ramesh collected honey all day. hear honey becomes the direct object.
b. Indirect – An indirect object answers the questions 'to whom', 'for whom', 'to what' and 'for what'. for example, in the sentence "Imran gave a gift to his mother." The verb is "gave" and in the sentence there are two objects "gift" and "mother".

14. What is the difference between Active and Passive voice?
Answer –
a. Active Voice – Sentences, where the subject does an action, are known to be in active voice.
b. Passive Voice – Sentences, in which the subject receives an action, are known to be in passive voice.

15. In English grammar, what are the several type of sentences?
Answer – A sentence is a collection of words that expresses an idea. all the sentences have a none or pronoun called the subject. Sentences always start from capital letters and end with a period, question mark or exclamation. The word "sentence" is from the Latin word which means "to feel".
Type of Sentence
a. Declarative Sentence – Declarative sentences end with a full stop, It is the most common type of sentence. example – I completed my work.
b. Interrogative Sentence – Interrogative sentences end with question marks. This type of sentence asks a question. example – Did you complete your homework?
c. Exclamatory Sentence – Exclamatory sentence ends with an exclamation mark. This sentence expresses a strong emotion. example – I come first in class!

d. Imperative Sentence – Imperative sentence can end with an exclamation mark or full stop. This sentence shows an order command, request, or advice. example – Complete your homework.

16. What are the Paragraphs?

Answer – A group of sentences forms a paragraph and paragraphs helps the reader to read the sections, because they break the text into different sections. .

If you're writing about your school, the first paragraph could consist of sentences describing the name, location, size, and other information. You can use sentences to express what you like about your school in the next paragraph.

UNIT 2 :Self-Management Skills Notes

1. Introduction to Self-Management

Definition: Self-management, also referred to as self-control, is the ability to control one's emotions, thoughts and behaviour effectively in different situations. It includes motivating oneself and working towards achieving personal and academic goals.

Importance:
- Students with strong self-management skills are better at coming to class on time, paying attention, obeying teachers/elders, and working with discipline.
- Helps a person to do better in studies or work.
- Needed for success in work and life by developing timeliness, discipline, problem-solving and work habits.

Core Aspects:
- Positive thinking: Believing that one can get things done and be happy.
- Result orientation: Dreaming big and achieving the desired results.
- Self-awareness: Knowing one's personality traits and using strengths effectively.

2. Motivation and Positive Attitude

Motivation: Derived from the word 'motive'. Directing behaviour towards a certain motive or goal is the essence of motivation.

Types of Motivation:
- Intrinsic Motivation: Comes from within. Activities are done for enjoyment, satisfaction, or because they are meaningful.
- Extrinsic Motivation: Comes from external rewards or incentives. For example, salary, recognition, prizes.

Positive Attitude: Makes a person happier, helps build relationships, improves chances of success, and aids in making better decisions.

Ways to Maintain Positive Attitude:
- Start the day with a morning routine and positive affirmations.
- Feed the mind with positivity (books, music, movies).
- Be proactive, focus on solutions instead of problems.
- Learn from failures and focus on the present.
- Work cheerfully towards goals.
- Exercise, yoga, meditation, and fresh air.
- Healthy diet, adequate sleep.
- Organise academic/work life and reduce stress.
- Take holidays to refresh.

Stress: A state of feeling upset, annoyed, and hopeless. It occurs when things do not work as expected. Managing stress is essential for a healthy balance.

Ways to Manage Stress:
- Stay positive and analyse situations calmly.

- Maintain an accomplishment sheet.
- Focus on the present instead of past.
- Talk to family and friends.
- Practise meditation and yoga.

3. Result Orientation

Definition: Result orientation is the ability to recognise important results and take steps to achieve them. It means focusing on the outcome rather than just the process.

How to Become Result Oriented:

1. Set clear goals.
2. Prepare an action plan (what, who, when, how, resources).
3. Use the right resources and tools.
4. Communicate with mentors and peers.
5. Make a calendar to track progress.
6. Work hard and believe in your dreams.

Goal Setting: Helps us understand what we want and how to measure success. Use SMART goals.

SMART Goals:

- S: Specific – goals should be clearly defined.
- M: Measurable – should be quantifiable to track progress.
- A: Action-oriented – include steps to achieve them.
- R: Realistic – must be attainable.
- T: Timely – must have deadlines.

Examples of Result-Oriented Goals:

- A student aims to score high marks.
- An athlete runs five miles daily.
- A traveller plans to reach a destination within a set time.

4. Self-Awareness

Definition: Self-awareness is about understanding one's own needs, desires, habits, traits, behaviours and feelings.

Steps towards Self-Awareness:

- Step 1: Become aware of one's emotions.
- Step 2: Make a habit of tracking one's feelings.
- Step 3: Expand awareness to all areas of life.

Personality: A cluster of thoughts, feelings, and behaviours that make a person unique.

Personality Traits: Relatively lasting patterns that distinguish individuals.

The Five Factor Model (Big Five Traits):

- Openness: Creative, curious, adventurous.
- Conscientiousness: Self-disciplined, responsible.
- Extraversion: Outgoing, talkative, energetic.
- Agreeableness: Kind, cooperative, considerate.

- Neuroticism: Anxious, self-doubtful, moody.

Personality Development: Shaped by temperament, character, environment, and culture.

5. Personality Disorders

Definition: Personality disorders are long-term unhealthy patterns of thoughts, feelings, and behaviours that cause distress and deviate from normal expectations.

Cluster A (Suspicious):
- Paranoid personality disorder: Distrust, suspicion.
- Schizoid personality disorder: Detached, aloof, emotionally cold.
- Schizotypal personality disorder: Odd beliefs, inappropriate emotions, avoids relationships.

Cluster B (Emotional & Impulsive):
- Antisocial personality disorder: Disregards social rules, aggressive, impulsive.
- Borderline personality disorder: Instability, emptiness, fear of abandonment, self-harm.
- Histrionic personality disorder: Overly dramatic, attention-seeking.
- Narcissistic personality disorder: Extreme self-importance, lack of empathy.

Cluster C (Anxious):
- Avoidant personality disorder: Socially inept, fears rejection, feels inadequate.
- Dependent personality disorder: Excessive reliance on others, lack of self-confidence.
- Obsessive-compulsive personality disorder: Extreme perfectionism, rigid adherence to rules.

Steps to Overcome Personality Disorders:
- Talk to someone, share feelings.
- Maintain physical health.
- Build confidence to handle difficulties.
- Engage in hobbies (music, dance, art).
- Stay positive, focus on challenges not problems.

MCQ

1. Self-management, also known as _____, is the ability to effectively control one's emotions, behavior and thoughts.

a. Self – control
b. Self – innovative
c. Self – Design
d. None of the above

Answer

a. Self – control

2. To manage oneself well, a person needs to develop the _____.

a. Positive Thinking
b. Result Orientation
c. Self Awareness
d. All of the above

Answer

d. All of the above

3. Motivation is derived from the word _____.

a. Motive b. Motion c. Motivation d. None of the above

Answer

a. Motive

4. Type of motivation in employability Skills?

a. Intrinsic Motivation b. Extrinsic Motivation

c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

5. _____ includes activities for which there is no apparent reward but one derives enjoyment and satisfaction in doing them.

a. Intrinsic Motivation b. Extrinsic Motivation

c. Both a) and b) d. None of the above

Answer

a. Intrinsic Motivation

6. _____ arises because of incentives or external rewards. Lack of motivation or incentives may lead to frustration.

a. Intrinsic Motivation b. Extrinsic Motivation

c. Both a) and b) d. None of the above

Answer

b. Extrinsic Motivation

7. A _____ attitude makes a person happier, and helps build and maintain relationships.

a. Negative attitude b. Positive attitude

c. Open attitude d. None of the above

Answer

b. Positive attitude

8. How we can maintain a positive attitude.

a. Start the day with a morning routine

b. Feed the mind with positivity

c. Be proactive, Focus on constructive and positive things

d. All of the above

Answer

d. All of the above

9. How we can maintain a positive outlook in the long run.

a. Physical exercise and fresh air & holidays with family and friends

b. Healthy diet & Adequate sleep

c. Organize academic life

d. All of the above

Answer

d. All of the above

10. How can you identify the stress?

a. Annoyed b. Hopeless c. Upset d. All of the above

Answer

d. All of the above

11. How can you manage stress?

a. Stay positive
b. Maintain an accomplishment
c. Keep your thoughts in present
d. All of the above

Answer

d. All of the above

12. How to become result oriented?

a. Set clear goals & Make a calendar
b. Prepare an action plan & word hard
c. Use the right resources and tools
d. All of the above

Answer

d. All of the above

13. What are the smart methods to set goals in self – management skills?

a. Specific & Measurable
b. Achievable & Realistic
c. Time bound
d. All of the above

Answer

d. All of the above

14. Examples of result – oriented goals?

a. A student may set a goal of scoring high marks in an exam
b. An athlete may run five miles a day
c. A traveler may try to reach a destination city within three hours.
d. All of the above

Answer

d. All of the above

15. If you are aware of your own values, likes, dislikes, strengths, and shortcomings as an individual. It denotes the fact that you are _____.

a. Self-Confident
b. Self-Control
c. Self Motivated
d. Self-Aware

Answer

d. Self-Aware

16. A _____ personality can lead to better performance, increased productivity and cordial relationships with others.

a. Positive
b. Negative
c. Both a) and b)
d. None of the above

Answer

. Positive

17. What are the different parameters of personality?

a. Openness & Agreeableness
b. Consciousness & Neuroticism
c. Extraversion
d. All of the above

Answer

d. All of the above

18. What are the common personality disorders?

a. Paranoid personality disorder
b. Antisocial personality disorder
c. Avoidant personality disorder
d. All of the above

Answer

d. All of the above

19. What are the steps to overcome from personality disorders?

a. Talk to someone b. Build confidence

c. Engage in hobbies d. All of the above

Answer

d. All of the above

20. Which of the following is not a parameter to describe an individual's personality?

a. Self-confidence b. Openness c. Neuroticism d. Agreeableness

Answer

c. Neuroticism

21. Which of the following is characterized by an extreme feeling of self importance?

a. Narcissistic personality disorder b. Borderline personality disorder

c. Dependent personality disorder d. None of the above

Answer

a. Narcissistic personality disorder

22. Ravi has feelings of emptiness, abandonment and suicide. What type of personality disorder is this?

a. Borderline b. Dependent c. Avoidant d. Obsessive

Answer

a. Borderline

23. Mona is helping her sister to overcome a personality disorder. What should she do?

a. Talk to her sister b. Engage her in hobbies

c. Help her build confidence d. All of the above

Answer

d. All of the above

LONG ANSWER QUESTIONS:

1. What is self awareness?

Answer – Self-awareness is about understanding one's own needs, desires, habits, traits, behaviors and feelings.

2. How to improve Self Management Skills?

Answer – To manage oneself well, a person needs to develop the following skills –

a. Positive Thinking: to think that one can get things done and be happy.

b. Result orientation: to dream big and achieve the desired or set results.

c. Self-awareness: to be aware of one's personality traits and make the best out of one's strengths.

3. What is motivation in Employability Skills?

Answer – The word 'motivation' comes from the word 'motive.' Motivation is defined as the act of directing one's behavior toward a specific motive or goal.

There are two types of Motivation –

a. Intrinsic Motivation – It includes actions for which there is no visible reward but which provide enjoyment and satisfaction.

b. Extrinsic Motivation – It arises because of incentives or external rewards. Lack of motivation or incentives may lead to frustration

4. What are the ways to maintain a positive attitude?

Answer – Following are some ways that can help one maintain a positive attitude.

a. Start the day with a morning routine
b. Feed the mind with positivity
c. Focus on constructive and positive things
d. Learn from failures
e. Move towards your goals and dreams.
f. Physical exercise and fresh air
g. Healthy diet
h. Organize academic life
i. Adequate sleep
j. Holidays with family and friends

5. What is stress and how to manage it?

Answer – Stress is a state of feeling upset, annoyed and hopeless.

Some of the ways to manage stress are given below.

a. Maintain a positive attitude and consider what is going wrong in a given situation. Once you've understood the situation, resolving it is simple.
b. Maintain an accomplishment sheet and enter even small achievements.
c. Whenever you feel negative thoughts are taking over, take a look at your accomplishment sheet.
d. Keep your thoughts in the present. Pondering over past issues makes us feel upset and helpless.
e. Talk to friends and family for comfort.
f. Practice meditation and yoga.

Health safety training
Online learning tools

6. What are the ways to become result oriented?

Answer –

a. Set clear goals: Setting clear and accurate goals is the first step one needs to take to meet the targets.

b. Prepare an action plan: An action plan describes the way a person or an organisation will meet the set objectives. It gives a detail of the steps to be taken to achieve the target.

c. Use the right resources and tools: One must evaluate the resources and tools needed to achieve those results and whether they are available.

Employability Skills Course
Self-management skills course

7. How can we set goals in daily life?

Answer – Goal setting helps us to understand what we want, how to achieve it and how do we measure our success.

a. Specific – Goals should be stated in specific terms. Vague goals are difficult to attain. Specific goals give us a concrete target. Hence, a goal should have a specific purpose.
b. Measurable – Goals should always be measurable. If we do not set our goals in measurable terms, it is difficult to assess whether we have achieved them or not.
c. Action-oriented – Goals do not just come true on their own. Effective goal setting should include action-based steps that one will follow to achieve the goal.
d. Realistic – There are few things more damaging to our sense of self-efficacy than setting ourselves up for failure. Goals must always be realistically attainable.
e. Timely – Goals must have deadlines. However, deadlines may change. But one must always set a deadline to get the job done within a specified time limit.

8. Give some examples of result oriented goals?

Answer – Examples of result – oriented goals are as follows
a. A student may set a goal of scoring high marks in an exam.
b. An athlete may run five miles a day.
c. A traveler may try to reach a destination city within three hours.

9. Steps towards self awareness?

Answer –
a. gaining a greater awareness of one's emotions
b. making a habit of tracking one's feelings.
c. expanding one's practice to areas of life beyond the person's feelings.

10. What do you mean by Personality?

Answer – Personality is a cluster of thoughts, feelings and behaviors that make a person unique and different from others.

11. What are the Big Five Factors of personality development?

Answer –
a. Openness: Individuals with openness to experience are, generally, creative, curious, active, flexible and adventurous. If a person is interested in learning new things, meeting new people and making friends, and likes visiting new places, the person can be called open-minded.
E learning Modules
b. Consciousness: Individuals, who listen to their conscience, are self-disciplined, do their work on time, take care of others before themselves and care about others' feelings.
c. Extraversion: Extroverts are individuals, who love interacting with people around and are, generally, talkative. A person, who can easily make friends and make any gathering lively, is confident and an extrovert.
d. Agreeableness: Individuals having such a trait are, generally, kind, sympathetic, cooperative, warm and considerate. They accommodate themselves in any situation. For example, people who help and take care of others are, generally, agreeable.
e. Neuroticism: Neuroticism is a trait, wherein, individuals show tendency towards anxiety, self-doubt, depression, shyness and other similar negative feelings. People, who have difficulty in meeting others and worry too much about things, show signs of neuroticism.

12. What are the common personality disorders?

Answer – Personality disorders involve long-term patterns of thoughts and behaviour that are unhealthy and rigid.

Suspicious

a. Paranoid personality disorder: The fear of others, especially friends, family members, and partners, is a defining feature of paranoid personality disorder. People who suffer from this disorder are sensitive to keeping emotions against others.

b. Schizoid personality disorder: The term 'schizoid' refers to a person's natural tendency to focus their attention on their inner life rather than the outside world. A person with schizoid personality disorder is detached and aloof, and prone to introspection and fantasy. The person shows little interest in forming personal relationships and seems to be emotionally cold.

c. Schizotypal personality disorder: People with this personality disorder feel that their ideas may influence other people or events. They frequently misunderstand actions. As a result, they have improper emotional reactions. They may avoid having emotional interactions on a regular basis.

Emotional and Impulsive

a. Antisocial personality disorder: Antisocial personality disorder is characterised by a disregard for social rules and obligations. They are obnoxious and aggressive, and they act rashly. They lack guilt and fail to learn from their mistakes. They may lie, steal, or harm others, and they may become alcoholics or drug addicts.

Online learning tools

b. Borderline personality disorder: Borderline personality disorder is characterised by a loss of self-worth, leading to feelings of emptiness and fears of abandonment.

c. Histrionic personality disorder: People with histrionic personality disorder frequently try to gain more attention by being overly dramatic.

d. Narcissistic personality disorder: People with narcissistic personality disorder believe that they are more important than others.

Anxious

a. Avoidant personality disorder: This form of disease affects people who are socially inexperienced, ugly, or inferior, and they are always afraid of being embarrassed, criticised, or rejected. They avoid social situations and frequently feel insufficient, inferior, or unattractive.

b. Dependent personality disorder: People with this illness have a low sense of self-esteem and a strong desire to be looked after. They require a great deal of assistance in making daily judgments and entrust crucial life decisions to others.

c. Obsessive-compulsive personality disorder: People with this disorder are strict in their commitment to rules and laws. They are characterised by an extreme of care for neatness, perfectionism, and strong attention to detail. If they're unable to achieve perfection, they get very uncomfortable.

13. Steps to overcome from personality disorders?

Answer –

a. Talk to someone. Most often, it helps to share your feelings.

b. Look after your physical health. A healthy body can help you maintain a healthy mind.

c. Build confidence in your ability to handle difficult situations.

d. Engage in hobbies, such as music, dance and painting. These have a therapeutic effect.

e. Stay positive by choosing words like 'challenges' instead of 'problems'.

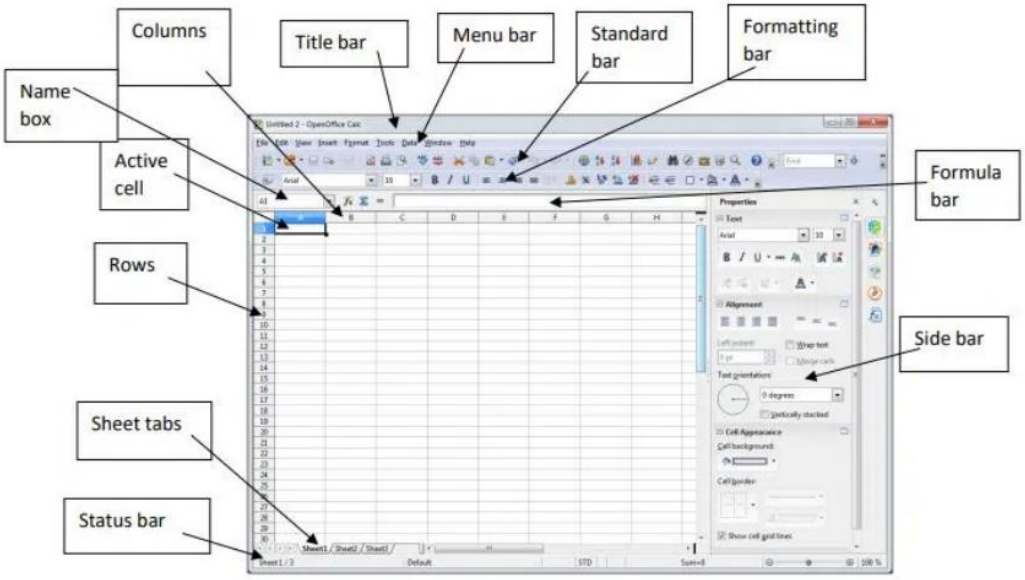
UNIT 3 :BASIC ICT SKILLS

SPREADSHEET APPLICATIONS

A spreadsheet program included in the free OpenOffice Calc is part of the OpenOffice package. The majority of the capabilities found in commercial spreadsheet tools are present in this user-friendly program. OpenOffice Calc is software that assists with formula computations and data analysis.

Spreadsheet programs have become very popular because of the following features:

- Built-in functions make calculations easier, faster, and more accurate
- Large volumes of data can be easily handled and manipulated
- Data can be exported to or imported from other software
- Data can be easily represented in pictorial form like graphs or charts
- Formulae are automatically recalculated whenever underlying data values are changed



CREATING A NEW WORKSHEET

To start OpenOffice Calc:

1. Click Start > All Programs > OpenOffice 4.1.5 > OpenOffice Calc
2. A spreadsheet workbook named Untitled1 opens up in an OpenOffice Calc application window

COMPONENTS OF A CALC SCREEN

- Title bar
- Menu bar
- Standard bar
- Formatting bar
- Name Box
- Worksheet tabs

CREATING A NEW WORKBOOK

The steps to create a new Calc workbook are:

- Select File > New > Spreadsheet
- Or click the New Document drop-down menu arrow on the Standard bar and select Spreadsheet
- Or press CTRL + N

SAVING A WORKBOOK

- Select Save option from the File menu
- Or click the Save icon on the Standard bar or Press CTRL + S
- If saving for the first time, a Save As dialog box will appear
- Type the file name and choose a location to save the file (file extension is .ods)

SELECTING CELLS

You can select range of cells in any one of the following ways:

Using the mouse

1. Click the cell you wish to start your selection from
2. Click and hold the left mouse button down, drag the mouse

Using the Keyboard

1. Place the cell pointer at one of the corner cells of the range to be selected
2. Press the SHIFT key and move to the diagonally opposite corner cell using the arrow keys
3. Release the SHIFT key when the required range has been selected

COPYING CELL CONTENTS

1. Select the cell(s) that contain(s) the data you want to copy
2. Select Copy option from the Edit menu, or click Copy button, or press CTRL + C
3. Click on the cell(s) where you want to paste the data
4. Select Paste option from the Edit menu, or click Paste button, or press CTRL + V

MOVING CELL CONTENTS

1. Select the cell that contains the data you want to cut
2. Select Cut option from the Edit menu, or click Cut button, or press CTRL + X
3. Click on the cell where you want to paste the data
4. Select Paste option from the Edit menu, or click Paste button, or press CTRL + V

INSERTING AND DELETING CELLS

To insert cells

1. Select the range of cells where you want to insert a block of cells
2. Select Cells option from the Insert menu
3. The Insert Cells dialog box appears
4. Select the appropriate option and click OK

To delete cells

1. Select the range of cells where you want to delete a block of cells
2. Select Delete Cells option from the Edit menu
3. The Delete Cells dialog box appears
4. Select the appropriate option and click OK

INSERTING ROWS AND COLUMNS

To insert rows

1. Select the row where you want to insert a new row
2. Select Rows option from the Insert menu, or right-click the row header and select Insert Rows
3. A new row is inserted above the selected row

To insert columns

1. Select the column where you want to insert a new column

2. Select Columns option from the Insert menu, or right-click the column header and select Insert Columns

DELETING ROWS AND COLUMNS

To delete rows

1. Select the row to be deleted
2. Right-click on the selected row header
3. Select Delete Rows option

To delete multiple rows, select them using the CTRL key, or by dragging the mouse while holding the left mouse button.

To delete columns

1. Select the column to be deleted
2. Right-click on the selected column header
3. Select Delete Columns option

To delete multiple columns, select them using the CTRL key, or by dragging the mouse while holding the left mouse button.

CHANGING ROW HEIGHT AND COLUMN WIDTH

Changing row height

- Dragging the divider below the row
- Double-clicking the divider to fit the row height to the cell contents
- Selecting Format > Row > Height, entering value, and clicking OK

Changing column width

- Dragging the divider to the right of the column header
- Double-clicking the divider to fit the column width to the cell contents
- Selecting Format > Column > Width, entering value, and clicking OK

FILTER AND SORTING

Filtering data

Filtering is a quick approach to locate and work with selected data based on criteria. Only rows or columns that satisfy the criteria are displayed. Different filtering methods in Calc:

- AutoFilter
- Standard Filter

AutoFilter

1. Select Data > Filter > AutoFilter
2. Drop-down arrows appear in each column heading
3. Click the drop-down arrow for desired column and select an item
4. Only rows meeting the criteria are displayed

- Select All to display all records again
- Select Top 10 to display highest 10 values

Standard Filter

1. Select Data > Filter > Standard Filter
2. The Standard Filter dialog box appears
3. Connect multiple conditions with logical AND or OR operators
4. Select options to filter records

Sorting data

Sorting on one column

1. Enter data in a worksheet
2. Select any cell in the column to sort
3. Click Sort option from Data menu
4. Select Ascending or Descending option and click OK

Sorting on multiple columns

1. Select any cell in the data range
2. Click Sort option from Data menu
3. Specify primary sort column and order
4. Add secondary sort criteria using Add Level button
5. Select options and click OK

FORMULAS AND FUNCTIONS

Formulas

Formulas compute results using arithmetic operations and always begin with = symbol. Formula components:

- Values: Numeric (45) or string ("Smiling")
- Cell Address: B4, A2:C6
- Functions: SUM, AVERAGE, MIN, MAX, etc.
- Operators: +, -, *, /, >, =, etc.
- Parenthesis: To control order of precedence (e.g., =(B2*B3)*2)

Mathematical Operators and Precedence

Operation	Operator	Precedence	Example	Result
Exponent	^	1	=2^3	8
Multiplication	*	2	=3*5	15
Division	/	2	=9/3	3
Addition	+	3	=12+15	27
Subtraction	-	3	=18-15	3

Order of evaluation:

1. Operations within brackets
2. Exponents
3. Division and multiplication (left to right)
4. Addition and subtraction (left to right)

Common Formula Errors

Error	Reason
###	Column not wide enough to display value
#DIV/0!	Division by zero
#VALUE!	Invalid argument (e.g., text where number expected)

Text Formulae

Text strings can be joined using concatenation (& operator). Example: "Keep"&" "&"Smiling" results in "Keep Smiling". Mathematical operations cannot be performed on strings.

Entering a Formula

1. Select the cell and enter formula directly or in Formula Bar
2. Press Enter key
3. Cell shows result; formula visible in Formula Bar when cell selected

Using Functions

Functions are predefined formulae that perform calculations using arguments.

Format: =function_name(Argument1;Argument2;Argument3;...)

SUM Function

Finds the sum of values. Example: =SUM(5,8,14,19) returns 46

AVERAGE Function

Finds the average of numbers. Example: =AVERAGE(3,6,9) returns 6

COUNT Function

Counts numeric values. Example: =COUNT(5,8,14,19) returns 4

MAX Function

Finds maximum value. Example: =MAX(74,102,134) returns 134

MIN Function

Finds minimum value. Example: =MIN(74,102,134) returns 74

PASSWORD PROTECTION

Using Tools menu

To protect worksheet or document:

1. Select Tools menu > Protect document > Choose Sheet or Document
2. For Sheet: Type password in Password box and Confirm text box
3. Click OK (password is case sensitive)

Protecting while Saving

1. Select Save As from File menu
2. Select Save with password check box
3. Click Save button
4. Enter password to open and confirm
5. Click More options to set additional permissions:
6. Open file read-only option
7. Password to allow editing

SAVING IN VARIOUS FORMATS

Saving in Microsoft Excel Format

1. Select Save As from File menu
2. Change Save as type to Microsoft Excel 97/2000/XP(xls)
3. Select location, type filename, and click Save

Saving in PDF format

1. Click Export Directly as PDF icon on Standard bar
2. Select location in Export dialog box
3. Type filename and click OK

Difference Between MS Excel and LibreOffice Calc

Feature	MS Excel	LibreOffice Calc
Developer	Microsoft	The Document Foundation
File Extension	.xlsx	.ods
Paid/Free	Paid (licensed)	Free (open source)
Online Support	Yes (Office 365, OneDrive)	Limited, but compatible
Interface	Ribbon-based	Menu + Toolbar-based

Shortcut Keys in Calc

- **Ctrl + C:** Copy
- **Ctrl + V:** Paste
- **Ctrl + X:** Cut
- **Ctrl + Z:** Undo
- **Ctrl + Y:** Redo
- **Ctrl + S:** Save
- **Ctrl + P:** Print
- **Ctrl + F:** Find
- **Ctrl + H:** Replace

MCQ:

1. ICT Stands for _____.

a. Informal and communication technology
b. Information and communication technology
c. Informal and commercial technology
d. Information and commercial technology

Answer

b. Information and communication technology

2. What are the different types of spreadsheets available?

a. Microsoft Excel b. LibreOffice Calc
c. Google Sheets d. All of the above

Answer

d. All of the above

3. A _____ is an arrangement of cells in a horizontal manner.

a. Row b. Column c. Cell d. None of the above

Answer

a. Row

4. A _____ is an arrangement of cells in a vertical manner.

a. Row b. Column c. Cell d. None of the above

Answer

b. Column

5. A _____ is a rectangle shaped box, where the row and column meet.

a. Row b. Column c. Cell d. None of the above

Answer

c. Cell

6. A _____ is a collection of cells in the form of a grid (a network of lines that intersect each other, making rectangles).

a. Worksheet b. Workbook c. Notebook d. None of the above

Answer

a. Worksheet

7. A _____is a spreadsheet that has one or more worksheets.

a. Worksheet b. Workbook c. Notebook d. None of the above

Answer

b. Workbook

8. A _____ is an electronic document used to store data in a systematic way and perform calculations just like an expense sheet.

a. spreadsheet b. worksheet c. workbook d. name box

Answer

a. spreadsheet

9. Which of the following functions can be performed with the help of spreadsheets?

i. Maintaining records

ii. Creating videos

iii. Analyzing data

iv. Performing financial calculations

v. Writing letters

a. i, ii, iii, iv, v b. i, iii, iv

c. i, iii, v d. iii, iv, v

Answer

b. i, iii, iv

10. You can enter a formula in a cell by starting with _____ symbol.

a. = equal to b. $ dollar c. # hash d. None of the above

Answer

a. = equal to

11. What are the different types of data in a spreadsheet?

a. Text b. Numbers c. Formula d. All of the above

Answer

d. All of the above

12. The extension of Spreadsheet (Open Office) is _____.

a. .XLS b. .ODF c. .TXT d. All of the above

Answer

b. .ODF

13. Shortcut key for printing the spreadsheet file is _____.

a. Ctrl + C b. Ctrl + V c. Ctrl + P d. Ctrl + PV

Answer

c. Ctrl + P

14. The correct order of steps for entering data in a spreadsheet is:

a. type the data, click the cell and press Enter.

b. click the cell, type the data and press Enter.

c. press Enter, click the cell and type the data.

d. click the cell, press Enter and type the data.

Answer

b. click the cell, type the data and press Enter.

15. What will you do to select an entire worksheet in a spreadsheet?

a. Click the File tab and select properties from the list

b. Click the grey row heading

c. Click the grey rectangle in the upper left corner of the spreadsheet

d. Click the grey column heading

Answer

c. Click the grey rectangle in the upper left corner of the spreadsheet

16. Ms Sharma is a teacher and has just started using a computer spreadsheet instead of her manual marks register. She has entered the marks of her students in various subjects in a spreadsheet. Now, she wants to find out the total marks for each student. What should she do?

a. She can use Sum function b. She can use + operator

c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

17. Ms Sharma has to find out the cell address of each cell with marks and type it. This requires some effort. Spreadsheet makes the task easier for Ms Sharma. Using a mouse, she can simply _____ the cell to be used in the formula instead of typing the cell addresses.

a. Drag the cell b. Click on the cell

c. Copy the cell d. None of these

Answer

a. Drag the cell

18. Which function is used to add multiple numbers in a spreadsheet.

a. =average() b. =sum() c. =min() d. None of the above

Answer

b. =sum()

19. Electronic spreadsheets have many options to make your content look neat and easy to read. This is called _____.

a. Font b. Formatting c. Alignment d. None of the above

Answer

b. Formatting

20. In case you want to give a different style or a bigger size to the heading, you can change the text style using the _____.

a. Font b. Alignment c. Indent d. None of the above

Answer

a. Font

21. In a spreadsheet, you can position the text in a cell to the left, right or center. This is called _____.

a. Font b. Alignment c. Indent d. None of the above

Answer

b. Alignment

22. What is the default alignment of numbers in a cell?

a. Left aligned b. Right aligned c. Center aligned d. Randomly aligned

Answer

b. Right aligned

23. What is the shortcut key to underline text in a spreadsheet?

a. Ctrl+b b. Ctrl+I c. Ctrl+l d. Ctrl+u

Answer

d. Ctrl+u

24. Which of the following features is used to perform addition in spreadsheets?

a. Format option b. Charts c. Graphs d. Formula

Answer

d. Formula

25. Which of the following signs define a formula?

a. + b. / c. = d. +

Answer

c. =

26. Ms Sharma has all her students' subject marks and their totals in a spreadsheet. If she wants to find three students with the highest marks, what will she do?

a. She will use Sort
b. She will use filter
c. She will use alignment
d. None of the above

Answer

a. She will use Sort

27. Ms Sharma wants to see the marks of students only from Class XII-A, what will she do?

a. She will use Sort
b. She will use filter
c. She will use alignment
d. None of the above

Answer

b. She will use filter

28. Ms Sharma is sharing the computer with other teachers and staff, she may want to protect her data. She may not want anyone else to open and make changes to her work. She can do this by protecting her spreadsheet using a _____.

a. Password
b. Track Change
c. Read Only
d. None of the above

Answer

a. Password

29. Which menu option will you use to sort data?

a. Tools
b. Data
c. Format
d. View

Answer

b. Data

30. Mr Gupta has a spreadsheet with a list of 500 items in his shop. A customer comes and asks for a particular item. How should he arrange the data so that he can find that item fast? What would Mr Gupta do? He will:

a. apply filters.
b. sort the data.
c. use a password.
d. format data.

Answer

b. sort the data.

31. Mr Verma shares the computer in his office with other colleagues. He wants to make sure no one sees the financial data he saves on the computer. What should he do?

a. Lock the computer in a cupboard
b. Change the password of his computer so that no one can use it
c. Apply password to the financial data sheet
d. Leave it as it is and hope that no one will open it
Information Technology Training

Answer

c. Apply password to the financial data sheet

32. Ms Sharma wants to explain the water cycle to her students. Mr Chaudhary wants to explain the working of a product to his clients and Dr Suman wants to give a lecture on her new research. All these people need to make an _____ for their audience.

a. Presentation b. Impress c. Power Point d. All of the above

Answer

d. All of the above

33. You have a summer vacation project on how to make biogas. You have done a lot of research, clicked pictures of biogas plants and taken videos of people who use biogas. Now, you have to make a presentation before the class. What would you use?

a. Chart paper b. Word document c. Presentation software d. Spreadsheet

Answer

c. Presentation software

34. What do you need to install on your computer to be able to run Impress?

a. Google b. Microsoft Office c. LibreOffice d. Apple iOS

Answer

c. LibreOffice

35. How many textboxes does the first slide of LibreOffice Impress have by default?

a. 1 b. 2 c. 3 d. 4

Answer

b. 2

36. Which is the correct step to save a presentation?

a. File>Save As>Type file name>Save b. File>Open>File name>Open

c. File>Template>Save as Template d. File>Close>Save>OK

Answer

a. File>Save As>Type file name>Save

37. Which is the correct step to close a presentation?

a. File>Save As>Type file name>Save b. File>Exit c. File>Close d. File>Export

Answer

b. File>Exit

38. Which is the correct step for printing a presentation?

a. File>Print b. File>Print>Handout c. File>Print>Handout>OK d. File>OK

Answer

c. File>Print>Handout>OK

39. If you have a blank slide with no textbox, you can insert a textbox using the option _____.

a. Insert>Textbox b. Insert> Table c. Insert> Format d. Insert> Style

Answer

a. Insert>Textbox

40. You can also change the size of the text by clicking on the _____.

a. Font Size b. Font Color c. Alignment d. All of the above

Answer

a. Font Size

41. To highlight a text in LibreOffice, you can make it _____ depending on the requirement.

a. Bold b. Underlined c. Italic d. All of the above

Information Technology Training

Answer

d. All of the above

42. The text can be aligned either _____.

a. Left or Right b. Center c. Justified d. All of the above

Answer

d. All of the above

43. Which menu option do you use to insert a slide?

a. Edit b. Insert c. Slide d. Tools

Answer

b. Insert

44. Which option will you use to change the colour of the text?

a. Font Color b. Font c. Highlight Color d. Format

Answer

a. Font Color

45. Which menu option will you click on to insert shapes and images?

a. Format b. Tools c. Edit d. Insert

Answer

d. Insert

46. What are the steps to insert a square shape in a presentation?

a. Insert > Shape> Line> Square b. Tools > Shape> Line> Square

c. Insert > Shape> Basic > Square d. Format > Text > Basic> Square

Answer

c. Insert > Shape> Basic > Square

47. What happens when you change the layout of a slide?

a. The format of the text changes

b. New slide is inserted

c. The arrangement of content (text, images, shapes) changes

d. The title gets aligned to the center of the slide

Answer

c. The arrangement of content (text, images, shapes) changes

LONG ANSWER QUESTIONS:

1. What is a spreadsheet?

Answer – A spreadsheet is an electronic document, which has rows and columns. It is used to store data in a systematic way and do calculations.

2. What are the different types of spreadsheet?

Answer – There are many types of spreadsheet available that have been created by different companies. Some of the most popular spreadsheets are –

a. Microsoft Excel

b. LibreOffice Calc

c. Google Sheets

3. What are the different components of a Spreadsheet?

Answer – The different components of a Spreadsheet are –

a. Row – Rows are the horizontal placement of cells.

b. Column – Columns are the vertical arrangement of cells.

c. Cell – A cell is a rectangular box where rows and columns intersect.

d. Name Box – The name box shows the position of the selected cell. The cell position is a combination of columns and rows.

e. Worksheet – A worksheet is a collection of cells in the form of a grid (a network of lines that intersect to form a rectangle). When you first open the spreadsheet, you'll see a blank worksheet named Sheet1.

f. Workbook – A workbook is a spreadsheet that has one or more worksheets.

4. What is the extension of libreoffice Calc file?

Answer – The libreoffice calc file extension is .ods.

5. What is a worksheet?

Answer – The worksheet in Calc is also referred to as spreadsheet. The spreadsheet can have many sheets. Each sheet can have many individual cells arranged in rows and columns. The sheet tab shows its default name as Sheet1, Sheet2, Sheet3, ….

6. What is the difference between Rows and Columns?

Answer – The sheet is divided into vertical columns and horizontal rows. The rows are numbered as 1,2,3,4,… and columns are numbered as A, B, C, D, …., Z,

7. What is the difference between cell and cell address?

Answer – The intersection of a row and column is called a cell. It is the basic element of a spreadsheet. It holds data, such as text, numbers, formulas and so on. A cell address is denoted by its column (letter) and row number. For example, D4, E9, Z89 are valid examples of cell addresses.

8. What is an Active cell?

Answer – In a spreadsheet, the cell is the place where we enter the data. The selected or activated cell is called an active cell.

9. Define Relative referencing, Mixed referencing and Absolute referencing.

Answer –

a) Relative referencing: When we drag the formula downwards in columns, the row number of the cell address gets one added. In the same way, when you drag the formula horizontally from the left to the right column, the column name of the cell address is added on to the next column letter.

b) Mixed referencing: As we have seen, when we drag the formula, row number or column name get changed in relative reference. To overcome this problem , In Mixed Referencing, the $ sign is used before row number or column name to make it constant. For example H2+C$14 (To make row 14 constant) or H2+$C14 (To make column C constant)

c) Absolute referencing: In Absolute referencing, a $ symbol is used before the column name as well as row number to make it constant in any formula. For example, C12, D5, etc. In this case, even if you drag your formula in any direction, the cell name remains constant.

Session 3: Working with Data and Formatting Text

10. What are the common formulas in Libreoffice?

Answer – The common formulas in LibreOffice is –

a. Addition
b. Subtraction
c. Average
d. Maximum and Minimum
e. Count

11. What do you mean by shorting in Libreoffice?

Answer – Shorting means arranging data in the spreadsheet according to the requirement. It can be done by depending on the value type stored in cells.

Example –

a. Alphabetically – A to Z or Z to A
b. Numerical – smallest to largest or largest to smallest

12. What is Filter in Spreadsheet?

Answer – Filtering helps to display relevant data from the spreadsheet. You can also temporarily remove irrelevant data.

13. How to protect a spreadsheet with password?

Answer – Protection helps the user to prevent data from others.

The steps to protect a spreadsheet are as follows –

Step 1 – Click on Tools and select Protect Spreadsheet
Step 2 – Type a password

Step 3 – Type the same password in the Confirm textbox
Step 4 – Click on Ok

14. What are the advantages of using a presentation?
Answer – Presentation software is being widely used to make digital presentations. It has many advantages, which are as follows.
a. Presentations have features like images, videos, animation and music.
b. Making changes in digital presentations is easy.
c. A digital presentation can be shown to a much larger audience by projecting on a Screen.
d. The presentation can be printed and distributed to the audience.

15. What are the different presentation software available?
Answer – There are a number of presentation software available, such as
a. LibreOffice Impress
b. Microsoft Office – PowerPoint
c. OpenOffice Impress
d. Google Slides
e. Apple Keynote

16. What are the steps to open a presentation file?
Answer – The steps to open a presentation are as follows –
Step 1: Open LibreOffice Impress
Step 2: Click on File
Step 3: Select Open from the drop – down
Step 4: Browse and select the folder where your file is saved

17. What is text Alignment?
Answer – Text alignment is a feature that allows users to horizontally align text in the presentation.
There are four different types of text alignment features,
a. Right Alignment
b. Left Alignment
c. Center Alignment
d. Justified Alignment

18. What are the various advanced features used in a digital presentation?
Answer – The various advanced features used in a digital presentation are –
a. Inserting shapes in presentation – You can insert different types of shapes in digital presentation like arrow, rectangle, circle, etc.
b. Inserting clipart and images in presentation – A picture speaks a thousand words. We use a lot of images in a presentation to make it simple and interesting.
c. Changing slide layout – The default layout of a LibreOffice Impress slide contains one textbox for the title and one for content. Layout helps to arrange the slide content in an organized way.

UNIT 4: Entrepreneurial Skills

1) Core ideas & definitions

- **Etymology:** The word *entrepreneur* comes from French **entreprendre** ("to undertake").
- **Entrepreneurship — key definitions (exact from book):**
  - **Cole:** "**A purposeful activity to initiate, maintain and aggrandize profit-oriented business.**"
  - **Stevenson & Jarillo:** "**The process by which individuals pursue opportunities without regard to the resources they currently control.**"
  - **Textbook composite definition:** "**An economic process, where an idea is generated or an opportunity is created, refined, developed and implemented, while being exposed to uncertainty to realize a profit by effective utilization of resources**."
- **Who is an entrepreneur? (book text):** "**An entrepreneur is someone, who has the capability to start an organization by organizing various resources required to make an opportunity purposeful. Entrepreneurs are not necessarily motivated by profit but may regard it as a standard for measuring achievement or success.**"
- **Disciplinary definitions of 'entrepreneur':**
  - **Economists:** brings resources, labor, materials and other assets into profit-making combinations.
  - **Psychologists:** driven by a psychological force, a desire to obtain or attain something.
  - **Sociologists:** actions determine social status and contribute to societal development.
  - **Management gurus:** has a vision and generates an action plan to achieve it.
- **Characteristics of entrepreneurship (from the text):** economic activity to create/develop/maintain a profit-oriented organisation; begins with identifying an opportunity; optimises resource utilisation; involves risk-taking by the enterprise and entrepreneur.
- **Entrepreneurship is both an art and a science:** requires stepwise procedures *and* the skill to digress when needed, yet remain profitable and growth-oriented (illustrated with music vs. sciences comparison).

---

2) Qualities of a successful entrepreneur (book list)

- **(a) Initiative:** Act early to seize opportunities; missed chances may not return.
- **(b) Willingness to take risks:** Every business involves uncertainty; entrepreneurs volunteer to take risks to succeed.
- **(c) Ability to learn from experience:**Don't repeat mistakes; learn quickly to avoid losses.
- **(d) Motivation:** Deep interest that drives you to complete tasks (example given of reading without stopping).
- **(e) Self-confidence:** Courage, enthusiasm, leadership; essential for success and to inspire others.

---

3) Types of entrepreneurs

- **Technical entrepreneurs:** use technical expertise to create machines/tools/methods; innovate to improve processes.
- **Non-technical entrepreneurs:** focus on services and non-manufacturing aspects (before/after production).
- **Professional entrepreneurs:** start, nurture to self-sustaining, then sell and start again (serial cycle).

- **IT entrepreneurs:** operate specifically in Information Technology (distinct from broader technical entrepreneurs in mech/electronics/civil).
- **Women entrepreneurs:** entrepreneurship undertaken by women; many have risen to the top.
- **Social entrepreneurs:** develop/fund/implement solutions benefiting society, culture, environment ("social innovator" used interchangeably).
- **Family business entrepreneurs:** run by a family and passed to the next generation; successor may inherit and/or modernise.
- **First-generation entrepreneurs:** have no entrepreneurship background; can start at any age.

---

4) Roles & functions of an entrepreneur (major functions)

(a) Identifying entrepreneurial opportunity: Sensing changing needs (food, fashion, education, etc.); requires imagination, creativity, and innovativeness.

(b) Turning ideas into action: Gather information on ideas/products/practices to meet market demand.

(c) Feasibility study: Assess market feasibility; anticipate problems; estimate inputs; prepare a **business plan/project report**.

(d) Resourcing: Ensure timely availability of money, machinery, raw materials, and workforce.

(e) Setting up an enterprise: Fulfil legal formalities; choose location; design premises; install machinery, etc.

(f) Managing the enterprise: Manage people, materials, finance; organise production; market offerings; ensure profits.

(g) Growth & development: After achieving results, set higher goals; continuously strive for excellence.

---

5) What motivates an entrepreneur? (Book specifics)

- **Standard of excellence:** sets high standards; works hard; shows innovativeness.
- **Uniqueness:** remains unique in what and how things are done.
- **Focus on long-term goals:** plans patiently for distant goals.
- **Need to influence:** sees ideas as revolutionary; expects substantial impact on the world.

---

6) Wage-employment vs entrepreneurship & basic definition

- **Wage-employed person:** works for an individual/organisation and gets paid.
- **Entrepreneur:** self-employed; takes risks; tries new ideas to improve the business.
- **Entrepreneurship (simple line in text):** "**the act of setting up and running a business, and taking risks in order to earn profits.**"

  Illustrative story:*Nutan* — wheelchair-bound, trained in a restaurant, started a tiffin service for 100 people, added a selectable menu (shows risk-taking, innovation, transition from wage-employment to entrepreneurship).

---

7) Identifying opportunities & risk-taking — practical ways

- **Community concerns:** start locally; spot inefficiencies; analyse how your idea helps; evaluate risks.
- **Personal experiences:** learn from experiences/challenges of self and successful entrepreneurs.

- **Research with others:** talk to like-minded people; join groups; discussions open doors and teach business nuances.
  Example given: Facebook began as Mark Zuckerberg's simple "connect college friends" idea → grew into a global social movement (used to illustrate spotting opportunities and taking the leap).

---

8) Startups — what the book says

- **Definition:** "**A startup is a company that is in the first stage of its operations.**" Differs from traditional businesses mainly in how they think about **growth**. Often financed by founders initially; later attracts outside investment.
- **Typical traits:** can start with minimum investment; many are tech/tech-enabled (e.g., online food delivery via an app partnering with restaurants).

---

9) Attitudes of an entrepreneur (Figure 4.1 — exact phrasing kept)

- **Taking initiative:** Ability to take charge and act before others.
- **Interpersonal skills:** Ability to work with others.
- **Perseverance:** Ability to continue even when it is difficult.
- **Organisational skills:** Ability to make optimum use of time, energy and resources to achieve goals.
- **Decisiveness:** Ability to make quick and profitable decisions.

---

10) Entrepreneurial competencies (Session 4)

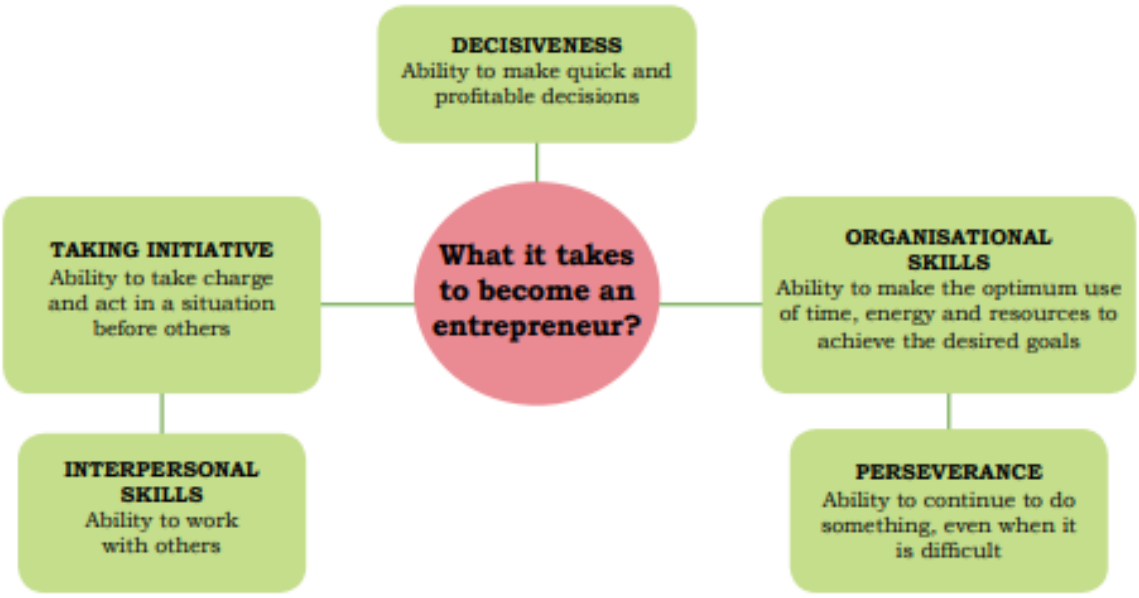


Figure 4.1: Attitudes of an entrepreneur

Competency = ability to do something well. The text maps attitudes ↔ competencies.

A) Decisiveness (process in Figure 4.3)

1. **Know yourself:** What problems matter to you? What are you good at?
2. **Identify opportunities:** Who do you want to help? What problems do they have? Where/when will the business work?
3. **Analyse the opportunity:** Why would it work? What could stop it? What do customers think? Who's the competition?

4. **Problem-solving:** Refine the idea; what makes it different? How will you use your strengths to make it succeed?
   Story (Ankur): shifted towel sales from village to city after market feedback; differentiated products/prices for two markets (classic opportunity identification + adaptation).
   B) Taking initiative (Figure 4.4)

- **Believe** you have the ability to act → **Realise** you are in control → **Take action** in that situation.
  Story (Elton): loved hairstyling; despite family pressure to take a software job, he took charge, convinced family, set up a salon and became well-known locally.
  C) Interpersonal skills — how to work well with people
- **Listening:** gives respect and builds trust.
- **Positive body language; Positive attitude; Stress management** (the figure lists these under interpersonal skills/attitudes).
- **Team diversity:** respect differences in looks, talk, beliefs; essential for growth.
  D) Perseverance
- **Not giving up; Working hard; Learning from failures** (as shown in attitudes-competencies figure).
  E) Organisational skills
- **Time management; Goal setting; Efficiency; Managing quality** (as per figure).

---

11) Extra clarifications the book makes

- **IT vs. Technical entrepreneurs:** technical work spans electronics/mechanical/civil; **IT entrepreneurs** innovate specifically within Information Technology.
- **"Opportunity"** in decisiveness: a chance or the right moment to do something (definition given explicitly).

MCQ:

1. Entrepreneurs discover an idea to seize on an opportunity, mobilize money, generate resources, and take calculated risks in order to expand the market or start a new business for _____.

a. Products b. Processes c. Services d. All of the above

Answer

d. All of the above

2. Entrepreneurship development refers to the process of enhancing entrepreneurial skills and knowledge through _____ programmes.

a. Structured training b. Institution building
c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

3. To determine the market potential, it is useful to _____ before programme formulation.

a. Need Assessment b. Need Survey c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

4. Training in entrepreneurship development needs to include support for_____ .

a. Entrepreneurship Orientation and awareness

b. Development of competencies

c. Improvement of business performance for growth and competitiveness

d. All of the above

Answer

d. All of the above

5. The _____ is a premier organization of the Ministry of Skill Development and Entrepreneurship, Government of India, engaged in training, consultancy and research to promote entrepreneurship and skill development.

a. National Institute for Entrepreneurship b. Small Business Development

c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

6. What are the training programs given by the Ministry of Skill Development and Entrepreneurship, Government of India, from the last 15 years.

a. Management Development Programmes

b. Entrepreneurship cum skill development programmes

c. Cluster Intervention

d. All of the above

Answer

d. All of the above

7. The word 'entrepreneur' is derived from the _____ entreprendre.

a. French word b. Austrian Word c. Greek Word d. None of the above

Answer

a. French word

8. The word 'entrepreneur' is derived from the French word entreprendre, which means 'to undertake'.

a. Undergoing b. Undertaking c. Underworking d. All of the above

Answer

b. Undertaking

9. Who told "The process by which individuals pursue opportunities without regard to the resources they currently control."

a. Stevenson b. Jarillo c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

10. Characteristics of entrepreneurship?

a. It is an economic activity done to create, develop and maintain a profit – oriented organization.

b. It begins with identifying an opportunity as a potential to sell and make profit in the market.

c. It deals with optimization in utilization of resources.

d. All of the above

Answer

d. All of the above

11. Qualities of a successful entrepreneur?

a. Initiative & Self confidence

b. Willingness to take risks & Hard work

c. Ability to learn from experience & Decision making ability

d. All of the above

Answer

d. All of the above

12. What are the different types of entrepreneurs?

a. Service entrepreneurs & Agricultural entrepreneurs

b. Business entrepreneurs & Non-technical entrepreneurs

c. Industrial entrepreneurs & Technical entrepreneurs

d. All of the above

Answer

d. All of the above

13. _____ entrepreneurs either create a new market for their services or provide a service in an existing market.

a. Service entrepreneurs b. Business entrepreneurs

c. Industrial entrepreneurs d. Agricultural entrepreneurs

Answer

a. Service entrepreneurs

14. _____ entrepreneurs, who undertake business and trading activities and are not concerned with the manufacturing work.

a. Service entrepreneurs b. Business entrepreneurs

c. Industrial entrepreneurs d. Agricultural entrepreneurs

Answer

b. Business entrepreneurs

15. _____l entrepreneur is, essentially, a manufacturer, who identifies the needs of customers and creates products or services to serve them.

a. Service entrepreneurs b. Business entrepreneurs

c. Industrial entrepreneurs d. Agricultural entrepreneurs

Answer

c. Industrial entrepreneurs

16. Agriculture has always been considered as a low-yielding entrepreneurship. _____ have now introduced new and innovative technology to maximize the yield, giving birth to agriculture entrepreneurship.

a. Service entrepreneurs
b. Business entrepreneurs
c. Industrial entrepreneurs
d. Agricultural entrepreneurs

Answer

d. Agricultural entrepreneurs

17. The Industrial Revolution gave birth to _____ entrepreneurs, who use their technical expertise to create and offer machines, tools and methods.

a. Technical entrepreneurs
b. Non-technical entrepreneurs
c. Professional entrepreneurs
d. IT entrepreneurs

Answer

a. Technical entrepreneurs

18. These entrepreneurs use their expertise in providing services to create a market for _____ entrepreneurs.

a. Technical entrepreneurs
b. Non-technical entrepreneurs
c. Professional entrepreneurs
d. IT entrepreneurs

Answer

b. Non-technical entrepreneurs

19. People who take up entrepreneurship in the field of Information Technology (IT) are called _____ entrepreneurs.

a. Technical entrepreneurs
b. Non-technical entrepreneurs
c. Professional entrepreneurs
d. IT entrepreneurs

Entrepreneurship course online

Answer

d. IT entrepreneurs

20. Individuals, who focus on developing solutions that benefit the society, are called _____ entrepreneurs.

a. Technical entrepreneurs
b. Non-technical entrepreneurs
c. Social entrepreneurs
d. IT entrepreneurs

Answer

c. Social entrepreneurs

21. When a family or an individual runs a business successfully and passes it onto the next generation, then such an entrepreneur is, generally, termed as family _____ entrepreneur.

a. Technical entrepreneurs b. Non-technical entrepreneurs
c. Social entrepreneurs d. Family business entrepreneurs

Answer

d. Family business entrepreneurs

22. Who are the 5th generation entrepreneurs?

a. Different age group without any entrepreneurship background
b. An individual who starts at the age of 50 years without any entrepreneurship background
c. Both a) and b)
d. None of the above
Entrepreneurship course online

Answer

c. Both a) and b)

23. What are the roles and functions of an entrepreneur?

a. Identifying entrepreneurial opportunity b. Turning ideas into action & Setting up an enterprise
c. Feasibility Study & Resourcing d. All of the above

Answer

d. All of the above

24. What motivates an entrepreneur?

a. Standard of excellence b. Uniqueness & Need to influence
c. Focus on long – term goals d. All of the above

Answer

d. All of the above

25. What are the ways by which aspiring entrepreneurs can identify new business opportunities and evaluate their potential and risks.

a. Community concerns b. Personal experiences
c. Research with others d. All of the above

Answer

d. All of the above

26. Startups come up with an innovative idea. At the same time, they like to get guidance and mentoring from an expert.

a. Startups b. Customer c. Both a) and b) d. None of the above

Answer

a. Startups

27. Salman tries new ideas as a tailor in a garment factory. Salman is a/an _____.

a. Entrepreneur b. Wage employed person c. Laborer d. Unskilled worker

Answer

b. Wage employed person

28. Prashant works for Surabhi, who is a businesswoman. Prashant is an _____.

a. Entrepreneur b. Wage employed person

c. Skilled worker d. Businessman

Answer

b. Wage employed person

29. What are the barriers that entrepreneurs may face while running their business?

a. Lack of adequate resources or raw material

b. Non – availability of skilled labor

c. Lack of requisite machinery and other infrastructure

d. All of the above

Answer

d. All of the above

30. What are the various schemes to help small – scale entrepreneurs gain capital

a. Government Mudra Yojana b. Credit Guarantee Scheme

c. Stand – Up India Scheme d. All of the above

Answer

d. All of the above

31. Mary believes that she will not have the ability to work with a team when she starts her business. The barrier she is facing is _____.

a. getting capital b. self-doubt c. risk taking

Answer

b. self-doubt

32. Sayed knows what business he wants to do but does not know what steps he has to take to get it running. The barrier stopping him is _____.

a. building the right team b. lack of plan c. risk taking

Answer

b. lack of plan

33. Harish has a rug business in India. He wants to start exporting rugs to Canada but does not know if they would sell there. The barrier he is facing is _____.

a. self-doubt b. lack of plan c. selecting the right business idea

Answer

a. self-doubt

34. What attitudes make a successful entrepreneur?

a. Decisiveness b. Taking initiative & Interpersonal Skills

c. Organizational Skills & Perseverance d. All of the above

Answer

d. All of the above

35. The ability to continue to do something, even when it is difficult, is called _____.

a. Initiative b. Organizational skills c. Perseverance

Answer

c. Perseverance

36. The ability to act in a situation before others do is _____.

a. Interpersonal skills b. Initiative c. Perseverance

Answer

b. Initiative

37. The ability to work with others is _____.

a. Organizational skills b. Interpersonal skills c. Initiative

Answer

a. Organizational skills

38. _____ is the ability to make profitable and quick decisions.

a. Decisiveness b. Initiative c. Interpersonal Skills d. Perseverance

Answer

a. Decisiveness

39. Identify the process where decisiveness can be developed.

a. Knowing yourself b. Problem solving c. Analyze the opportunity d. All of the above

Answer

d. All of the above

40. _____ is the ability to take charge and act in a situation before others.

a. Decisiveness b. Initiative c. Interpersonal Skills d. Perseverance

Answer

b. Initiative

41. Identify the process where Initiative can be developed in the entrepreneur.

a. Realize that you can take action b. Believe you can take action

c. Take action d. All of the above

Answer

d. All of the above

42. _____ means dealing with relationships. It is between two or more persons.

a. Decisiveness b. Initiative c. Interpersonal Skills d. Perseverance

Answer

c. Interpersonal Skills

43. Entrepreneurs want to develop their entrepreneurship skills, and suggest some ways to do that.

a. Listening b. Body Language c. Positive Attitude d. All of the above

Answer

d. All of the above

44. That state of mental pressure or tension is called _____.

a. Stress b. Headache c. Mental disturbed d. None of the above

Answer

a. Stress

45. How you can deal with stress.

a. Taking a walk in nature

b. Doing a physical activity like running, swimming, etc.

c. Practicing deep breathing exercises & Practicing meditation or yoga

d. All of the above

Answer

d. All of the above

46. Perseverance is the ability to continue to do something, even when it is difficult.

a. Perseverance b. Meditation c. Stress d. None of the above

Answer

a. Perseverance

47. How can you build perseverance in your daily life?

a. Not giving up when there is a difficult situation

b. Working hard to achieve the goal

c. Analyzing failures and learning from them

d. All of the above

Answer

d. All of the above

48. Time management is the process of planning and following a conscious control of time spent on specific activities. Time management includes_____.

a. Planning well b. Setting goals c. Setting deadlines d. All of the above

Answer

d. All of the above

49. A goal is somewhere we want to reach. It is the aim or result that we want from an activity. Goal makes us to _____ to complete the work.

a. Work harder b. Motivates us c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

LONG ANSWER QUESTIONS:

1. What training needs entrepreneurship to develop business.

Answer – Training in entrepreneurship development needs to include –

a. Entrepreneurship Orientation and awareness

b. Development of competencies

c. Improvement of business performance for growth and competitiveness

2. What do you mean by entrepreneur?

Answer – The word "entrepreneur" is derived from the French word "entreprendre," which means 'to undertake'.

3. What are the different characteristics of entrepreneurship?

Answer – Some of the characteristics of entrepreneurship are –

a. It is an economic activity done to create, develop and maintain a profit-oriented organization.

b. It all starts with recognizing an opportunity as a way to sell and profit in the market.

c. It deals with optimization in utilization of resources.

d. It is the ability of an enterprise and an entrepreneur to take risks.

4. What is the quality of a successful entrepreneur?

Answer – There are many qualities needed to successfully run an enterprise.

a. Initiative – An entrepreneur must be able to take initiative and take advantage of an opportunity.

b. Willingness to take risks – There is a certain amount of risk in any firm. It is not necessary that every business shall earn a profit. The entrepreneur takes up risks and starts a business.

c. Ability to learn from experience – It is possible for an entrepreneur to make mistakes. Once an error is committed, it must be ensured that it is not repeated in future otherwise it may lead to heavy losses.

d. Motivation – When you're inspired to do anything, you won't stop until you've finished it.

e. Self-confidence – A person must have faith in himself or herself in order to succeed in life.

f. Hard work – In life, there is no replacement for hard work. Every day, while running a business, one or more problems may develop. The entrepreneur must be vigilant in order to recognize and resolve difficulties as soon as possible.

g. Decision making ability – An entrepreneur must make a number of decisions when running a business. As a result, the individual must be capable of making appropriate and timely decisions.

5. What are the different types of entrepreneurs?

Answer – There are many types of entrepreneurs, which include the following.

a. Service entrepreneurs – These entrepreneurs either develop a new market for their services or serve an existing one. They recognize an opportunity and turn it into a service.

b. Business entrepreneurs – These are entrepreneurs who engage in commercial and marketing operations but do not engage in production. A business entrepreneur identifies a product's market potential.

c. Industrial entrepreneurs – An industrial entrepreneur is simply a manufacturer that recognises client needs and develops products or services to meet those demands.
d. Agricultural entrepreneurs – Agriculture has always been thought of as a low-returning business. Agriculturists have now implemented new and innovative technology in order to increase productivity.
e. Technical entrepreneurs – The Industrial Revolution gave birth to technical entrepreneurs, and they use their technical expertise to build and sell machinery, tools, and procedures.
f. Non – technical entrepreneurs – These entrepreneurs use their expertise in providing services to create a market for technical entrepreneurs.
g. Professional entrepreneurs – Such an entrepreneur starts a business, nurtures it and makes it reach a point of self-sustenance.
h. IT entrepreneurs – IT entrepreneurs are people who start businesses in the field of information technology (IT).
i. Women entrepreneurs – Women entrepreneurs, as the term suggests, are women who start their own business.
j. Social Entrepreneurs – Social entrepreneurs are those that focus on generating solutions that benefit society.
k. Family business entrepreneurs – When a family or a person successfully operates a firm and passes it on to the next generation, such entrepreneur is referred to as a family business entrepreneur.
l. First generation entrepreneurs – First-generation entrepreneurs are those who have never worked in the business world before. They can be of various ages and come from various backgrounds.

6. What is the role of an entrepreneur?
Answer – Some of the major role of entrepreneur are –
a. Identifying entrepreneurial opportunity – Entrepreneur must have imagination, creativity, and innovation.
b. Turning ideas into action – Entrepreneurs must be able to bring their ideas to life. To meet market demand, they should be able to gather information on ideas, products, and methods.
c. Feasibility study – Entrepreneurs conduct studies to assess the market feasibility of a proposed product or service.
d. Resourcing – To establish a successful business, an entrepreneur requires a variety of resources such as money, machinery, raw materials, and labor.
e. Setting up an enterprise – Some legal requirements may be required of the entrepreneur. In addition, the person must try to find a good place, design the premises, install machinery, and do a variety of other tasks.
f. Managing the enterprise – The person is responsible for managing the labor, materials, finances, and the organization of goods and services production.

7. What motivates an entrepreneur?
Answer – Some of the qualities that motivate an entrepreneur

a. Standard of excellence – An entrepreneur always sets high goals for himself and works to meet them by working hard and being creative.
b. Uniqueness – one of the most important qualities is to remain unique in everything the person does and the way it is done.
c. Focus on long-term goals – An entrepreneur is more concerned with what will be accomplished in the distant future than with what will be accomplished in the near future.
d. Need to influence – The entrepreneur perceives one's ideas as revolutionary and expects them to influence the world in a substantial way.
e. Identifying opportunities and risk taking – Finding the perfect business opportunity for you is a major concern for young entrepreneurs.

8. What are environmental barriers?
Answer – One of the biggest barriers that entrepreneurs face is environmental factors.
a. Lack of adequate resources or raw material
b. Non-availability of skilled labor
c. Lack of requisite machinery and other infrastructure
d. Unavailability of monetary resources on time

9. What are personal barriers?
Answer – Any new enterprise is risky because people are always afraid of what could go wrong, and these worries create personal boundaries.
Some of the personal barriers are –
a. Self doubt
b. Forming a team and teamwork
Session 3: Entrepreneurial Attitudes

10. Which attitudes make a person successful entrepreneur?
Answer – The attitudes that make a successful entrepreneur are –
a. Decisiveness – Ability to make quick and profitable decisions
b. Taking Initiative – Ability to take charge and act in a situation before others
c. Organizational Skills – Ability to make the optimum use of time, energy and resources to achieve the desired goals.
d. Interpersonal Skills – Ability to work with others
e. Perseverance – Ability to continue to do something, even when it is difficult.

11. What is decisiveness?
Answer – Decisiveness is defined as the ability to make quick and profitable decisions.
The following steps can help you develop your decisiveness.
a. Knowing yourself – You must first choose what is important to you and the type of problem you want to solve.
b. Identify opportunities – It is important to identify the best way and location for a business to operate.

c. Analyze the opportunity – What do customers think of the idea and what is the competition for that idea.
d. Problem solving – You might make changes to your idea after researching the clients and the market.

12. What do you mean by interpersonal skills?
Answer – Interpersonal means dealing with relationships. It is between two or more persons.
Some of the ways to develop interpersonal skills are –
a. Listening – It is important to listen when someone talks.
b. Body Language – When working with people or listening, our facial expressions, gestures, and postures are important.
c. Positive attitude – People with a positive attitude are hopeful and see the best even in difficult situations.
d. Stress management – When various personalities work together in a team, it can lead to tension for everyone. In such situations, stress management becomes important for everyone's health and well-being.

13. What is perseverance?
Answer – Perseverance is the capacity to keep going even when things are challenging.
Perseverance can be developed in a variety of ways.
a. Not giving up when there is a difficult situation
b. Working hard to achieve the goal
c. Analyzing failures and learning from them

14. Why organizational skills are important?
Answer – The capacity to make the best use of one's time, energy, and resources in order to achieve one's objectives is referred to as organizational skills.
Organizational Skills include the following –
a. Time management – Time management is the process of planning and following a conscious control of time spent on specific activities.
b. Goal Setting – A goal is somewhere we want to reach. It is the aim or result that we want from an activity. Goal makes us work harder, motivates us to complete what we start and achieve the target.
c. Efficiency – Efficiency is the ability to do things well, successfully, without wasting time.
d. Managing quality – Managing quality in a business means setting and maintaining a standard of excellence for products or services being provided to customers.

UNIT 5 :Green Skills

Definition of Green Jobs

- *Green jobs are defined as jobs that help bring about and maintain transition to environmentally sustainable forms of production and consumption.*
- **United Nations Environment Programme (UNEP) Definition**: Green jobs are works in agriculture, research and development (R&D), manufacturing, and services that substantially contribute to **preserving or restoring environmental quality**.

Definition of Green Collar Worker

- *A green collar worker is one employed in the environmental sectors of the economy.*
- **Examples**: Green building architect, environmental consultant, solar panel installer, organic farmer, waste management manager, environmental lawyer.

Benefits of Green Jobs

1. Increase efficiency of energy and raw materials.
2. Reduce greenhouse gas (GHG) emissions.
3. Minimise waste and pollution.
4. Protect and restore ecosystems.
5. Support adaptation to climate change.

Examples of Green Jobs by Sector

1. Agriculture

- Organic farming/gardening, avoids soil contamination and water pollution.
- Jobs: organic farmers, integrated pest managers, agriculture tourism guides.

2. Transportation

- Cleaner fuels: Compressed Natural Gas (**CNG**), ethanol, biodiesel.
- **Electric Vehicle Programme** by **Energy Efficiency Services Limited (EESL)** under the **Ministry of Power**.
- Jobs: EV mechanics, e-rickshaw drivers, charging station operators.

3. Water Conservation

- Rainwater harvesting, bamboo drip irrigation, cycle-run pumps.
- Jobs: water quality tester, conservation specialist.

4. Renewable Energy (Solar & Wind)

- Jobs: Solar Photovoltaic (**PV**) installer, solar lighting technician, roofer, wind turbine technician.

5. Eco-Tourism

- Jobs: eco-tour guides, eco-tourism entrepreneurs.

6. Green Buildings & Construction

- Focus on eco-friendly design, sustainable materials.
- Jobs: green building architect, landscaper, green component technician.

7. Solid Waste Management

- **Solid waste**: any discarded solid material from domestic, industrial, agricultural sources.
- Jobs: waste segregation staff, composting managers, e-waste recyclers, waste auditors.

8. Appropriate Technology

- Small-scale, eco-friendly technologies adapted to local needs.
- Examples: solar lamps, biogas plants, cycle-powered pumps.

- Jobs: biogas technician, rainwater harvesting installer, water filtration worker.

Role of Government in Green Jobs

- Reduce carbon dioxide (**CO_2**) emissions in the automobile sector.
- Promote non-fossil fuel energy: solar, hydro, wind.
- Encourage **green buildings** in construction.
- Ministry of Skill Development and Entrepreneurship (**MSDE**) has established **Skill Council for Green Jobs (SCGJ)**.
- *Make in India* campaign encourages growth of the green economy and green employment.

Importance of Green Jobs

1. Limiting Greenhouse Gas (GHG) Emissions

- **Greenhouse gases (GHG)**: Carbon dioxide (**CO_2**), methane (**CH_4**), nitrous oxide (**N_2O**), ozone (**O_3**), chlorofluorocarbons (**CFCs**).
- Causes: burning fossil fuels, vehicle emissions, refrigerants, agriculture.
- Solutions: renewable energy, alternate fuels (CNG, ethanol, biodiesel).

---

2. Minimising Waste and Pollution

- Reuse scrap material (paper mills, plastics).
- Improve quality control (reduce defective products).
- Waste exchange (waste of one industry becomes input of another).
- Manage electronic waste (**e-waste**) – mobiles, laptops, televisions.
- Use eco-friendly materials: banana leaf plates, biodegradable packaging.

---

3. Protecting and Restoring Ecosystems

- **Ecosystem**: A community of living and non-living beings interacting with each other.
- Problems: deforestation, natural disasters, climate change.
- Measures: avoid deforestation, plant trees, restore soil health, conserve wetlands, stabilise slopes with vegetation.

---

4. Adapting to Climate Change

- Some climate damage is irreversible; must adapt.
- Example: farmers use drought-resistant crop varieties in dry areas.
- **National Action Plan on Climate Change (NAPCC)** launched by the **Government of India** with 8 Missions:

1. National Solar Mission
2. National Mission for Enhanced Energy Efficiency
3. National Mission on Sustainable Habitat
4. National Water Mission
5. National Mission for Sustaining the Himalayan Ecosystem
6. National Mission for a Green India
7. National Mission on Sustainable Agriculture

8. National Mission on Strategic Knowledge for Climate Change

---

Quick Summary Table (Full Forms Included)

Term / Programme	Definition / Full Form
Green Jobs	Jobs that help transition to eco-sustainable production & consumption.
Green Collar Worker	Person employed in environmental sectors (e.g., solar installer).
EESL	Energy Efficiency Services Limited – implements EV programme under Ministry of Power.
MSDE	Ministry of Skill Development and Entrepreneurship.
SCGJ	Skill Council for Green Jobs – works in renewable energy, waste management, sustainable development.
NAPCC	National Action Plan on Climate Change – 8 missions (Solar, Energy Efficiency, Habitat, Water, Himalayas, Green India, Agriculture, Knowledge).
GHGs	Greenhouse Gases – CO_2, CH_4, N_2O, O_3, CFCs.
HEPA	High Efficiency Particulate Air filter.

MCQ:

1. Nowadays, we are experiencing unpredictable weather conditions due to _____.

a. Climatic Change
b. Environmental Deterioration
c. Both a) and b)
d. None of the above

Answer

c. Both a) and b)

2. What are the various ways of controlling and minimizing the pollution levels?

a. Plant trees near our houses and other places
b. Use eco – friendly mode of transport
c. Use cloth bags instead of plastic bags
d. All of the above

Answer

d. All of the above

3. _____ is the most popular plant which absorbs harmful pollutants from the air.

a. Areca Palm b. Rubber Absorb c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

4. A way we can contribute to the environment is by encouraging _____.

a. Green Jobs b. Green Product c. Green Surface d. All of the above

Answer

. Green Jobs

5. In which sector or industry Green Jobs are required.

a. Traditional Sectors b. Manufacturing and Construction Sector

c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

6. According to the United Nations Environment Program (UNEP), green jobs or green collar jobs are works in _____ that contribute substantially to preserving or restoring environmental quality.

a. Agricultural b. Research and Development

c. Manufacturing and service d. All of the above

Answer

d. All of the above

7. The Ministry of Skill Development and Entrepreneurship has set up the _____. It is working towards developing competencies of people in the domain of renewable energy, sustainable development and waste management.

a. Skill Council of Green Jobs b. Green Council of Development

c. Both a) and b) d. None of the above

Answer

a. Skill Council of Green Jobs

8. The Indian government is also providing impetus for green jobs as part of restructured growth strategies under its _____ campaign.

a. Make in India b. Make in Japan

c. Make in USA d. All of the above

Answer

a. Make in India

9. Green jobs help to _____.

a. Increase the efficiency of energy and raw material b. Control waste and pollution

c. Reduce greenhouse gas emissions d. All of the above

Answer

d. All of the above

10. _____ is the process by which we can grow plants and crops in an environmentally friendly way.

a. Organic Gardening b. Organic Farming

c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

11. _____are free from chemical residues of synthetic fertilizers, and hence, are good for our health.

a. Organic fruits b. Organic Vegetable

c. Both a) and b) d. None of the above

Answer

c. Both a) and b)

12. 15 – 20 group members in the village involved in eco – friendly farming are called _____.

a. FIG (Framer Interest Groups) b. SID (Social Interest Groups)

c. FEC (Farmer Eco Group) d. None of the above

Answer

a. FIG (Framer Interest Groups)

13. The _____ set up by the Indian government, can be utilized to provide support activities like technology dissemination, training awareness to the local youth and farmers for collection, storage and reuse of agro-waste.

a. Farmer Vigyan Kendras b. Krishi Vigyan Kendras

c. Indian Vigyan Kendras d. None of the above

Answer

b. Krishi Vigyan Kendras

14. CNG Stands for _____.

a. Common Natural Gas b. City Natural Gas

c. Compressed Natural Gas d. None of the above

Answer

c. Compressed Natural Gas

15. The new biofuel Policy announced by the Government of India on _____ and taken initiatives for enhanced use of biomass.

a. 12 August 2018 b. 10 August 2018

c. 12 August 2019 d. 10 August 2019

Answer

b. 10 August 2018

16. The_____, under the Ministry of Power, has launched an 'electric vehicle programme', which aims towards offering a comprehensive solution to facilitate the adoption of disruptive technology in India.

a. EESL (Energy Efficiency Services Limited) b. EESL (Electric Efficiency Services Limited)

c. EESL (Essential Energy Services Limited) d. None of the above

Answer

a. EESL (Energy Efficiency Services Limited)

17. Skill training of entrepreneurs in handling and maintenance of e-rickshaws can create more green jobs in the transportation sector.

a. E-rickshaws b. Auto rickshaws c. Both a) and b) d. None of the above

Answer

a. E-rickshaws

18. _____ can be installed in all houses and buildings and helps to increase the water table.

a. Water Harvesting System b. Electricity Run Pumps

c. Water westing System d. None of the above

Answer

a. Water Harvesting System

19. Some common green jobs in the agriculture sector are_____.

a. Water quality testing b. Water Conservation

c. Water Management d. All of the above

Answer

d. All of the above

20. Using _____ for drip irrigation is an eco-friendly way to irrigate the land. It uses water efficiently without any wastage. It is cheaper to build, and after 2–3 years it can be added to the soil as manure.

a. Bamboo Channels b. Plastics Channels c. Cast Channels d. None of the above

Answer

a. Bamboo Channels

21. _____ is intended to provide an experience to visitors to understand the importance of conserving resources, reducing waste, enhancing the natural environment and reducing pollution.

a. Eco – tourism b. Eco – Vehicle c. Eco – Visitor d. None of the above

Answer

a. Eco – tourism

22. Some common green jobs in the constructions are_____.

a. Gardening & Landscape b. Maintenance of Green components

c. Water Management d. All of the above

Answer

d. All of the above

23. The Ministry of Urban Development, Government of India, has classified solid waste in _____ based on the waste material.

a. 10 categories b. 14 categories c. 25 categories d. 100 categories

Answer

b. 14 categories

24. Solid waste management system includes collection_____.

a. Segregation
b. Transportation
c. Processing and disposal of waste
d. All of the above

Answer

d. All of the above

25. Some common green jobs in waste management are_____.

a. E-waste recycling
b. Solid waste management & waste reduction
c. Waste auditing & waste control
d. All of the above

Answer

d. All of the above

26. Examples of appropriate technology are _____.

a. Bike powered or hand powered water pumps
b. Solar lamps in street lights
c. Solar buildings
d. All of the above

Answer

d. All of the above

27. Rita's children fall sick very often. They have a cold and cough every month. What should Rita do?

a. Grow organic food in kitchen garden
b. Use air purifiers with HEPA filters
c. Keep them inside the house all the time
d. Paint her house with non-VOC paint

Answer

a. Grow organic food in kitchen garden

28. Jaanvi's children have grown up. They have a lot of clothes that are too small for them now. What should Jaanvi do with these old clothes?

a. Throw them in the trash can
b. Burn them
c. Keep using them even though they are too small
d. Donate them or make bags out of them

Answer

d. Donate them or make bags out of them

29. Some of the greenhouse gasses are _____.

a. Ozone and chlorofluorocarbons
b. Nitrous oxide & Methane
c. Carbon dioxide
d. All of the above

Answer

d. All of the above

30. _____gases can trap heat from the earth and prevent it from escaping into outer space. This causes the earth to heat, leading to 'global warming'.

a. Ozone and chlorofluorocarbons
b. Nitrous oxide
c. Methane
d. All of the above

Answer

d. All of the above

31. What are the different ways where we can minimize the waste production in the manufacturing plants and factories.

a. Reusing scrap material
b. Ensuring quality control & use eco friendly material
c. Waste exchange & Managing e-waste
d. All of the above

Answer

d. All of the above

32. The Government of India launched the National Action Plan on Climate Change (NAPCC) in June 2008 to deal with climate change and related issues. The NAPCC comprises eight missions in specific areas of _____.

a. Solar energy and Enhanced energy efficiency
b. Habitat, Water & Sustaining Himalayan ecosystem
c. Forestry, Agriculture and Strategic knowledge for climate change
d. All of the above

Answer

d. All of the above

33. There is a garbage bin in your canteen and it, generally, gets filled beyond capacity every day. What should you do?

(i) Ask the canteen management to get a bigger bin
(ii) Suggest ways to reduce the amount of garbage collected
(iii) Throw garbage on the floor and walk away
(iv) Talk to your friends about the problem

a. (i), (ii), (iii)
b. (i), (ii)
c. (i), (iii), (iv)
d. (i), (iv)

Answer

b. (i), (ii)

34. In a steel factory, a lot of utensils are being made. The manager finds a number of defective pieces, which have to be discarded. How can the person minimise the waste?

a. Give it to kabaadiwala or scrapdealer
b. Dump it in a landfill site
c. Send it back to the production line — to be melted
d. Sell it in market

Answer

c. Send it back to the production line — to be melted

LONG ANSWER QUESTIONS:

1. What is Green Jobs?

Answer – According to the United Nations Environment Program (UNEP), Green jobs, also known as green collar jobs, are jobs in agricultural, administrative, research and development, manufacturing, and service industries that contribute significantly to the preservation or restoration of environmental quality

2. What are the Benefits of Green Jobs?

Answer – The greening of the economy presents a major opportunity to start new businesses, develop new markets and lower energy costs.

Green jobs that contribute to protecting the environment and reducing carbon footprint are becoming a key economic driver of the twenty-first century.

Benefits of Green Jobs –

a. increase the efficiency of energy and raw material.
b. reduce greenhouse gas emissions.
c. control waste and pollution.
d. protect and restore ecosystems.
e. support adaptation to the effects of climate change

3. What do you mean by Farmer Interest Group?

Answer – Farmers' cooperatives are one of the most successful ways to reduce agricultural risk and boost small and marginal farmers' livelihoods.

Farmers are organized into 15–20 member groups at the village level (called Farmer Interest Groups or FIGs) and their associations are built up to a federating point, i.e. Farmer Producer Organizations (FPOs).

Farmers' cooperatives (FPOs) are farmer cooperatives whose members are mostly small or marginal farmers.

4. What is the BIOFUEL policy?

Answer – The Government of India's new biofuel policy, which was announced on August 10, 2018 focusses on initiatives for enhanced use of biomass so as to improve the availability of ethanol through starch and sugar-based feedstock, develop ethanol technologies and increase the production of biodiesel for blending.

Biofuels, such as BioCNG, Bio-methanol, and other biofuels made directly or indirectly from organic material, can create green jobs.

5. What is environmentally friendly drip irrigation?

Answer – There are green jobs in water harvesting and conservation. Drip irrigation with bamboo channels is an environmentally friendly approach to feed the ground. It efficiently uses water without wasting any.

It is less expensive to construct, and after 2–3 years the bamboo rots added to the soil as manure.

6. What are the different sectors where Green Jobs can be used.

Answer – Green jobs can be used in different sectors like –

a. Agriculture
b. Transportation
c. Water conservation
d. Solar and Wind energy
e. Eco – tourism
f. Building and construction
g. Solid waste management

7. What is appropriate technology?

Answer – Appropriate technology is small-scale, environmentally friendly technology that is customized to meet local needs. Bike- or hand-powered water pumps, solar streetlights, solar structures, and other examples of appropriate technology

8. What are the ways to reduce the amount of waste in Industries?

Answer – Some of the ways where we can reduce the waste in the Industries are –

a. Reusing scrap material – For example, In paper mills, damaged rolls are returned to the beginning of the production process and used as raw material. Off-cuts and scrap from the production of plastic objects are repurposed into new products.

b. Ensuring quality control – If product quality is maintained, the number of rejected products will decrease, resulting in less waste. Automated monitoring technology is use, which can aid in the early detection of production issues.

c. Waste exchange – This is where the waste product of one process becomes the raw material for another. It represents the way of reducing waste disposal through re-use.

d. Managing e-waste – With advanced technology, we have also encountered problems in managing e-waste such as outdated phones, laptops, and television sets.

e. Use of eco – friendly material – Scientists have discovered a variety of environmentally friendly materials, such as banana leaves and disposable paper plates, among others. These should be made widely available and their use needs to be encouraged.

9. What is NAPCC?

Answer – In June 2008, the Indian government launched the National Action Plan on Climate Change (NAPCC) to address climate change and related challenges. Solar energy, improved energy efficiency, habitat, water, sustaining Himalayan ecosystems, forestry, agriculture, and strategic knowledge for climate change are among the eight purposes of the NAPCC.

PART B : SUBJECT SPECIFIC SKILLS

UNIT 1 : Python Programming

This unit provides a comprehensive guide to Python libraries and programming concepts essential for data analysis, artificial intelligence, and machine learning. It emphasizes practical, hands-on exercises, team discussions, web searches, and case studies to help students master advanced techniques.

---

1. Python Libraries

Python libraries are pre-written code collections that simplify programming tasks by providing reusable functions and methods. They are essential for data science and analytics. Two of the most important libraries are **NumPy** and **Pandas**.

1.1 NumPy Library

- **Definition:** Numerical Python, a powerful library for numerical computing and array processing.
- **Key Feature:** Arrays have a "rank," which indicates the number of dimensions.
- **Use Case:** Enables efficient numerical operations, mathematical computations, and array manipulation.

1.2 Pandas Library

- **Application in AI:**
  - Used to load datasets, display summary statistics, and perform group-wise analysis.
  - Example: Analyze marketing campaigns (campaign type, budget, engagement metrics, sales).
- **Functionality:**
  - Provides robust data manipulation and aggregation.
  - Enables easy analysis of complex datasets and supports visualizations.

1.2.1 Pandas Data Structures

Pandas provides two main data structures:

1. **Series:** A one-dimensional labeled array that can hold any data type.
   - Creation Example: A Series can be created from scalar values, lists, or dictionaries.
2. **DataFrame:** A two-dimensional, labeled data structure with columns of potentially different types (similar to spreadsheets or SQL tables).

Creating DataFrames:

- **From NumPy Arrays:**

```
import pandas as pd
import numpy as np
array1 = np.array([90,100,110,120])
array2 = np.array([50,60,70])
array3 = np.array([10,20,30,40])
marksDF = pd.DataFrame([array1, array2, array3], columns=['A','B','C','D'])
print(marksDF)
```

Output:

```
AB     C     D
090.0100.0110.0120.0
150.060.070.0NaN
```

210.020.030.040.0

- **From Dictionary of Arrays/Lists:**

```
import pandas as pd
data = {'Name': ['Varun','Ganesh','Joseph','Abdul','Reena'], 'Age':[37,30,38,39,40]}
df = pd.DataFrame(data)
print(df)
```

Output:

```
Name  Age
0   Varun   37
1  Ganesh   30
2  Joseph   38
3   Abdul   39
4   Reena   40
```

- **From List of Dictionaries:**

```
import pandas as pd
listDict = [{'a':10,'b':20},{'a':5,'b':10,'c':20}]
df = pd.DataFrame(listDict)
print(df)
```

Output:

```
ab     c
010.020.0NaN
15.010.020.0
```

1.2.2 Dealing with Rows and Columns

- **Adding a New Column:**

```
ResultSheet = {'Rajat': pd.Series([90,91,97],index=['Maths','Science','Hindi']),
            'Amrita': pd.Series([92,81,96],index=['Maths','Science','Hindi'])}
Result = pd.DataFrame(ResultSheet)
Result['Fathima'] = [89,78,76]
```

Output:

```
Rajat  Amrita  Fathima
Maths      90     92      89
Science    91     81      78
Hindi      97     96      76
```

- **Adding a New Row (DataFrame.loc[]):**

```
Result.loc['English'] = [90,92,89]
```

Output:

```
Rajat  Amrita  Fathima
Maths       90     92      89
Science     91     81      78
Hindi       97     96      76
English     90     92      89
```

- **Changing Data Values:**
  Result.loc['Science'] = [92,84,90,72,96,88]
  Output:

	Rajat	Amrita	Fathima
Maths	90	92	89
Science	92	84	90
Hindi	97	96	76
English	90	92	89

1.2.3 Deleting Rows or Columns

- Use drop() method:
  - axis=0 → Delete rows.
  - axis=1 → Delete columns.

  Example:

```
# Delete Hindi row
Result2 = Result.drop('Hindi', axis=0)
print(Result2)
```

Output:

	Rajat	Amrita	Fathima
Maths	90	92	89
Science	92	84	90
English	90	92	89

```
# Delete 'Rajat' and 'Fathima' columns
Result3 = Result.drop(['Rajat', 'Fathima'], axis=1)
print(Result3)
```

Output:

	Amrita
Maths	92
Science	84
Hindi	96
English	92

1.2.4 Attributes of DataFrames

- df.index → Index of the DataFrame.
- df.columns → Column labels.
- df.shape → Dimensions (rows, columns).
- df.head(n) → First n rows.
- df.tail(n) → Last n rows.

---

2. Import and Export Data between CSV Files and DataFrames

CSV Files

- CSV stands for **Comma-Separated Values**, a simple format for storing tabular data.

- Easy to read/write and widely used in data science.
- Pandas simplifies importing/exporting CSV data into DataFrames.

2.1 Importing a CSV File

- Use pd.read_csv("filename.csv").

Google Colab Steps:

1. Open Google Colab and create a new notebook.
2. Click the folder icon.
3. Upload the CSV file.
4. Execute:

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

Python IDE:

```
import pandas as pd
df = pd.read_csv('C:/PANDAS/studentsmarks.csv', sep=",", header=0)
```

2.2 Exporting a CSV File

- Use df.to_csv().

```
df.to_csv(path_or_buf='C:/PANDAS/resultout.csv', sep=',')
df.to_csv("resultout.csv", index=False)  # In Google Colab
```

---

3. Handling Missing Values

Two strategies:

1. Drop rows or columns with missing values.
2. Fill missing values with estimates.

Functions:

- isnull(): Detect missing values.
- dropna(): Remove rows/columns with missing values.
- fillna(value): Replace missing values with a given value.

Example:

```
marks.isnull().sum().sum()   # Total missing values
drop = marks.dropna()        # Drop missing
FillZero = marks.fillna(0)   # Fill missing with zero
```

---

4. Case Study: Handling Missing Student Marks

A dataset of student marks contains missing values due to absence or events.

```
import pandas as pd
import numpy as np

ResultSheet = {
 'Maths': pd.Series([90,91,97,89,65,93], index=['Heena','Shefali','Meera','Joseph','Suhana','Bismeet']),
```

'Science': pd.Series([92,81,np.NaN,87,50,88],
index=['Heena','Shefali','Meera','Joseph','Suhana','Bismeet']),
'Hindi': pd.Series([81,71,67,82,np.NaN,89],
index=['Heena','Shefali','Meera','Joseph','Suhana','Bismeet'])
}
marks = pd.DataFrame(ResultSheet)

- Detect missing values:
  print(marks.isnull())
  print(marks.isnull().sum().sum()) # Output: 3
- Drop missing:
  drop = marks.dropna()
- Fill missing:
  FillZero = marks.fillna(0)

---

5. Practical Activity: Linear Regression

Dataset: USA_Housing.csv.

Steps:

1. Load dataset.
2. Check for missing values.
3. Perform Exploratory Data Analysis (EDA).
4. Split data into training and testing sets.
5. Train model and make predictions.
   Code:
   from sklearn.linear_model import LinearRegression
   model = LinearRegression()
   model.fit(X_train, y_train)
   predictions = model.predict(X_test)

---

6. Exercises

A. Objective Questions

1. Primary data structure in Pandas: **Series**
2. fillna(0): Fills missing values with zero.
3. Library for data management: **Pandas**
4. Read CSV: pd.read_csv("filename.csv")
5. df.shape: Returns rows and columns.
6. Export CSV: to_csv().

B. Short Answer

- **DataFrame:** 2D table-like structure with rows and columns.
- **Create Series:**
  import pandas as pd
  data = {'a':1, 'b':2, 'c':3}
  series = pd.Series(data)
- Missing value strategies: dropna() or fillna().

- head(n): Returns first n rows.
- **NumPy role:** Efficient numerical computation, array handling.
- isnull(): Detects missing values.

C. Long Answer

1. **Import/Export Data:**
   import pandas as pd
   df = pd.read_csv('data.csv')
   df.to_csv('output.csv', index=False)
2. **Handle Missing:** Use dropna() or fillna().
3. **Linear Regression:** Predict outcomes based on variables.
4. **NumPy vs Pandas:**
   - NumPy arrays: Homogeneous, fast math operations.
   - Pandas: Heterogeneous, data manipulation.
5. **Adding Rows/Columns:**
   df['new'] = [value1, value2]
   df.loc[len(df)] = [value1, value2]
6. **Attributes:** index, columns, shape.

D. Case Study

1. Replace missing marks with mean:
   df.fillna(df.mean(), inplace=True)
2. Aggregate sales:
   total_sales = df.groupby('Product')['Sales'].sum()
3. Marketing dataset analysis:
   grouped = df.groupby('Campaign Type').agg({'Sales':'mean','Engagement':'mean'})
4. Clean employee data:
   df.drop(['Irrelevant Column'], axis=1)
   df.fillna(0, inplace=True)

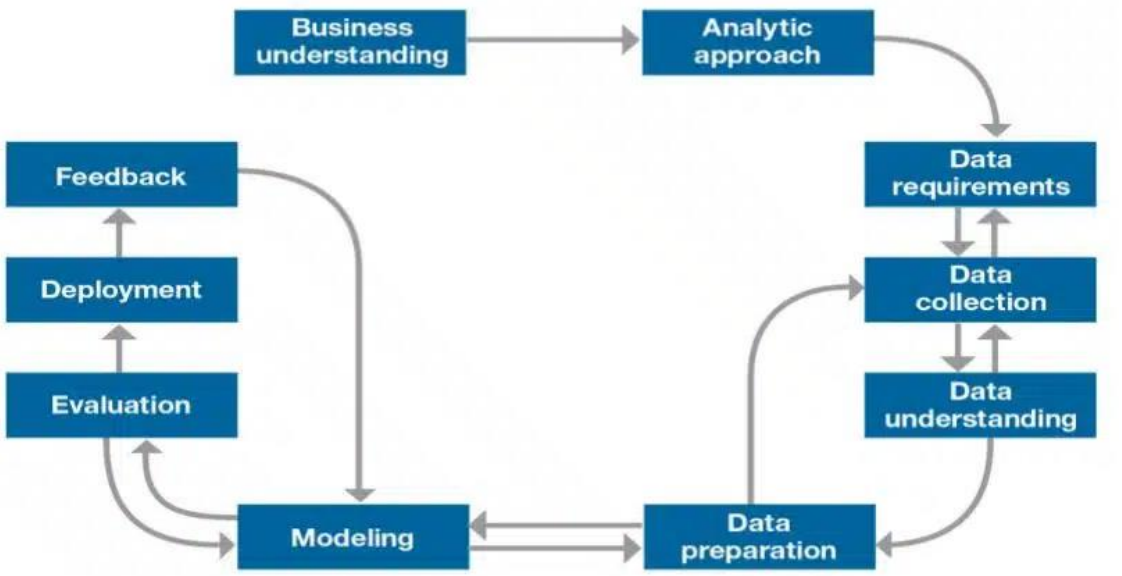
Unit 2: Data Science Methodology: An Analytical Approach to Capstone Project

A data science methodology is a structured approach to solving problems. **A methodology gives the data scientist a framework for designing an AI project.** The framework will help the team to decide on the methods, processes, and strategies that will be employed to obtain the correct output required from the AI project.

Definition: Data Science Methodology is a process with a prescribed sequence of iterative steps that data scientists follow to approach a problem and find a solution.

Data Science Methodology which was **introduced by John Rollins, a Data Scientist at IBM Analytics**. It consists of 10 steps. The technique is broken down into **five modules**, each of which covers two stages and explains why each is necessary.

1. From Problem to Approach
2. From Requirements to Collection
3. From Understanding to Preparation
4. From Modelling to Evaluation
5. From Deployment to Feedback



1. Business understanding

In the first stage, we must understand the problem and try to comprehend what is exactly required in the business. This is also known as problem scoping and defining. The term can use the 5W1H Problem Canvas to deeply understand the issue. This stage also involves using the Design Thinking Framework.

To solve a problem, it is crucial to understand the customer's needs. This can be achieved by asking relevant questions and engaging in discussions with all stakeholders.

2. Analytic Approach

In this stage, the data scientist identifies and collects the questions or clarification from the stakeholders which is required for analysis. In this stage data scientist involves asking more questions to stakeholders so that the AI project team can decide on the correct approach to solve the problem.

To solve a particular problem, there are four main types of data analytics.

1. Descriptive Analytics
2. Diagnostic Analytics
3. Predictive Analytics

4. Prescriptive Analytics

Descriptive Analytics: Descriptive analytics summarizes the past data to identify trends and patterns. Descriptive analytics use tools like graphs, charts, and statistical measures like mean, median, mode to understand the data. For example: To calculate the average marks of students in an exam or analyzing sales data from the previous year.

Diagnostic Analytics: Diagnostic analytics understand the reason behind why some things have happened. Diagnostic analytics analyze past data using techniques like root cause analysis, hypothesis testing, correlation analysis, etc. For example, if the sales of a company dropped, diagnostic analysis will help to find the cause for it by analyzing questions like "Is it due to poor customer service?" or "low product quality?"

Predictive Analytics: This analytics uses the past data to make predictions about future events or trends, using techniques like regression, classification, clustering, etc. The main purpose is to foresee future outcomes and make informed decisions. For example, a company can use predictive analytics to forecast its sales, demand, inventory, customer purchase patterns, etc., based on previous sales data.

Prescriptive Analytics: Prescriptive analytics is a data-driven approach in machine learning and statistical algorithms to recommend actions that can improve business outcomes. The techniques used in prescriptive analytics are optimization, simulation, decision analysis, etc. For example, to design the right strategy to increase the sales during festival season by analyzing past data and thus optimize pricing, marketing, production, etc.

We can summarize each of these analytics as given in Table

	Descriptive Analytics	**Diagnostic Analytics**	**Predictive Analytics**	**Prescriptive Analytics**
Focus	Questions on summarizing historical data	Questions on understanding why certain events occurred	Questions on predicting future outcomes based on historical data patterns	Questions on determining the best course of action
Purpose	Identify patterns, trends, and anomalies in past data	Uncover root causes and factors contributing to specific outcomes	Forecast future events or behaviors	Recommend specific actions or interventions based on predictive insights. May indirectly influence classification through recommendations

- **Selecting the right AI technique:**
- **Classification:** Used when predicting a category (e.g., spam or not spam).
- **Regression:** Used when predicting a numeric value (e.g., house price).
- **Clustering:** Grouping similar items (e.g., customer segments).
- **Anomaly Detection**: Spotting unusual events (e.g., fraud).
- **Recommendation Systems:** Suggesting items (e.g., movies, products).

3. Data requirements

In data requirements, the 5W1H questioning method is used to identify the data requirements and wants to find the purpose of data. Data requirements understand the steps involved in the processes that create, read, update, or delete data and determine the correct use of data.

Key Considerations:

What type of data do we need?

- **Numerical Data:** e.g., prices, age, temperature
- **Categorical Data:** e.g., gender, city, product type
- **Textual Data:** e.g., reviews, tweets, comments
- **Multimedia Data:** e.g., images, audio, video (for advanced AI tasks)

What format should the data be in?

- **CSV**: Simple and widely used for tabular data
- **JSON/XML**: Common for web data and APIs
- **Excel**: Often used for business reports
- **SQL Databases**: Used for structured enterprise data

From where will the data be sourced?

- Surveys
- Observations
- Sensors/IoT devices
- Interviews
- Government or public datasets (e.g., data.gov)
- Open data platforms (e.g., Kaggle, UCI ML Repository)
- Organizational databases

How much data is enough?

- Consider **sample size**, **data completeness**, and **distribution**.
- Assess whether the data will allow **reliable model training and validation**.

What about data quality?

- Assess whether the data is **accurate, up-to-date, unbiased**, and **relevant** to the problem.

Types of Data (based on structure):

Data Type	Description	Examples
Structured Data	Organized in rows and columns	Excel sheets, SQL databases
Semi-structured Data	Has tags or markers, but not fixed schemas	JSON, XML
Unstructured Data	No predefined structure	Images, videos, social media posts

<u>Determining the specific information needed for our analysis or project includes:</u>

- Identifying the types of data required, such as numbers, words, or images.
- Considering the structure in which the data should be organized, whether it is in a table, text file, or database.
- Identifying the sources from which we can collect the data, and
- Any necessary cleaning or organization steps required before beginning the analysis.

4. Data collection

Data collection is a process where the data is collected from different sources; it is a fundamental step in data science. Data requirements are decision-makers deciding whether the data collected from different sources requires more or less data. There are mainly two sources of data collection:

- **Primary data source:** Primary data is raw and unprocessed data that is collected from the original source, like direct observation, experimentation, surveys, interviews, or other methods.
- **Secondary data source:** Secondary data is ready-to-use data. Secondary data sources refer to the data that is already stored in different areas, like web scraping, databases, social media data, satellite data, etc.

Challenges in Data Collection:

Accessibility issues: Data behind paywalls or protected APIs

Privacy concerns: Especially for personal or sensitive data

Data inconsistency: Varying formats across sources

Incomplete or outdated data: May not reflect current realities

5. Data Understanding

Data understanding is a process where we want to understand if the collected data can solve the problem or not. We also want to check the relevance of the data and want to identify that the data can address the specific problem or question that is going to be evaluated.

Common Issues Discovered in This Phase:

- Missing or incomplete data
- Duplicate records
- Outliers that may skew the analysis
- Incorrect formats (e.g., dates stored as text)

6. Data preparation

This stage covers all the activities to build the set of data that will be used in the modelling step. Data is transformed into a state where it is easier to work with. Data preparation includes

- Cleaning of data (dealing with invalid or missing values, removal of duplicate values and assigning a suitable format)
- Combine data from multiple sources (archives, tables and platforms)
- Transform data into meaningful input variables

Feature Engineering:

Creating new variables from existing data that might be more meaningful.

Examples:

From “Date of Purchase” → extract “Day of Week”, “Month”, “Season”

From “Price” and “Area” → derive “Price per Square Foot”

Useful for improving model accuracy and performance.

7. AI modelling

AI modeling is a method of creating algorithms or models that can learn and make intelligent decisions without human intervention. The modeling stage uses the initial version of the dataset prepared and focuses on developing models according to the analytical approach previously defined. Data modeling focuses on developing models that are either descriptive or predictive.

Descriptive Modeling: It is a concept in data science and statistics that focuses on summarizing and understanding the characteristics of a dataset without making predictions or decisions. This includes summarizing the main characteristics, patterns, and trends that are present in the data.

Common Descriptive Techniques:

Summary Statistics: This includes measures like:

- Mean (average), Median, Mode
- Standard deviation, Variance
- Range (difference between the highest and lowest values)
- Percentiles (e.g., quartiles)

Visualizations: Graphs and charts to represent the data, such as:

- Bar charts
- Histograms
- Pie charts
- Box plots
- Scatter plots

Predictive modeling: It involves using data and statistical algorithms to identify patterns and trends in order to predict future outcomes or values. It relies on historical data and uses it to create a model that can predict future behavior or trends or forecast what might happen next. It involves techniques like regression, classification, and time-series forecasting, and can be applied in a variety of fields, from predicting exam scores to forecasting weather or stock prices.

Types of Modeling Approaches:

- **Supervised Learning**
  - The model is trained on a labeled dataset, meaning the input data is paired with the correct output.
  - Used for:
  - Classification: Predicting categories (e.g., spam vs. non-spam)
  - Regression: Predicting continuous values (e.g., house prices)
- **Unsupervised Learning**
  - The model finds hidden patterns or structures in unlabeled data.
  - Used for:
  - Clustering: Grouping similar items (e.g., customer segmentation)
  - Dimensionality Reduction: Simplifying datasets while retaining structure
- **Other Approaches**
  - Semi-supervised Learning: Combination of labeled and unlabeled data
  - Reinforcement Learning: Learning through trial and error (e.g., robotics, game AI)

8. Evaluation

Evaluation in an AI project cycle is the process of assessing how well a model performs after training. It involves using test data to measure metrics like accuracy, precision, recall, or F1 score. This helps determine if the model is reliable and effective before deploying it in real-world situations.

Model evaluation can have two main phases.

First phase – Diagnostic measures

It is used to ensure the model is working as intended. If the model is a predictive model, a decision tree can be used to evaluate the output of the model, check whether it is aligned to the initial design or requires any adjustments.

Second phase – Statistical significance test

This type of evaluation can be applied to the model to verify that it accurately processes and interprets the data. This is designed to avoid unnecessary second guessing when the answer is revealed.

9. Deployment

Deployment refers to the stage where the trained AI model is made available to the users in real-world applications. Once the model is evaluated and the data scientist is confident it will work, it is deployed and put to the ultimate test.

10. Feedback

The last stage in the data science methodology is feedback. Feedback from the users will help to refine the model and assess it for performance and impact. Feedback from users can be received in many ways.

Model Validation

Model validation is a process that evaluates the performance and reliability of a model. Model Validation offers a systematic approach to measure its accuracy and reliability, providing insights into how well it generalizes to new, unseen data. The benefits of Model Validation include

- Enhancing the model quality.
- Reduced risk of errors
- Prevents the model from overfitting and underfitting.

Model Validation Techniques

The commonly used Validation techniques are Train-test split, K-Fold Cross Validation, Leave One out Cross Validation, Time Series Cross Validation etc. Let's discuss the Train test split and K-Fold Cross Validation.

Train Test Split

The train-test split is a technique for evaluating the performance of a machine learning algorithm. It can be used for classification or regression problems and can be used for any supervised learning algorithm. The procedure involves taking a dataset and dividing it into two subsets, The first subset is used to train the model and is referred to as the training dataset. The second subset is used to test the model.

- **Train Dataset:** Used to fit the machine learning model.
- **Test Dataset:** Used to evaluate the fit machine learning model.

How to Configure the Train-Test Split

The parameter is used for the size of the train and test datasets, normally represented as percentages. For example, if 67% of data is allocated for training, then 33% is reserved for testing. The training and testing split depends on the project goal.

Common split percentages include:

● Train: 80%, Test: 20%

- Train: 70%, Test: 30%
- Train: 67%, Test: 33%

K-Fold Cross Validation

K-Fold cross-validation is a technique that splits a dataset into subsets, or folds, to evaluate the model's performance. For example, suppose you have 100 data points you want to evaluate using K-Fold cross-validation.

- **Step 1:** Divide the 100 data points into 5 equal parts (folds), each containing 20 data points.
- **Step 2:** Use the 1st fold as the test set and the remaining 4 folds as the training set.
- **Step 3:** Use the 2nd fold as the test set, and the remaining 4 will be the training set.
- **Step 4:** Continue the above steps until each fold has been used as the test set once.

Use the performance metric like accuracy and F1 score to find the final average of these metrics to get the overall model performance.

Difference between Train-Test Split and Cross Validation

Train-Test Split	**Cross Validation**
Normally applied on large datasets	Normally applied on small datasets
Divides the data into training data set and testing dataset.	Divides a dataset into subsets (folds), trains the model on some folds, and evaluates its performance on the remaining data.
Clear demarcation on training data and testing data.	Every data point at some stage could be in either testing or training data set.

MODEL PERFORMANCE – EVALUATION METRICS

Evaluation metrics are used to check the performance and effectiveness of the machine learning model. Evaluation metrics help to compare different models to identify the best-performing one for a specific task. The evaluation matrix is categorized into classification problems and regression problems.

- **Classification Problems:** The target variable is divided into distinct classes. Metrics include accuracy, precision, recall, F1-score, and AUC-ROC.
- **Regression Problems:** The target variable is continuous. Metrics include mean squared error (MSE), mean absolute error (MAE), and R-squared.

Evaluation Metrics for Classification

1. **Confusion Matrix**

A Confusion Matrix is used to evaluate the performance of a classification model. It summarizes the predictions against the actual outcomes. It creates an N X N matrix, where N is the number of classes or categories that are to be predicted. Suppose there is a problem, which is a binary classification, then N=2 (Yes/No). It will create a 2×2 matrix.

- **True Positives:** It is the case where the model predicted Yes and the real output was also yes.
- **True Negatives:** It is the case where the model predicted No and the real output was also No.
- **False Positives:** It is the case where the model predicted Yes but it was actually No.
- **False Negatives:** It is the case where the model predicted No but it was actually Yes.

The Confusion Matrix		Reality	
		Yes	No
Prediction	Yes	True Positive (TP)	False Positive (FP)
	No	False Negative (FN)	True Negative (TN)

2. **Precision and Recall**

   Precision measures "What proportion of predicted Positives is truly Positive?"

   Precision = (TP)/(TP+FP).

   Precision should be as high as possible.

   Recall measures "What proportion of actual Positives is correctly classified?"

   Recall = (TP)/(TP+FN)

3. **F1-score**

   A good F1 score means that you have low false positives and low false negatives, so you're correctly identifying real threats, and you are not disturbed by false alarms. An F1 score is considered perfect when it is 1, while the model is a total failure when it is 0.

   F1 = 2* (precision * recall)/(precision + recall)

4. **Accuracy**

   Accuracy = Number of correct predictions / Total number of predictions

   Accuracy = (TP+TN)/(TP+FP+FN+TN)

Evaluation Metrics for Regression

1. **MAE (Mean Absolute Error)**

   Mean Absolute Error is a sum of the absolute differences between predictions and actual values. A value of 0 indicates no error or perfect predictions

2. **MSE (Mean Square Error)**

   Mean Square Error (MSE) is the most commonly used metric to evaluate the performance of a regression model. MSE is the mean(average) of squared distances between our target variable and predicted values.

$$MSE = \frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}$$

3. **RMSE (Root Mean Square Error)**

   Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). RMSE is often preferred over MSE because it is easier to interpret since it is in the same units as the target variable.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

QUESTION BANK:

Multiple Choice Questions

1. What is the first step in Data Science Methodology?
   a. Data Understanding b. Business Understanding
   c. Data Collection d. Evaluation
2. Which of the following techniques helps to determine why an event happened?
   a. Descriptive Analytics b. Predictive Analytics
   c. Diagnostic Analytics d. Prescriptive Analytics
3. Which is NOT a type of Data Analytics?
   a. Descriptive Analytics b. Diagnostic Analytics
   c. Cognitive Analytics d. Predictive Analytics
4. What is used in Descriptive Analytics to summarize past data?
   a. Decision Trees b. Root Cause Analysis
   c. Summary Statistics d. Regression
5. Which of the following is NOT a characteristic of predictive analytics?
   a. Forecasting future events b. Using historical data
   c. Recommending actions based on data d. Using techniques like regression
6. In which step of the Data Science Methodology is data cleaned and transformed for analysis?
   a. Data Collection b. Data Preparation
   c. Data Understanding d. Evaluation
7. Which of the following is a key component of Feature Engineering?
   a. Splitting data into training and testing sets
   b. Creating new features to improve model performance
   c. Identifying data sources
   d. Evaluating model performance
8. Which data type does NOT follow a predefined structure?
   a. Structured data b. Semi-structured data
   c. Unstructured data d. Categorical data
9. Which of the following is an example of a primary data source?
   a. Books b. Social media posts
   c. Surveys d. Online databases
10. Which technique is used to assess the model's performance by splitting the data into two sets?
    a. K-Fold Cross Validation b. Train-Test Split
    c. Data Understanding d. Data Collection
11. Which approach helps to identify patterns based on historical data and is used for forecasting?
    a. Predictive Analytics b. Descriptive Analytics
    c. Diagnostic Analytics d. Prescriptive Analytics
12. In which phase do data scientists evaluate whether the data collected is relevant to the problem being solved?
    a. Data Collection b. Data Preparation
    c. Data Understanding d. Business Understanding
13. What is a Confusion Matrix used for in machine learning?
    a. To display the distribution of data
    b. To evaluate the performance of a classification model

c. To split data into training and testing sets
d. To train machine learning models

14. Which of the following metrics is NOT used in evaluating classification models?
    a. Precision
    b. Recall
    c. F1-Score
    d. Mean Squared Error
15. Which type of analysis helps to identify the best course of action to achieve a desired outcome?
    a. Predictive Analytics
    b. Descriptive Analytics
    c. Diagnostic Analytics
    d. Prescriptive Analytics
16. Which stage in the Data Science Methodology involves gathering raw data from various sources?
    a. Data Preparation
    b. Data Collection
    c. Data Understanding
    d. Model Deployment
17. In the Data Science Methodology, what is the purpose of the 'Evaluation' stage?
    a. To analyze data quality
    b. To assess the model's performance
    c. To clean and preprocess the data
    d. To define the business problem
18. Which metric is used to measure the proportion of correctly predicted positive instances in a classification model?
    a. Recall
    b. Precision
    c. F1-Score
    d. Accuracy
19. Which type of model focuses on understanding relationships within data without predicting future outcomes?
    a. Predictive Modeling
    b. Descriptive Modeling
    c. Classification Modeling
    d. Regression Modeling
20. In K-Fold Cross Validation, what does the 'k' represent?
    a. The number of data features
    b. The number of algorithms used
    c. The number of data splits
    d. The number of training epochs

Long Answer Questions

1. What is Data Science Methodology, and why is it important?
2. Explain the steps involved in Data Science Methodology.
3. What are the different types of analytics, and how do they differ from each other?
4. What is Feature Engineering, and why is it crucial for machine learning models?
5. Describe the differences between structured, semi-structured, and unstructured data. Provide examples of each.
6. What is the importance of the "Evaluation" phase in the Data Science Methodology?
7. Explain the concept of K-Fold Cross-Validation and its advantages over simple train-test split.
8. What are the different evaluation metrics used for classification models?
9. What is the role of "Business Understanding" in Data Science?
10. Explain the differences between "Descriptive Modeling" and "Predictive Modeling."
11. How does Data Preparation impact the performance of machine learning models?
12. What are some common challenges faced in the "Data Collection" phase of Data Science?
13. Why is "Model Validation" important, and what techniques are commonly used?
14. What are the key considerations when deploying a machine learning model in a real-world scenario?
15. What is "Feedback" in Data Science Methodology, and why is it important?

Part B – Answers

Multiple Choice Answers

1. **b. Business Understanding**
2. **c. Diagnostic Analytics**
3. **c. Cognitive Analytics**
4. **c. Summary Statistics**
5. **c. Recommending actions based on data**
6. **b. Data Preparation**
7. **b. Creating new features to improve model performance**
8. **c. Unstructured data**
9. **c. Surveys**
10. **b. Train-Test Split**
11. **a. Predictive Analytics**
12. **c. Data Understanding**
13. **b. To evaluate the performance of a classification model**
14. **d. Mean Squared Error**
15. **d. Prescriptive Analytics**
16. **b. Data Collection**
17. **b. To assess the model's performance**
18. **b. Precision**
19. **b. Descriptive Modeling**
20. **c. The number of data splits**

LONG ANSWERED QUESTIONS WITH ANSWER:

1. What is Data Science Methodology and why is it important?

Answer:

Data Science Methodology is a systematic approach used by data scientists to tackle complex data-related problems and derive actionable insights. It provides a structured framework for addressing business or research problems, from defining the problem to deploying and maintaining machine learning models. The importance of Data Science Methodology lies in its ability to organize the entire data science process, ensuring that each stage is tackled systematically, reducing inefficiencies and risks of errors. It helps in identifying the correct problem to solve, gathering the appropriate data, selecting the right algorithms, and evaluating the performance of the model to ensure the solution is effective.

2. Explain the steps involved in Data Science Methodology.

Answer:

The Data Science Methodology consists of several iterative steps that guide data scientists through the process of solving a data problem. These steps are:

Business Understanding: This is the first step where the data scientist understands the business problem that needs to be solved. This involves engaging with stakeholders to identify the business objectives and constraints.

Analytic Approach: In this step, the data scientist formulates an analytical strategy to solve the problem. Questions like "What type of data do I need?" and "What approach should I use?" are considered.

Data Requirements: This stage identifies the data needed for analysis, including the type of data, format, and sources from which it can be collected.

Data Collection: Data is gathered from various primary or secondary sources. This could include surveys, databases, websites, or sensors.

Data Understanding: The collected data is explored and analyzed to understand its quality, structure, and relevance to the problem.

Data Preparation: This step involves cleaning and transforming the data into a format suitable for analysis, including handling missing values, duplicates, and feature engineering.

Modelling: Machine learning models are built based on the prepared data, using algorithms such as regression, classification, clustering, etc.

Evaluation: The performance of the model is evaluated using metrics like accuracy, precision, recall, or F1-score. This ensures the model meets business objectives.

Deployment: The model is deployed into real-world applications or business processes, where it is used to make predictions or decisions.

Feedback: Post-deployment, feedback is gathered from users and the model's performance is monitored for improvements or adjustments.

3. What are the different types of analytics, and how do they differ from each other?

Answer:

There are four main types of analytics:

Descriptive Analytics:

This type of analytics focuses on summarizing historical data to understand what happened in the past. It uses statistical methods and visualization tools (e.g., bar charts, histograms) to identify trends and patterns. Example: Analyzing last quarter's sales data to understand the overall performance.

Diagnostic Analytics:

Diagnostic analytics seeks to understand why something happened. It typically involves methods like root cause analysis, hypothesis testing, and correlation analysis to uncover the factors behind certain events. Example: Investigating why sales dropped in a particular region by examining customer feedback, marketing campaigns, and competitor activity.

Predictive Analytics:

Predictive analytics uses historical data and statistical models to predict future outcomes. It helps businesses anticipate future trends or events. Example: Forecasting future sales based on historical data and seasonal trends.

Prescriptive Analytics:

This type of analytics provides recommendations for the best course of action to achieve a desired outcome. It uses optimization and simulation techniques to suggest decision-making strategies. Example: Recommending the optimal price for a product during a holiday sale to maximize revenue.

4. What is Feature Engineering and why is it crucial for machine learning models?

Answer:

Feature Engineering is the process of creating new features (variables) from raw data to improve the performance of machine learning models. This process involves selecting, modifying, or generating new features based on domain knowledge or data insights.

For example, in predicting house prices, features such as the "age of the house" or "price per square foot" can be derived from existing data, like the year built and square footage.

Feature Engineering is crucial because it directly impacts the model's accuracy. By transforming raw data into meaningful input variables, data scientists can enhance the model's ability to capture complex

relationships in the data. Additionally, well-engineered features can help in reducing noise and improving model interpretability.

5. Describe the differences between structured, semi-structured, and unstructured data. Provide examples of each.

Answer:

Structured Data:

This type of data is highly organized and fits into predefined models or tables, typically in relational databases. Structured data can easily be stored, queried, and analyzed. Examples include Excel sheets, SQL databases, and transaction records.

Semi-structured Data:

Semi-structured data has some level of organization but does not fit into a strict schema. It might contain tags or markers that separate different pieces of data but lacks a fixed structure. Examples include JSON files, XML files, and email content.

Unstructured Data:

Unstructured data does not have a predefined format or structure, making it more difficult to analyze. It can include text, images, videos, and audio files. Examples include social media posts, images, audio files, and web pages.

6. What is the importance of the "Evaluation" phase in the Data Science Methodology?

Answer:

The "Evaluation" phase is crucial as it allows data scientists to assess whether the model developed during the "Modeling" phase effectively solves the business problem. It involves testing the model on a separate dataset (usually the test set) to evaluate its performance.

During this phase, key performance metrics such as accuracy, precision, recall, F1-score, and AUC (Area Under the Curve) are calculated, depending on the type of model and the problem being solved. This helps in determining if the model meets the requirements, and if not, adjustments are made to improve it.

Without proper evaluation, there is a risk of deploying a model that performs poorly or fails to generalize to unseen data.

7. Explain the concept of K-Fold Cross-Validation and its advantages over simple train-test split.

Answer:

K-Fold Cross-Validation is a technique used to assess the performance of machine learning models by splitting the dataset into 'k' equal parts or folds. The model is trained on 'k-1' folds and tested on the remaining fold. This process is repeated 'k' times, with each fold serving as the test set once.

Advantages:

More reliable estimate: Since every data point gets used for both training and testing, the model's performance is evaluated more thoroughly, leading to a more accurate estimate of its generalization ability.

Reduces bias: Using multiple folds reduces the likelihood of over fitting or bias caused by a single train-test split.

Works well with smaller datasets: Cross-validation allows efficient use of limited data, as every observation is used for both training and testing.

8. What are the different evaluation metrics used for classification models?

Answer:

The main evaluation metrics for classification models are:

Accuracy:

The proportion of correct predictions out of all predictions.

Formula: (TP + TN) / (TP + TN + FP + FN).

Precision:

The proportion of true positive predictions out of all positive predictions made by the model.

Formula: TP / (TP + FP).

Recall (Sensitivity):

The proportion of true positive predictions out of all actual positives.

Formula: TP / (TP + FN).

F1-Score:

The harmonic mean of precision and recall, providing a balance between the two metrics.

Formula: 2 * (Precision * Recall) / (Precision + Recall).

Confusion Matrix:

A table that shows the actual vs. predicted values for each class, helping to visualize the performance of the classification model.

9. What is the role of "Business Understanding" in Data Science?

Answer:

"Business Understanding" is the first and most important step in the Data Science Methodology. This phase focuses on clearly defining the business problem that needs to be addressed and understanding the goals of the project.

The goal is to bridge the gap between the business objectives and the data science solution. By understanding the context, scope, and constraints of the problem, data scientists can formulate a relevant analytical approach, determine the type of data needed, and align the project's goals with the business outcomes.

Inadequate business understanding may result in wasted resources and the development of models that do not address the key issues or objectives.

10. Explain the differences between "Descriptive Modelling" and "Predictive Modelling."

Answer:

Descriptive Modelling:

Descriptive modelling aims to understand the characteristics or structure of data without making predictions. It focuses on summarizing and describing historical data. Examples include clustering and association analysis, where patterns in the data are described rather than predicted.

Example: Segmenting customers into groups based on purchasing behaviour.

Predictive Modelling:

Predictive modelling uses historical data to forecast future outcomes. It involves using algorithms like regression, classification, or time-series forecasting to predict future values or behaviours.

Example: Predicting whether a customer will churn based on their purchase history.

11. How does Data Preparation impact the performance of machine learning models?

Answer: Data Preparation is one of the most critical and time-consuming steps in Data Science. It directly affects the model's performance, as the quality of the input data determines how well the machine learning model can learn from it.

Key activities in Data Preparation include:

Handling missing data – Missing values are imputed or removed to ensure the model receives complete and accurate information.

Feature scaling – Normalizing or standardizing numerical features ensures that all variables contribute equally to the model.

Encoding categorical variables – Converting categorical data into a format suitable for algorithms (e.g., one-hot encoding or label encoding).

Outlier detection – Identifying and dealing with outliers that might distort model performance.

A well-prepared dataset enables the model to learn more effectively, resulting in better accuracy and generalizability.

12. What are some common challenges faced in the "Data Collection" phase of Data Science?

Answer:

The "Data Collection" phase can present several challenges, including:

Data Accessibility: Not all relevant data may be readily available or easily accessible. Some data might be locked behind paywalls or subject to privacy concerns.

Data Quality: Collected data may contain errors, inconsistencies, or missing values, requiring significant cleaning and preprocessing.

Data Privacy: Ensuring that data collection complies with regulations like GDPR or HIPAA, especially when dealing with sensitive information.

Data Volume: Large datasets can be challenging to store, process, and analyze. They may also require specialized tools and infrastructure.

Data Relevance: Collecting the right data that aligns with the business problem is crucial. Irrelevant data can introduce noise and hinder model performance.

13. Why is "Model Validation" important, and what techniques are commonly used?

Answer:

Model Validation is crucial to ensure that the machine learning model performs well not just on the training data but also on unseen data (i.e., it generalizes well). Without proper validation, models may overfit to training data and perform poorly on real-world data.

Common validation techniques include:

Train-Test Split: Dividing the data into two sets—one for training the model and one for testing its performance.

K-Fold Cross-Validation: Dividing the dataset into 'k' parts, training the model on 'k-1' folds, and testing on the remaining fold. This process is repeated 'k' times to get multiple performance metrics.

Leave-One-Out Cross-Validation (LOOCV): A special case of cross-validation where each data point is used for testing once while training the model on the rest of the data.

These methods help ensure that the model performs robustly on new data and is not overfitting.

14. What are the key considerations when deploying a machine learning model in a real-world scenario?

Answer:

When deploying a machine learning model, several key considerations must be taken into account:

Scalability: The model should be able to handle large amounts of data and scale as required by the business.

Integration: The model must be integrated with existing business processes or systems, ensuring that predictions or decisions are incorporated into workflows.

Performance Monitoring: Post-deployment, continuous monitoring of the model's performance is essential to ensure it provides accurate results and meets business needs.

Maintenance: The model may require periodic updates or retraining as new data is collected or business conditions change.

Feedback Loops: Incorporating feedback from users and stakeholders allows for continuous improvement and adjustments to the model.

15. What is "Feedback" in Data Science Methodology, and why is it important?

Answer:

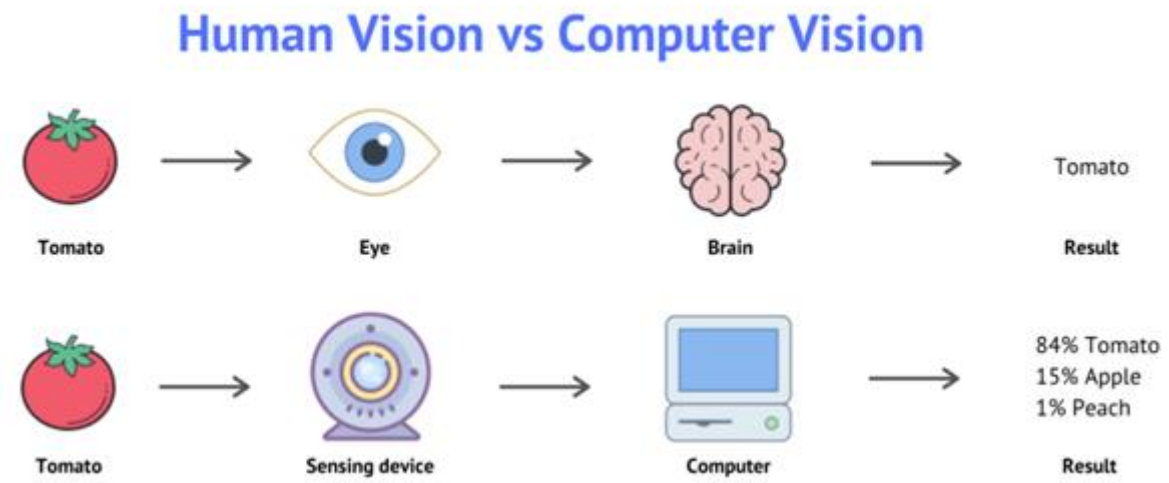
Feedback in Data Science refers to the process of collecting insights from users, stakeholders, and system performance after the deployment of the model. Feedback helps identify any shortcomings in the model and provides valuable data for improving it.

The importance of feedback lies in its role in model refinement. By continuously receiving and acting on feedback, data scientists can fine-tune models to ensure they remain relevant and accurate over time, addressing any shifts in data or business objectives.

UNIT 3 :Making Machines See

1. Introduction to Computer Vision (CV)

- **Definition**: A branch of Artificial Intelligence (AI) that trains machines to see, interpret, and understand the visual world through **digital images** and **videos**.
- **Human Vision Analogy**:
  - Human eyes → **Cameras (Sensing Devices)**
  - Optic nerve → **Data transmission**
  - Visual cortex → **Algorithms & AI models**
- CV derives meaningful information from digital images, videos and other visual input and makes recommendations or takes actions accordingly.



- **Purpose**: Derive meaningful information from visual input, make decisions, and automate visual tasks.
- **Core Technologies**:
  - **Image Recognition** – Identifying objects/scenes.
  - **Neural Networks** – Especially **CNNs** (Convolutional Neural Networks).
  - **Feature Extraction** – Shapes, colors, textures, edges.
  - **Model Training & Evaluation** – Learning patterns, measuring accuracy.

WORKING OF COMPUTER VISION

- **Definition**
  - A digital image is a picture stored on a computer as a sequence of numbers that computers can understand.
- Ways to Create Digital Images
  - **Design software →** e.g., Paint, Photoshop.
  - **Digital cameras →** capturing photos directly.
  - **Scanners →** converting printed images into digital form.

Representation of Digital Images

1. **Pixels (Picture Elements)**
   - Smallest unit of an image.
   - Each pixel represents a **specific color value**.

- Collection of pixels → forms the complete image.

2. **Digitization Process**
- Converts an image into a **grid of pixels**.
- Each pixel is assigned a **numerical value**.

3. **Resolution**
- Determined by the **number of pixels** in the image.
- **Higher resolution** → more detail & closer to the original scene.

4. **Pixel Values**
- **Monochrome images** (Black & White):
  - Value **0 = Black**
  - Value **255 = White**
  - Range: **0–255**

Binary Art: Recreating Images with 0s and 1s

Steps of Digital Image Processing

Step 1: Choose an Image

- Select any image to work with.
- Free sources: **Pixabay, Unsplash, Pexels**.

Step 2: Resize the Image

- Resize to smaller dimensions (**200–300 px** recommended).
- Tool: **https://imageresizer.com/**
- Save resized image.

Step 3: Convert to Grayscale

- Convert image into **shades of gray** (1 channel).
- Tool: **https://pinetools.com/grayscale-image**
- Download grayscale version.

Step 4: Extract Pixel Values

- Convert grayscale image into **numerical pixel values** (e.g., 0 for black, 1 for white).
- Tool: **Boxentriq Pixel Value Extractor**.

Step 5: Copy Pixel Values

- Copy the extracted values from the tool.

Step 6: Paste into Word Document

- Paste values into **MS Word / Google Docs**.

Step 7: Adjust Font Size

- Select all values → set font size to **1**.
- 0s and 1s will **recreate the grayscale image**.

Representation of Coloured Images (RGB Model)

1. **RGB Model**

- Each pixel represented by **Red, Green, Blue channels**.
- Each channel: value **0–255**.

2. **Binary Representation**
   - **1 byte = 8 bits** → possible values = 28=2562^8 = 256.
   - **0 = 00000000 (all bits 0)**
   - **255 = 11111111 (all bits 1)**
3. **Colour Range**
   - By combining intensities of R, G, B → over **16 million colours** possible.

2. Stages of the Computer Vision Process

1. Image Acquisition

Definition

- **Image acquisition** = Initial stage of computer vision.
- Involves **capturing digital images/videos** → provides **raw data** for further analysis.

Methods of Image Acquisition

1. **Digital Cameras** – Directly capture images/videos.
2. **Scanners** – Convert physical photos/documents into digital form.
3. **Design Software** – Create images artificially (e.g., Photoshop, Paint).
4. **Specialized Imaging (Scientific/Medical)**
   - **MRI (Magnetic Resonance Imaging)**
   - **CT (Computed Tomography)**

Factors Affecting Image Quality

1. **Device Capabilities** – Higher resolution devices capture more details.
2. **Lighting Conditions** – Poor lighting → low-quality images.
3. **Angle of Capture** – Different angles → different perspectives.

Applications

- **Medical field** → MRI/CT scans for diagnosis & treatment.
- **Scientific research** → Detailed internal imaging of tissues/structures.
- **General use** → Photos, documents, digital media.

2. Preprocessing

Definition

- **Preprocessing** = Enhancing the quality of an acquired image before analysis.
- Goal:
  - Remove unwanted disturbances (noise).
  - Highlight key features.
  - Ensure consistency across dataset.

Common Preprocessing Techniques

a) Noise Reduction

- Removes unwanted elements (blurriness, spots, distortions).
- Makes images clearer for analysis.

- **Example:** Removing grainy effect in low-light photos.

b) Image Normalization

- Standardizes pixel values to a fixed range (e.g., **0–1** or **–1 to 1**).
- Ensures uniform scaling → improves learning of AI models.
- **Example:** Scaling pixel values from 0–255 → 0–1.

c) Resizing / Cropping

- Adjusts size/aspect ratio of images.
- Ensures all images have the **same dimensions** for analysis.
- **Example:** Resizing all images to **224×224 pixels** before neural network input.

d) Histogram Equalization

- Enhances **brightness & contrast** of an image.
- Redistributes pixel intensity → reveals hidden details.
- **Example:** Making a low-contrast photo appear sharper.

Main Goals of Preprocessing

✓ Remove noise (unwanted disturbances).

✓ Highlight important features.

✓ Ensure **consistency & uniformity** across dataset.

3. Feature Extraction

Definition

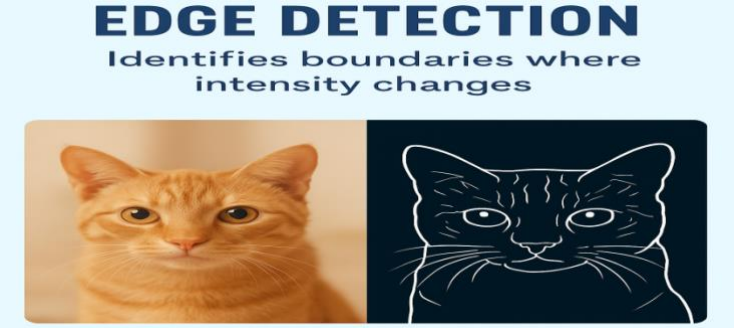
- **Feature extraction** = Identifying and extracting **important visual patterns/attributes** from a pre-processed image.
- Helps in representing images with **meaningful information** instead of raw pixel values.

Factors Affecting Feature Extraction

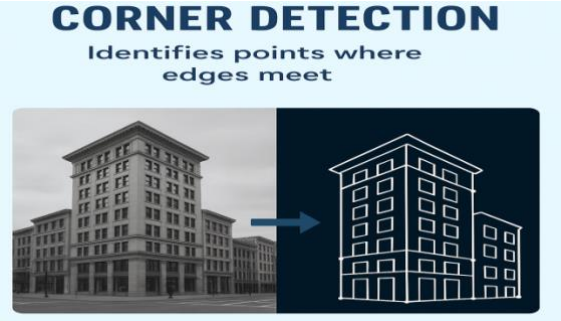
- **Complexity of image** (simple vs detailed).
- **Computational resources** (speed, memory).
- **Application requirements** (e.g., face recognition vs medical imaging).

Common Feature Extraction Techniques

1. **Edge Detection**
   - Identifies **boundaries** between regions with significant change in intensity.
   - Useful for outlining shapes/objects.



2. **Corner Detection**
   - Finds points where **two or more edges meet**.
   - Focuses on **sharp changes in gradients** → corresponds to **corners/junctions** in objects.



3. **Texture Analysis**
   - Extracts features like **smoothness, roughness, repetition, regularity**.
   - Helps in classifying surfaces (e.g., grass vs water).

4. **Colour-based Feature Extraction**
   - Measures **colour distribution** within an image.
   - Helps discriminate objects/regions based on **colour characteristics**.

4. Detection & Segmentation

Definition

- **Detection & segmentation** = Tasks in computer vision to **identify and isolate objects/regions of interest** in an image.
- Widely used in **autonomous driving, medical imaging, object tracking**.

Categories of Tasks

1. **Single Object Tasks**

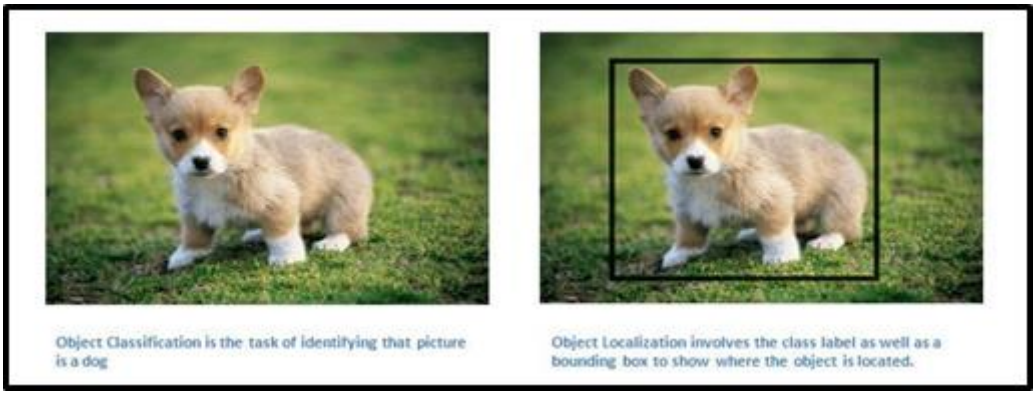
Focus on analyzing one object in the image.

i) Classification

- Determines the **category/class** of a single object.
- **Algorithms:**
  - **Supervised:** KNN (K-Nearest Neighbor).
  - **Unsupervised:** K-means clustering.

ii) Classification + Localization

- Classifies the object **AND** predicts its location with a **bounding box**.



2. Multiple Object Tasks

Handles images containing **many objects or classes**.

i) Object Detection

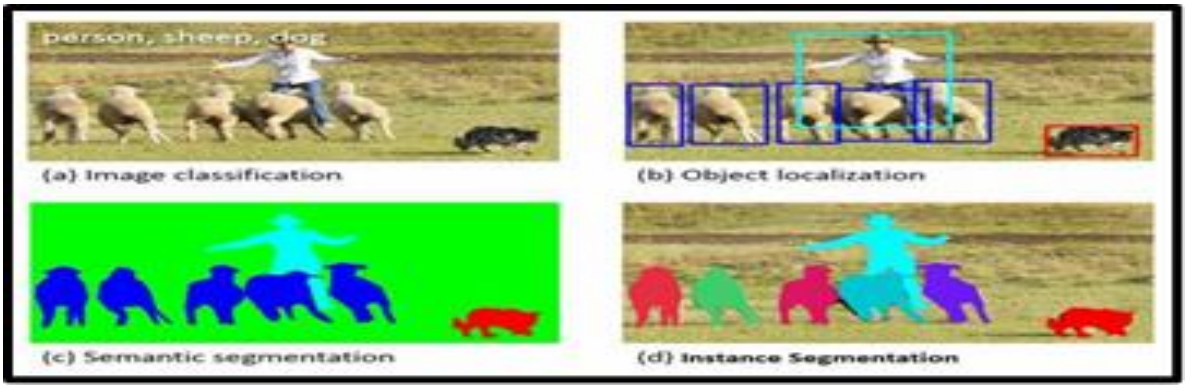
- Identifies &**locates multiple objects** in an image.
- Draws **bounding boxes** + assigns **class labels**.
- **Difference from classification:**
  - *Classification:* classifies the **whole image**.
  - *Detection:* finds and classifies **all objects** inside image.
- **Popular Algorithms:**
  - **R-CNN (Region-based CNN)**
  - **R-FCN (Region-based Fully Convolutional Network)**
  - **YOLO (You Only Look Once)**
  - **SSD (Single Shot Detector)**

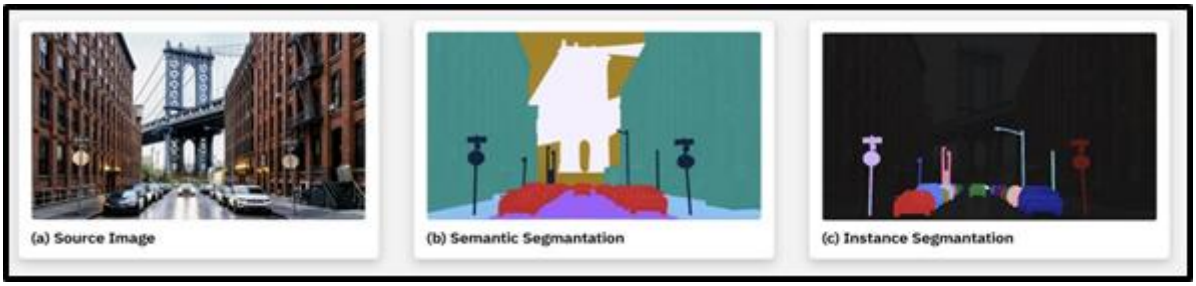
ii) Image Segmentation

- Groups **pixels with similar characteristics** → assigns a class to each pixel.
- Creates a **mask** for each object = pixel-wise identification.
- **Techniques Used:** Edge Detection, Brightness discontinuities, etc.

Types of Image Segmentation:

- **a) Semantic Segmentation**
  - Classifies pixels into classes.
  - **Objects of same class are NOT differentiated.**
  - Example: All animals in an image labelled as "animal."
- **b) Instance Segmentation**
  - Classifies pixels into classes **and differentiates each object instance**.
  - Example: Two dogs in same image are labelled separately.





5. High-Level Processing

Definition

- **Final stage** of computer vision.
- Focuses on **interpreting & extracting meaningful information** from detected objects/regions.
- Provides a **deeper understanding** of visual content → supports decision-making.

Key Tasks in High-Level Processing

1. **Object Recognition** – Identifying & categorizing objects.
2. **Scene Understanding** – Analysing overall scene (context, environment).
3. **Context Analysis** – Inferring relationships between elements in a scene.
4. **Insight Extraction** – Deriving useful knowledge from complex visual data.

Techniques Used

- **Machine Learning (ML)**&**Deep Learning** algorithms.
- Sophisticated models for **classification, reasoning, and inference**.

Applications

- **Autonomous Driving** – Detecting traffic signs, pedestrians, vehicles.
- **Medical Diagnostics** – Identifying diseases from X-rays, MRIs, CT scans.
- **Surveillance & Security** – Recognizing suspicious activity.
- **Smart Devices** – Facial recognition, gesture control.

3. Applications of Computer Vision

Application	Description	Example/Use Case
Facial Recognition	Detects and identifies human faces.	Face book tagging, phone face unlock
Healthcare	Identifies diseases, tumors, abnormalities through medical imaging.	Cancer detection via MRI/CT scans
Self-Driving Vehicles	Analyzes surroundings, detects objects, reads traffic signals.	Tesla Autopilot, Waymo cars
Optical Character Recognition (OCR)	Extracts text (printed/handwritten) from images/documents.	Scanned bills, invoices, e-documents
Machine Inspection	Detects product defects, flaws, and ensures quality control in manufacturing.	Automated assembly line checks

Application	Description	Example/Use Case
3D Model Building	Creates 3D models from real objects.	AR/VR, robotics, 3D scene reconstruction
Surveillance	Monitors CCTV footage, detects suspicious behavior, prevents crime.	Smart city security cameras
Fingerprint & Biometrics	Identifies/validates user through biometric traits.	Aadhaar authentication, smartphone unlock

4. Challenges in Computer Vision

1. Reasoning & Analytical Issues

- CV requires not only **image identification** but also **accurate interpretation**.
- Systems must define attributes within visual content.
- Without robust **reasoning skills**, extracting meaningful insights becomes difficult → limits system effectiveness.

2. Difficulty in Image Acquisition

- Image quality is affected by:
  - **Lighting variations**
  - **Different perspectives & scales**
  - **Occlusion** (when objects overlap/cover each other)
- Complex scenes with multiple objects are harder to analyze.
- High-quality data is crucial for accurate results.

3. Privacy & Security Concerns

- **Surveillance systems** (e.g., facial recognition) raise ethical issues.
- Possible **infringement of privacy rights**.
- Requires careful handling due to **regulatory scrutiny** and **public debate**.

4. Duplicate & False Content

- Vulnerabilities in CV algorithms may lead to:
  - **Fake or manipulated content** (deepfakes, frauds).
  - **Duplicate data** → misinformation.
- **Data breaches** can worsen the issue, damaging trust & reputation.

5. Future Trends

- **Improved Accuracy** – Enhanced recognition in complex scenes.
- **Real-Time Processing** – Faster with minimal latency for live use.
- **Integration with AI** – Combining CV with NLP, robotics, IoT.
- **Expanded Applications** – Healthcare, agriculture, environmental monitoring, AR/VR.

Computer Vision – Question Bank (MCQ, True/False & Assertion-Reason)

Part A: Multiple Choice Questions (1–25)

1. The field of study that helps to develop techniques to help computers “see” is _____.

a) Python b) Convolution

c) Computer Vision d) Data Analysis

2. A digital image is stored on a computer in the form of _____.

a) Pictures b) Sequence of numbers

c) Audio waves d) Pixels only

3. Task of taking an input image and outputting a class label that best describes the image is _____.

a) Image classification b) Image localization

c) Image identification d) Image prioritization

4. Which of the following is not a way to create digital images?

a) Using design software b) Capturing with a camera

c) Using a scanner d) Speaking into a microphone

5. 1 byte = _____ bits.

a) 10 b) 8 c) 2 d) 1

6. A computer sees an image as a series of _____.

a) Colours b) Pixels c) Objects d) All of the above

7. Which process is related to capturing images using a camera, scanner, or other devices?

a) Image Acquisition b) Preprocessing

c) Feature Extraction d) Detection

8. Which algorithm may be used for supervised learning in Computer Vision?

a) KNN b) K-means c) K-fold d) KEAM

9. Choose the incorrect statement related to preprocessing stage:

a) It enhances the quality of acquired image.

b) Noise reduction and Image normalization are applied.

c) Histogram equalization adjusts brightness/contrast.

d) Edge and corner detection are preprocessing techniques.

10. Which stage standardizes pixel values across images for consistency?

a) Noise Reduction b) Image Normalization

c) Resizing d) Histogram Equalization

11. High-level processing in computer vision involves:

a) Object recognition and scene understanding b) Capturing the raw image

c) Removing noise d) Extracting pixel values

12. Which feature extraction technique identifies boundaries where pixel intensity changes sharply?

a) Edge Detection b) Corner Detection

c) Texture Analysis d) Colour Extraction

13. Which type of segmentation differentiates between objects of the same class?

a) Semantic Segmentation b) Instance Segmentation

c) Edge-based Segmentation d) Texture-based Segmentation

14. The main difference between classification and detection is:

a) Classification works only on grayscale images

b) Detection only identifies colours

c) Classification labels the whole image, detection finds multiple objects

d) Detection requires no algorithm

15. RGB colour model can represent over _____ possible colours.

a) 16 million b) 256 c) 255 d) 65,000

16. Which preprocessing method improves visibility in low-contrast images?

a) Noise Reduction b) Histogram Equalization

c) Normalization d) Cropping

17. Object detection algorithms such as YOLO, SSD, and R-CNN are used for:
    a) Converting text to speech
    b) Identifying and locating objects with bounding boxes
    c) Translating languages d) Standardizing datasets
18. In grayscale images, pixel values range from:
    a) 1–255 b) 0–100 c) 0–255 d) 1–1000
19. The main concern with duplicate and false content in CV is:
    a) Reduced resolution
    b) Generation of deepfakes and misinformation
    c) High cost of training
    d) Low storage requirements
20. MRI and CT scans are examples of _____.
    a) Image Acquisition techniques b) High-level processing
    c) Image Normalization d) Texture Analysis
21. A student resizes all images in a dataset to 224×224 before training a neural network. Which preprocessing method is this?
    a) Normalization b) Resizing/Cropping
    c) Noise Reduction d) Segmentation
22. If two dogs in the same image are labelled separately though both belong to “dog” class, which segmentation is applied?
    a) Semantic Segmentation b) Instance Segmentation
    c) Edge Detection d) Classification
23. A self-driving car identifying pedestrians, road signs, and vehicles simultaneously is performing:
    a) Classification only b) Object Detection
    c) Image Normalization d) Histogram Equalization
24. During a security event, a person with a red scarf is identified by the system. Which feature extraction technique was used?
    a) Edge Detection b) Texture Analysis
    c) Corner Detection d) Colour-based Extraction
25. Which of the following is a key challenge in Computer Vision?
    a) High accuracy of CNNs b) Privacy and security concerns in surveillance
    c) Over 16 million colour combinations d) Availability of scanners and cameras

Answer Key (MCQs)

1-c, 2-b, 3-a, 4-d, 5-b
6-b, 7-a, 8-a, 9-d, 10-b
11-a, 12-a, 13-b, 14-c, 15-a
16-b, 17-b, 18-c, 19-b, 20-a
21-b, 22-b, 23-b, 24-d, 25-b

Part B: True/False Questions (26–30)

26. A digital image is always stored in the form of pixels only. (True/False)
27. Higher resolution images contain more pixels and provide greater detail. (True/False)
28. Image normalization scales pixel values to a consistent range such as 0–1. (True/False)
29. Object detection is the same as image classification. (True/False)
30. Histogram Equalization enhances brightness and contrast in images. (True/False)

Answer: 26. True 27. True 28. True 29. False 30. True

Part C: Assertion-Reason Questions (31–35)

a) Both A and R are true, and R is the correct explanation of A.

b) Both A and R are true, but R is not the correct explanation of A.

c) A is true, R is false.

d) A is false, R is true.

31. Assertion (A): Edge detection identifies regions where pixel intensity changes sharply.
Reason (R): It helps in highlighting object boundaries within an image.

32. Assertion (A): Instance segmentation assigns pixel-wise labels to all objects in an image.
Reason (R): Semantic segmentation does not differentiate between multiple objects of the same class.

33. Assertion (A): Computer vision systems face challenges related to privacy and security.
Reason (R): Facial recognition in surveillance systems can raise ethical concerns.

34. Assertion (A): Texture analysis extracts features such as smoothness and repetition.
Reason (R): It focuses on analysing changes in image brightness.

35. Assertion (A): MRI and CT scans are examples of image acquisition techniques in scientific fields.
Reason (R): They are used for detecting pixel intensity changes in natural scenes.

Answers: 31. A 32. **B** 33. **A** 34. **C** 35. **C**

Short Questions

1. Define Computer Vision. How is it related to Artificial Intelligence?

Answer: Computer Vision is a branch of AI that enables machines to interpret and analyze visual data using digital images and deep learning models.

2. What is a digital image and how is it stored in a computer?

Answer: A digital image is represented as a grid of pixels, where each pixel holds a numerical value corresponding to its brightness or color.

3. How does image resolution affect the quality of an image?

Answer: Higher resolution means more pixels, resulting in sharper and more detailed images.

4. Differentiate between image classification and object detection with an example.

Answer: Classification labels the entire image as one category (e.g., "cat"), whereas detection identifies and locates objects within the image (e.g., "cat" and "dog" with bounding boxes).

5. Name two algorithms used for object detection and mention one use case.

Answer: R-CNN and YOLO. They are used in surveillance to detect suspicious objects in live video feeds.

6. What is image preprocessing and why is it necessary in computer vision?

Answer: It enhances image quality by removing noise, normalizing pixels, and improving contrast, helping models analyze images accurately.

7. Which feature extraction method highlights object boundaries in an image and why is it important?

Answer: Edge detection highlights sharp intensity changes, helping to identify object shapes and boundaries.

8. How does histogram equalization improve image quality?

Answer: It spreads pixel intensity values evenly to enhance brightness and contrast, making hidden details visible.

9. Write any two real-life applications of Computer Vision.

Answer: (i) Facial recognition on social media. (ii) Medical imaging for detecting tumors or diseases.

10. Distinguish between semantic segmentation and instance segmentation with examples.

Answer: Semantic segmentation groups pixels of the same class together (e.g., "all cars"), whereas instance segmentation differentiates each individual object (e.g., "Car 1," "Car 2").

11. What is image acquisition? Mention one factor that affects its quality.

Answer: Image acquisition is capturing digital images using cameras or scanners. Quality is affected by lighting conditions.

12. Which supervised learning algorithm is used for image classification and why?

Answer: K-Nearest Neighbour (KNN) because it classifies images based on similarity to labeled examples.

13. Why do lighting variations create challenges in computer vision?

Answer: They alter pixel intensity, making object detection and recognition less reliable.

14. Why does computer vision raise privacy concerns in surveillance systems?

Answer: It can track individuals continuously, risking unauthorized monitoring and misuse of personal data.

15. What is high-level processing in computer vision and where is it applied?

Answer: It interprets detected objects and scenes to derive insights, used in autonomous vehicles and medical diagnostics.

16. What problem is posed by duplicate and false content in computer vision?

Answer: Manipulated images/videos spread misinformation and can harm reputations.

17. How do single-object detection tasks differ from multi-object detection tasks?

Answer: Single-object tasks locate and classify one object, while multi-object tasks detect and label several objects simultaneously.

18. Why are Convolutional Neural Networks (CNNs) preferred for feature extraction?

Answer: CNNs automatically learn and extract hierarchical image features, reducing manual effort and improving accuracy.

19. Which computer vision application can help urban planners create detailed city models?

Answer: 3D model building using computer vision and aerial imagery.

20. How does image resolution affect printed photographs in a student project?

Answer: Higher resolution produces clearer, detailed prints, while low resolution causes blurriness when enlarged.

Computer Vision - Long Answer Questions (4 Marks Each)

1. Explain the stages of the Computer Vision process with suitable examples.

Answer: Stages include:

1. Image Acquisition – Capturing images using cameras or scanners.
2. Preprocessing – Enhancing image quality (noise reduction, normalization).
3. Feature Extraction – Identifying edges, corners, colors, textures.
4. Detection/Segmentation – Locating and labeling objects within an image.

5. High-Level Processing – Interpreting and making decisions.
Example: Autonomous vehicles use CV to capture road scenes, detect objects, and take actions.

2. What is image segmentation? Explain semantic and instance segmentation with examples.

Answer: Image segmentation is the process of classifying pixels into meaningful regions.
- Semantic segmentation: Groups all pixels of the same class together (e.g., all cars marked as 'car').
- Instance segmentation: Differentiates each individual object even if they belong to the same class (e.g., Car 1, Car 2).

3. Describe four real-world applications of Computer Vision.

Answer: 1. Facial recognition in social media.
2. Medical imaging to detect tumors or diseases.
3. Self-driving cars to detect vehicles, pedestrians, and signals.
4. 3D model building for robotics and AR/VR.

4. What are the main challenges faced by Computer Vision systems?

Answer: 1. Reasoning and analytical issues in interpreting visual data.
2. Difficulty in image acquisition due to lighting, occlusion, and perspective.
3. Privacy and security concerns from surveillance.
4. Spread of duplicate or false content causing misinformation.

5. Differentiate between classification, detection, and localization tasks in Computer Vision with examples.

Answer: 1. Classification – Assigns a single label to the entire image (e.g., image contains a dog).
2. Detection – Identifies multiple objects and draws bounding boxes (e.g., detecting dog and cat in the same image).
3. Localization – Finds exact location of a single object in the image (e.g., drawing a box around a dog).

6. Explain how preprocessing techniques improve Computer Vision models.

Answer: 1. Noise reduction – Removes blurriness and artifacts.
2. Normalization – Standardizes pixel value ranges for consistency.
3. Resizing/Cropping – Ensures uniform input dimensions.
4. Histogram equalization – Improves brightness and contrast for better detection.

7. How does high-level processing in Computer Vision help in decision-making?

Answer: High-level processing interprets detected objects and scenes to:
1. Understand object relationships.
2. Assist medical diagnosis by analyzing patterns.
3. Guide autonomous vehicles by recognizing road conditions.
4. Enhance surveillance by detecting suspicious activity.

8. Explain why convolutional neural networks (CNNs) are widely used in Computer Vision.

Answer: 1. Automatically extract hierarchical features from images.
2. Reduce manual feature engineering efforts.
3. Provide high accuracy in recognition and detection tasks.
4. Work effectively on large datasets with deep learning.

9. How can Computer Vision be used to build detailed 3D city models?

Answer: 1. Aerial images and topographic data are captured.
2. Computer Vision algorithms identify buildings, roads, and vegetation.
3. Segmentation and detection create accurate maps.
4. 3D reconstruction tools generate city-scale digital models.

10. Discuss the ethical concerns related to Computer Vision applications.

Answer: 1. Privacy invasion due to facial recognition in surveillance.
2. Security risks from deepfakes and manipulated media.
3. Regulatory challenges in using personal visual data.
4. Risk of bias in AI models affecting fairness and decisions.

COMPETENCY BASED QUESTIONS:

Q1. A printing company is preparing large posters from digital photos. A designer wants to know how image resolution will affect the clarity of the posters. Explain.

Ans: Higher resolution images have more pixels, resulting in sharper and clearer prints. Low-resolution images become blurry when enlarged.

Q2. Riya is scanning old family photos to create a digital album. She wonders if scanning at a higher resolution is necessary. What would you advise her?

Ans: Scanning at higher resolution captures finer details, ensuring better quality when viewed or printed.

Q3. A museum uses AI cameras to monitor visitor movement. Which feature extraction method should be applied to detect sudden crowd congestion in a narrow corridor?

Ans: Corner detection, as it identifies points of abrupt movement or intersections where congestion occurs.

Q4. A company wants to classify fabrics (smooth, rough, patterned) using AI. Which feature extraction method should they use and why?

Ans: Texture analysis, because it identifies surface patterns and variations in pixel intensity.

Q5. A fruit packaging plant uses AI to separate defective apples from good ones on a conveyor belt. Which segmentation technique should be used?

Ans: Instance segmentation, because it isolates individual apples for inspection and sorting.

Q6. A city government wants to classify satellite images into land use categories such as water, forest, and urban areas. Which segmentation technique is suitable?

Ans: Semantic segmentation, as it groups pixels into broad categories without distinguishing individual instances.

Q7. An interior design firm wants to generate 3D models of building interiors using floor plans and photographs. Which application of computer vision can help?

Ans: 3D model building using computer vision.

Q8. A self-driving car company wants to improve its system for detecting pedestrians and vehicles. Which computer vision capability should be enhanced?

Ans: Object detection using algorithms like YOLO or R-CNN to identify and classify multiple objects in real time.

Q9. A social media platform detects videos where people appear to say things they never said. What is this issue called?

Ans: False or fake content (deepfake manipulation).

Q10. A news agency uses AI to verify whether viral images are real or doctored. Which major challenge of computer vision does this represent?

Ans: Detecting duplicate or manipulated content to prevent misinformation.

UNIT 4: AI WITH ORANGE DATA MINING TOOL

Overview

This unit introduces **Orange Data Mining**, a powerful **open-source, component-based visual programming tool** designed for **Data Science, Machine Learning, Computer Vision, and Natural Language Processing (NLP)**.

Students will gain hands-on experience building workflows by **dragging and connecting widgets**, enabling them to visualize data, train models, and interpret results. The unit focuses on real-world applications of AI concepts, encouraging creativity, group collaboration, and practical problem-solving.

Learning Objectives

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

- Gain a **comprehensive understanding** of the Orange Data Mining tool and its capabilities.
- Explore **data visualization, preprocessing, modeling, and evaluation** through practical exercises.
- Apply Orange to **Computer Vision** tasks like image clustering and **NLP** tasks like text preprocessing and sentiment analysis.
- Build workflows to **analyze and interpret data effectively**.
- Demonstrate AI problem-solving skills in real-world scenarios using Orange.

Prerequisites

Students should have:

- A basic understanding of **Data Science and Machine Learning concepts**.
- Awareness of **Computer Vision and NLP terminology**.
- Knowledge of basic algorithms like **classification, clustering, and evaluation metrics**.

Key Concepts

- **Data Mining:** The process of extracting meaningful insights and identifying patterns from large datasets.
- **Orange Data Mining Tool:** An open-source visual programming platform with a drag-and-drop interface for AI tasks.
- **Widgets:** Individual components used to perform tasks such as loading datasets, training models, evaluating performance, or visualizing results.
- **Applications:** Data Science (tabular data analysis), Computer Vision (image analytics), and NLP (text processing and analysis).

Beneficiaries of Orange Data Mining Tool

- **Data Analysts & Scientists:** Simplifies advanced analysis for professionals without coding expertise.
- **Researchers:** Helps with research data exploration and hypothesis validation.
- **Educators & Students:** Makes AI and data science concepts easier to teach and learn.
- **Businesses:** Supports decision-making through visual analytics and predictions.
- **Open-Source Community:** Offers free and customizable access for all users.

Components of Orange Tool

1. **Blank Canvas:** Workspace to design analysis workflows by dragging and connecting widgets.
2. **Widgets:** Visual components that perform tasks like loading data, visualization, model training, etc.
3. **Connectors:** Lines that connect widgets, showing the data flow between components.

Default Widget Categories

- **Data Widgets:** Import and manage datasets (e.g., File, Data Table, SQL Table).
- **Transform Widgets:** Clean, preprocess, and transform data.
- **Visualize Widgets:** Create scatter plots, box plots, histograms, and heatmaps.
- **Model Widgets:** Train ML models (classification, clustering, regression).
- **Evaluate Widgets:** Assess model performance with metrics like accuracy and F1-score.
- **Unsupervised Widgets:** Perform clustering, dimensionality reduction, and association mining.

Key Domains of AI in Orange

1. Data Science with Orange

- Load datasets, visualize data, and train models.
- Example: **Iris Flower Classification**
  - Load the **Iris dataset** using the File widget.
  - Explore data using Data Table and Scatter Plot.
  - Train a decision tree classifier and use **Predictions** widget to classify new samples.
  - Evaluate model performance using **Test & Score** and **Confusion Matrix**.
- Concepts covered: Cross-validation, accuracy, precision, recall, and F1-score.

---

2. Computer Vision with Orange

- Analyze and cluster images using **Image Analytics Add-On**.
- Example: **Cat vs. Dog Image Clustering**
  - Import images using **Import Images** widget.
  - Generate image embeddings with **Image Embedding**.
  - Compute image similarity with **Distance** widget.
  - Visualize clusters using **Hierarchical Clustering** and **Image Viewer**.
- Applications: Image classification, object recognition, and clustering tasks.

---

3. NLP with Orange

- Process text data using the **Text Add-On**.
- Example Workflow:

1. Load or create text data using **Corpus/Corpus Creator**.
2. Visualize text frequency with **Word Cloud**.
3. Preprocess text: Tokenize, lowercase, remove punctuation, stopwords, and apply stemming/lemmatization.
4. Analyze cleaned text for sentiment or topic modeling.

- Outcome: Understand text structure and extract meaningful insights.

Hands-On Practice Ideas

- **Data Science Project:** Classify fruits and vegetables based on nutritional information.
- **Computer Vision Project:** Group animals and birds into clusters by visual similarities.
- **Text Analysis Project:**Analyze themes and frequent words in a story/article using NLP tools.

Evaluation Metrics

Metric	Purpose
Accuracy	Measures overall correctness of predictions.
Precision	Proportion of true positives among predicted positives.
Recall	Proportion of correctly identified true positives.
F1-Score	Harmonic mean of precision and recall for balanced performance.
Confusion Matrix	Table showing TP, TN, FP, FN for each class.

Exercises

A. Multiple Choice Questions

1. Widget for calculating accuracy, precision, recall, and F1 score:
   b. Test and Score
2. Widget for detailed model classification breakdown:
   c. Confusion Matrix
3. Widget for text normalization:
   Preprocess Text
4. Cross-validation is also known as:
   Rotation Estimation
5. In Word Cloud, frequent words appear larger.
   True
6. Widget for converting images into numerical representation:
   Image Embedding
7. Widget for computing similarities between image embeddings:
   Distance
8. Lines that link widgets together:
   Connectors

B. Short Answer Questions

1. **List three main components of Orange.**
   - Blank Canvas
   - Widgets
   - Connectors
2. **What is Cross-Validation?**
   A model evaluation technique that trains and tests a model on multiple data splits for better accuracy estimation.
3. **Purpose of Corpus Viewer widget:**
   To browse text data and highlight search terms.
4. **Use of Preprocess Text widget:**
   To clean and standardize text for NLP tasks (tokenization, stopword removal, etc.).

C. Competency-Based Questions

1. Which open-source visual programming software is used for machine learning and data mining?
   Orange Data Mining Tool

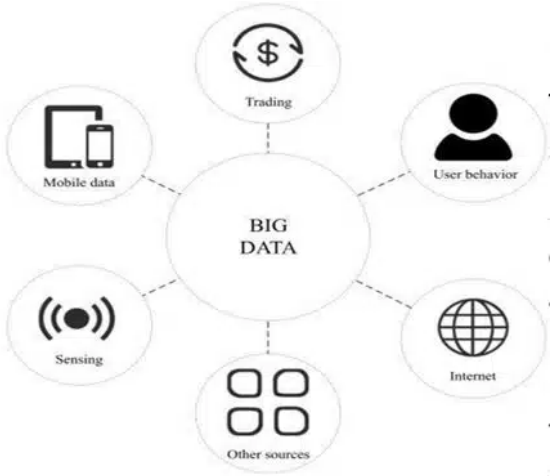
2. Add-on needed for image analytics:
   Image Analytics Add-On
3. Add-on needed for text processing:
   Text Add-On
4. Widget to identify classification errors and class-wise performance:
   Confusion Matrix

UNIT 5: INTRODUCTION TO BIG DATA AND DATA ANALYTICS

What is Big Data?

To understand Big Data, let us first understand small data.

- **Small Data:** Small data refers to datasets that are easily comprehensible by people as they are easily accessible, informative, and actionable. This will help the business to make useful information and make better choices in everyday tasks. For example, a small store might track daily sales to decide what products to restock.
- **Big Data:** Big Data refers to extremely large and complex datasets that regular computer programs and databases cannot handle. It comes from three main sources: transactional data (online purchases), machine data (sensor readings), and social data (social media posts). To analyze and use Big Data effectively, special tools and techniques are required. For example, companies like Amazon and Netflix use Big Data to recommend products or shows based on users' past activities.



Types of Big Data

There are three different types of data:

a. Structured Data

b. Semi Structured Data

c. Unstructured Data

Aspect	Structured Data	Semi-Structured Data	Unstructured Data
Definition	Quantitative data with a defined structure	A mix of quantitative and qualitative properties	No inherent structures or formal rules
Data Model	Dedicated data model	May lack a specific data model	Lacks a consistent data model
Organization	Organized in clearly defined columns	Less organized than structured data	No organization exhibits variability over time
Accessibility	Easily accessible and searchable	Accessible but may be harder to analyze	Accessibility depends on the specific data format
Examples	Customer information, transaction records, product directories	XML files, CSV files, JSON files, HTML files, semi-structured documents	Audio files, images, video files, emails, PDFs, social media posts

Advantages and Disadvantages of Big Data:

<u>Advantages</u>

- **Enhanced Decision Making:** Organizations can make data-driven decisions based on the insights derived from big data.

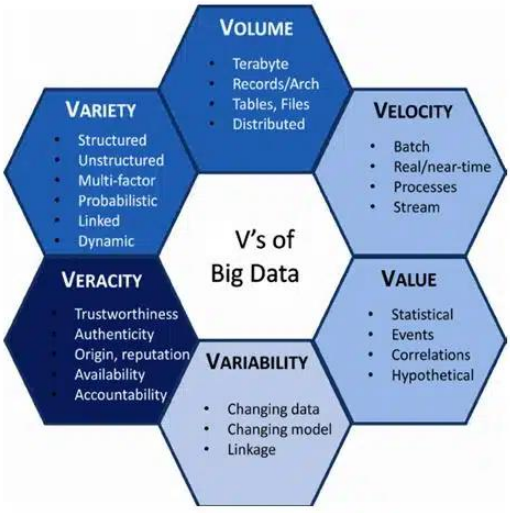
- **Improved efficiency and productivity:** Big data analysis can help the organization to improve the efficiency and productivity of the business.
- **Better Customer Insights:** Big data can help the organization to gain a deeper understanding of customer behavior, preferences, and needs.
- **Competitive Advantage:** Leveraging big data analytics provides organizations with a competitive edge by enabling them to uncover market trends, identify opportunities, and stay ahead of competitors.
- **Innovation and Growth:** Big Data fosters innovation by facilitating the development of new products, services, and business models based on insights derived from data analysis, driving business growth and expansion.

Disadvantages

- **Privacy and Security Concerns:** It is difficult to collect, store and analysis the big data. Managing big data is also having a security and privacy concerns due to unauthorized access, data breaches, and misuse of personal information.
- **Data Quality Issues:** Ensuring the accuracy, reliability, and completeness of data can be challenging, as Big Data often consists of unstructured and heterogeneous data sources, leading to potential errors and biases in analysis.
- **Technical Complexity:** Implementing and managing Big Data infrastructure and analytics tools require specialized skills and expertise, leading to technical challenges and resource constraints for organizations.
- **Regulatory Compliance:** Organizations face challenges in meeting data protection laws like GDPR (General Data Protection Regulation) and The Digital Personal Data Protection Act, 2023. These laws require strict handling of personal data, making compliance essential to avoid legal risks and penalties.
- **Cost and Resource Intensiveness:** The cost of acquiring, storing, processing, and analyzing Big Data, along with hiring skilled staff, can be high. This is especially challenging for smaller organizations with limited budgets and resources.

Characteristics of Big Data

The “characteristics of Big Data” refer to the defining attributes that distinguish large and complex datasets from traditional data sources. These characteristics are commonly described using the **“3Vs” framework: Volume, Velocity, and Variety**. The 6Vs framework provides a holistic view of Big Data, emphasizing not only its volume, velocity, and variety but also its veracity, variability, and value.



- **Velocity:** Velocity refers to the speed at which data is generated, delivered, and analyzed. For example: Google alone generates more than 40,000 search queries per second.

- **Volume:** Every day a huge volume of data is generated as the number of people using online platforms has increased exponentially. Such a huge volume of data is considered Big Data. According to the surveys conducted in varuious organization estimates, 328.77 million terabytes of data are created each day.
- **Variety:** Big data encompasses data in various formats, including structured, unstructured, semi-structured, or highly complex structured data. These can range from simple numerical data to complex and diverse forms such as text, images, audio, videos, and so on.
- **Veracity:** Veracity is a characteristic in Big Data related to consistency, accuracy, quality, and trustworthiness. Not all data that undergoes processing holds value. Therefore, it is essential to clean data effectively before storing or processing it.
- **Value:** The goal of big data analysis lies in extracting business value from the data. Hence, the business value derived from big data is perhaps its most critical characteristic. Without obtaining valuable insights, the other characteristics of big data hold little significance.
- **Variability:** This refers to establishing if the contextualizing structure of the data stream is regular and dependable even in conditions of extreme unpredictability. It defines the need to get meaningful data considering all possible circumstances.

Big Data vs Traditional Data:

Feature	Traditional Data	Big Data
Volume	Limited (gigabytes)	Massive (terabytes to zettabytes)
Variety	Limited to structured formats	Includes all formats
Velocity	Low (processed in batches)	High (often real-time)
Flexibility	Less flexible	Highly adaptable and dynamic

Big Data Analytics

Data analytics involves analysing datasets to uncover insights, trends, and patterns. Technologies commonly used in data analytics include statistical analysis software, data visualization tools, and relational database management systems (RDBMS). Big data analytics uses advanced analytic techniques against huge, diverse datasets that include structured, semi-structured, and unstructured data from different sources and in various sizes from terabytes to zettabytes.

Types of Analytics:

1. **Descriptive Analytics:**
- Summarizes past events and trends.
- *Example:* Monthly sales reports.
2. **Diagnostic Analytics:**
- Explains why something happened.
- *Example:* Why did customer churn increase?
3. **Predictive Analytics:**
- Uses historical data to forecast future outcomes.
- *Example:* Predicting product demand or customer behavior.
4. **Prescriptive Analytics:**
- Recommends actions based on analysis.
- *Example:* Suggesting the best marketing strategy.

Big Data Analytics emerges as a consequence of four significant **global trends**:

1. **Moore's Law:** Moore's Law has enabled the handling and analysis of massive datasets, driving the evolution of Big Data Analytics.
2. **Mobile Computing:** Due to smartphones and mobile devices, the vast amount of real-time data can be collected from anywhere.
3. **Social Networking:** Social media platforms use data sharing, interactions, and user-generated content for collecting massive datasets, which helps to analyze the data easily.
4. **Cloud Computing:** Cloud computing allows organizations to access hardware and software resources remotely via the Internet, which reduces the investments in software and hardware.

Working on Big Data Analytics

Big data analytics involves collecting, processing, cleaning, and analyzing enormous datasets to improve organizational operations. The working process of big data analytics includes the following steps –

Step 1: Gather data

Each company has a unique approach to data collection. Organizations can now collect data from various sources, including cloud storage, mobile apps, and IoT sensors.

Step 2: Process Data

Once data is collected and stored, it must be processed properly to get accurate result on analytical queries. Various processing options are available:

- Batch processing which looks at large data blocks over time.
- Stream processing looks at small batches of data at once, shortening the delay time between collection and analysis for quicker decision-making.

Step 3: Clean Data

All data should be clean for better results. Cleaning data helps to eliminate duplicate or irrelevant data.

Step 4: Analyze Data

Getting big takes time to analyse or make usable; advanced analytics processes can turn big data into big insights.

Mining Data Streams:

A data stream is a continuous, real-time flow of data generated by various sources like sensors, satellite images, the internet, and web traffic, etc. Mining data streams refers to the process of extracting meaningful patterns, trends, and knowledge from a continuous flow of real-time data. For instance, a sudden spike in searches for "election results" on a particular day might indicate that elections were recently held in a region or highlight the level of public interest in the results.

Future of Big Data Analytics

- **Real-Time Analytics:** It will allow businesses to process real-time data for decision-making, such as live monitoring customer behavior or tracking supply chain activities.
- **Development of Advanced Models in Predictive Analytics:** Predictive analytics will help to integrate more sophisticated machine learning algorithms to enable the forecasting of trends and behaviors with greater precision.
- **Quantum Computing:** Quantum computers will be able to solve complex problems much faster than classical computers.

QUESTION BANK

MCQs:

1. What is the first step in Data Science Methodology?
   a. Data Understanding b. Business Understanding
   c. Data Collection d. Evaluation
2. Which of the following techniques helps to determine why an event happened?
   a. Descriptive Analytics b. Predictive Analytics
   c. Diagnostic Analytics d. Prescriptive Analytics
3. Which is NOT a type of Data Analytics?
   a. Descriptive Analytics b. Diagnostic Analytics
   c. Cognitive Analytics d. Predictive Analytics
4. What is used in Descriptive Analytics to summarize past data?
   a. Decision Trees b. Root Cause Analysis
   c. Summary Statistics d. Regression
5. Which of the following is NOT a characteristic of predictive analytics?
   a. Forecasting future events b. Using historical data
   c. Recommending actions based on data d. Using techniques like regression
6. In which step of the Data Science Methodology is data cleaned and transformed for analysis?
   a. Data Collection b. Data Preparation
   c. Data Understanding d. Evaluation
7. Which of the following is a key component of Feature Engineering?
   a. Splitting data into training and testing sets
   b. Creating new features to improve model performance
   c. Identifying data sources
   d. Evaluating model performance
8. Which data type does NOT follow a predefined structure?
   a. Structured data b. Semi-structured data
   c. Unstructured data d. Categorical data
9. Which of the following is an example of a primary data source?
   a. Books b. Social media posts
   c. Surveys d. Online databases
10. Which technique is used to assess the model's performance by splitting the data into two sets?
    a. K-Fold Cross Validation b. Train-Test Split
    c. Data Understanding d. Data Collection
11. Which approach helps to identify patterns based on historical data and is used for forecasting?
    a. Predictive Analytics b. Descriptive Analytics
    c. Diagnostic Analytics d. Prescriptive Analytics
12. In which phase do data scientists evaluate whether the data collected is relevant to the problem being solved?
    a. Data Collection b. Data Preparation
    c. Data Understanding d. Business Understanding
13. What is a Confusion Matrix used for in machine learning?
    a. To display the distribution of data
    b. To evaluate the performance of a classification model
    c. To visualize regression errors
    d. To split data into training and testing sets

Answers

1. b. Business Understanding
2. c. Diagnostic Analytics
3. c. Cognitive Analytics
4. c. Summary Statistics
5. c. Recommending actions based on data
6. b. Data Preparation
7. b. Creating new features to improve model performance
8. c. Unstructured data
9. c. Surveys
10. b. Train-Test Split
11. a. Predictive Analytics
12. c. Data Understanding
13. b. To evaluate the performance of a classification model

Assertion-Reasoning Questions

Choose the correct option:

a) Both A and R are true, and R is the correct explanation of A.

b) Both A and R are true, but R is not the correct explanation of A.

c) A is true, but R is false.

d) A is false, but R is true.

1. **Assertion (A):** Big Data refers to large, complex datasets that traditional data-processing applications cannot handle.
   Reason (R): Traditional data-processing systems are limited in their ability to manage and analyze Big Data due to its size, complexity, and variety.
2. **Assertion (A):** Hadoop is a key tool used for Big Data analytics.
   Reason (R): Hadoop allows distributed storage and processing of large datasets across multiple machines, making it scalable for Big Data applications.
3. **Assertion (A):** Unstructured data is the easiest type of data to analyze in Big Data applications.
   Reason (R): Unstructured data, such as text, images, and videos, does not follow a specific format or structure, making it more challenging to process and analyze.
4. **Assertion (A):** The “Volume” characteristic of Big Data refers to the amount of data generated.
   Reason (R): As data volume increases, the amount of computational resources required for processing also increases, which can lead to challenges in managing Big Data.
5. **Assertion (A):** Real-time analytics allows businesses to make decisions based on historical data.
   Reason (R): Real-time analytics processes data as it is generated, providing immediate insights that allow businesses to make timely decisions.
6. **Assertion (A):** The “Variety” of Big Data refers to the different formats of data, such as structured, semi-structured, and unstructured.
   Reason (R): The variety of data requires specialized tools and techniques for effective processing, as each data type has different structures and storage needs.
7. **Assertion (A):** Big Data analytics can only be applied to structured data.
   Reason (R): Structured data is highly organized, which makes it easier to analyze compared to unstructured or semi-structured data.

8. **Assertion (A):** Veracity in Big Data refers to the speed at which data is generated and analyzed.
   Reason (R): Veracity is concerned with the accuracy and quality of data, rather than its speed, ensuring that data is reliable for analysis.
9. **Assertion (A):** Predictive analytics is a type of Big Data analytics that uses historical data to forecast future trends.
   Reason (R): By analyzing past trends and behaviors, predictive analytics can anticipate future outcomes, enabling businesses to plan accordingly.
10. **Assertion (A):** The "Velocity" characteristic of Big Data is concerned with the variety of data sources.
    Reason (R): Velocity refers to the speed at which data is generated and must be processed in real-time or near-real-time, not the variety of data sources.

Answers

1. a) Both A and R are true, and R is the correct explanation of A.
2. a) Both A and R are true, and R is the correct explanation of A.
3. d) A is false, but R is true.
4. a) Both A and R are true, and R is the correct explanation of A.
5. d) A is false, but R is true.
6. a) Both A and R are true, and R is the correct explanation of A.
7. d) A is false, but R is true.
8. d) A is false, but R is true.
9. a) Both A and R are true, and R is the correct explanation of A.
10. d) A is false, but R is true.

LONG ANSWER QUESTIONS:

1. What is Big Data, and how does it differ from small data?
2. Explain the 3V's of Big Data and their significance.
3. What are the main types of Big Data, and how do they differ from each other?
4. Describe the advantages and disadvantages of using Big Data.
5. What are the 6V's of Big Data, and how do they provide a more holistic view of Big Data?
6. What is the role of Hadoop in Big Data processing?
7. Explain the differences between structured, semi-structured, and unstructured data.
8. What is Data Mining, and how does it apply to Big Data?
9. How does Big Data Analytics work, and what are its key components?
10. What are some common types of Big Data Analytics?
11. What is Data Stream Mining, and how is it used?
12. What challenges are associated with the "Veracity" of Big Data?
13. How can Big Data be used in the healthcare industry?
14. What is the significance of "Cloud Computing" in Big Data Analytics?
15. How do the "Volume" and "Velocity" aspects of Big Data affect the analysis process?
16. What role do Machine Learning algorithms play in Big Data analytics?
17. How does "Batch Processing" differ from "Stream Processing" in Big Data analytics?
18. What are the ethical concerns associated with Big Data?
19. What tools are commonly used in Big Data analytics, and how do they help?
20. What are the future trends in Big Data Analytics?

Answers:

What is Big Data, and how does it differ from small data?

Answer: Big Data refers to vast, complex datasets that traditional database systems cannot handle due to their size, speed, or structure. Unlike small data, which is manageable and easily understood, Big Data requires specialized tools and techniques for analysis. It often includes transactional data, machine-generated data, and social media data, making it harder to analyze using conventional methods. The large scale and complexity of Big Data make it a significant resource for businesses and researchers looking to extract valuable insights.

2. Explain the 3V's of Big Data and their significance.

Answer: The 3V's of Big Data are Volume, Velocity, and Variety.

- **Volume** refers to the massive amount of data generated, often in the range of petabytes or even exabytes daily.
- **Velocity** describes the speed at which data is generated and must be processed, requiring real-time analysis.
- **Variety** encompasses the different types of data, including structured (e.g., databases), semi-structured (e.g., XML files), and unstructured data (e.g., social media posts, videos).

These three characteristics distinguish Big Data from traditional data and require advanced analytics tools to extract meaningful insights.

3. What are the main types of Big Data, and how do they differ from each other?

Answer: The main types of Big Data are:

- **Structured Data:** Data that is highly organized and stored in databases or spreadsheets, typically in rows and columns (e.g., customer information).
- **Semi-structured Data:** Data that does not have a fixed schema but has some form of organization, such as XML files or JSON files.
- **Unstructured Data:** Data that lacks any predefined structure, including text, images, videos, and social media posts.

Each type requires different methods for analysis and storage, with unstructured data being the most challenging to process.

4. Describe the advantages and disadvantages of using Big Data.

Answer:

Advantages:

- **Enhanced decision-making:** Big Data enables organizations to make informed, data-driven decisions based on vast datasets.
- **Improved efficiency:** By analyzing large volumes of data, businesses can optimize operations and reduce inefficiencies.
- **Better customer insights:** Big Data allows for the personalization of services and targeted marketing.
- **Competitive advantage:** Organizations can uncover trends and predict future outcomes.

Disadvantages:

- **Privacy and security concerns:** The collection and analysis of personal data raise ethical issues and data protection risks.
- **Data quality issues:** Ensuring the accuracy and consistency of Big Data is challenging, as it often includes unstructured and heterogeneous data.
- **Technical complexity:** Big Data requires specialized infrastructure, tools, and expertise, which can be costly and resource-intensive.
- **Compliance challenges:** Companies must adhere to regulations such as GDPR, which can be complex and costly to implement.

5. What are the 6V's of Big Data, and how do they provide a more holistic view of Big Data?

Answer: The 6V's of Big Data expand upon the 3V framework by adding three more characteristics:

- **Volume:** The sheer amount of data generated every day.
- **Velocity:** The speed at which data is produced and needs to be processed.
- **Variety:** The different types of data (structured, semi-structured, unstructured).
- **Veracity:** The accuracy and trustworthiness of the data, ensuring its suitability for analysis.
- **Value:** The insights and benefits derived from analyzing the data.
- **Variability:** The inconsistencies or unpredictability in data flows, requiring systems to adapt.

Together, these dimensions highlight the complexity of managing and analyzing Big Data effectively.

6. What is the role of Hadoop in Big Data processing?

Answer: Hadoop is a distributed file system and software framework used to store and process large datasets across clusters of computers. It allows for the storage of vast amounts of data in a fault-tolerant manner, making it scalable and efficient. Hadoop's MapReduce programming model enables parallel processing, allowing data to be analyzed in chunks across multiple nodes. This is especially useful for handling Big Data, where traditional data processing tools are insufficient.

7. Explain the differences between structured, semi-structured, and unstructured data.

Answer:

- **Structured Data** is highly organized and easily searchable, typically stored in relational databases with a fixed schema (e.g., customer names, addresses).
- **Semi-structured Data** does not have a fixed schema but contains tags or markers to separate elements (e.g., XML, JSON).
- **Unstructured Data** lacks any predefined structure and is more difficult to process and analyze (e.g., text, images, videos, audio files).

These differences impact how the data is stored, accessed, and analyzed.

8. What is Data Mining, and how does it apply to Big Data?

Answer: Data mining is the process of discovering patterns, trends, and relationships in large datasets. In the context of Big Data, it involves applying machine learning algorithms, statistical models, and data processing tools to extract meaningful insights. Data mining can identify customer preferences, predict trends, and detect anomalies, all of which are valuable for business decision-making.

9. How does Big Data Analytics work, and what are its key components?

Answer: Big Data Analytics involves collecting, storing, cleaning, processing, and analyzing large datasets to uncover insights and trends. Key components include:

- **Data collection:** Gathering data from various sources (social media, sensors, transactions).
- **Data storage:** Using distributed storage solutions like Hadoop or cloud-based storage systems.
- **Data processing:** Cleaning and organizing data to make it ready for analysis.
- **Data analysis:** Applying techniques such as machine learning, predictive analytics, and statistical modeling to extract insights.
- **Data visualization:** Presenting the findings in a visual format, such as graphs and charts, to help stakeholders make informed decisions.

10. What are some common types of Big Data Analytics?

Answer: The common types of Big Data Analytics include:

- **Descriptive Analytics:** Summarizes past data to identify patterns and trends.
- **Diagnostic Analytics:** Analyzes historical data to understand the causes behind specific outcomes.
- **Predictive Analytics:** Uses historical data to forecast future trends or events.
- **Prescriptive Analytics:** Recommends actions based on data insights to achieve desired outcomes.

11. What is Data Stream Mining, and how is it used?

Answer: Data Stream Mining refers to the real-time processing and analysis of continuous streams of data. Unlike traditional data mining, which analyzes static datasets, stream mining analyzes data as it arrives, without storing it completely. It is used in applications like monitoring social media feeds, detecting fraud in financial transactions, or tracking sensor data in IoT devices.

12. What challenges are associated with the "Veracity" of Big Data?

Answer: Veracity in Big Data refers to the trustworthiness, quality, and accuracy of the data. Challenges include:

- **Data inconsistencies:** Large volumes of data may contain errors, duplications, or missing values, making it difficult to trust the insights derived from them.
- **Data bias:** Inaccurate or biased data sources can lead to misleading conclusions.
- **Data cleaning issues:** Ensuring that data is properly cleaned and formatted is a time-consuming process, especially when dealing with unstructured data.

13. How can Big Data be used in the healthcare industry?

Answer: Big Data analytics in healthcare can be used to predict disease outbreaks, personalize patient care, and improve medical research. For example, predictive analytics can help hospitals forecast patient admissions and optimize resource allocation. By analyzing patient data, doctors can offer personalized treatments, improving outcomes. Big Data can also help detect fraud and improve drug development by identifying trends in clinical trial data.

14. What is the significance of "Cloud Computing" in Big Data Analytics?

Answer: Cloud computing provides scalable and cost-effective infrastructure for storing and processing Big Data. With cloud services, organizations can access powerful computing resources on-demand, without the need for large upfront investments in hardware. This allows businesses to analyze vast datasets quickly and efficiently, while also enabling collaboration across multiple locations. Additionally, cloud-based tools offer flexibility, security, and reliability for Big Data analytics.

15. How do the "Volume" and "Velocity" aspects of Big Data affect the analysis process?

Answer:

- **Volume** affects the storage and processing of data, as larger datasets require specialized infrastructure and tools like distributed file systems (e.g., Hadoop). Handling high volumes also requires more computational power to process data efficiently.
- **Velocity** refers to the speed at which data is generated and needs to be processed. Real-time or near-real-time analytics are required to make quick decisions based on up-to-date data, such as monitoring social media feeds or tracking financial transactions.

16. What role do Machine Learning algorithms play in Big Data analytics?

Answer: Machine Learning algorithms are essential for analyzing large datasets by automatically detecting patterns and making predictions. In Big Data analytics, they can be used for classification, clustering, regression, and anomaly detection. These algorithms enable businesses to forecast trends, personalize customer experiences, and detect fraud, among other tasks. They can also learn from new data over time, improving accuracy and efficiency.

17. How does "Batch Processing" differ from "Stream Processing" in Big Data analytics?

Answer:

- **Batch Processing** involves collecting large amounts of data and processing them in blocks or batches over time. This method is more suitable for less time-sensitive tasks like analyzing historical data.

- **Stream Processing**, on the other hand, processes data in real-time or near-real-time as it is generated. This is used in scenarios where immediate analysis is required, such as monitoring live data streams from sensors or social media feeds.

18. What are the ethical concerns associated with Big Data?

Answer: Ethical concerns around Big Data include:

- **Privacy issues:** Collecting personal data raises concerns about how that data is used and whether it is adequately protected.
- **Data misuse:** There is a risk of using data for purposes other than what it was intended for, such as targeting vulnerable populations for marketing or surveillance.
- **Bias and discrimination:** Algorithms based on biased data may result in discriminatory practices, such as denying certain groups access to services or opportunities.

19. What tools are commonly used in Big Data analytics, and how do they help?

Answer: Common tools include:

- **Hadoop:** A framework for storing and processing large datasets using a distributed file system.
- **Tableau:** A data visualization tool that helps present Big Data insights through graphs and charts.
- **R and Python:** Programming languages widely used for statistical analysis and machine learning in Big Data.
- **Spark:** A data processing engine that supports real-time analytics and machine learning.

These tools enable businesses to handle, process, analyze, and visualize Big Data effectively.

20. What are the future trends in Big Data Analytics?

Answer: Future trends include:

- **Real-time analytics:** With the increasing speed of data generation, real-time analytics will allow businesses to make immediate decisions based on live data.
- **AI and Machine Learning integration:** These technologies will continue to enhance predictive analytics, enabling more accurate forecasts.
- **Quantum computing:** This promises to accelerate data processing and enable the analysis of even larger datasets.
- **Data privacy regulations:** As Big Data usage grows, more stringent regulations will be implemented to protect user privacy and ensure ethical data practices.

Competency-Based Questions

1. A retail company wants to optimize its inventory management using Big Data. How would you use Big Data analytics to solve this problem?
2. A healthcare organization is looking to use Big Data to predict disease outbreaks in specific regions. What approach would you take to apply Big Data analytics in this case?
3. You are working for an e-commerce company that wants to personalize its marketing strategies based on customer data. How would you leverage Big Data analytics to achieve this?
4. An insurance company is dealing with fraudulent claims and wants to identify patterns in fraudulent activities using Big Data. How would you approach this problem?
5. A financial institution is looking to manage risks and optimize its investment strategies using Big Data. How would you apply Big Data analytics to improve the institution's decision-making process?

Answers:

1. A retail company wants to optimize its inventory management using Big Data. How would you use Big Data analytics to solve this problem?

Answer:

To optimize inventory management using Big Data, I would first gather data from various sources, such as sales transactions, customer preferences, and historical purchase patterns. By using **predictive analytics**, I could forecast future demand for each product, helping the company to maintain the right inventory levels. I would also analyze **seasonal trends** and **regional variations** to ensure the inventory aligns with customer behavior and market fluctuations. Additionally, **descriptive analytics** can be applied to identify patterns in past sales, highlighting which products are overstocked or understocked. Finally, **real-time data** from customer interactions or web searches can be analyzed to adjust inventory dynamically.

2. A healthcare organization is looking to use Big Data to predict disease outbreaks in specific regions. What approach would you take to apply Big Data analytics in this case?
Answer:
To predict disease outbreaks using Big Data, I would collect data from a variety of sources, including hospital records, social media, online health searches, and environmental factors (e.g., weather conditions). **Predictive analytics** would be employed to identify early warning signs and detect patterns that may signal an impending outbreak. For example, by analyzing historical data on past outbreaks, weather trends, and mobility data, I could predict the likelihood of an outbreak occurring in specific regions. **Machine learning models** could be used to refine predictions over time, continuously learning from new data and improving the accuracy of future forecasts.
Additionally, **real-time data streams** from sensors, health reports, and news feeds can help monitor and respond to outbreaks as they occur.

3. You are working for an e-commerce company that wants to personalize its marketing strategies based on customer data. How would you leverage Big Data analytics to achieve this?
Answer:
To personalize marketing strategies, I would first analyze customer behavior data, including past purchases, browsing history, and demographic information. Using **cluster analysis** and **segmentation techniques**, I could identify distinct customer groups with similar preferences or behaviors. **Predictive analytics** would then be used to forecast future purchases, and personalized recommendations would be made to each group based on their past activity and preferences. Additionally, **sentiment analysis** on customer reviews and social media posts could provide valuable insights into how customers feel about certain products, which can be integrated into marketing campaigns. **A/B testing** could be used to fine-tune personalized offers and promotions for maximum engagement.

4. An insurance company is dealing with fraudulent claims and wants to identify patterns in fraudulent activities using Big Data. How would you approach this problem?
Answer:
To detect fraudulent claims, I would first collect data on past claims, customer profiles, and any known fraudulent activities. **Anomaly detection** algorithms could be used to identify claims that deviate from typical patterns, such as unusually high claim amounts or claims made in suspicious circumstances. Using **machine learning models**, I could train the system to recognize characteristics of fraudulent claims based on historical data, and continuously improve the model as more data becomes available. I would also analyze **social media and external databases** to cross-check customer information, helping to identify inconsistencies or patterns indicative of fraud. The use of **predictive**

modeling would help the insurance company proactively flag potentially fraudulent claims before they are processed.

5. A financial institution is looking to manage risks and optimize its investment strategies using Big Data. How would you apply Big Data analytics to improve the institution's decision-making process?
Answer:
To optimize investment strategies, I would first gather diverse financial data, including stock market trends, historical performance, economic indicators, and sentiment analysis from news and social media. Using **predictive analytics**, I would forecast potential market movements, helping the institution anticipate risks and identify profitable investment opportunities. I would apply **descriptive analytics** to analyze past investment performance, detecting patterns that indicate high returns or risk factors. **Real-time data analytics** could be leveraged to track market fluctuations and adjust strategies dynamically. Furthermore, by incorporating **machine learning algorithms**, I could continuously refine investment models based on changing market conditions, providing more accurate recommendations for optimizing investment portfolios and minimizing risks.

UNIT 6: Understanding Neural Networks

What is a neural network?

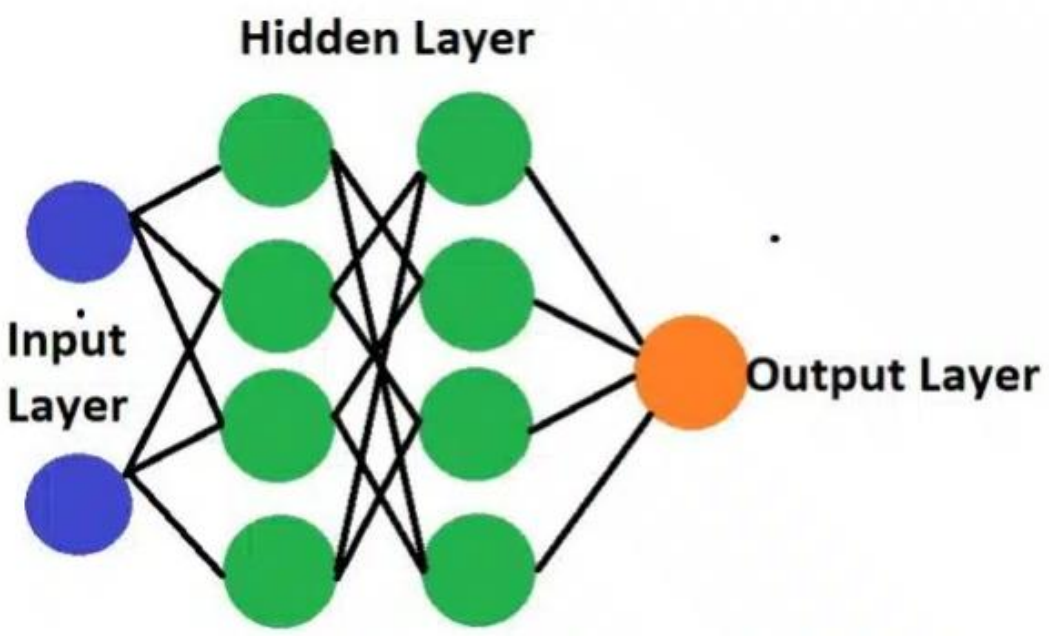
Artificial neural networks help to generate the best possible outcome without making any changes anywhere, as the human brain does. Neural networks can extract data features automatically without the need for a programmer. A network of neurons is called a "neural network.". For example, auto replying to the emails, suggesting email replies, spam filtering, Facebook image tagging, showing items of our interest in the e-shopping web portals, and many more. One of the best-known examples of a neural network is Google's search algorithm.

Definition: A neural network is a machine learning program, or model, that makes decisions in a manner similar to the human brain, by using processes that mimic the way biological neurons work together to identify phenomena, weigh options and arrive at conclusions.

Parts of a Neural Network

Every neural network comprises layers of interconnected nodes — an input layer, hidden layer(s), and an output layer.

1. **Input Layer:** The input layer is the first layer of a neural network and helps to receive data from the outside world.
2. **Hidden Layers:** This is the second layer of the neural network, which processes the input data using multiple hidden layers. Each hidden layer contains nodes or artificial neurones. These nodes are interconnected with each other. It can be multi-layer.
3. **Output Layer:** The last layer of the neural network is the output layer, which helps to generate the output and final predictions or outcomes of the neural network.



An Artificial Neural Network (ANN) with two or more hidden layers is known as a Deep Neural Network. The process of training deep neural networks is called Deep Learning. The term "deep" in deep learning refers to the number of hidden layers (also called depth) of a neural network. A neural network that only has three layers is just a basic neural network.

Components of a neural network

The key components of a neural network are as follows:

What do you mean by Neurons in AI?

- Neurones are also called nodes.
- Neurones are fundamental building blocks of a neural network.
- Neurones can receive inputs from the external sources.
- Each neurone computes a weighted sum of its input.

What are weights in neurons?

- Weights represent the strength of connections between neurones.
- Each connection (synapse) has an associated weight.
- Weights convey the importance of that feature in predicting the final output.
- During training, neural networks learn optimal weights to minimise error.

Why we use activation functions?

- Activation functions in neural networks are like decision-making.
- The activation function decides whether a neurone should be activated or not for sending a signal based on input received.
- There are different types of activation functions, like the Sigmoid Function, Tanh Function, ReLU (Rectified Linear Unit), etc.
- Activation functions allow non-linearities in the model, which can help to understand complex patterns in data and decision-making.

What is the purpose of bias?

- Bias terms are constants added to the weighted sum before applying the activation function.
- They allow the network to shift the activation function horizontally.
- Bias helps account for any inherent bias in the data.

What is the purpose of connections in neural networks?

- Connections represent the synapses between neurons.
- Each connection has an associated weight.
- Biases (constants) are also associated with each neuron, affecting its activation threshold.

What is the purpose of Learning Rule?

- Neural networks learn by adjusting their weights and biases.
- The learning rule specifies how these adjustments occur during training.
- Back propagation, a common learning algorithm, computes gradients and updates weights to minimize the network's error.

What do you mean by propagation Functions?

Propagation helps to move data from one layer to another layer where one becomes the input of the next layer and another layer becomes the output layer. The most common types are "forward propagation," where the data flows from the input layer to the output layer to make predictions, and "backpropagation," where the error signal is calculated and used to update the network weights during training, which allows the machine to learn from its mistakes.

How the Neural Network works?

A neural network is a type of machine learning model work similar to human brain. The neural network contain layers or nodes, each node perform some simple mathematical operation. The layers in neural network are –

- **Input Layer:** Take input from the real world
- **Hidden Layer:** Process the inputs using weights, biases and activation function.
- **Output Layer:** Generate the final output.

Lets understand how Neural Network perform a task

1. **Input Layer:** Each node takes an input value (x)
2. **Weights and Biases:** Each input is multiplies by a weight(w) and a bias(b).
   ∑wixi + bias = (????1 x ????1) + (????2 x ????2) + (????3 x ????3) + ????
   f(x) = 1 , if ∑w1x1 + b
3. **Activation function:** The result (z) passes through an activation function to determine the output (a).
   a = 1 if z >= 0
   a = 0 if z < 0
4. **Outputs** from one layer become inputs to the next layer.
   Lets see one example
   We have converted the example of the textbook to the restaurant analogy.
   Suppose a neural network as a restaurant where:
   Ingredients = **Input values (????)**
   Chefs = **Nodes/Neurons**
   Recipe = **Weights (????) and Bias (????)**
   Taste Test = **Activation function**
   Serving = **Output**
   Let us see a simple problem.
   CASE I: Let the features be represented as x1,x2 and x3.
   Input Layer:
   Feature 1, x1 = 2 (2 tomatoes)
   Feature 2, x2 = 3 (3 onions)
   Feature 3, x3 = 1(1 carrot)
   Hidden Layer:
   Weight 1, w1 = 0.4 (0.4 teaspoons of salt per tomato)
   Weight 2, w2 = 0.2 (0.2 teaspoons of salt per tomato)
   Weight 3, w3 = 0.6 (0.6 teaspoons of salt per tomato)
   bias = 0.1 (extra seasoning)
   threshold = 3.0
   Output: Using the formula: (Suppose the total seasoning added by chefs can be calculated)
   ∑wixi + bias = w1x1 + w2x2 + w3x3 + bias
   = (0.42) + (0.23) + (0.6*1) + 0.1
   = 0.8 + 0.6 + 0.6 + 0.1
   = 2.1
   Now, we apply the threshold value (Activation function: The chefs taste the dish)
   If output > threshold, then output = 1 (active)
   If output < threshold, then output = 0 (inactive)
   In this case:
   Output (2.1) < threshold (3.0)
   So, the output of the hidden layer is:
   Output = 0
   This means that the neurone in the hidden layer is inactive, meaning the dish does not pass; the output is 0.
   CASE II
   Let's say we have another neuron in the output layer with the following weights and bias:

w1 = 0.7

w2 = 0.3

bias = 0.2

The output of the hidden layer (0) is passed as input to the output layer:

Output = w1 x x1 + w2 x x2 + bias

= 0.7 x 0 + 0.3 x 0 + 0.2

= 0.2

Let's assume the threshold value for the output layer is 0.1:

Output (0.2) > threshold (0.1)

So, the final output of the neural network is:

Output = 1

Types of Neural Networks

Neural networks are classified into different types based on their purposes. The most common types of neural networks are:

- Standard Neural Network (Perceptron)
- Feed Forward Neural Network (FFNN)
- Convolutional Neural Network (CNN)
- Recurrent Neural Network (RNN)
- Generative Adversarial Network (GAN)

1. Standard Neural Network (Perceptron)

Standard Neural Network which is known as Perceptron is a fundamental concept in machine learning. In standard Neural Network does't have a hidden layer. The perceptron created by Rosenblatt in 1958, Perceptron is a simple neural netowrk with single layer of input node connected to a layer of output nodes. It is uses threshold logic Unit as aftifical neurons.

Application: Useful in spam detection and basic decision-making.

2. Feed Forward Neural Network (FFNN)

FFNN is also known as multilayer perceptions. FFNN has an input layer, one or more hidden layers, and an output layer. In FFNN, data flow from the input layer to the output layer. FFNN uses activation functions and weights to process information; FFNN is capable of handling noisy data.

Application: used in tasks like image recognition, natural language processing (NLP), and regression.

3. Convolutional Neural Network (CNN)

CNNs are basically used in visual data. CNNs can extract the features from the image, which can help to recognise patterns and objects in the image. CNNs use three-dimensional data for image classification and object recognition tasks.

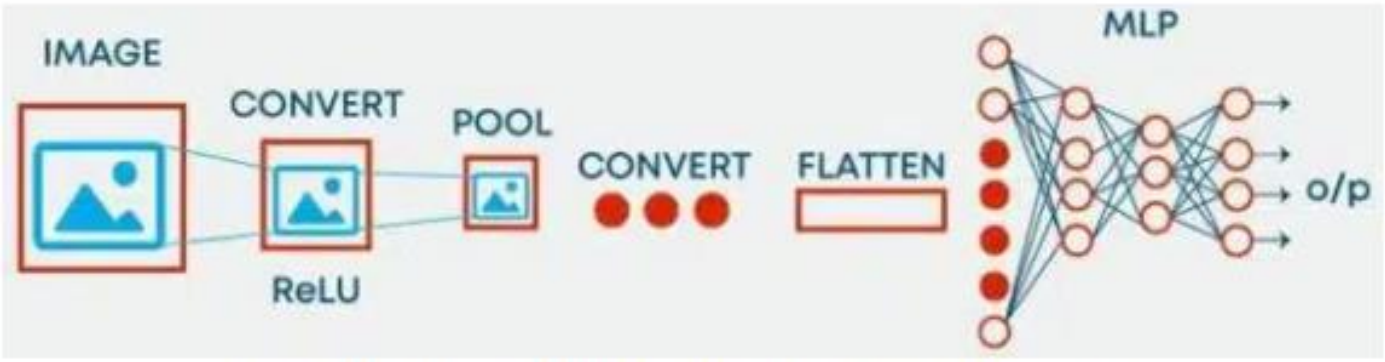


Image Source: https://www.mygreatlearning.com/blog/types-of-neural-networks/

Application: Dominant in computer vision for tasks such as object detection, image recognition, style transfer, and medical imaging

4. Recurrent Neural Network (RNN)

A recurrent neural network is designed for sequential data; this neural network deals with data that has a sequence, like a sentence or time series. RNNs have a connection loop that can go back to the previous layer, which is known as a feedback connection.

Application: Used in NLP for language modelling, machine translation, chatbots, as well as in speech recognition, time series prediction, and sentiment analysis.

5. Generative Adversarial Network (GAN)

GANs consist of two neural networks – a generator and a discriminator – trained simultaneously. The generator creates new data instances, while the discriminator evaluates them for authenticity. They are used for unsupervised learning and can generate new data samples. GANs are used for generating realistic data, such as images and videos.

Application: Widely employed in generating synthetic data for various tasks like image generation, style transfer, and data augmentation.

Future of NN and its impact on society

Neural networks improve themselves; they are advanced algorithms that can improve the productivity of the various industries like manufacturing, finance, and healthcare. Neural networks also have a capability for analysing large amounts of data and generating recommendations for individuals or for industry. Neural networks help in economic growth, and they can create new job opportunities in the field, like data science and artificial intelligence.

Neural networks have some concerns about ethical issues such as data privacy, bias in algorithms, and job displacement. It is critical to the programmer to handle these uses carefully and ensure that the neural networks benefit society as a whole.

MCQs

1. What is a neural network?

A) A set of interconnected resistors
B) A machine learning model inspired by the human brain
C) A network of physical neurons in the brain
D) A rule-based expert system
Answer: B

2. What are the three main layers of a neural network?

A) Input, Processing, Storage
B) Input, Hidden, Output
C) Training, Testing, Validation
D) Nodes, Edges, Weights
Answer: B

3. What is the primary function of the input layer in a neural network?

A) Store data permanently
B) Process and classify data
C) Receive and pass input data to the next layer
D) Generate output predictions
Answer: C

4. What is the role of weights in a neural network?
   A) Control activation functions
   B) Determine the importance of input features
   C) Define the number of layers in the network
   D) Store training data
   Answer: B

5. Which function is used to determine whether a neuron is activated?
   A) Cost function
   B) Activation function
   C) Loss function
   D) Training function
   Answer: B

6. Which of the following is NOT a common activation function?
   A) Sigmoid
   B) ReLU
   C) Tanh
   D) Linear Regression
   Answer: D

7. What is the main purpose of bias in a neural network?
   A) Improve the learning rate
   B) Offset the weighted sum to help with non-linearity
   C) Remove errors from training data
   D) Increase network complexity
   Answer: B
8. What is backpropagation?
   A) A technique to remove redundant neurons
   B) A method to adjust weights by minimizing errors
   C) A way to generate random predictions
   D) A visualization tool for deep learning models
   Answer: B

9. Which neural network is best suited for image recognition?
   A) Feedforward Neural Network
   B) Convolutional Neural Network (CNN)
   C) Recurrent Neural Network (RNN)
   D) Generative Adversarial Network (GAN)
   Answer: B

10. What is the distinguishing feature of Recurrent Neural Networks (RNNs)?
    A) Their ability to process sequential data
    B) Their efficiency in recognizing images
    C) Their use of filters to extract spatial features
    D) Their application in unsupervised learning
    Answer: A

11. What does a deep neural network (DNN) contain?
    A) One hidden layer
    B) More than three layers including input and output layers
    C) Only an input and output layer
    D) A fixed number of neurons
    Answer: B

12. Which algorithm is commonly used for optimization in neural networks?
    A) K-means clustering
    B) Gradient descent
    C) Decision trees
    D) Support vector machines
    Answer: B

13. In neural networks, what is overfitting?
    A) A network learning patterns that are too complex
    B) A network that learns only simple relationships
    C) A model that generalizes well
    D) A model with too few neurons
    Answer: A

14. What is the primary advantage of using CNNs over FNNs?
    A) CNNs can process sequential data
    B) CNNs are more efficient in recognizing spatial patterns
    C) CNNs do not require activation functions
    D) CNNs do not use weights and biases
    Answer: B

15. Which component helps a neural network “learn”?
    A) Neurons
    B) Activation functions
    C) Weights and biases
    D) Hidden layers
    Answer: C

16. What is dropout in neural networks?
    A) A method to remove outliers in training data
    B) A regularization technique to prevent overfitting
    C) A function to speed up training
    D) A way to optimize activation functions
    Answer: B

17. What is the purpose of an epoch in training neural networks?
    A) A single pass through the entire training dataset
    B) The process of tuning hyperparameters
    C) A measure of network complexity
    D) The number of neurons in a layer
    Answer: A

18. What is the primary application of Generative Adversarial Networks (GANs)?
    A) Predicting stock prices
    B) Generating synthetic data
    C) Language translation
    D) Time series forecasting
    Answer: B

19. Which tool is commonly used for building neural networks?
A) TensorFlow B) Excel C) Tableau D) SQL
Answer: A

20. What is a perceptron?
A) A basic unit of deep learning B) A single-layer neural network
C) A recurrent neural network D) A convolutional layer
Answer: B

21. What does the loss function in a neural network measure?
A) The network's accuracy
B) The error between predicted and actual values
C) The number of training examples
D) The total number of neurons
Answer: B

22. Which type of neural network is best for natural language processing?
A) CNN B) FNN C) RNN D) Perceptron
Answer: C

23. What is the key feature of a feedforward neural network?
A) It has cyclic connections B) Data flows only in one direction
C) It uses backpropagation D) It operates without activation functions
Answer: B

24. What type of dataset is best suited for RNNs?
A) Images B) Time-series data C) Tabular data D) Static graphs
Answer: B

25. Which of the following best describes the term "training data"?
A) The dataset used for final evaluation B) The data used to adjust weights and biases
C) A dataset stored in a neural network D) The output of a trained model
Answer: B

Short Answer types Questions with answer:

1. What is a neural network, and how does it work?
Answer: A neural network is a machine learning model inspired by the structure of the human brain. It consists of interconnected nodes (neurons) organized in layers: an input layer, one or more hidden layers, and an output layer. Each neuron receives input, processes it using weights and an activation function, and passes the result to the next layer. The network is trained using algorithms like backpropagation, adjusting weights to minimize error and improve predictions over time.

---

2. Explain the importance of activation functions in neural networks.

Answer: Activation functions introduce non-linearity to a neural network, allowing it to model complex patterns in data. Without them, the network would behave like a linear regression model, unable to capture intricate relationships. Common activation functions include:

- Sigmoid: Maps inputs between 0 and 1, useful for binary classification.
- ReLU (Rectified Linear Unit): Helps avoid vanishing gradient issues by setting negative values to zero.
- Tanh: Similar to sigmoid but outputs between -1 and 1, allowing stronger gradients for learning.

---

3. What are the key components of a neural network?
Answer: A neural network consists of:

1. Neurons (Nodes): The fundamental units that process and transmit information.
2. Weights: Values assigned to connections between neurons to determine importance.
3. Bias: A constant added to a neuron's weighted sum before activation.
4. Layers:

- Input layer: Receives raw data.
- Hidden layers: Process data through transformations.
- Output layer: Produces final predictions.

5. Activation Functions: Decide whether a neuron should be activated based on inputs.
6. Loss Function: Measures the network's prediction error.
7. Optimizer: Adjusts weights to reduce loss.

---

4. Explain the difference between a feedforward neural network and a recurrent neural network.
Answer:

- Feedforward Neural Network (FNN): Data flows in one direction, from input to output. There are no loops or feedback connections, making it suitable for static data like image classification.
- Recurrent Neural Network (RNN): Designed for sequential data, RNNs use feedback loops to retain information from previous steps, making them ideal for time-series analysis, speech recognition, and language modeling.

---

5. How does backpropagation work in training a neural network?
Answer: Backpropagation is an optimization algorithm used to train neural networks. It consists of:

1. Forward Propagation: Inputs pass through the network, and an output is generated.
2. Loss Calculation: The difference between the predicted and actual values is measured using a loss function.
3. Backward Propagation: The error is propagated back through the layers to adjust weights using gradient descent.
4. Weight Update: The optimizer updates the weights to reduce the loss. This process repeats until the network achieves satisfactory accuracy.

---

6. What are convolutional neural networks (CNNs), and why are they used in image processing?
Answer: CNNs are a type of neural network designed for image recognition tasks. They use convolutional layers to extract spatial features, pooling layers to reduce dimensionality, and fully connected layers for classification. CNNs efficiently detect patterns such as edges, shapes, and textures, making them ideal for applications like facial recognition, object detection, and medical imaging.

---

7. What are generative adversarial networks (GANs), and how do they work?

Answer: GANs are a type of neural network used for generating synthetic data. They consist of two networks:

- Generator: Creates new data samples.
- Discriminator: Evaluates if the generated data is real or fake.

Both networks train simultaneously, with the generator improving its output to fool the discriminator. GANs are used for image generation, style transfer, and data augmentation.

---

8. What are some common challenges in training deep neural networks?

Answer:

1. Overfitting: The network memorizes training data instead of generalizing.
2. Vanishing/Exploding Gradients: Gradient values become too small (vanish) or too large (explode), hindering learning.
3. High Computational Cost: Training deep networks requires substantial processing power and memory.
4. Hyperparameter Tuning: Selecting optimal parameters like learning rate and batch size can be complex.
5. Lack of Training Data: Deep networks require large datasets for accurate predictions.

---

9. What is the significance of the learning rate in neural networks?

Answer: The learning rate controls how much the weights are updated during training.

- Too high: The network may not converge and overshoot the optimal solution.
- Too low: Training becomes slow, and the network may get stuck in local minima.

Optimizers like Adam and SGD help adjust the learning rate dynamically.

---

10. How do neural networks handle non-linearity?

Answer: Neural networks introduce non-linearity through activation functions like ReLU, Sigmoid, and Tanh. Without these functions, the network would only be capable of learning linear relationships, limiting its ability to capture complex patterns in data.

---

11. What is dropout, and why is it used in neural networks?

Answer: Dropout is a regularization technique where random neurons are ignored during training. This prevents overfitting by ensuring that the network does not become too reliant on specific neurons, improving generalization on unseen data.

---

12. How do Recurrent Neural Networks (RNNs) process sequential data?

Answer: RNNs retain information from previous inputs using a feedback loop. They maintain a hidden state that updates at each time step, allowing them to recognize patterns over time. This makes RNNs useful for tasks like speech recognition and time-series forecasting.

---

13. How does batch normalization improve neural network training?

Answer: Batch normalization normalizes the input of each layer, reducing internal covariate shift. This stabilizes training, speeds up convergence, and allows for higher learning rates, improving model accuracy.

---

14. What is transfer learning, and why is it useful?

Answer: Transfer learning involves using a pre-trained model on a new task. Instead of training from scratch, an existing model trained on a large dataset (e.g., ImageNet) is fine-tuned for a specific application, reducing training time and improving performance with limited data.

---

15. What is the role of pooling layers in CNNs?
Answer: Pooling layers reduce the spatial dimensions of feature maps, decreasing computational complexity and making the network more robust to variations in input images. Common types include max pooling and average pooling.

Competency-Based Question on Neural Networks Application

Question 1: You are working as a data scientist for a healthcare company that wants to build an AI model to detect pneumonia from chest X-rays. Which type of neural network would you choose, and why?

Answer:

A Convolutional Neural Network (CNN) would be the best choice because it is specifically designed for image processing tasks. CNNs use convolutional layers to extract important spatial features such as edges, textures, and patterns from medical images. Pooling layers help reduce dimensionality, making computation more efficient. This enables the model to accurately detect pneumonia in chest X-rays by learning patterns in medical imaging datasets.

Question 2: You are building a predictive text system for a smartphone keyboard. Which type of neural network would you use and how would it work?

Answer:

A Recurrent Neural Network (RNN) would be the best choice because it is designed for sequential data processing. RNNs retain memory of previous words in a sequence through feedback loops, allowing them to predict the next word based on context. An advanced variant, Long Short-Term Memory (LSTM) networks, can handle long-term dependencies better, improving accuracy in text prediction. The model would be trained on large datasets of text to learn word relationships and sentence structures.

Question 3: A company has developed an AI hiring system that uses a neural network to screen job applications. However, it has been found that the system favors certain demographics over others. What could be causing this issue, and how can it be addressed?

Answer:

The issue is likely due to algorithmic bias, which occurs when the training data used for the neural network is imbalanced or reflects human biases. If the historical hiring data includes biases against certain demographics, the neural network will learn and replicate those biases.

To address this:

1. Use Diverse and Representative Data: Ensure training data includes candidates from all demographics.
2. Regular Audits: Continuously test the model for biased outcomes and retrain as needed.
3. Fairness Constraints: Implement fairness-aware learning techniques to reduce bias.
4. Transparency: Make hiring criteria explainable and interpretable to avoid hidden biases.

Question 4: A company has developed a neural network model that performs exceptionally well on training data but fails on real-world test data. What problem is the model facing, and how can it be resolved?

Answer:

The model is suffering from overfitting, meaning it has memorized the training data instead of learning general patterns. This reduces its ability to perform well on unseen data.

To prevent overfitting:

1. Use Dropout: Randomly deactivate neurons during training to prevent dependency on specific features.
2. Increase Training Data: More diverse data helps the model generalize better.
3. Regularization (L1/L2): Penalizes large weight values, preventing the model from becoming too complex.
4. Cross-Validation: Splitting data into multiple sets ensures the model performs consistently.

Question 5: You are tasked with building a temperature conversion model that predicts Fahrenheit values based on Celsius input. How would you implement this using a neural network?

Answer:

To build this model using a neural network in TensorFlow/Keras, follow these steps:

1. **Prepare Training Data:**
- Create a dataset of Celsius values (X) and corresponding Fahrenheit values (Y).
- Example: X = [-40, -10, 0, 8, 15, 22, 38], Y = [-40, 14, 32, 46, 59, 72, 100].
2. **Define the Neural Network:**
- Use a simple feedforward neural network with one input layer, one hidden layer, and one output layer.
- model = tf.keras.Sequential([tf.keras.layers.Dense(units=1, input_shape=[1])])
3. **Compile the Model:**
- Use Mean Squared Error as the loss function and Adam optimizer for training.
- model.compile(loss='mean_squared_error', optimizer=tf.keras.optimizers.Adam(0.1))
4. **Train the Model:**
- model.fit(X, Y, epochs=500, verbose=False)
5. **Make Predictions**:
- Convert new values: print(model.predict([100])) should approximate 212°F.

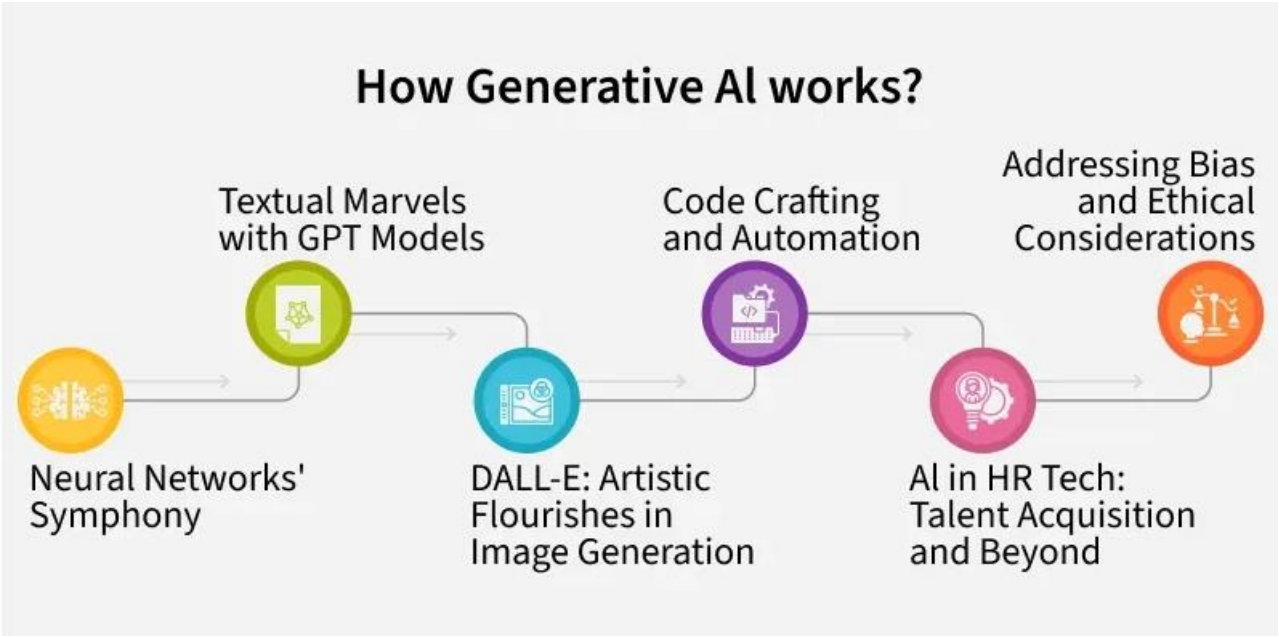
UNIT-7 Generative AI

Introduction to Generative AI

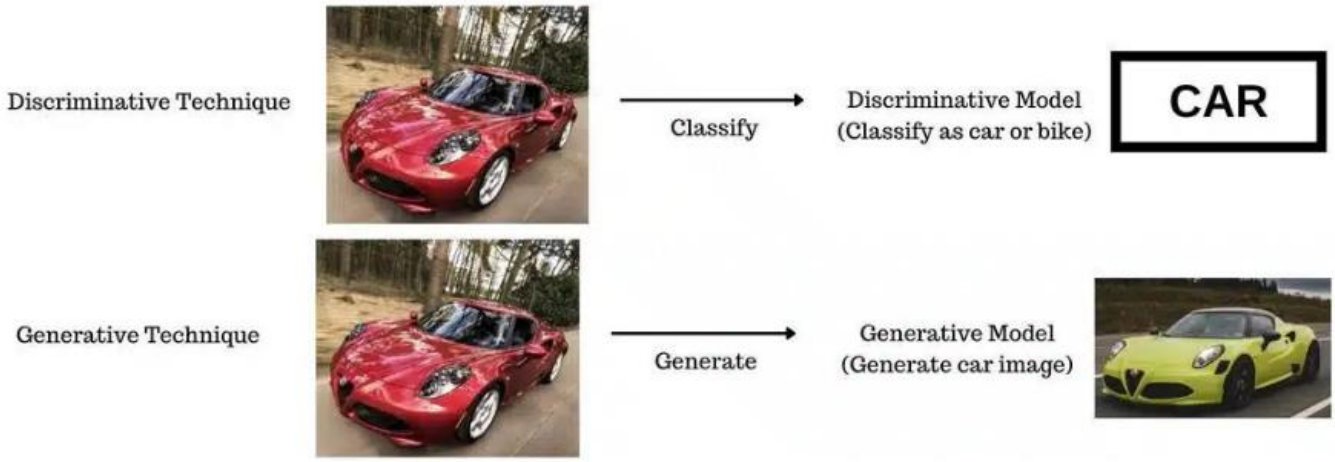
Generative AI is a subset of artificial intelligence that helps to create new content such as text, images, audio, and video. Generative AI uses machine learning algorithms to learn from existing datasets and generate similar content. Generative AI is used in various fields, like in education, art and design, film and animation, music, gaming, healthcare, etc. But there are some challenges of generative AI; this technology can be used for creating fake images, misusing content, etc. Examples of generative AI tools are ChatGPT, Gemini, Copilot, and DALL-E, etc.

Working of Generative AI

Generative AI uses neural networks to create new content based on existing data. The generative AI can generate text, images, audio, and videos based on the previous data. Generative Adversarial Networks and Variational Auto encoders are the most popular models in the field of Generative AI. These two models are powerful model and play an important role in generating text, images, audio, and video file.



1. Generative Adversarial Networks (GANs): A Generative Adversarial Network is a deep learning architecture that consists of two neural networks, a generator and a discriminator, to compete against each other to generate more authentic data based on a training dataset. The generator helps to generate new data samples such as images or text and discriminator helps to differentiate between real and fake data. GANs are applied in various domains like image generation, style transfer, and data augmentation.

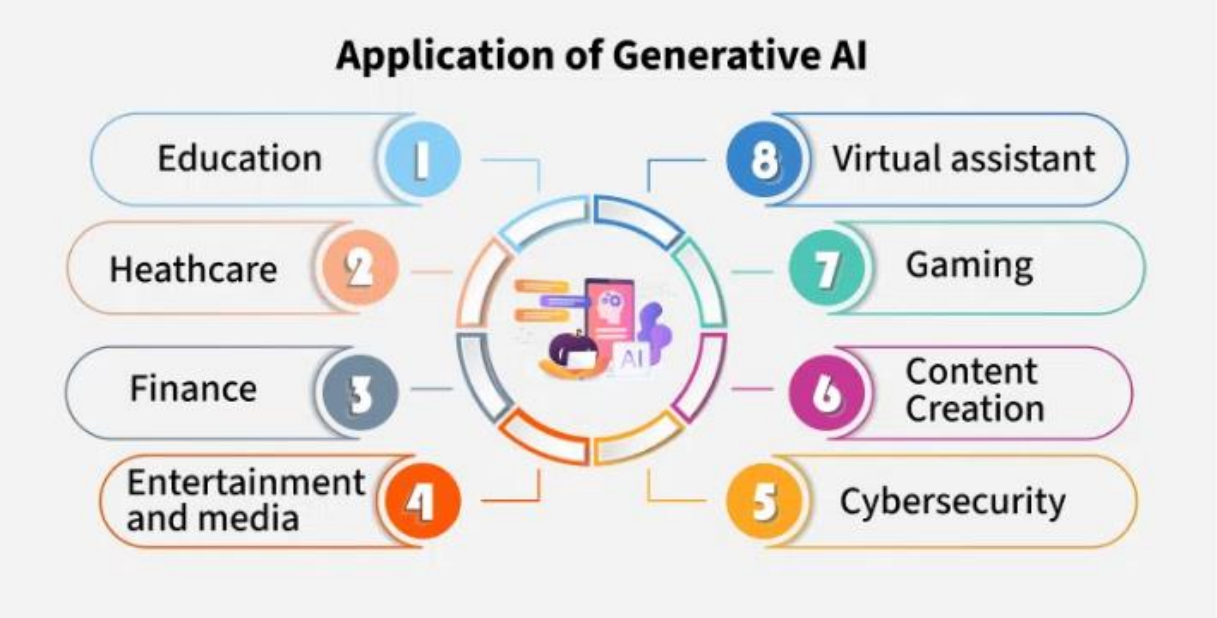


2. Variational Autoencoders (VAEs): Variational autoencoders are a machine learning algorithm that is designed to generate data uniquely. Variational autoencoders can improve the quality of the image, you can create an image using text. Variational autoencoders consist of two parts: an encoder and a decoder. The encoder understands the data and converts it into a hidden space (lower-dimensional space) called a latent space, and the decoder translates the information back from this hidden space into its original form. Variational autoencoders are used in different applications like data generation, detecting anomalies in data and filling missing information.

Differences between Generative AI and Discriminative AI

	Generative AI	**Discriminative AI**
Purpose	It is used for generate new content like text, image, audio and video.	It is used for categorizes data into specific group based on the existing data.
Models	Use tricks to guesses based on the patterns to create new things.	Find rules to separate and recognize patterns
Training Focus	Tries to understand what makes data unique and how to create new similar data	Focuses on learning how to draw lines or make rules
Application	create new artworks, generate new ideas for stories, and find unusual patterns in data.	Powers things like facial and speech recognition and helps make decisions like whether an email is spam or not.
Examples of Algorithms used	Naïve Bayes, Gaussian discriminant analysis, GAN, VAEs, LLM, DBMs, Autoregressive models	Logistic Regression, Decision Trees, SVM, Random Forest

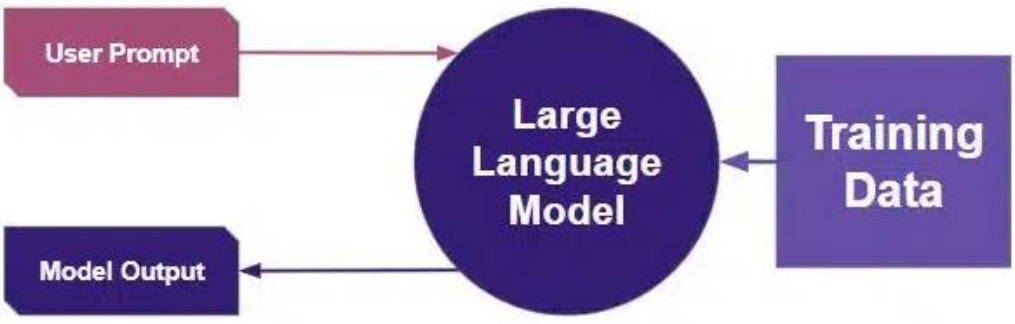
Applications of Generative AI



- **Image Generation:** It involves creating new images based on patterns learned from existing datasets. These models analyse the characteristics of input images and generate new ones.
- **Text Generation:** Text generation is when computers write sentences that sound like people wrote them. It involves creating written content that mimics human language patterns.
- **Video Generation:** It involves creating new videos by learning from existing ones, including animations and visual
  effects. These models learn from videos to create realistic and unique visuals.
- **Audio Generation:** Audio generation involves computers producing new sounds, such as music or voices, based on sounds they have heard.

LLM- Large Language Model

A Large Language Model is a deep learning algorithm that can generate and classify text, answer the questions and translate text from one language to another language. LLMs are called large because this type of algorithm is trained on a large dataset of text and code.



Transformers in LLMs

Transformers is a type of neural network that processes sequential data like text or image, transformers are a fundamental of LLMs enabling efficient and effective learning of complex language patterns and relationships within vast amounts of text data.

Some leading Large Language Models (LLMs) are:

- **OpenAI's GPT-4o:** OpenAI GPT work on machine learning principles. It is a pretrained transformer model that can understand and generate content in the form of text and images.

- **Google's Gemini 1.5 Pro:** Gemini has multimodal capabilities, which can process and analyse complex audio inputs, generate new code, suggest improvements etc.
- **Meta's LLaMA 3.1:** Meta LLaMA is an open-source and powerful language model. It is a large language model (LLMs) that is used for reading handwriting, creating graphs and charts, and use in research and commercial purposes.
- **Anthropic's Claude 3.5:** Claude is a powerful and intelligent AI model that has more natural and engaging dialogue; it can generate text and write code.
- **Mistral AI's Mixtral 8x7B:** Mistral AI models are powerful language processing tools that are used in chatbots, customer support, language translation, and content creation.

Applications of LLMs:

- **Text Generation:** AI applications can generate meaningful text based on the user input, like dialogue generation, story writing, content creation and poetry generation. Some other examples include.
- Can translate the natural language
- Code writing
- Autocompleting text and generating continuation for sentences or paragraphs
- Examil auto-completion
- **Audio Generation:** LLMs are not capable for generating audio signals directly, but they are capable of generating text to speech. LLMs enable TTS systems to synthesise natural-sounding speech from text inputs.
- **Image Generation:** LLMs can generate new images using text prompts. LLMs understand the visual content to produce textual descriptions and relevant images.
- **Video Generation:** LLMs can generate video based on a script or using textual descriptions. LLMs are used for generating subtitles, captions, or scene summaries for video.

Limitations of LLM:

- LLMs are expensive and slow, LLMs need lots of computing power to run on computers.
- LLMs can generate incorrect or misleading information.
- LLMs cannot adapt to new situations easly, LLMs are not capable of understanding real things.

Risks associated with LLM:

- LLMs learn from the internet, they can adopt any harmful biases.
- LLMs can disclose the personal information of the user.
- If the LLMs trained using sensitive information, then accidentally reveal the confidential details.

Future of Generative AI

Generative AI will address complex challenges in fields like healthcare and education, enhance NLP tasks like multilingual translation, and expand in multimedia content creation. Collaboration between humans and AI will deepen, emphasizing AI's role as a supportive partner across domains.

- **Ethical and Social Implications of Generative AI:** Generative AI, with its ability to create realistic content such as images, videos, and text, brings about a multitude of ethical and social considerations. An ethical and social implication of generative AI becomes crucial for ensuring responsible development and deployment.
- **Deepfake Technology:** The emergence of deepfake AI technology, such as DeepFace Lab and FaceSwap, raises concerns about the authenticity of digital content. Deepfake algorithms can generate compelling fake images, audio, and videos. Examples: Deepfake AI tools, such as DeepArt's style transfer algorithms, can seamlessly manipulate visual content, creating deceptiv and misleading media.
- **Bias and Discrimination:** Generative AI models, exemplified by Clearview AI's facial recognition algorithms, have demonstrated biases that disproportionately affect certain demographic groups. Examples: The AI-powered hiring platform developed by HireVue has faced criticism for perpetuating recruitment bias.
- **Plagiarism:** Presenting AI-generated content as one's work, whether intentionally or unintentionally, raises ethical questions regarding intellectual property rights and academic integrity.
- **Transparency:** Transparency in the use of generative AI is paramount to maintaining trust and accountability. Disclosing the use of AI-generated content, particularly in academic and professional settings, is essential to uphold ethical standards and prevent instances of academic dishonesty.

Points to Remember:

- Be cautious and transparent when using generative AI.
- Respect copyright and avoid presenting AI output as your own.
- Consult your teacher/institution for specific guidelines.

MCQs:

1. What is Generative AI primarily used for?

a) Classifying data

b) Generating new content similar to training data

c) Performing mathematical calculations

d) Optimizing database performance
Answer: b) Generating new content similar to training data

2. Which of the following is NOT a type of Generative AI application?
a) Image generation
b) Video generation
c) Data classification
d) Audio generation
Answer: c) Data classification

3. What distinguishes Generative AI from Discriminative AI?
a) Generative AI creates new data, while Discriminative AI classifies existing data
b) Generative AI identifies patterns in labeled data
c) Discriminative AI generates new data samples
d) Discriminative AI does not use machine learning
Answer: a) Generative AI creates new data, while Discriminative AI classifies existing data

4. Which of the following is an example of Generative AI?
a) ChatGPT
b) Decision Trees
c) Linear Regression
d) Naïve Bayes Classifier
Answer: a) ChatGPT

5. What is the role of Large Language Models (LLMs) in AI?
a) Creating and classifying text
b) Sorting numerical datasets
c) Translating programming languages
d) Running optimization algorithms
Answer: a) Creating and classifying text

6. Generative Adversarial Networks (GANs) consist of which two components?
a) Generator and Discriminator
b) Encoder and Decoder
c) Classifier and Regressor
d) Training and Testing Modules
Answer: a) Generator and Discriminator

7. Which AI model is best suited for creating realistic images?
a) GANs
b) SVM
c) Decision Trees
d) Logistic Regression
Answer: a) GANs

8. What is the primary function of a Discriminative AI model?
a) Generate new data
b) Differentiate between data categories
c) Create images
d) Write human-like text
Answer: b) Differentiate between data categories

9. What does a Variational Autoencoder (VAE) do?
a) Translates languages
b) Encodes and decodes data into latent space
c) Identifies spam emails
d) Performs clustering
Answer: b) Encodes and decodes data into latent space

10. Which company developed Gemini AI?
a) OpenAI b) Google
c) Microsoft d) Meta
Answer: b) Google

11. Deepfakes are created using which technology?
a) GANs b) Reinforcement Learning
c) Decision Trees d) Clustering
Answer: a) GANs

12. Which ethical concern is associated with Generative AI?
a) Data bias b) Computational efficiency
c) Reduced storage capacity d) Increased energy consumption
Answer: a) Data bias

13. What is the main drawback of LLMs?
a) Lack of computational power b) Generation of factually incorrect information
c) Inability to process text d) No real-world applications
Answer: b) Generation of factually incorrect information

14. Which AI technique is used in music generation?
a) Voicebox b) Decision Trees
c) Support Vector Machines d) K-Nearest Neighbors
Answer: a) Voicebox

15. Which tool is NOT an example of Generative AI?
a) Canva b) ChatGPT
c) Google Sheets d) DALL-E
Answer: c) Google Sheets

16. Which Generative AI tool specializes in video creation?
a) Lumiere b) Perplexity AI
c) MidJourney d) SVM
Answer: a) Lumiere

17. What do Transformers do in LLMs?
a) Process large amounts of text data efficiently b) Convert images into text
c) Generate deepfake videos d) Predict stock market trends
Answer: a) Process large amounts of text data efficiently

18. What is one way to prevent biases in AI-generated content?
a) Use diverse training data b) Reduce dataset size
c) Only use labeled data d) Avoid deep learning models
Answer: a) Use diverse training data

19. What is the main concern with AI-generated text in academic writing?
a) Transparency b) Storage issues
c) Speed of text generation d) Programming complexity
Answer: a) Transparency
20. Which model is optimized for text-based prompts?
a) Gemini-Pro b) Gemini-Vision
c) GANs d) DBMs
Answer: a) Gemini-Pro

21. Which company created LLaMA?
a) Google b) OpenAI
c) Meta d) Microsoft
Answer: c) Meta

22. What does “latent space” refer to in AI?
a) Compressed data representation b) The output layer of neural networks
c) Memory storage d) The final decision boundary
Answer: a) Compressed data representation

23. Which AI model is best for detecting anomalies in data?
a) VAEs b) Decision Trees
c) Clustering Algorithms d) Linear Regression
Answer: a) VAEs

24. How do chatbots use Generative AI?
a) To generate responses dynamically b) To classify customer queries
c) To analyze user preferences d) To perform sentiment analysis
Answer: a) To generate responses dynamically

25. Which AI model is best for summarizing text?
a) GPT-4 b) CNN
c) Random Forest d) Decision Trees
Answer: a) GPT-4

26. What is an example of AI-generated plagiarism?
a) Using AI-generated text without citation b) Writing an essay manually
c) Summarizing an article d) Using a spell-checker
Answer: a) Using AI-generated text without citation

27. What type of learning is used in Generative AI?
a) Unsupervised b) Supervised
c) Reinforcement d) Semi-Supervised
Answer: a) Unsupervised

28. Which is a disadvantage of AI-generated content?
a) Lack of originality b) Increased efficiency
c) Faster content creation d) Lower resource usage
Answer: a) Lack of originality

29. What is the purpose of Gemini API?
a) Creating chatbots b) Making images
c) Processing video files d) Running SQL queries
Answer: a) Creating chatbots

30. Which AI model is used in face recognition?
a) CNNs b) LSTMs
c) Decision Trees d) GANs
Answer: a) CNNs

31. What is a major risk of using Generative AI in media?
a) Overproduction of content b) Creation of misinformation and deepfakes
c) Increased server storage requirements d) Slower internet speeds
Answer: b) Creation of misinformation and deepfakes

32. Which is an example of AI-generated video modification?
a) Deepfake b) Decision Tree Algorithm
c) Logistic Regression d) Neural Style Transfer
Answer: a) Deepfake

33. What makes Generative AI useful in education?
a) It can generate personalized learning content b) It replaces human teachers
c) It prevents plagiarism d) It reduces internet usage
Answer: a) It can generate personalized learning content

34. Which AI tool helps generate marketing copy?
a) ChatGPT b) Tableau
c) Microsoft Excel d) AutoML
Answer: a) ChatGPT
35. What is the primary ethical concern of AI-generated job applications?
a) Biased hiring decisions b) Slow application processing
c) Increased hiring costs d) Lack of job opportunities
Answer: a) Biased hiring decisions

36. How do LLMs process human language effectively?
a) By using deep learning and transformers b) By translating text into binary code
c) By manually tagging words in a dataset d) By relying on human input for responses
Answer: a) By using deep learning and transformers

37. Which AI tool can create realistic voiceovers?
a) Voicebox
b) Gemini API
c) OpenAI Codex
d) Canva
Answer: a) Voicebox

38. How can companies prevent Generative AI from spreading false information?
a) Implement AI-generated content verification systems
b) Limit AI-generated content
c) Disable AI usage
d) Stop using neural networks
Answer: a) Implement AI-generated content verification systems

39. What does multimodal AI refer to?
a) AI that can process and generate multiple types of data (text, image, audio)
b) AI that only works with numerical data
c) AI that does not use neural networks
d) AI that processes data but does not generate content
Answer: a) AI that can process and generate multiple types of data (text, image, audio)

40. What is the main application of LLaMA models?
a) Text generation and NLP tasks
b) Image classification
c) Self-driving cars
d) Spam detection
Answer: a) Text generation and NLP tasks

41. Why is transparency important in AI-generated content?
a) To ensure accountability and credibility
b) To make AI processing slower
c) To reduce the number of AI-generated texts
d) To make it easier for AI to generate responses
Answer: a) To ensure accountability and credibility

42. What is an effective way to detect deepfake videos?
a) AI-based detection tools
b) Manual checking only
c) Avoiding all online videos
d) Using image compression techniques
Answer: a) AI-based detection tools

43. Why is data diversity important in AI model training?
a) To reduce bias and improve accuracy
b) To decrease the model's computational power
c) To make AI models work slower
d) To prevent AI from generating text
Answer: a) To reduce bias and improve accuracy

44. What is a key benefit of AI-generated content in business?
a) Faster content creation for marketing
b) Eliminates the need for human employees
c) Increases hardware costs
d) Reduces the internet's processing speed
Answer: a) Faster content creation for marketing

45. How can Generative AI improve customer service?
a) By creating AI-powered chatbots for instant responses
b) By eliminating the need for customer support teams
c) By generating long wait times
d) By reducing the number of customers
Answer: a) By creating AI-powered chatbots for instant responses

46. What is an example of AI-generated plagiarism?
a) Using AI-generated text without attribution
b) Writing essays by hand
c) Using a calculator to solve math problems
d) Watching online tutorials
Answer: a) Using AI-generated text without attribution

47. Which AI tool can generate realistic human-like conversations?
a) Claude 3.5 b) Canva
c) Microsoft Excel d) Tableau
Answer: a) Claude 3.5

48. What should be done to ensure ethical AI usage in the workplace?
a) Implement AI guidelines and transparency policies
b) Ban all AI tools
c) Reduce AI model training
d) Remove AI from the hiring process
Answer: a) Implement AI guidelines and transparency policies

49. What is one concern of using Generative AI in news writing?
a) The spread of misinformation b) Faster reporting
c) Increased reader engagement d) Lower writing costs
Answer: a) The spread of misinformation
50. How can AI-generated images be distinguished from real ones?
a) Using AI detection software b) Analyzing the image manually
c) Avoiding online images d) Decreasing screen brightness
Answer: a) Using AI detection software

Short Answer Questions:

1. **What is Generative AI, and how does it work?**
Answer:
Generative AI is a branch of artificial intelligence that creates new content, such as text, images, audio, and video, based on training data. It learns patterns from existing datasets using machine learning models like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). These models analyze data and generate new samples that resemble the training data, making AI capable of creative tasks like content writing, music composition, and deepfake creation.

2. **How do Generative AI models differ from Discriminative AI models?**

Answer:

Generative AI models aim to learn the data distribution and generate new content, while Discriminative AI models focus on distinguishing between different classes of data.

- Generative models: Create new data samples (e.g., GANs, VAEs, LLMs).
- Discriminative models: Classify data into predefined categories (e.g., Logistic Regression, SVM, Decision Trees).

For example, Generative AI can create an image of a cat, while Discriminative AI can classify whether an image contains a cat or not.

---

3. **What are Generative Adversarial Networks (GANs) and how do they function?**

Answer:

GANs are a type of deep learning model consisting of two neural networks:

- Generator: Creates new data samples that resemble real data.
- Discriminator: Evaluates and differentiates between real and generated data.

These two networks compete with each other in an adversarial process. Over time, the Generator improves, producing increasingly realistic outputs. GANs are widely used for image generation, deepfakes, and style transfer.

---

4. **What is a Variational Autoencoder (VAE), and how is it used in Generative AI?**

Answer:

A Variational Autoencoder (VAE) is a neural network used in Generative AI to learn data representations efficiently. It consists of:

- Encoder: Compresses input data into a latent space (a condensed representation).
- Decoder: Reconstructs the original data from this latent space.

Unlike GANs, VAEs focus on capturing the underlying structure of data rather than just generating realistic images. They are useful in tasks like image synthesis, anomaly detection, and missing data reconstruction.

---

5. **What are Large Language Models (LLMs), and how do they contribute to Generative AI?**

Answer:

Large Language Models (LLMs) are deep learning models trained on massive text datasets to perform various Natural Language Processing (NLP) tasks. They can generate human-like text, translate languages, answer questions, and assist in creative writing. Examples include GPT-4, Gemini, and Claude 3.5. LLMs use transformer architectures, enabling them to process and understand contextual relationships in text efficiently.

---

6. **How do transformers improve the performance of Large Language Models (LLMs)?**

Answer:

Transformers are neural network architectures designed to process large sequences of text efficiently. They use mechanisms like:

- Self-attention: Allows the model to focus on important words in a sentence while generating output.
- Positional encoding: Helps retain word order information.
- Parallel processing: Improves speed by handling multiple words simultaneously.

These features enable LLMs like GPT and Gemini to generate coherent and contextually relevant text.

7. **What are some real-world applications of Generative AI?**

Answer:

Generative AI has diverse applications, including:

- Image Generation: Tools like DALL-E create realistic images from text prompts.
- Text Generation: ChatGPT and Gemini generate human-like responses in conversations.
- Audio Generation: AI models like Voicebox generate synthetic voices and music.
- Video Creation: AI tools like Lumiere generate videos based on text descriptions.
- Medical Research: AI-generated synthetic data aids in disease prediction and diagnosis.

---

8. **What are deepfakes, and why are they considered a major ethical concern?**

Answer:

Deepfakes are AI-generated videos or images that manipulate real visuals to create highly realistic yet fake content. They use GANs to replace faces, voices, or actions in media. Ethical concerns include:

- **Misinformation and Fraud**: Deepfakes can spread false information, leading to public deception.
- **Privacy Violations**: Fake videos can damage personal reputations.
- **Political Manipulation**: Altered media can influence elections and public opinion.

To counteract deepfakes, AI detection tools and strict regulations are essential.

---

9. **How can biases in Generative AI models be addressed?**

Answer:

Bias in Generative AI arises when models are trained on imbalanced or non-representative data. Solutions include:

- **Diverse Training Data**: Ensuring datasets include varied demographics and perspectives.
- **Algorithm Auditing**: Regularly evaluating AI outputs for unintended biases.
- **Human Oversight:** Involving human reviewers to assess AI-generated content.
- **Transparency**: Making AI decision-making processes understandable and open to scrutiny.

---

10. **What is the Gemini API, and how can it be used to create AI-powered chatbots?**

Answer:

The Gemini API is an AI-powered tool developed by Google that allows developers to create advanced chatbots. Steps to build a chatbot using Gemini API:

1. Obtain an API Key: Register on Google AI Studio and acquire an API key.
2. Set Up Python Environment: Install the Gemini API package.
3. Initialize the API: Connect to the Gemini model using authentication.
4. Develop a Chat System: Implement user interactions using generative AI.
5. Deploy the Chatbot: Use it for customer support, education, or research applications.

---

11. **What are some limitations of Generative AI?**

Answer:

Despite its capabilities, Generative AI has limitations:

- High Computational Cost: Requires powerful hardware for processing.
- Data Dependency: Quality depends on the training dataset.
- Bias and Misinformation: Can generate false or biased information.

- Lack of Originality: AI-generated content may lack genuine creativity.

To improve AI-generated outputs, continuous model refinement and ethical AI practices are necessary.

---

12. **What are the risks associated with Large Language Models (LLMs)?**

Answer:

Key risks of LLMs include:

- Data Privacy Issues: Models may memorize and leak sensitive data.
- Bias and Ethical Concerns: AI-generated content may reflect societal biases.
- Misinformation: AI can produce incorrect but confident-sounding answers.
- Security Threats: AI models can be exploited for cybercrimes and deepfake scams.

Proper regulation, transparency, and ethical AI development can help mitigate these risks.

---

13. **What ethical considerations should be made when using Generative AI in content creation?**

Answer:

- Plagiarism Prevention: Clearly label AI-generated content and credit original authors.
- Data Transparency: Disclose training data sources to ensure credibility.
- Regulatory Compliance: Follow copyright laws and fair usage policies.
- Bias Mitigation: Ensure content does not perpetuate stereotypes or discrimination.

---

14. **How is Generative AI used in personalized marketing and advertising?**

Answer:

Generative AI helps in:

- Content Generation: AI creates personalized ad copy and visuals.
- Customer Interaction: Chatbots assist customers with tailored recommendations.
- A/B Testing: AI optimizes ad campaigns by generating multiple variations.
- Predictive Analytics: AI analyzes consumer behavior for targeted marketing.

---

15. **How can AI-generated media be identified and prevented from spreading misinformation?**

Answer:

- AI Detection Tools: Use deepfake detection software.
- Watermarking: Mark AI-generated content with metadata.
- Fact-Checking: Cross-check AI-generated information with credible sources.
- Public Awareness: Educate users on identifying AI-generated misinformation.

Assertion and reasoning Based Questions:

1**. Assertion (A):** Generative AI can create new and unique content, such as text, images, and music.
Reason (R): Generative AI models, such as GANs and VAEs, learn patterns from existing data to generate similar but new content.

a) Both A and R are true, and R is the correct explanation of A.

b) Both A and R are true, but R is not the correct explanation of A.

c) A is true, but R is false.

d) A is false, but R is true.

Answer: a) Both A and R are true, and R is the correct explanation of A.

Explanation: Generative AI models analyze training data and generate new outputs that resemble it. Techniques like GANs and VAEs allow AI to create realistic images, text, and music by learning the structure of input data.

2**. Assertion (A):** Large Language Models (LLMs) like GPT-4 and Gemini are trained on vast amounts of text data to generate human-like responses.
Reason (R): LLMs use supervised learning only to process and generate text.
a) Both A and R are true, and R is the correct explanation of A.
b) Both A and R are true, but R is not the correct explanation of A.
c) A is true, but R is false.
d) A is false, but R is true.
Answer: c) A is true, but R is false.
Explanation: While LLMs like GPT-4 and Gemini are trained on vast text datasets, they primarily use unsupervised learning and transformer-based architectures, not just supervised learning, to understand and generate text.

---

3. **Assertion (A):** Deepfake technology can be used to create highly realistic fake videos.
Reason (R): Deepfakes use discriminative models to generate realistic images and videos.
a) Both A and R are true, and R is the correct explanation of A.
b) Both A and R are true, but R is not the correct explanation of A.
c) A is true, but R is false.
d) A is false, but R is true.
Answer: c) A is true, but R is false.
Explanation: Deepfake technology relies on Generative Adversarial Networks (GANs), which are generative models, not discriminative ones. GANs generate realistic media by refining outputs through adversarial training.

---

4. **Assertion (A):** Generative AI can sometimes generate biased or misleading information.
Reason (R): The training data used for AI models may contain biases, which the AI can learn and replicate.
a) Both A and R are true, and R is the correct explanation of A.
b) Both A and R are true, but R is not the correct explanation of A.
c) A is true, but R is false.
d) A is false, but R is true.
Answer: a) Both A and R are true, and R is the correct explanation of A.
Explanation: AI models learn from large datasets that may include historical biases. If these biases exist in the training data, the AI model may unintentionally generate biased or misleading content.

---

5. **Assertion (A):** AI-generated text and images must always be credited to the AI model that created them.
Reason (R): AI-generated content is legally recognized as having independent intellectual property rights.
a) Both A and R are true, and R is the correct explanation of A.
b) Both A and R are true, but R is not the correct explanation of A.
c) A is true, but R is false.

d) A is false, but R is true.
Answer: c) A is true, but R is false.
Explanation: While AI-generated content should be credited to maintain ethical transparency, AI does not have independent intellectual property rights. Currently, copyright laws attribute ownership to the user or company that developed and trained the AI, not the AI itself.

Competency Based Questions

1**. Scenario**:
A digital marketing agency wants to automate content creation for its clients using Generative AI. They plan to use AI for generating promotional images, writing social media posts, and even creating short video advertisements.
Question:
What are some advantages and ethical concerns the agency should consider while using Generative AI for content creation?
Answer:
Advantages:

- Efficiency: AI can generate high-quality content quickly, reducing manual effort.
- Cost-effectiveness: Automating content creation reduces the need for large creative teams.
- Customization: AI can tailor content for different audiences based on data analysis.

Ethical Concerns:

- Plagiarism and Copyright Issues: AI-generated content may resemble existing copyrighted material.
- Bias in Content: If AI is trained on biased data, it may produce misleading or unfair content.
- Misinformation: AI can generate fake or misleading ads, harming brand reputation.

To mitigate risks, the agency should disclose AI use, verify content authenticity, and ensure compliance with copyright laws.

2**. Scenario:**
A news organization is considering using a Large Language Model (LLM) like ChatGPT to generate automated news reports based on real-time data.
Question:
How can the news organization ensure that AI-generated news articles are accurate, unbiased, and ethical?
Answer:
To ensure responsible AI-generated news, the organization should:

- Use Verified Sources: Train AI models on credible, fact-checked news databases.
- Human Review: Have journalists review AI-generated content before publishing.
- Bias Detection: Implement AI bias detection tools to avoid misrepresentation.
- Transparency: Clearly label AI-generated articles to inform readers.
- Regular Updates: Continuously update training data to reflect current and diverse perspectives.

By following these steps, the organization can balance AI efficiency with journalistic integrity.

---

3. **Scenario:**
A startup is developing a healthcare chatbot using Generative AI to assist patients by providing symptom-based recommendations.
Question:

What are the potential benefits and risks of using Generative AI in healthcare, and how can the startup address these risks?

Answer:

Benefits:

- 24/7 Availability: Chatbots provide instant support anytime.
- Cost Reduction: AI reduces the need for human agents for initial consultations.
- Data-Driven Insights: AI can analyze patient interactions to suggest improvements.

Risks and Solutions:

- Inaccuracy in Diagnoses: AI-generated recommendations should not replace medical professionals; the chatbot must clearly state its limitations.
- Data Privacy Issues: Implement strong encryption and compliance with HIPAA/GDPR to protect patient data.
- Bias in Medical Advice: Train AI on diverse, high-quality medical data to prevent biased recommendations.

By incorporating human oversight and secure AI practices, the chatbot can enhance healthcare accessibility without compromising patient safety.

---

4. **Scenario:**

A social media company is considering the use of Generative Adversarial Networks (GANs) to create personalized content recommendations and advertisements for users.

Question:

How can the company balance AI-driven personalization with ethical considerations such as privacy and misinformation?

Answer:

Balancing AI-driven personalization with ethics:

- User Consent: Obtain explicit permission before collecting and analyzing user data.
- Misinformation Control: Use fact-checking AI models to prevent false content from being recommended.
- Bias-Free Advertising: Ensure AI models do not favor specific demographics unfairly.
- Transparency: Allow users to opt-out of AI-based content recommendations.

By implementing these responsible AI guidelines, the company can enhance user experience while maintaining trust and compliance with ethical standards.

---

5. **Scenario:**

A university is introducing an AI-powered tool to help students generate academic essays and research summaries.

Question:

What strategies should the university implement to ensure AI is used responsibly in academic settings?

Answer:

To ensure ethical AI use in education, the university should:

- **AI Usage Guidelines:** Clearly define when and how students can use AI tools.
- **Plagiarism Detection**: Use AI-generated content detectors to prevent academic dishonesty.
- **Critical Thinking Emphasis**: Encourage students to fact-check AI-generated summaries and use AI as a learning aid rather than a replacement for independent thought.

- **Citations and Transparency**: Require students to cite AI sources when using AI-generated content in assignments.

UNIT-8 Data Story Telling

STORY TELLING is defined as the **art of narrating** stories to **engage an audience**.

Three forms of storytelling:-**Visual, Oral and Written**

It **originated** in the **ancient times with visual stories**, such as **cave drawings**, and **then** shifted to **oral traditions**, . Later, **words formed** into **narratives** that included **written, printed and typed stories**.

WHY STORYTELLING IS SO POWERFUL?

- ⇨ Create **engaging experiences** that **transport** the audience **to another space and time**
- ⇨ Establish **a sense of community belongingness** and identity
- ⇨ Make **information** more **compelling**
- ⇨ **Draw lessons** from past **and to reimaging the future by** affecting necessary changes
- ⇨ Can be entertaining, engaging, thought provoking, and inspiring change
- ⇨ Describes **a specific interpretation–communicating the important messages most effective ways**

WHY STORYTELLING IS CROSS-CULTURAL?

- ⇨ Enhances **global networking** by **increasing the awareness** about the **cultural differences**
- ⇨ Enhancing **cross-cultural understanding**
- ⇨ Story telling is an **integral part** of **indigenous cultures**.
- ⇨ **shapes, empowers and connects** people by doing **away with judgment** or **critic**
- ⇨ Facilitates openness **for** embracing differences**.**
- ⇨ **Engage** the audience **across boundaries and cultures**

ELEMENTS OF DATA STORY TELLING

1. **Characters:** The characters are the people or animals or some things or objects which are featured in a story. They perform the actions and drive the story.
2. **Plot/setting:** Setting refers to the time or location in which the story takes place. Plot refers to the sequence of the events of the story.
3. **Conflict:** It is the problem or the situation the characters are dealing with. It drives the story forward which makes the story engaging and a key element for the characters.
4. **Resolution:** It is the end of the story where the characters arrive at a particular situation to resolve the conflict. It is the stage after climax which is the peak or height of any story.
5. **Insights:** The ability to have a clear, deep, and sometimes sudden understanding of a complicated problem or situation.

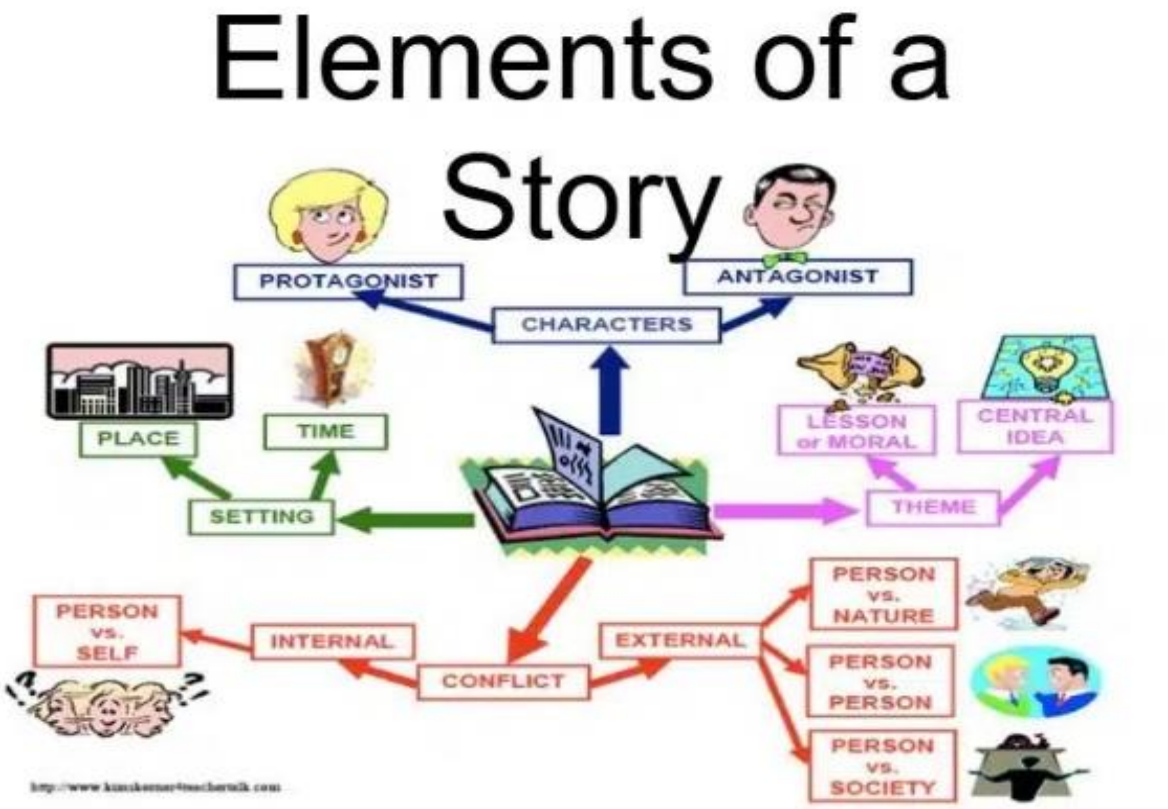


Image Source: https://slideplayer.com/slide/8259494/

THE STEPS INVOLVED IN TELLING EFFECTIVE DATA STORIES ARE GIVEN BELOW:

- ⇨ Understanding the audience
- ⇨ Choosing the right data and visualisations
- ⇨ Drawing attention to key information
- ⇨ Developing a narrative
- ⇨ Engaging your audience

Datastorytellingis a **structured approach** for **communicating insights drawn from data**. Three key elements: **data, visuals, and narrative**.

Narrative is accompanied **with data**, helps **to explain** the **audience what's happening** in the data and **why a particular insight has been generated**

Visuals are applied to data, they can **enlighten the audience** to **the insights** that they **wouldn't perceive without the charts** or graphs.

Narrative and visuals are merged together, they can **engage or even entertain** an audience. When you **combine** the **right visuals and narrative with the right data**, you have a data story that can **influence and drive change**

The **importance of a narrative comes** from the fact that **it explains what is going on** within the data set. It **offers a context and meaning, relevance and clarity**. A narrative shows the audience **where to look and what not to miss** and also**keeps the audience engaged**.

STEPS THAT CAN ASSIST IN FINDING COMPELLING STORIES IN THE DATASETS ARE AS FOLLOWS:

Step1: **Get the data and organize it**.

Step2: Visualize the data.

Step3: **Examine data relationships**.

Step4: Createa simple narrative embedded with conflict.

DATA STORY TELLING HAS ACQUIRED A PLACE OF IMPORTANCE BECAUSE:

⇨ It makes **information memorable** and **easiert or etainin the long run**.

⇨ It helps to **standardize communications and spread results**.

⇨ Stories that **incorporate data and analytics are more convincing** than those based e**ntirely on anecdotes** or personal experience.

⇨ It is an **effective tool to transmit human experience**. **Narrative** is the way we simplify and **make sense of a complex world**. It supplies **context, insight, interpretation**—all the things that make **data meaningful, more relevant and interesting**.

⇨ **No matter how impressive an analysis**, or how **high-quality the data**, it is not going to compel change unless the **people involved understand** what is **explained through a story.**

MCQs:

1. What makes storytelling a powerful tool across cultures?

   a) Use of technology
   b) Sense of community and identity
   c) Detailed data visualizations
   d) Statistical analysis

   Answer: b) Sense of community and identity

2. How does storytelling enhance cross-cultural understanding?

a) By using complex language
b) By reducing data dependency
c) By doing away with judgment and criticism
d) By eliminating differences
Answer: c) By doing away with judgment and criticism

3. Why are stories more impactful than just data in driving change?
a) They contain more statistics
b) They provide a clear narrative
c) They engage and entertain the audience
d) They require less explanation
Answer: c) They engage and entertain the audience

4. What is one reason data storytelling is considered compelling?
a) Data on its own is always unclear
b) Visuals alone are enough to drive action
c) Data becomes engaging through narrative
d) Data loses meaning when visualized
Answer: c) Data becomes engaging through narrative

5. What is the first step in telling an effective data story?
a) Developing a narrative
b) Drawing attention to key information
c) Understanding the audience
d) Visualizing the data
Answer: c) Understanding the audience

8. Which of the following is NOT a step in effective data storytelling?
a) Developing a narrative
b) Creating complex visualizations
c) Drawing attention to key information
d) Choosing the right data
Answer: b) Creating complex visualizations

9. What is a major goal of storytelling with data?
a) Confusing the audience with too much information
b) Entertaining the audience with irrelevant data
c) Connecting data with context for effective communication
d) Showing data as isolated points
Answer: c) Connecting data with context for effective communication

10. Why is narrative important when presenting data?
a) It reduces the need for visuals
b) It helps explain the meaning behind the data
c) It removes the need for data entirely
d) It replaces statistical analysis
Answer: b) It helps explain the meaning behind the data

11. In data storytelling, what helps uncover compelling stories in the data?
a) Disjointed data presentation
b) Examining data relationships
c) Ignoring narrative elements
d) Presenting raw numbers only
Answer: b) Examining data relationships

13. What tool helps standardize communications and spread results in data storytelling?

a) Graphs and charts
b) Stories with data and analytics
c) Written reports
d) Anecdotes from personal experience
Answer: b) Stories with data and analytics

14. Which element is NOT essential in a good data story?
a) Data b) Narrative c) Visuals d) Complex algorithms
Answer: d) Complex algorithms

15. Why are visuals important in data storytelling?
a) They can replace the need for narrative
b) They provide an emotional connection
c) They highlight insights not visible in raw data
d) They reduce the need for statistical analysis
Answer: c) They highlight insights not visible in raw data

17. Which of the following is true about storytelling and data?
a) Data stories are purely statistical
b) Stories provide context, making data more meaningful
c) Data stories eliminate the need for visuals
d) Stories and data are unrelated
Answer: b) Stories provide context, making data more meaningful

18. What helps keep an audience engaged in data storytelling?
a) Using technical jargon
b) Explaining only the statistics
c) Combining narrative and visuals
d) Presenting raw data
Answer: c) Combining narrative and visuals

19. What is the last step in effective data storytelling according to the document?
a) Engaging the audience
b) Choosing the right data
c) Developing a narrative
d) Visualizing the data
Answer: a) Engaging the audience

20. Why do data stories often lead to action compared to raw statistics?
a) They avoid using data altogether
b) They reduce ambiguity by connecting data to a narrative
c) They focus solely on visuals
d) They are more complex than statistics
Answer: b) They reduce ambiguity by connecting data to a narrative

22. What drives change more effectively than just data according to the document?
a) Personal anecdotes
b) Data stories that integrate narrative and visuals
c) Complex algorithms
d) Simplified numerical charts
Answer: b) Data stories that integrate narrative and visuals

23. Why are stories that include data more convincing than anecdotal stories?
a) They provide a standardized format
b) They eliminate the need for human interpretation

c) They merge personal experience with analytics
d) They present factual evidence along with narrative
Answer: d) They present factual evidence along with narrative

24. What should you do after organizing data for a story?
a) Develop a narrative
b) Visualize the data
c) Engage the audience
d) Ignore relationships between data points
Answer: b) Visualize the data

25. What element is essential in simplifying complex data for an audience?
a) Charts alone
b) Data relationships
c) Disjointed statistics
d) High-level programming
Answer: b) Data relationships

26. Why is storytelling considered an effective tool for transmitting human experience?
a) It simplifies and makes sense of complex information
b) It relies solely on facts
c) It is focused on data accuracy only
d) It eliminates the need for personal connection
Answer: a) It simplifies and makes sense of complex information

27. Which of the following is a step in finding compelling stories in data?
a) Ignoring data points
b) Organizing and examining data relationships
c) Avoiding data visualization
d) Eliminating conflict from the narrative
Answer: b) Organizing and examining data relationships

29. What helps the audience better understand data stories?
a) Randomly selected data
b) Personal anecdotes only
c) Properly merged narrative, visuals, and data
d) Data without visual aids
Answer: c) Properly merged narrative, visuals, and data

30. What is the key benefit of using data storytelling in education, according to the document?
a) It provides more statistics
b) It helps students engage with data in a meaningful way
c) It eliminates the need for data analysis
d) It focuses solely on text-based narratives
Answer: b) It helps students engage with data in a meaningful way

LONG ANSWERED QUESTION:

1. Explain how storytelling enhances cross-cultural understanding and why it is considered a powerful tool for global networking. Provide examples of how stories create a sense of community and identity.

Answer: Storytelling is a powerful tool because it creates engaging experiences that transcend cultural boundaries, transporting audiences to different spaces and times. It fosters a sense of community, belongingness, and identity by sharing common experiences and emotions across different cultures. Storytelling enhances cross-cultural understanding by providing a medium through which people can

communicate beyond language and cultural differences, opening up a space for shared understanding without judgment.

In global networking, storytelling plays a crucial role by bridging cultural gaps, making it easier for people from different backgrounds to connect on a personal and emotional level. Stories draw lessons from the past and provide a vision for the future, which can be especially important in professional environments. For example, storytelling in marketing or business helps brands connect with diverse audiences by tapping into universally understood human emotions, such as hope, fear, or joy, and making their messages resonate across cultures.

---

2. Discuss the concept of data storytelling. How does it differ from presenting raw data? Explain the elements involved in creating a compelling data story and how each contributes to making information engaging and actionable.

Answer: Data storytelling is a structured approach to communicating insights derived from data by combining three key elements: data, visuals, and narrative. It transforms raw data into a coherent story that engages the audience, making it more than just a collection of facts and figures. Unlike raw data, which is often dry and difficult to interpret, data storytelling provides context, clarity, and relevance, enabling the audience to understand the significance of the data.

The key elements of data storytelling include:

1. **Data:** It forms the foundation of the story, providing the factual basis for the insights being shared.
2. **Visuals**: Charts, graphs, and other visual aids help to illuminate the data and make it easier for the audience to perceive patterns or trends that may not be obvious in raw numbers.
3. **Narrative**: This ties the data and visuals together, explaining what the data means and why it matters. A good narrative makes the data relatable and provides a call to action.

By combining these elements, data storytelling not only informs but also engages, entertains, and drives change, making it more actionable than raw data alone.

---

3. What steps should be followed to create an effective data story? Explain each step in detail, using examples of how it would apply to a classroom scenario or another practical setting.

Answer: Creating an effective data story involves several key steps:

1. **Understanding the Audience**: The first step is to understand the needs, preferences, and prior knowledge of the audience. In a classroom scenario, a teacher must assess students' existing knowledge before introducing new data concepts. This helps in tailoring the story to make it more relatable and impactful.
2. **Choosing the Right Data**: Not all data is equally relevant. In a classroom, a teacher might choose data that reflects the students' performance trends over a semester to highlight areas of improvement.
3. **Drawing Attention to Key Information**: In a data story, it is important to focus on the most relevant and actionable insights. A teacher could highlight the fact that a majority of students showed improvement after a new teaching method was introduced.
4. **Developing a Narrative**: The narrative connects the data to a story. For instance, the teacher might explain how changes in teaching methods, based on initial student feedback, led to better understanding and improved results.
5. **Engaging the Audience**: The final step involves keeping the audience engaged. In the classroom, a teacher could use interactive charts or ask students to contribute their interpretations of the data, making the session more participatory.

---

4. How does narrative help reduce ambiguity and make data more relatable to an audience? Discuss the role of narrative in transforming data from mere statistics into an engaging and impactful story.

Answer: A narrative provides context, meaning, and relevance to raw data, transforming it from mere statistics into a story that resonates with the audience. Data on its own can be ambiguous, leaving room for multiple interpretations. A well-crafted narrative, however, explains what the data represents and why it is important, guiding the audience toward the correct interpretation.

Narratives also help make data more relatable by humanizing the information. For instance, instead of presenting a series of numbers showing declining student performance, a teacher could explain how distractions in online learning environments contributed to the decline, making the data easier to understand and emotionally engaging. This kind of storytelling not only clarifies the data but also evokes empathy, prompting the audience to consider the data from a personal perspective and encouraging action.

---

5. Analyze the results of the pre- and post-polls on student interest in Science as presented in the document. How can this data be turned into a compelling story to drive change in teaching methods?

Answer: The pre-poll results showed that 11% of students felt "bored" with Science, while only 19% were "excited." After a month of implementing new teaching methods, the post-poll results showed significant improvement: only 12% of students felt "bored," and 38% were "excited." This data can be transformed into a compelling story by focusing on the problem (lack of student interest), the intervention (the teacher's new methods), and the resolution (improved student engagement).

To make the data compelling, the story could highlight how the initial feedback prompted the teacher to adjust their methods to include more interactive elements, and how this change led to a noticeable increase in student excitement and engagement. This data story emphasizes the importance of adapting teaching methods based on student feedback and shows the positive impact such changes can have.

---

6. What role do visuals play in data storytelling? Explain how combining visuals with narrative can help drive change or influence decision-making.

Answer: Visuals in data storytelling serve to illuminate patterns and trends that may not be obvious in raw data. They provide an immediate, clear representation of the data, helping the audience quickly grasp key insights. Charts, graphs, and other visual aids can transform complex datasets into easily digestible formats, making the data more accessible.

When combined with a strong narrative, visuals enhance the emotional and intellectual impact of the data story. For example, a graph showing a drop in student interest in Science, followed by another graph showing improvement after changes in teaching methods, visually reinforces the success of the intervention. This combination of visuals and narrative makes the data story more persuasive and actionable, driving decision-makers to implement similar changes in other contexts.

---

7. Why is it important to examine data relationships in storytelling? How does uncovering relationships between data points enhance the story being told?

Answer: Examining data relationships is crucial because it helps uncover hidden patterns, trends, and connections that may not be immediately apparent when looking at individual data points. These relationships provide deeper insights into the causes and effects behind the data, which in turn makes the story more compelling and informative.

For example, in a classroom scenario, analyzing the relationship between teaching methods and student performance can reveal which specific methods had the most impact. If a teacher notices that students

who engaged in group discussions performed better on tests, this relationship provides valuable insights that can shape future teaching strategies. Uncovering such connections makes the data story more robust, guiding more informed decision-making.

---

8. How does storytelling with data help standardize communication and improve retention of information? Provide examples of how data storytelling can be used in educational or business settings.

Answer: Storytelling with data helps standardize communication by providing a consistent narrative framework that everyone can understand, regardless of their level of expertise. This consistency ensures that the key message remains the same, even when presented to different audiences. Furthermore, storytelling enhances retention because stories are more memorable than raw data. The human brain is wired to remember narratives, especially those that evoke emotions or provide a relatable context.

In education, data storytelling can be used to track student progress over time. A teacher could present a narrative showing how different teaching methods impacted test scores, making it easier for parents and administrators to understand the data. In business, data storytelling can be used in performance reviews, where visuals and narratives illustrate how employee contributions have influenced company goals, leading to better retention and more informed decisions.

---

9. What challenges can arise when presenting disjointed data without a narrative? How can these challenges be mitigated through effective storytelling techniques?

Answer: When data is presented as disjointed charts and graphs without a narrative, the audience may struggle to understand the relevance or significance of the information. This can lead to confusion, misinterpretation, or a lack of engagement. Disjointed data often lacks context, making it difficult for the audience to see how the individual data points relate to one another or to the larger picture.

These challenges can be mitigated through effective storytelling techniques, such as creating a cohesive narrative that ties the data together and highlights key insights. By explaining the “why” behind the data and using visuals to support the narrative, the presenter can guide the audience through the data in a logical and engaging way, ensuring that the key message is clear and understood.

---

10. Discuss the significance of using conflict in data storytelling. How can conflict enhance the narrative and drive audience engagement?

Answer: Conflict is a key element of storytelling, and it plays an important role in data storytelling as well. Conflict creates tension and curiosity, which naturally engages the audience and keeps them invested in the story. In data storytelling, conflict often arises from a problem or challenge that needs to be addressed, such as declining student performance or ineffective business strategies.

By presenting conflict, the storyteller sets the stage for a resolution, making the data more compelling and relatable. For example, a data story about a company’s declining sales might present the conflict as a failure to adapt to market trends. The narrative would then focus on how data-driven decisions helped the company overcome the challenge, leading to increased sales. This kind of story not only engages the audience but also demonstrates the power of data in solving real-world problems.

KENDRIYA VIDYALAYA SANGATHAN

DELHI REGION

ARTIFICIAL INTELLIGENCE (SUBJECT CODE-843)

CLASS XII (SESSION 2025-2026)

Max. Time:2 Hours **Max.Marks:50**

General Instructions:

1. Please read the instructions carefully.
2. This Question Paper consists of **21 questions** in two sections : Section A & Section B.
3. Section A has Objective type questions whereas Section B contains Subjective type questions.
4. **Out of the given (5+16=)21 questions, a candidate has to answer (5+10 =)15 questions in the allotted (maximum) time of 2 hours.**
5. Allquestionsofaparticularsectionmustbeattemptedinthecorrectorder.
6. **SECTION A-OBJECTIVE TYPE QUESTIONS (24MARKS):**
   - **i.** This section has 05 questions.
   - **ii.** Marks allotted are mentioned against each question/part.
   - **iii.** There is no negative marking.
   - **iv.** Do as per the instructions given.
7. **SECTIONB–SUBJECTIVE TYPE QUESTIONS (26 MARKS):**
   - **i.** This section has 16 questions.
   - **ii.** A candidate has to do 10questions.
   - **iii.** Do as per the instructions given.
   - **iv.** Marks allotted are mentioned against each question/part.

SECTIONA:OBJECTIVETYPEQUESTIONS

Q.1	**Answerany4outofthegiven 6 questions on Employability Skills (1x4=4 marks)**	
i.	Identify the direct object and indirect object in the sentence : “He offered a coffee to me”. a. He, coffee b. He, me c. Coffee, me d. Coffee, he	**1**
ii.	"She completed all her assignments on time because she wanted to earn a bonus from her manager." The sentence given above is an example of: a. Intrinsic motivation b. Extrinsic motivation c. External motivation d. Lack of motivation	**1**
iii.	**Statement1**: Personality development occurs by the ongoing interaction of temperament, character and environment. **Statement2**:Culture does not play any role in shaping personalities. a. Both Statement 1 and Statement 2 are correct b. Both Statement 1 and Statement2 are incorrect c. Statement 1 is correct but Statement 2 is incorrect	**1**

	d. Statement 2 is correct but Statement 1 is incorrect	
iv.	Match the following: A. To Copy text — 1.Ctrl+v B. To Cut text — 2.Ctrl+c C. To Paste text — 3.Ctrl+x a. A -2; B- 3; C–1 b. A-3; B-1; C-2 c. A-1; B-3; C –2 d. A-1; B-2; C–3	**1**
v.	Aamir has been an entrepreneur for 15 years. His business has a culture, where people working with him are free to give feedback on his work and share new ideas for the business. Which attitude can you witness in Aamir? a. Barrier b. Perseverance c. Competitiveness d. Interpersonal skills	**1**
vi.	Jaanvi's children have grown up. They have a lot of clothes that are too small for them now. What should Jaanvi do with these old clothes? a. Throw them in the trash can. b. Donate them or make bags out of them. c. Keep using them even though they are too small. d. Burn them.	**1**

Q2	**Answer any 5 out of the given 6 questions (1x5= 5marks)**	
i.	Business Sponsors defines the problem and project objectives from a _____perspective. a. Economic b. Feedback c. Business d. Data Collection	**1**
ii.	What is the primary goal of data storytelling? To confuse the audience with complex charts To entertain the audience with anecdotes To communicate insights and findings effectively using data To hide information from the audience	**1**
iii.	Match the following: Which option should be taken? — A. Clustering Which group? — B. Regression Which category? — C. Anomaly Detection Is this unusual? — D. Classification How much or how many? — E. Recommendation	**1**

	1→D, 2→B ,3→E, 4→A ,5→C 1→C, 2→D, 3→B, 4→E, 5→A 1→A, 2→B, 3→C, 4→E, 5→D 1→E, 2→A, 3→D, 4→C, 5→B	
iv.	Choose the incorrect statement related to preprocessing stage of computer vision a. It enhances the quality of acquired image b. Noise reduction and Image normalization is often employed with images c. Techniques like histogram equalization can be applied to adjust the distribution of pixel intensities d. Edge detection and corner detection are ensured in images.	**1**
v.	The main difference between classification and detection is: a) Classification works only on grayscale images b) Detection only identifies colours c) Classification labels the whole image, detection finds multiple objects d) Detection requires no algorithm	**1**
vi.	Which layer of a neural network directly produces predictions? A) Input LayerB) Hidden Layer C) Output LayerD) Bias Layer	**1**
Q.3	**Answer any 5 out of the given 6 questions (1x5= 5marks)**	
i.	_____empowers computer vision systems to extract valuable insights and drive intelligent decision-making in various applications, ranging from autonomous driving to medical diagnostics. a. Low level processing b. High insights c. High-level processing d. None of the above	**1**
ii.	A city government collected data on traffic accidents at intersections to identify high-risk areas and prioritize safety improvements. They analyzed the data to identify patterns and trends in accident occurrence and severity. Subsequently, they developed a data storytelling report to present their findings to city officials and propose targeted interventions. What is the primary purpose of the city government's data storytelling report? A) To analyze public transportation usage. B) To identify high-risk areas for traffic accidents. C) To assess air quality levels in the city. D) To evaluate the performance of road maintenance crews.	**1**
iii.	Which type of neural network is commonly used for image recognition? A. Feedforward neural network. B. Convolutional neural network. C. Recurrent neural network. D. Perceptron.	**1**
iv.	Match the following and choose the correct options: i. Descriptive approach A. Statistical Analysis ii. Diagnostic approach B. Current Status iii. Predictive approach C. How to solve it?	**1**

	iv. Prescriptive approach D. Probabilities of action a.(i)—A , (ii)—B, (iii) – C , (iv)—D b. (i)—B , (ii)—A, (iii) – D , (iv)—C c. (i)—D , (ii)—B, (iii) – A , (iv)—C d. (i)—A , (ii)—C, (iii) – B , (iv)—D	
v.	In a decision-making neural network with inputs X1=1, X2=0, X3=1 and weights W1=5, W2=2, W3=4, bias = -3, what is the output sum before activation? A) 11 B) 6 C) 3 D) -3	**1**
vi.	If a neural network memorizes the training data instead of learning patterns, this is called: A) Regularization B) Overfitting C) Forward Propagation D) Undertraining	**1**

Q.4	**Answerany5outofthegiven6 questions(1x5= 5marks)**	
i.	Which technique is commonly used for analyzing large datasets to discover patterns and relationships? a)Linear regression b) Data mining c) Decision trees d) Naive Bayes	**1**
ii.	Identify the following icons: (a) (b)  	**1**
i.	In k-fold cross validation, what does k represent? a. number of subsets b. number of experiments c. number of folds d. all of the above	**1**
v.	What is the role of activation functions in neural networks? A. They determine the learning rate of the network. B. They introduce non-linearity, allowing the network to learn complex patterns. C. They control the size of the neural network. D. They define the input features of the network.	**1**
.	Why are biases included in neurons? A) To eliminate weights B) To allow shifting of decision boundaries C) To fix activation to 0 D) To stop backpropagation	**1**

.	Given the following list of data types, categorize each as Structured, Unstructured, or Semi-Structured. Which of the following is incorrect. a. A customer database with fields such as Name, Address, Phone Number, and Email. -unStructured b. A JSON file containing product information with attributes like name, price, and specifications. Semi-Structured c. Audio recordings of customer service calls. Unstructured d. A sales report in Excel format with rows and columns. Structured	**1**

Q.5	**Answerany5 outofthegiven6 questions(1x5= 5 marks)**	
i.	Which machine learning algorithms are commonly used in generative AI? a) Support Vector Machines (SVM) b) Decision Trees c) Generative Adversarial Networks (GANs) and VariationalAutoencoders (VAEs) d) Random Forests	**1**
ii.	True or False Real-time analytics involves processing and analyzing data as it is generated, without any delay.	**1**
iii.	Which of the following is **correct**? Testingdataistheoneonwhichwetrainandfitourmodelbasicallytofitthe parameters Testing data is used to evaluate the model's performance. Testing data provides a measure of how well the model generalizes to new, unseen data. 1)and3) only b. 1)and2)only c.1), 2)and3) d. 2)and3) only	**1**
iv.	Which term describes the process of extracting useful information from large datasets? a) Data analytics b) Data warehousing c) Data integration d) Data virtualization	**1**
v.	True or False Data storytelling involves presenting raw data without any analysis or interpretation.	**1**
vi.	Which generative AI model is known for its application in text generation? a) Meta AI's Voice box b) Google's Lumiere c) Open AI's Chat GPT d) Stability AI	**1**

SECTIONB:SUBJECTIVETYPEQUESTIONS

Answer any 3 out of the given 5 questions on Employability Skills (2x3=6marks)
Answer each question in 20 – 30 words.

Q.6	In the acronym "RESPECT" for active listening, what does the "R" stand for? Explain	**2**
Q.7	Rajesh is a middle-aged man who frequently checks if the doors are locked, even when he has just locked them. He feels anxious if he doesn't check multiple times. Suggest one way in which he can address his behavior.	**2**

Q.8	What is the primary purpose of a worksheet in spreadsheet software?	**2**
Q.9	What is the aim of the **Credit Guarantee Fund Scheme for Micro and Small Enterprises**?	**2**
Q.10	How do green jobs contribute to community health?	**2**

Answerany4 outofthegiven6 questionsin20–30words each(2x4 =8marks)

Q. 11	Write a short note on the steps done during Data Preparation.	**2**
Q.12	Write any four applications of computer vision.	**2**
Q.13	Describe Freytag's Pyramid and its application in data storytelling.	**2**
Q.14	Why is Big Data Analytics important in modern industries and decision-making processes?	**2**
Q.15	You are a data scientist working for a healthcare company, tasked with developing a neural network model to predict the likelihood of a patient having a heart attack based on their medical history. The dataset you have been provided with contains various features such as age, gender, family medical history, blood pressure, cholesterol levels, and previous medical conditions. Your task is to determine which type of neural network model would you build so that you can accurately predict whether a patient is at risk of having a heart attack.	**2**
Q.16	How can "Creative Horizons" ensure their AI-generated content's originality and copyright compliance?	**2**

Answerany3 outofthegiven5 questionsin50–80 wordseach (4x3=12 marks)

Q.17	The Red Fort is hosting a grand cultural event, and keeping everyone safe is top priority! A state-of-the-art security system utilizes different "FEATURE EXTRACTION" to analyseslive video feeds and identify potential issues. Identify the feature extraction technique that can be used in the following situation. A large bag is left unattended near a crowded entrance. A person tries to climb over a wall near a blind spot. A group of people starts pushing and shoving in a congested area. A wanted person with a distinctive red scarf enters the venue.	**4**
Q.18	Given the data: **x y** 44 47 46 48 48 55 50 58 52 49 Using the regression equation y=0.7x+17.8y = 0.7x + 17.8y=0.7x+17.8, calculate the RMSE (Root Mean Square Error).	**4**
Q.19	Explain any four feature the 6 V's related to Big data.	**4**
Q.20	Discuss the different types of neural networks, including feedforward neural networks, convolutional neural networks, and recurrent neural networks. Provide examples of real- world applications where each type is used.	**4**

Q.21	A research team is using Generative AI to generate synthetic data for training medical imaging algorithms to detect rare diseases. However, there are concerns about the potential biases and inaccuracies in the generated data. What steps can the research team take to mitigate biases and ensure the accuracy and reliability of the synthetic data generated by Generative AI for medical imaging applications?	**4**

ARTIFICIAL INTELLIGENCE (SUBJECT CODE-843)

MARKING SCHEME FOR CLASS XII (SESSION 2025-2026)

Max.Time:2Hours **Max.Marks:50**

SECTION A: OBJECTIVE TYPE QUESTIONS

Q.No.	QUESTION	Marks
Q.1		
i.	c. Coffee, me	**1**
ii.	b.Extrinsic motivation	**1**
iii.	c. Statement 1 is correct but Statement 2 is incorrect	**1**
iv.	A → 2; B → 3;C→ 1	**1**
v.	d. Interpersonal skills	**1**
vi.	Donate them or make bags out of them.	**1**

Q.2		
i.	c. Business	**1**
ii.	c. To communicate insights and findings effectively using data	**1**
iii.	1→E,2→A,3→D,4→C, 5→B	**1**
iv.	Edge detection and corner detection are ensured in images.	**1**
v.	c) Classification labels the whole image, detection finds multiple objects	**1**
vi.	C) Output Layer	**1**

Q. 3		
i.	c. High-level processing	**1**
ii.	B) To identify high-risk areas for traffic accidents.	**1**
iii.	Convolutional neural network.	**1**
iv.	b. (i)—B , (ii)—A, (iii) – D , (iv)—C	**1**
v.	B) 6	**1**
vi.	B) Overfitting	**1**

Q.4		
i.	b)Datamining	**1**
ii.	a. Narrative b. Visuals	**1**
iii.	d. all of the above	**1**
iv.	B. They introduce non-linearity, allowing the network to learn complex patterns.	**1**
v.	B) To allow shifting of decision boundaries	**1**
vi.	a. A customer database with fields such as Name, Address, Phone Number, and Email. -unStructured	**1**

Q. 5		
i.	c) Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs)	**1**

ii.	(True)	**1**
iii.	d. 2)and3) only	**1**
iv.	a) Data analytics	**1**
v.	(False)	**1**
vi.	OpenAI's ChatGPT	**1**

SECTION B: SUBJECTIVE TYPE QUESTIONS

Q.No.	QUESTION	Marks
Q.6	**"R" stands for *Rapport.*** In active listening, building *Rapport* is crucial as it establishes a sense of trust and understanding between the speaker and the listener. This encourages open communication, where both parties feel valued and respected. (1 mark for the word '***Rapport***'; and 1 mark for explanation)	**2**
Q.7	Rajesh could benefit from **Cognitive Behavioral Therapy (CBT)**. CBT can help him ❖ recognize and ❖ challenge the irrational fears that lead to his compulsive checking behavior, ❖ teaching him coping strategies to reduce anxiety. (Any2oftheabove–1markeach)	**2**
Q.8	A worksheet is used to enter and organize data, perform calculations, and create charts within a grid of rows and columns. It allows users to analyze and manipulate data for various tasks. (1markeach)	**2**
Q.9	The aim of the Credit Guarantee Fund Scheme is to provide collateral-free loans to micro and small enterprises, facilitating access to capital for business expansion and sustainability. (Any 2;1markeach)	**2**
Q.10	Green jobs improve community health by reducing pollution and promoting cleaner energy sources, leading to healthier living environments and decreased incidence of health problems related to pollution. (Any 2;1markeach)	**2**
Q.11	The most time-consuming stage is Data Preparation. Here data is transformed into a state where it is easier to work with. Feature Engineering is also a part of Data Preparation. Data preparation includes ● cleaning of data (dealing with invalid or missing values, remove duplicates and give a suitable format)	**2**

	● combine data from multiple sources (archives, tables and platforms) ● transform data into meaningful input variables	
Q.12	1) Facial recognition: Popular social media platforms like Facebook uses facial recognition to detect and tag users. 2) Healthcare: Helps in evaluating cancerous tumours, identify diseases or abnormalities. Object detection & tracking in medical imaging 3) Surveillance: Live footage from CCTV cameras in public places helps to identify suspicious behaviour, identify dangerous objects, and prevent crimes by maintaining law and order. 4) Fingerprint recognition and biometrics: Detects fingerprints and biometrics to validate a user's identity.	**2**
Q.13	Freytag's Pyramid is a narrative structure with five stages: exposition, rising action, climax, falling action, and resolution. In data storytelling, it serves as a framework to structure the presentation of data and insights, guiding the audience through a compelling narrative journey from introduction to resolution.	**2**
Q.14	Big Data Analytics is important in modern industries and decision-making processes because it: 1. Enables Data-Driven Decisions: By analyzing vast and diverse datasets, organizations can make informed decisions based on insights and trends. 2. Improves Efficiency and Productivity: Identifying inefficiencies and optimizing resource allocation helps streamline processes.	**2**
Q.15	This problem is that of binary classification so, a feed-forward neural network with multiple hidden layers can be effective.	**2**
Q.16	The agency should implement checks to verify the uniqueness of the content, use AI models trained on copyright-compliant datasets, and provide proper attribution or licenses for AI-generated outputs, ensuring compliance with copyright laws and ethical guidelines.	**2**
Q.17	Texture Analysis Edge detection Corner detection Colour based extraction	**4**

Q.18	Calculate predicted values using the regression equation: For x = 44: y=0.7(44)+17.8=51.8y = 0.7(44) + 17.8 = 51.8y=0.7(44)+17.8=51.8 For x = 46: y=0.7(46)+17.8=53.2y = 0.7(46) + 17.8 = 53.2y=0.7(46)+17.8=53.2 For x = 48: y=0.7(48)+17.8=54.6y = 0.7(48) + 17.8 = 54.6y=0.7(48)+17.8=54.6 For x = 50: y=0.7(50)+17.8=56.0y = 0.7(50) + 17.8 = 56.0y=0.7(50)+17.8=56.0 For x = 52: y=0.7(52)+17.8=57.4y = 0.7(52) + 17.8 = 57.4y=0.7(52)+17.8=57.4 Calculate errors: Error for (44, 47): 47−51.8=−4.847 - 51.8 = -4.847−51.8=−4.8 Error for (46, 48): 48−53.2=−5.248 - 53.2 = -5.248−53.2=−5.2 Error for (48, 55): 55−54.6=0.455 - 54.6 = 0.455−54.6=0.4 Error for (50, 58): 58−56.0=2.058 - 56.0 = 2.058−56.0=2.0 Error for (52, 49): 49−57.4=−8.449 - 57.4 = -8.449−57.4=−8.4 Calculate squared errors and mean: Squared errors: 23.04,27.04,0.16,4.00,70.5623.04, 27.04, 0.16, 4.00, 70.5623.04,27.04,0.16,4.00,70.56 Mean squared error (MSE): 23.04+27.04+0.16+4.00+70.565=24.16\frac{23.04 + 27.04 + 0.16 + 4.00 + 70.56}{5} = 24.16523.04+27.04+0.16+4.00+70.56 =24.16 Calculate RMSE: RMSE = MSE=24.16≈4.91\sqrt{MSE} = \sqrt{24.16} \approx 4.91MSE=24.16 ≈4.91	**4**
Q.19	The 6 V's of Big Data are: I. Volume: Refers to the massive amount of data generated daily, ranging from terabytes to exabytes. For example, 328.77 million terabytes of data are created every day. II. Velocity: Describes the speed at which data is generated, delivered, and analyzed. For instance, Google processes over 40,000 search queries per second. III. Variety: Indicates the different forms of data, including structured (e.g., databases), semi-structured (e.g., XML files), and unstructured data (e.g., videos, images, and social media posts). IV. Veracity: Focuses on the accuracy, quality, and trustworthiness of data. It involves ensuring that data is reliable and suitable for analytical models by addressing inconsistencies or inaccuracies. V. Value: Refers to the insights and business benefits that can be extracted from Big Data. Without deriving value, the other characteristics hold little significance. VI. Variability: Highlights the inconsistencies or unpredictability in data flow, requiring systems to adapt and extract meaningful insights even in dynamic conditions	**4**
Q.20	Feedforward neural networks are the simplest type of neural network and are used for general-purpose learning tasks such as classification and regression. Convolutional neural networks (CNNs) are specialized for image	**4**

	recognition tasks and are used in applications like facial recognition and object detection. Recurrent neural networks (RNNs) are designed for sequential data processing and are used in applications such as speech recognition and language translation.	
Q.21	The research team can take the following steps: • Ensure that the training data used to train the Generative AI model represent a diverse range of demographics, including underrepresented groups, to mitigate biases. • Validate the synthetic data generated by the model against real medical imaging data to ensure accuracy and reliability. • Involve domain experts, such as medical professionals and imaging specialists • Continuously refine and improve the Generative AI model based on feedback from domain experts and the performance of the medical imaging algorithms trained on the synthetic data.	**4**