

# CBSE Board Examination – 2024

## ARTIFICIAL INTELLIGENCE

### Solved Paper

### Class– 10<sup>th</sup>

Maximum Marks: 50

Time allowed: 2 hours

#### GENERAL INSTRUCTIONS:

- (i) Please read the instructions carefully.
- (ii) This question paper consists of 21 questions in two sections: Section A & Section B.
- (iii) Section A has objective type questions whereas Section B contain subjective type questions. Out of the given (5 + 16 z) 21 questions, a candidate has to answer (5 + 10 15 questions in the allotted (maximum) time of 2 (v)
- (iv) All questions of a particular section must be attempted in the correct order.
- (v) Section A — Objective type questions (24 marks) : (a) This section has 05 questions.  
(b) Marks allotted are mentioned against each question/part.  
(c) There is no negative marking.  
(d) Do as per the instructions given.
- (vi) Section B — Subjective type questions (26 marks) :  
(a) This section has 16 questions.  
(b) A candidate has to do 10 questions.  
(c) Do as per the instructions given.  
(d) Marks allotted are mentioned question/part.

#### SECTION– A

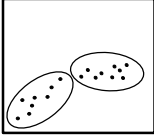
##### (Objective Type Questions)

1. Answer any 4 out of the given 6 questions :  $4 \times 1 = 4$

- (i) SMART method can be used to set goals to make you successful in your career and personal life. What does 'A' in SMART stand for?  
(A) Abrupt                      (B) Accountable  
(C) Achievable                (D) Admirable
- (ii) Which of the following is not a key element of self-management skills?  
(A) Prioritising your work  
(B) Not taking feedback  
(C) Goal setting  
(D) Staying updated about new practices
- (iii) Which of the following is a quality of successful entrepreneurs?  
(A) Hard working            (B) Resistance to change  
(C) Lazy                        (D) Less-confident
- (iv) The most important software in any computer is the \_\_\_\_\_. This is the software that starts working as soon as we switch on a computer.  
(A) Web Browsers            (B) Operating System  
(C) Office Software          (D) Designing Software
- (v) Which of the following types of communication takes place when one individual addresses a large gathering?  
(A) Written communication  
(B) Public communication  
(C) Small group communication  
(D) Interpersonal communication
- (vi) **Assertion (A):** Organic farming technique is an example of a green skill that is essential for sustainable agriculture.

**Reason (R):** Organic farming technique priorities environment friendly and sustainable practices such as using natural fertilizers, avoiding synthetic pesticides and promoting soil health.

- (A) Both (A) and (R) are true and (R) is the correct explanation for (A).
  - (B) Both (A) and (R) are true and (R) is not the correct explanation of (A).
  - (C) (A) is true, but (R) is false.
  - (D) (A) is false, but (R) is true.
2. Answer any 5 out of given 6 questions:  $5 \times 1 = 5$
- (i) Which of the following contributes to the efficiency of an AI project?  
(A) High Model Complexity  
(B) Relevant and Authentic Training Data  
(C) Minimal Preprocessing  
(D) Limited Hardware Resources
  - (ii) This real life application of NLP is used to provide an overview of a news item or blog post, while avoiding redundancy from multiple sources and maximising the diversity of content obtained. Which is this application?  
(A) Chatbot  
(B) Virtual Assistant  
(C) Sentiment Analysis  
(D) Automatic Summarisation
  - (iii) Which of the following represents a machine that is smart but not considered Artificial Intelligence (AI) enabled ?  
(A) A robotic vacuum cleaner that can navigate and clean floors autonomously.  
(B) A chatbot that engages in natural language conversations and answers questions.

- (C) A smartphone with facial recognition for unlocking the device.  
 (D) A digital alarm clock that rings at a set time every morning.
- (iv) Which of the following words represents an example of a lemma resulting from lemmatisation for “caring” in context to Natural Language Processing (NLP)?  
 (A) Care (B) Cared  
 (C) Cares (D) Car
- (v) Intrapersonal Intelligence is a concept that:  
 (A) Measures an individual’s ability to understand others’ emotions and feelings.  
 (B) Assesses one’s proficiency in mathematics and logical reasoning.  
 (C) Describes the level of self-awareness someone has, starting from realizing weaknesses, strengths, to recognizing their own feelings.  
 (D) Evaluates an individual’s spatial navigation and visualization skills.
- (vi) For Data Science, usually the data is collected in the form of tables. These tabular datasets can be stored in different formats. Which of the following formats is not used for storing data in a tabular format?  
 (A) CSV (B) Website  
 (C) SQL (D) Spreadsheet
3. Answer any 5 out of given 6 questions:  $5 \times 1 = 5$
- (i) \_\_\_\_\_ is one of the parameter for evaluating a model’s performance and is defined as the fraction of positive cases that are correctly identified.  
 (A) Precision (B) Accuracy  
 (C) Recall (D) F1
- (ii) In the AI project cycle, which of the following represents the correct order of steps?  
 (A) Data Exploration, Problem Scoping, Modelling, Evaluation, Data Acquisition.  
 (B) Problem Scoping, Data Acquisition, Data Exploration, Modelling, Evaluation.  
 (C) Modelling, Data Acquisition, Evaluation, Problem Scoping, Data Exploration.  
 (D) Data Acquisition, Data Exploration, Problem Scoping, Modelling, Evaluation.
- (iii) \_\_\_\_\_ is a concept to unify statistics, data analysis, machine learning and their related methods in order to understand and analyse actual phenomena with data.  
 (A) Computer Vision  
 (B) Natural Language Processing  
 (C) Data Science  
 (D) Computer Science
- (iv) In computer vision which of the following tasks is used for multiple objects?  
 (A) Classification  
 (B) Classification + Localisation  
 (C) Instance Segmentation  
 (D) Localisation
- (v) In spam email detection, which of the following will be considered as “False Negative”?  
 (A) When a legitimate email is accurately identified as not spam.  
 (B) When a spam email is mistakenly identified as legitimate.  
 (C) When an email is accurately recognised as spam.  
 (D) When an email is inaccurately labelled as important.
- (vi) Which of the following applications is not associated with Natural Language Processing (NLP)?  
 (A) Sentiment Analysis  
 (B) Speech Recognition  
 (C) Spam Filtering in emails  
 (D) Stock Market Analysis
4. Answer any 5 out of the given 6 questions:  $5 \times 1 = 5$
- (i) **Statement 1:** Confusion matrix is an evaluation metric.  
**Statement 2:** Confusion Matrix is a record which helps in evaluation.  
 (A) Both Statement 1 and Statement 2 are correct.  
 (B) Both Statement 1 and Statement 2 are incorrect.  
 (C) Statement 1 is correct and Statement 2 is incorrect.  
 (D) Statement 2 is correct and Statement 1 is incorrect.
- (ii) Which form of unsupervised learning does the following diagram indicate?  
 (A) Clustering  
 (B) Regression  
 (C) Reinforcement learning  
 (D) Classification
- 
- (iii) Bag of Words is a \_\_\_\_\_ model which helps in extracting features out of the text which can be helpful in machine learning algorithms.  
 (A) Data Science (DS)  
 (B) Virtual Reality (VR)  
 (C) Natural Language Processing (NLP)  
 (D) Computer Vision (CV)
- (iv) Which of the following represents an example of a recommendation system?  
 (A) An online clothing store that offers a wide variety of clothing options.  
 (B) A search engine that retrieves relevant web pages based on user queries.  
 (C) An e-commerce website that displays customer reviews and ratings for products.  
 (D) A music streaming platform that suggests songs and playlists based on user listening history.
- (v) Name any two search engines.
- (vi) What is the primary need for evaluating an AI model’s performance in the AI Model Development process?  
 (A) To increase the complexity of the model.  
 (B) To visualize the data.  
 (C) To assess how well the chosen model will work in future.  
 (D) To reduce the amount of data used for training.
5. Answer any 5 out of the given 6 questions :  $5 \times 1 = 5$
- (i) **Assertion (A):** The term used to refer to the number of pixels in an image is resolution.  
**Reason (R):** Resolution in an image denotes the total number of pixels it contains, usually represented as height x width.  
 (A) Both (A) and (R) are true and (R) is the correct explanation for (A).  
 (B) Both (A) and (R) are true and (R) is not the correct explanation for (A).

- (C) (A) is true, but (R) is false.  
 (D) (A) is false, but (R) is true.
- (ii) When a machine possesses the ability to mimic human traits, i.e., make decisions, predict the future, learn, and improve on its own, it is said to have:  
 (A) Computational Skills  
 (B) Learning Capability  
 (C) Artificial Intelligence  
 (D) Cognitive Processing
- (iii) **Statement 1:** To evaluate a model's performance, we need either precision or recall.  
**Statement 2:** When the value of both Precision and Recall is 1, the F1 score is 0.  
 (A) Both statement 1 and statement 2 are correct.  
 (B) Both statement 1 and statement 2 are incorrect.  
 (C) Statement 1 is correct, but statement 2 is incorrect.  
 (D) Statement 1 is incorrect, but statement 2 is correct.
- (iv) The concept of \_\_\_\_\_ is used to apply face filters on various social media platforms.  
 (A) NLP  
 (B) Computer Vision  
 (C) Data Science  
 (D) Block chain Technology
- (v) The 4 W's Problem Canvas helps in identifying the key elements related to the given problem.  
 Which of the following is NOT one of the blocks of the Problem Canvas?  
 (A) When (B) Where  
 (C) What (D) Why
- (vi) Which domain of AI is used for interacting with virtual assistants such as Siri and Alexa ?  
 (A) Machine Learning (ML)  
 (B) Computer Vision (CV)  
 (C) Natural Language Processing (NLP)  
 (D) Technical Vision (TV)

## SECTION- B

### (Subjective Type Questions)

Answer any 3 out of given 5 questions on Employability Skills. Answer each question in 20-30 words.  $3 \times 2 = 6$

- Give any two examples of how individual choices and behaviours can contribute in achieving sustainable development.
- List any two common misconceptions about entrepreneurship.
- What is the importance of time management in effectively dealing with stress? Provide any one strategy for improving time management skills to reduce stress.
- Mention any two measures that individuals or organisations can take to protect their data from theft and viruses.
- The method of communication that you choose could affect the relationship with your peers, superiors and customers. Write the four factors on the basis of which you can choose the right method of communication. Answer any 4 out of given 6 questions in 20-30 words each.  $4 \times 2 = 8$

- Differentiate between Machine Learning (ML) and Deep Learning (DL).
- What are the primary differences between Script-bots and Smart-bots ?
- What do you mean by Evaluation of an AI model? Also explain the concept of over fitting with respect to AI model Evaluation.
- For a healthcare organisation's objective of predicting disease outbreaks and efficiently allocating resources through the analysis of medical records, would you recommend using supervised learning or unsupervised and efficiently allocating resources through the analysis of medical learning as the preferred machine learning approach? Explain your choice briefly.
- What role does data play in AI based applications ? Name any two sources of online data collection for building any AI based application.
- Differentiate between grayscale and RGB images.  
 Answer any 3 out of given 5 questions in 50-80 words each.  $3 \times 4 = 12$
- What are Neural networks? Briefly explain all the layers of a neural network.
- Give any four examples of applications of AI that we see around us.
- Consider the following two documents:  
 Document 1 : ML and DL are part of AI.  
 Document 2: DL is a subset of ML.  
 Implement all four steps of the Bag of Words (BoW) model to create a document vector table. Depict the outcome of each step.
- Consider the following graphs (Figure 1 and Figure 2) that demonstrate the two types of Supervised Learning Models of Artificial Intelligence. Identify and explain each model giving suitable examples of each.

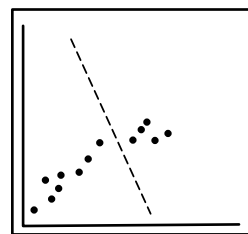


Figure 1

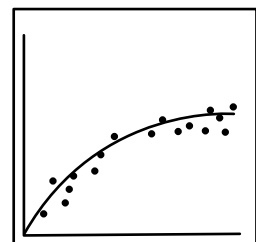


Figure 2

- A binary classification model has been developed to classify news articles as either "Fake News" or "Real News". The model was tested on a dataset of 500 news articles, and the resulting confusion matrix is as follows:

Confusion Matrix		Reality	
		Yes	No
Predicted	Yes	45	15
	No	20	420

- How many total cases are True Negative in the above scenario?
- Calculate Precision, Recall and F1-Score.

# ANSWERS

## SECTION– A

(20 × 1 = 20)

**1. (i) Option (C) is correct.**

*Explanation:* In the SMART method, “A” stands for Achievable. This means that the goals you set should be realistic and attainable within the resources, constraints, and time frame available to you. Setting achievable goals ensures that they are within reach and can be successfully accomplished, which contributes to the overall success in both your career and personal life.

**(ii) Option (B) is correct.**

*Explanation:* Self-management skills typically involve the ability to prioritize work, set goals, and stay updated about new practices. However, the option “Not taking feedback” is counter productive to effective self-management. Feedback is crucial for self-improvement, learning, and growth. It provides valuable insights into one’s performance, helps identify areas for improvement, and fosters personal and professional development. Therefore, not taking feedback would hinder rather than enhance self-management skills.

**(iii) Option (A) is correct.**

*Explanation:* Successful entrepreneurs are often characterized by their strong work ethic and determination to achieve their goals. They are willing to put in the effort, time, and energy required to turn their visions into reality. Hard work is a fundamental aspect of entrepreneurship, as it often involves overcoming challenges, persevering through setbacks, and continuously striving for success. Every other option would never help in moulding a successful entrepreneur.

**(iv) Option (B) is correct.**

*Explanation:* The operating system (OS) is the most important software in any computer. This system software manages the computer’s hardware and provides a platform for other software applications to run. The operating system boots up and starts working as soon as you switch on the computer, handling tasks such as managing memory, controlling input and output devices, and providing a user interface for interaction with the computer. Examples of operating systems include Windows, macOS, Linux, and others.

**(v) Option (B) is correct.**

*Explanation:* Public communication occurs when one individual addresses a large gathering or audience. This type of communication involves delivering a message to a group of people, such as in public speaking, lectures, presentations, or speeches. It differs from interpersonal communication, which involves interactions between individuals, and small group communication, which involves

communication within a small group of people. Written communication, on the other hand, involves exchanging information through written text, such as emails, letters, reports, or memos.

**(vi) Option (A) is correct.**

*Explanation:* Both the assertion and reason are true and the reason provided (R) correctly explains why organic farming is considered a green skill essential for sustainable agriculture. Organic farming aligns with principles of environmental sustainability by employing methods that minimize harm to the environment and promote long-term soil health and biodiversity. Therefore, option (a) is the correct choice.

**2.(i) Option (B) is correct.**

*Explanation:* Having relevant and authentic training data contributes significantly to the efficiency of an AI project. Training data forms the foundation upon which machine learning models are built, and the quality of this data directly impacts the performance and accuracy of the models. Relevant data ensures that the model learns from examples that are representative of the problem domain, while authentic data reflects real-world scenarios, enabling the model to generalize well to unseen data.

**(ii) Option (D) is correct.**

*Explanation:* Automatic summarization is the real-life application of natural language processing (NLP) used to provide an overview of a news item or blog post while avoiding redundancy from multiple sources and maximizing the diversity of content obtained. Automatic summarization techniques analyze the input text and generate a condensed version that captures the key information and main points, making it easier for users to quickly understand the content without reading the entire document.

**(iii) Option (D) is correct.**

*Explanation:* While all options involve some form of “smartness,” option (d) does not involve Artificial Intelligence (AI) technology. A digital alarm clock operates based on a predefined set of instructions (e.g., ringing at a set time), without the need for learning, adaptation, or decision-making capabilities characteristic of AI.

**(iv) Option (A) is correct.**

*Explanation:* In the context of natural language processing and lemmatization, a lemma is the base or dictionary form of a word. When lemmatizing the word “caring,” the lemma would be “care.” Lemmatization reduces words to their base or dictionary form, which helps in standardizing words for analysis, retrieval, or processing. Therefore, “care” (option a) is the correct lemma resulting from lemmatization for “caring.”

**(v) Option (C) is correct.**

*Explanation:* Intrapersonal intelligence, as proposed

by Howard Gardner's theory of multiple intelligences, refers to the ability to understand oneself. It involves self-awareness, recognizing one's own emotions, strengths, weaknesses, motivations, and goals. Individuals with high intrapersonal intelligence typically have a deep understanding of themselves and are able to effectively regulate their emotions and actions based on this understanding. This type of intelligence is crucial for personal development, self-reflection, and achieving personal goals.

**(vi) Option (B) is correct.**

*Explanation:* Websites are not typically used for storing data in tabular format. Websites are platforms for displaying information or providing interactive services to users. However, data stored on websites may often be organized in databases, which could use tabular formats like SQL databases or other structured formats. But the website itself is not a format for storing tabular data.

The other options, CSV (Comma-Separated Values), SQL (Structured Query Language), and spreadsheet (such as Excel), are commonly used formats for storing data in tabular form.

**3.(i) Option (C) is correct.**

*Explanation:* Recall, also known as sensitivity or true positive rate, is the parameter for evaluating a model's performance that measures the fraction of positive cases that are correctly identified by the model. It calculates the proportion of true positive cases (correctly identified positive cases) out of all actual positive cases.

**(ii) Option (B) is correct.**

*Explanation:* Here's a breakdown of each step:

1. **Problem Scoping:** This involves clearly defining the problem statement, objectives, and scope of the AI project. It's crucial to understand what problem the AI system is meant to solve and what success looks like.
2. **Data Acquisition:** Once the problem is scoped, the next step is to gather the relevant data needed to train and test the AI model. This may involve collecting data from various sources, such as databases, APIs, or sensors.
3. **Data Exploration:** After acquiring the data, it's important to explore and analyze it to understand its characteristics, quality, and patterns. This step helps in identifying any data preprocessing steps that may be necessary before modelling.
4. **Modelling:** Once the data is prepared, the modelling phase involves building and training the AI model using machine learning or deep learning techniques. This step aims to develop a model that can effectively address the problem defined earlier.
5. **Evaluation:** After training the model, it's essential to evaluate its performance using appropriate metrics. This step helps assess how well the model generalizes to new, unseen data and whether it meets the project objectives.

So, option (b) represents the correct order of steps in the AI project cycle.

**(iii) Option (C) is correct.**

*Explanation:* Data Science is the concept that unifies statistics, data analysis, machine learning, and related methods to understand and analyze real-world phenomena using data. It encompasses a wide range of techniques and methodologies aimed at extracting insights, making predictions, and informing decision-making processes from various types of data.

**(iv) Option (C) is correct.**

*Explanation:* Instance segmentation in computer vision is used for multiple objects. Instance segmentation involves not only identifying the objects in an image but also distinguishing between individual instances of the same object. This means that each object instance is uniquely segmented and labeled within the image. This is particularly useful when there are multiple instances of the same class of objects present in an image, allowing for precise localization and differentiation between them.

**(v) Option (B) is correct.**

*Explanation:* A "False Negative" in spam email detection occurs when a spam email is mistakenly identified as legitimate/not spam. This means that the spam filter failed to detect the spam and allowed it to pass through to the user's inbox, which can potentially result in the user being exposed to unwanted or harmful content.

**(vi) Option (D) is correct.**

*Explanation:* Stock market analysis is not typically associated with Natural Language Processing (NLP). Instead, it involves analyzing financial data, market trends, and economic indicators to make investment decisions. While NLP can be used in various applications such as sentiment analysis, speech recognition, and spam filtering in emails, it is not commonly employed in the context of stock market analysis.

**4. (i) Option (A) is correct.**

*Explanation:*

- **Statement 1:** Confusion matrix is indeed an evaluation metric commonly used in machine learning to assess the performance of classification models.
- **Statement 2:** Confusion matrix is also a record that summarizes the performance of a classification model by tabulating the number of true positive, false positive, true negative, and false negative predictions.

Both statements are correct in describing the nature and purpose of a confusion matrix. Therefore, option (a) is the correct choice.

**(ii) Option (A) is correct.**

*Explanation:* The diagram indicates a grouping or clustering of data points without explicit labels or categories, it represents the process of clustering, which is a form of unsupervised learning.

It can't be a regression problem as its not predicting values. The given image doesn't depict an environment where an agent is interacting and learning to maximize rewards, hence its not reinforcement learning. In the classification problem we are trying to find a perfect boundary that separates the classes, here that is not the case.

**(iii) Option (C) is correct.**

**Explanation:** Bag of Words (BoW) is a technique used in Natural Language Processing (NLP) to represent text data as numerical feature vectors. It involves creating a vocabulary of unique words from the corpus of text documents and then representing each document as a vector of word counts or presence/absence of these words. These feature vectors can then be used in machine learning algorithms for various NLP tasks such as text classification, sentiment analysis, and document clustering.

**(iv) Option (D) is correct.**

**Explanation:** A recommendation system is designed to suggest items or content to users based on their preferences, behaviour, or historical interactions. In this case, a music streaming platform recommending songs and playlists based on a user's listening history exemplifies a recommendation system. This system analyzes past listening behaviour to provide personalized recommendations, enhancing user experience and engagement.

**(v) Two well-known search engines are:**

1. Google
2. Bing

**(vi) Option (C) is correct.**

**Explanation:** Evaluating the performance of an AI model allows developers to understand how effectively the model solves the problem it was designed for. It helps in assessing the model's accuracy, reliability, and generalization capabilities, providing insights into its effectiveness in real-world scenarios. By evaluating the model's performance, developers can make informed decisions about its deployment, optimization, and potential improvements.

**5.(i) Option (A) is correct.**

**Explanation:** Both the assertion and the reason are true. Resolution indeed refers to the number of pixels in an image, and it is typically represented as the product of its height and width in pixels. Therefore, option (a) is the correct choice as the reason (R) correctly explains assertion (A).

**(ii) Option (C) is correct.**

**Explanation:** When a machine possesses the ability to mimic human traits such as making decisions, predicting the future, learning, and improving on its own, it is referred to as having Artificial Intelligence (AI). AI involves the development of computer systems that can perform tasks that typically require human intelligence, including problem-solving, pattern recognition, and decision-making. Therefore, option (c) is the correct choice.

**(iii) Option (B) is correct.**

**Explanation:**

- **Statement 1:** This statement is incorrect. Precision and recall are both important metrics for evaluating a model's performance, especially in binary classification tasks, but they are not the only metrics available. Other metrics such as accuracy, F1 score, and specificity are also commonly used.
- **Statement 2:** This statement is also incorrect. When both precision and recall have a value of 1 (perfect

scores), the F1 score will also have a value of 1, not 0. The F1 score is the harmonic mean of precision and recall and is calculated using the formula  $F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$ . If precision and recall are both 1, then  $F1 = 2 * (1 * 1) / (1 + 1) = 1$ .

**(iv) Option (B) is correct.**

**Explanation:** Computer vision is the concept used behind the face filters on various social media platforms. Computer vision involves the development of algorithms and techniques to enable computers to interpret and understand visual information from digital images or videos. In the context of social media platforms, computer vision algorithms analyze facial features in real-time to apply filters, effects, or augmentations to users' faces, enhancing their photos or videos.

**(v) Option (B) is correct.**

**Explanation:** The 4 W's Problem Canvas typically consists of the following blocks:

1. What: Describes the problem or opportunity.
2. Why: Identifies the reasons or motivations behind the problem.
3. Who: Specifies the stakeholders or people affected by the problem.
4. When: Indicates the timing or duration of the problem.

"Where" is not typically included as one of the blocks in the Problem Canvas.

**(vi) Option (C) is correct.**

**Explanation:** Interacting with virtual assistants such as Siri and Alexa involves understanding and processing natural language inputs from users. Natural Language Processing (NLP) is the domain of artificial intelligence (AI) that deals with the interaction between computers and humans through natural language. NLP enables virtual assistants to understand spoken or written language, interpret user queries, and generate appropriate responses.

**SECTION- B**

6. Individual choices like reducing waste and using sustainable transportation can minimize environmental impact, while adopting energy-efficient practices at home can conserve resources, contributing to sustainable development.
7. Two common misconceptions about entrepreneurship are:
  1. Entrepreneurship is all about taking big risks.
  2. Entrepreneurs are born, not made.
8. Effective time management helps prioritize tasks, reduce procrastination, and create a sense of control, reducing stress. One strategy is the "Pomodoro Technique," where work is broken into intervals with short breaks in between to maintain focus and productivity.
9. Two measures individuals or organizations can take to protect their data are:

1. Implementing strong password policies and using multi-factor authentication.
  2. Regularly updating software and installing antivirus programs to safeguard against viruses and malware.
10. The four factors to consider when choosing the right method of communication are:
1. **Audience:** Consider the preferences, expectations, and communication styles of your peers, superiors, and customers.
  2. **Message Complexity:** Determine the complexity and sensitivity of the message to decide whether it's best communicated verbally, in writing, or face-to-face.
  3. **Urgency:** Evaluate the urgency of the message and choose a communication method that ensures timely delivery and response.
  4. **Context:** Take into account the context of the communication, including the physical environment, cultural norms, and technological capabilities, to ensure effective communication and relationship-building.
11. Machine Learning (ML) involves algorithms that learn from data to make predictions, while Deep Learning (DL) is a subset of ML that uses artificial neural networks with multiple layers for learning.
12. Script-bots follow predefined scripts or instructions, while Smart-bots utilize artificial intelligence and machine learning algorithms to adapt and learn from interactions, making them more flexible and capable of handling diverse situations.
13. Evaluation of an AI model refers to assessing its performance and effectiveness in solving a particular task or problem. This process involves testing the model using data that it hasn't seen before, measuring its accuracy, precision, recall, or other relevant metrics, and comparing the model's predictions or outcomes against the ground truth.
- Overfitting is a phenomenon in machine learning where a model learns to perform extremely well on the training data but fails to generalize well to new, unseen data. This occurs when the model captures noise or random fluctuations in the training data rather than the underlying patterns, resulting in poor performance on test or validation data. Overfitting often happens when the model is too complex or when it's trained with insufficient data. To address overfitting, techniques such as regularization, cross-validation, and early stopping can be used.
14. For the healthcare organization's objective of predicting disease outbreaks and efficiently allocating resources through the analysis of medical records, I would recommend using supervised learning as the preferred machine learning approach. Supervised learning is suitable when there are labeled data available, which in this case would include historical medical records with information about disease outbreaks and resource allocations. By training a supervised learning model on this labeled data, the model can learn patterns and relationships between various factors (such as symptoms, demographics, geographic locations) and disease outbreaks, enabling accurate predictions and informed resource allocation decisions.
15. Data plays a crucial role in AI-based applications as it serves as the foundation for training, testing, and improving machine learning models. The quality, quantity, and diversity of data directly impact the performance and accuracy of AI systems. Two sources of online data collection for building AI-based applications are:
1. **Web scraping:** Extracting data from websites by automatically retrieving and parsing information from web pages.
  2. **Application programming interfaces (APIs):** Accessing data from online services or platforms through APIs, which provide a structured way to interact with and retrieve data from various sources such as social media platforms, weather services, or financial databases.
16. Grayscale images contain only shades of gray, ranging from black to white, representing the intensity of light at each pixel. RGB images, on the other hand, are composed of three colour channels (red, green, and blue), allowing each pixel to have a combination of different colours, resulting in a wider range of colours and the ability to represent full-colour images.
17. Neural networks are computational models inspired by the structure and functioning of the human brain, used in machine learning. They consist of interconnected nodes arranged into layers:
1. **Input Layer:** Receives input data where data is fed into the network.
  2. **Hidden Layers:** Perform computations and feature extraction, using weighted connections and activation functions.
  3. **Output Layer:** Produces the final output, representing predictions or classifications. Hidden layers allow neural networks to learn complex patterns from data through iterative optimization processes.
- Each layer's nodes (neurons) apply transformations to the input data using weights and activation functions, allowing the network to learn complex patterns and make predictions.
18. **Four examples of AI applications commonly seen around us include:**
1. Virtual assistants like Siri and Alexa, which use natural language processing to understand and respond to user queries.
  2. Recommendation systems used by streaming services and e-commerce platforms to suggest content or products based on user preferences.
  3. Autonomous vehicles, which employ computer vision and machine learning algorithms to navigate and make driving decisions.
  4. Medical diagnosis systems that analyze the patient data to assist healthcare professionals in identifying diseases and recommending

treatments. These applications demonstrate AI's versatility in improving efficiency and decision-making across various domains.

19. **Step 1:** Tokenization - Split each document into individual words:

**Document 1:** [ML, and, DL, are, part, of, AI]

**Document 2:** [DL, is, a, subset, of, ML]

**Step 2:** Build a vocabulary - Create a set of unique words from all documents:

	ML	and	DL	are	part	of	AI	is	a	subset
Doc 1	1	1	1	1	1	1	1	0	0	0
Doc 2	1	0	1	0	0	1	0	1	1	1

20. As described the graphs in Figure 1 and Figure 2.

Figure 1 is Classification Model and Figure 2 is Regression Model.

We will explain two types of supervised learning models in artificial intelligence along with suitable examples:

- 1. Classification Model:** Classification models are used to predict categorical labels or classes for input data points. In Figure 1, if this graph represents a classification model, it might depict distinct clusters or regions representing different classes.

**Example:** Classifying emails as spam or not spam based on features such as sender, subject line, and content. By training a classification model on a labeled dataset of emails (spam or non-spam), the model can learn to classify new emails into the appropriate category.

- 2. Regression Model:** Regression models are used to predict continuous numerical values based on input features. In Figure 2, if this graph represents a regression model, it might depict a scatter plot of data points where the goal is to find a line (or curve) that best fits the data.

**Example:** Predicting house prices based on features such as square footage, number of bedrooms, and location. Given historical data of house prices and corresponding features, a regression model can be trained to predict the price of a new house.

21. (a) To find the total True Negatives (TN), we look at the confusion matrix where the predicted label is

**Vocabulary:** [ML, and, DL, are, part, of, AI, is, a, subset]

**Step 3:** Count word occurrences - Count the frequency of each word in each document:

Document 1: [1, 1, 1, 1, 1, 1, 1, 0, 0, 0]

Document 2: [1, 0, 1, 0, 0, 1, 0, 1, 1, 1]

**Step 4:** Create document vectors - Represent each document as a vector using word frequencies:

**Document Vector Table:**

	ML	and	DL	are	part	of	AI	is	a	subset
Doc 1	1	1	1	1	1	1	1	0	0	0
Doc 2	1	0	1	0	0	1	0	1	1	1

"No" and the actual label is also "No".

From the matrix, the True Negatives (TN) are 420.

Precision, Recall, and F1-Score are calculated as follows:

From the confusion matrix we have:

- TP (True Positives) = 420
- FP (False Positives) = 15
- FN (False Negatives) = 20

- (i) To find Precision:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Precision} = 420 / (420 + 15)$$

$$= 420 / 435$$

$$= 0.9655$$

- (ii) To find Recall:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Recall} = 420 / (420 + 20)$$

$$= 420 / 440$$

$$= 0.9545$$

- (iii) To find F1 score:

$$\text{F1-Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

$$\text{F1-Score} = 2 * (0.9655 * 0.9545) / (0.9655 + 0.9545)$$

$$= 0.9600$$

Therefore,

Precision = 0.9655,

Recall = 0.9545, and

F1-Score = 0.9600.

