CBSE | DEPARTMENT OF SKILL EDUCATION

ARTIFICIAL INTELLIGENCE

QUESTION BANK – CLASS 10

CHAPTER 1: INTRODUCTION TO AI: FOUNDATIONAL CONCEPTS

One (01) Mark Questions

Fill in the Blanks:

- 1. The basis of decision making depends upon the availability of ______and how we experience and understand it. (information/data/conditions/ past experience/ knowledge/awareness.)
- 2. A machine can also become intelligent if it is trained with _____ which helps them achieve their tasks (data)

True/False:

- 1. A machine is artificially intelligent when it can accomplish tasks by itself. (True)
- 2. Is a smart washing machine an example of an Artificially Intelligent devices? (False)
- 3. Platforms like Netflix, Amazon, Spotify, YouTube etc. show us recommendations on the basis of what we like. (True)

Direct Question:

1. What do you understand by linguistic Intelligence?

Linguistic intelligence means intelligence to understand and interpret human natural language and try to extract meaning out of it.

OR

Linguistic Intelligence refers to the ability of an individual to understand both written and spoken language and the additional ability to write and speak the language too.

OR

This is the intelligence of language and communication. It includes the ability to speak, articulate, and express, and convey one's thoughts and feelings to the outside world in one or more languages. This can be at an oral and written level. It also includes the ability to listen to and to understand other people.

2. What do you understand by Interpersonal Intelligence?

Understanding human emotions, feelings and influenced by them is known as interpersonal intelligence.

OR

Interpersonal intelligence is the ability to communicate with others by understanding other people's feelings, being influenced by the person.

OR

Interpersonal intelligence refers to the ability of a person to relate well with people and manage relationships. It enables people to understand the needs and motivations of those around them, which helps strengthen their overall influence.

OR

*Inter*personal intelligence refers to the ability to understand social situations and the behavior of other people.

3. Define Artificial Intelligence.

A machine is artificially intelligent when it can accomplish tasks by itself - collect data, understand it, analyze it, learn from it, and improve it.

OR

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 artificial intelligence.

OR

Artificial Intelligence is a way of making a computer, a computer-controlled robot, or a software think intelligently, in a similar manner to how intelligent humans think.

OR

AI is a form of intelligence; a type of technology and a field of study. AI theory and development of computer systems (both machines and software) are able to perform tasks that normally require human intelligence.

OR

Artificial Intelligence works to implement human intelligence in machines: creating systems that understand, think, learn, and behave like humans.

OR

Artificial Intelligence or AI for short, refers to any technique that enables computers to mimic human intelligence. An artificially intelligent machine works on algorithms and data fed to it and gives the desired output.

4. Mention two types of machines which have evolved with time.

Television/Mobile Phones/ Ceiling Fans/ Microwave ovens/ Headphones / Speakers/ Harvesters/ Refrigerators/Air Conditioners etc.

(1 mark for any two right answers)

5. What do you understand by mathematical and logical reasoning?

A person's ability to regulate, measure, and understand numerical symbols, abstraction and logic.

OR

Reasoning is based on previous established facts. To establish a new fact or truth one has to put it to the test of reasoning. If the new fact coincides with the previously established facts, it is called logical or rational. It is the ability of a person to regulate, measure and understand numerical symbols, abstraction and logic.

Two (02) Mark Questions

1. Mention four examples of artificially intelligent applications in our smartphones.

Phone Smart Lock / Snapchat filter / Shopping websites / Netflix / YouTube / Face Detection / Google Maps / Emotions recognition / Google assistant / Natural language recognition / image detection / beauty filters etc. (2 marks for any four right examples)

2. How does a machine become Artificially Intelligent?

A machine becomes intelligent by training with data and algorithm. AI machines keep updating their knowledge to optimize their output.

OR

Machines also become intelligent once they are trained with some information which helps them achieve their tasks. AI machines also keep updating their knowledge to optimize their output.

3. Mention four examples of machines that are not AI but confused with AI.

0r

Mention four examples of machines that are smart but not AI.

Automatic gates in shopping malls / remote control drones/ a fully automatic washing machine/ Air Conditioner/ Refrigerator/ Robotic toy cars/ Television etc.

4. How does learning and adapting help an AI machine in improvising itself?

An artificially intelligent machine collects real time data and tries to figure out new patterns in it. Machines learn in a similar way human being; by supervision or by observation and respond according to past experiences in similar scenarios. A machine learns from its mistakes. The more the machine gets trained on data, the more accurate result it gives.

For example:

Any virtual assistant initially trained with few basic instructions, but with time, the machine captures the data fed by the user, may be the wake-up time of the user, sleeping time, dinner time and so on. Later in time, the machine gives reminders of similar things on the basis of data and adapts these new commands.

OR

Just as humans learn how to walk and then improve this skill with the help of their experiences, an AI machine too gets trained first on the training data and then optimizes itself according to its own experiences which makes AI different from any other technological device/machine.

5. Pick the odd one out and justify your answer:

a. Snap Chat Filter

c. Chatbot d. Image search Option

Ans: Chatbot (1 marks), as it is NLP based, the other three are Computer vision based (1marks for justification).

b. Face Lock in Phone

6. Explain how AI works in the following areas (any two):

a. Google Search Engine b. Voice Assistants c. E-commerce websites

a. Google Search Engine:

With the help of AI, Google Search Engine has been turned into Intelligent search which is a new network of systems that produces direct answers. It uses voice and image searches and has incorporated deep learning to fasten the searches with more accuracy.

b. Voice assistant:

AI is being used in voice assistants to recognize words spoken by the user. NLP has capabilities like "Speech-to-Text" convert the natural language of the user into text for further processing. As the digital assistant answers more and more queries, it "learns" using ML algorithms. The more tasks it performs, its ML algorithms help it "learn" from the tasks and the preferences of the user. As a result, the digital assistant improves its performance over time.

c. E-commerce website:

With the use of big data, AI in E-Commerce is impacting customer choices by recording the data of previous purchases, searched products, and online browsing habits. Product recommendations provide multiple benefits for E-commerce retailers including: Higher number of returning customers.

7. How has AI changed the gaming world?

AI has changed the world of gaming by making the game more intelligent by providing them the ability to learn using machine learning algorithms. Games these days try to understand human patterns and give responses on the basis of it and also give new difficulty levels.

OR

AI has changed the gaming world in terms of feel and emotions. Some video games react to player skill level. Depending on how well you do, adaptive AI ratchets the game's difficulty level up and down to give you a greater challenge when you need it or to prevent you from rage-quitting in frustration. AI can also adapt to your playing style by making the game more exciting.

8. Why training with information/Data is important in Artificial Intelligent devices?

Similar to human beings, AI devices need experience to give better results and improve in every next iteration. For giving better results, the machine should be trained with some real data. The more the amount of accurate data, the better predictions will be made by the machine. Hence, data is very important in AI devices.

OR

The AI devices need to be trained with information / Big data to produce the best possible accurate results. All of AI's learning happens only through this data. So, it makes sense to have as big a dataset as is required to include variety, subtlety, and nuance that makes the model viable for practical use. Before training, the model is just a theorist.

4 Mark Questions

1. What is Intelligence? Explain in brief any three types of intelligence that are mainly perceived by human beings?

Intelligence is the 'ability to perceive or infer information, and to retain it as knowledge to be applied towards adaptive behavior within an environment or context.'

OR

Intelligence is the ability to interact with the world (speech, vision, motion, manipulation), ability to model the world and to reason about it, ability to learn, ability to make decisions and to adapt.

OR

Intelligence has been defined in many ways: It involves abstract reasoning, mental representation, problem solving, and decision making, the ability to learn, emotional knowledge, creativity, and adaptation to meet the demands of the environment effectively.

As per major researches, there are mainly 9 types of Intelligence;

- (i) **Mathematical Logical Intelligence:** A person's ability to regulate, measure, and understand numerical symbols, abstraction and logic
- (ii) **Linguistic Intelligence:** Language processing skills both in terms of understanding orimplementation in writing or speech.
- (iii) **Spatial Visual Intelligence:** It is defined as the ability to perceive the visual world and the relationship of one object to another.
- (iv) **Kinesthetic Intelligence:** Ability that is related to how a person uses his limbs in a skilled manner.
- (v) **Musical Intelligence**: As the name suggests, this intelligence is about a person's ability to recognize and create sounds, rhythms, and sound patterns
- (vi) **Intrapersonal Intelligence:** Describes the level of self-awareness someone has starting from realizing weakness, strength, to recognizing his own feelings
- (vii) **Existential Intelligence:** An additional category of intelligence relating to religious andspiritual awareness.
- (viii) **Naturalist Intelligence:** An additional category of intelligence relating to the ability to process information on the environment around us.
- (ix) **Interpersonal Intelligence:** Interpersonal intelligence is the ability to communicate with others by understanding other people's feelings and the influence of the person.

2. Differentiate between what is AI and what is not AI with the help of an example?

AI	Machine	Not AI machine				
1.	AI machines are trained with data and algorithm.	1.	Smart machines which are not AI, do not require training data, they work on			
2.	AI machines learn from mistakes and		algorithms only.			
	experience. They try to improvise on	2.	Smart machines work on fixed			
	their next iterations.		algorithms and they always work with			
3.	AI machines can analyses the situation and can take decisions		the same level of efficiency, which is programmed into them.			

CBSE Question Bank – AI – Class 10 – Chapter 1 Foundational Concepts

AI Machine			Not AI machine				
	accordingly.	3.	Machines which are not AI cannot take				
4.	AI based drones capture the real-time		decisions on their own.				
	data during the flight, processes it in	4.	An automatic door in a shopping mall,				
	real-time, and makes a human-		seems to be AI-enabled, but it is built				
	independent decision based on the		with only sensor technology.				
	processed data.						

(Any other valid examples of AI and Non-AI machine can be considered.)

While we see a lot of AI applications around us, there still exist a lot of them which are smart but not intelligent.

An AI enabled machine should not only recognize, but should also do something with its gathered information. Artificial intelligence" must mean a human-made interface with the power to reason and integrate knowledge. AI must demonstrate at least some of the following behaviors associated with human intelligence: planning, learning, reasoning, problem solving, knowledge representation, perception, motion, manipulation and, to a lesser extent, social intelligence, and creativity.

Most IOT items are ordinary things outfitted with sensors and connected to the Internet. For example, sensors in your office can recognize shadows or movements, but that doesn't make them an example of artificial intelligence. A fully automatic washing machine can work on its own, but it requires human intervention to select the parameters of washing and to do the necessary preparation for it to function correctly before each wash, which makes it an example of automation, not AI.

3. How can AI be integrated with non-AI technologies? Explain with the help of an example.

Today's world is changing with the adoption of IOT (Internet of Things). IOT is helping in prominently capturing a tremendous amount of data from multiple sources. The convergence of AI (Artificial Intelligence) and IOT can redefine the way industries, business, and economies function. AI enabled IoT creates intelligent machines that simulate smart behavior and supports decision making with little or no human interference.

While IOT provides data, artificial intelligence acquires the power to unlock responses, offering both creativity and context to drive smart actions. Here are some examples:

- Ex. 1: Self-driving Cars: Tesla's self-driving cars are the best example of IoT and AI working together. With the power of AI, self-driving cars predict the behavior of pedestrians and cards in various circumstances. For example, they can determine road conditions, optimal speed, weather and getting smarter with each trip.
- Ex. 2: Robots in Manufacturing: Manufacturing is one of the industries that already embraced new technologies like IoT, artificial intelligence, facial recognition, deep learning, Robots and many more. Robots employed in factories are turning smarter with the support of implanted sensors, which facilitates data transmission. Moreover, as the robots are provisioned with artificial intelligence algorithms, they can learn from newer data. This approach not only saves time and cost but also makes the manufacturing process better over time.

- Ex.3: Weather forecasting System: In a weather forecasting system, where IOT temperature sensor and humidity sensors collect data from the physical world, AI tries to figure out patterns from previous data collected and tries to interpret and give accurate predictions of upcoming day weather.
- Ex.4: Smart Drones: Initially drones were only able to capture photographs, these were not AI drones. As the scientist used to analyze the data captured through drones. Now the drones are incorporated with AI, which helps them to make decisions also on the basis of the picture they capture.

*Any one example from above or any other matching example can be given

4. Read the given scenario and answer the questions that follow:

A farmer keeps rabbits in three large hutches that stand in a row in his backyard. Each of the hutches is painted different colours – red, yellow and green. Until recently, the number of rabbits in the green hutch was twice as large as the number of rabbits in the yellow hutch. Then, one day, the farmer took five rabbits out of the left-side hutch and gave them away to the local school's pet corner. He also took half of the rabbits that remained in the left-side hutch and moved them to the red hutch.

a. What was the colour of the left-side hutch? Justify your answer with explanation.

The answer is yellow.

Explanation: As we already know at the outset the number of rabbits in the green hutch was twice as large as the number of rabbits in the yellow hutch. This means that the number of rabbits in the green hutch was an even number. After the farmer removed five rabbits from the left side hutch, then the number of rabbits that remained there also became an even number. This is proven by the fact that it was divisible by 2. Therefore, before those five were removed, the left side hutch contained an uneven number of rabbits hence the left side hutch cannot be the green one, but based on the given information, it cannot be the red one. Hence it is yellow.

5. A scenario is given to you below. Read it and answer the questions that follow:

Late one night, a car ran over a pedestrian in a narrow by street and drove away without stopping. A policeman who saw the vehicle leave the scene of the accident reported it moving at very high speed. The accident itself was witnessed by six bystanders. They provided the following conflicting accounts of what had happened:

- It was a blue car driven by a man;
- The car was moving at high speed and its headlights were turned off;
- The car did have license plates; it wasn't going very fast;
- It was a Toyota and its headlights were turned off;
- The car didn't have license plates; the driver was a woman;
- It was a grey Ford.

When the car and its driver were finally apprehended, it turned out that only one of the six eyewitnesses gave a fully correct description. Each of the other five provided one true and one false piece of information. Keeping that in mind, can you determine the following:

a.	What was the car's brand?	Ans: FORD
b.	What was the colour of the car?	Ans: BLUE
C.	Was the car going fast or slow?	Ans: FAST
d.	Did it have license plates?	Ans: NO
e.	Were its headlights turned on?	Ans: NO
f.	Was the driver a man or a woman?	Ans: WOMAN

Explanation: Out of the statements of 6 bystanders, the third statement becomes false as the policeman who saw the vehicle leave the scene of the accident reported it moving at very high speed. Then eliminating all false statements of bystanders, the above results can be extracted.

6. A firefighter has to get to a burning building as quickly as he can. There are three paths that he can take. He can take his fire engine over a large hill (5 miles) at 8 miles per hour. He can take his fire engine through a windy road (7 miles) at 9 miles per hour. Or he can drive his fire engine along a dirt road which is 8 miles at 12 miles per hour. Which way should he choose? (speed=distance/time)

To reach the destination quickly, the fire fighter has to calculate the time required on the basis of given data. Driving his fire engine 5 miles at 8 miles per hour takes 37.5 minutes. Driving his fire engine 7 miles at 9 miles per hour takes about 47 minutes. Driving his fire engine 8 miles at 12 miles per hour takes 40 minutes So he should choose to drive his fire engine over the hill.)

7. A thief has just found a pair of ancient treasure caves. One of the caves is filled with unbelievable treasure and the other has a fire breathing monster that will eat anyone who opens that cave. One cave has a black door decorated with diamonds and the other cave has a brown door decorated with sapphires. Each of the doors has an engraved description on top. The descriptions say:

a. Black Door: Monster is here.

b. Brown Door: Only One Door speaks the truth.

Which door should the thief open?

The treasure is in the Black door.

Explanation: Let us look at the description on the Brown door. It can be correct or wrong.

Scenario 1: The description on the Brown door is true. Then the description on the Black door has to be false. That means that the inscription on the Black door is false and the cave with black door contains the treasure!

Scenario 2: The description on the Brown door is false. Then either both the descriptions are false or both are true. Both cannot be true as that is impossible and not consistent. That means that both descriptions are false.

8. How intelligent robots are helping us in accomplishing dangerous jobs?

Robots let humans avoid some hurtful work:

- (i) Lifting up heavy material at the construction site.
- (ii) Stirring and mixing metals or liquids at a high temperature.
- (iii) Collecting and packaging of radioactive waste.
- (iv) Working in contaminated and dusty environments.

9. How AI helps in giving you personalized experience online?

AI based recommendations: AI uses advanced machine learning algorithms to analyze browser history, page clicks, social interactions (likes, shares), past purchases, the duration for which a page was viewed, location, etc. to gauge customer interests and preferences. AI can help deliver product recommendations based on frequently bought items, or related products. It can even help customize web pages and elements to suit a customer's needs. For instance, Netflix does intense behavior analysis based on behavior and demographic data to determine the content that will resonate with their customers.

Chatbots and Automated Messaging: AI-powered chatbots and messaging agents can enhance the customer experience across channels. They can answer simple queries, engage customers, efficiently handle multiple interactions,

Automated Service Interactions: AI-driven programs can send automated messages to customers regarding a pending service, a part replacement, or a regular order.

Curating Select Products: Amazon has come up with the concept of the Amazon 4-star retail store. Products that have received a multitude of 4-star ratings will be offered in this physical store. Amazon will use its product recommendation engine to identify trending products and customers' favorites and bring them to a brick and mortar setting.

CBSE | DEPARTMENT OF SKILL EDUCATION

ARTIFICIAL INTELLIGENCE

QUESTION BANK – CLASS 10

CHAPTER 2: INTRODUCTION TO AI: BASICS OF AI

One (01) Mark Questions

Fill in the blanks:

- 1. One of the major sources of data for many major companies is the device which all of us have in our hands all the time______ (Smartphone/ Mobile Phones)
- 2. The world of Artificial Intelligence revolves around _____ (Data)

True/False:

3. All the apps collect some kind of data. (True)

Direct Questions:

4. What do you understand by Machine Learning?

Machine Learning is a subset of Artificial Intelligence which enables machines to improve at tasks with experience (data). The intention of Machine Learning is to enable machines to learn by themselves using the provided data and make accurate Predictions/ Decisions.

OR

Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

OR

Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals: learn from experience.

5. What do you understand by Deep Learning?

Deep Learning is the most advanced form of Artificial Intelligence. In Deep Learning, the machine is trained with huge amounts of data which helps it in training itself around the data. Such machines are intelligent enough to develop algorithms for themselves.

OR

Deep learning is an artificial intelligence (AI) function that imitates the workings of the human brain in processing data and creating patterns for use in decision making.

OR

Deep learning is a subset of machine learning where artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data.

6. What are the three domains of AI?

- Data Science/ Big Data
- Computer Vision
- Natural Language Processing (NLP)

7. Name any two examples of Data science?

(Any two out of the following)

Price Comparison Websites/ Website Recommendations/ Fraud and Risk detection/ Internet search/ Personalized healthcare recommendations / Optimizing Traffic routes in real-time / image tagging.

8. Name any two examples of Computer vision? (Any two out of the following)

> Self-Driving cars/ Autonomous vehicles Face Lock in Smartphones/ MedicalImaging/ Facial recognition /Security Systems / Waste Management / Satellite imaging.

9. Name any two examples of Natural Language Processing? (Any two out of the following)

Email filters/Smart assistants/ Sentiment Analysis/Automatic Summarization/Search results / Language translation / Digital phone calls.

10. Name any two examples of Machine Learning?

(Any two out of the following)

Virtual Personal Assistants, Recommendation systems like Netflix, Face Apps, Online Fraud Detection

New Addition

MCQ (Correct answers are highlighted)

- **11.** Snapchat filters use _____ and _____ to enhance your selfie with flowers, cat ears etc.
 - a) machine learning and deep learning
 - b) data and image processing
 - c) augmented reality and machine learning
 - d) NLP and computer vision

12. Based on the image below, choose the correct domain or domains of AI required for it:



- a) Data
- b) NLP
- c) Computer Vision
- d) Both (a) and (b)
- **13.** Rock paper and scissors game is based on the following domain:
 - a) Data for AI
 - b) Natural Language Processing
 - c) Computer Vision
 - d) Image processing
- **14.** Select a game which is based on Data Science domain of AI:
 - a) Rock Paper and Scissors b) Mystery Animal
 - c) Emoji Scavenger Hunt d) Pokémon
- **15.** Identify the domain of AI in the following image:



- a) Data Science
- b) Natural Language Processing
- c) Computer Vision
- d) Rule Based

Two (02) Mark Questions

1. What is Data science? Give an example of it.

Data sciences is a domain of AI related to data systems and processes, in which the system collects numerous data, maintains data sets and derives meaning/sense out of them. The information extracted through data science can be used to make a decision about it.

OR

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.

OR

Data Sciences, it is a concept to unify statistics, data analysis, machine learning and their related methods in order to understand and analyses actual phenomena with data.

For example: a company that has petabytes of user data may use data science to develop effective ways to store, manage, and analyze the data.

CBSE QUESTION BANK – AI – CLASS 10 – CHAPTER 2 INTRODUCTION TO AI: BASICS 3 OF AI

2. What is Computer Vision? Give an example of it.

Computer Vision, abbreviated as CV, is a domain of AI that depicts the capability of a machine to get and analyze visual information and afterwards predict some decisions about it. The entire process involves image acquiring, screening, analyzing, identifying and extracting information.

OR

Computer vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos.

OR

The Computer Vision domain of Artificial Intelligence, enables machines to see through images or visual data, process and analyze them on the basis of algorithms and methods in order to analyze actual phenomena with images.

For Example: - Self-Driving cars/ Automatic Cars, Face Lock in Smartphones

3. What is Natural Language Processing? Give an example of it.

Natural Language Processing, abbreviated as NLP, is a branch of artificial intelligence that deals with the interaction between machine/computers and humans using the natural language. *Natural language* refers to language that is spoken and written by people, and natural language processing (NLP) attempts to extract information from the spoken and written word using algorithms.

OR

Natural Language Processing, or NLP, is the sub-field of AI that is focused on enabling machine/computers to understand and process human languages. AI is a subfield of Linguistics, Computer Science, Information Engineering, and Artificial Intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

OR

In NLP, we teach machines how to understand and communicate in human language. Natural language refers to speech analysis in both audible speeches, as well as text of a language. NLP systems capture meaning from an input of words (sentences, paragraphs, pages, etc.)

For Example: Email filters, Smart assistants: - Apple's Siri and Amazon's Alexa

4. Where do we collect data from?

Data can be collected from various sources like -

- Surveys
- Sensors
- Observations
- Web scrapping (Internet)
- Interviews
- Documents and records.
- Oral histories

5. Why do we need to collect data?

Data to a machine is similar to food for human being to function. The world of Artificial Intelligence revolves around Data. Every company whether small or big is collecting data from as many sources as possible. Data is called the New Gold today. It is through data collection that a business or management has the quality information they need to make informed decisions from further analysis, study, and research. Data collection allows them to stay on top of trends, provide answers to problems, and analyze new insights to great effect.

6. What is data mining? Explain with example.

Data mining is the process of analyzing large data sets and extracting the useful information from it. Data mining is used by companies to turn raw data into useful information. It is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information

OR

Data mining is an automatic or semi-automatic technical process that analyses large amounts of scattered information to make sense of it and turn it into knowledge. It looks for anomalies, patterns or correlations among millions of records to predict results, as indicated by the SAS institute, a world leader in business analytics.

Example:

Price Comparison websites- They collect data about a product from different sites and then analyze trends out of it and show up the most appropriate results.

Data mining is also known as Knowledge Discovery in Data (KDD) To be moved to chapter no. 3

7. What do you understand by Data Privacy?

The world of Artificial Intelligence revolves around Data. Proper and ethical handling of own data or user data is called data privacy. It is all about the rights of individuals with respect to their personal information.

Data privacy or information privacy is a branch of data security concerned with the proper handling of data – consent, notice, and regulatory obligations. More specifically, practical data privacy concerns often revolve around: Whether or how data is shared with third parties

8. Is data which is collected by various applications ethical in nature? Justify your

Yes, most of the times, the data collected by various applications is ethical in nature as the users agree to it by clicking on allow when the application asks for various permissions. They ask for our data for various facilities like - to show us personalized recommendations and advertisements and to make their app more accurate and efficient.

No, the data collected by various applications is not always ethical in nature. Sometimes, we just share our data to non – trusted third party applications without reading what happens to our data. This may lead to unethical use of our data. If one does not want to share his/her data with anyone, he/she can opt for alternative applications which are of similar usage and keep the data private. For example, an alternative to WhatsApp is the Telegram app which does not collect any data from us.

Note: This is an open-ended question, so both the answers yes/no will be considered right with correct justification.

Artificial Intelligence Machine Learning Deep Learning

9. Fill in the blanks for the image given below:

Three (03) Mark Questions

1. What do you understand by AI bias? Discuss in detail with some examples.

AI bias is the underlying prejudice in data that's used to create AI algorithms, which can ultimately result in discrimination and other social consequences.

AI Bias can creep into algorithms in several ways. AI systems learn to make decisions based on training data, which can include biased human decisions or reflect historical or social inequities, even if sensitive variables such as gender, race, or sexual orientation are removed. Amazon stopped using a hiring algorithm after finding it favored applicants based on words like "executed" or "captured" that were more commonly found on men's resumes, for example. Another source of bias is flawed data sampling, in which groups are over- or underrepresented in the training data.

For Example

- Majorly, all the virtual assistants have a female voice. It is only now that some companies have understood this bias and have started giving options for male voices but since the virtual assistants came into practice, female voices are always preferred for them over any other voice. Can you think of some reasons for this?
- If you search on Google for salons, the first few searches are mostly for female salons. This is based on the assumption that if a person is searching for a salon, in all probability it would be a female. Do you think this is a bias? If yes, then is it a Negative bias or Positive one?

2. What do you understand by Data Privacy? Discuss in detail with some examples.

Data privacy, sometimes also referred to as information privacy, is an area of data protection that concerns the proper handling of sensitive data including, notably, personal data but also other confidential data, such as certain financial data and intellectual property data, to meet regulatory requirements as well as protecting the confidentiality and immutability of the data. It focuses on how to collect, process, share, archive, and delete the data in accordance with the law.

Privacy, in the broadest sense, is the right of individuals, groups, or organizations to control who can access, observe, or use something they own, such as their bodies, property, ideas, data, or information.

Control is established through physical, social, or informational boundaries that help prevent unwanted access, observation, or use. For example:

- A physical boundary, such as a locked front door, helps prevent others from entering a building without explicit permission in the form of a key to unlock the door or a person inside opening the door.
- A social boundary, such as a members-only club, only allows members to access and use club resources.
- An informational boundary, such as a non-disclosure agreement, restricts what information can be disclosed to others.

Privacy of information is extremely important in this digital age where everything is interconnected and can be accessed and used easily. The possibilities of our private information being extremely vulnerable are very real, which is why we require data privacy.

3. What do you understand by AI, ML & DL? How are they different from each other?

a) Artificial Intelligence (AI)

AI is incorporating human intelligence to machines. Whenever a machine completes tasks based on a set of rules that solve problems (algorithms), such an "intelligent" behavior is what is called artificial intelligence.

b) Machine Learning (ML)

ML is a subset of AI that uses statistical learning algorithms to build smart systems. The ML systems can automatically learn and improve without explicitly being programmed.

c) Deep Learning (DL)

In Deep Learning, the machine is trained with huge amounts of data which helps it in training itself around the data. Such machines are intelligent enough to develop algorithms for themselves.



How they differ?

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

4. Why do apps collect data in our phone?

One of the major sources of data for many major companies is the device which all of us have in our hands all the time: Smartphones. Smartphones have nowadays become an integral part of our lives. Most of us use smartphones more than we interact with people around us. For the facilities that smartphones provide us, Apps need a lot of data which is collected from the user like details about your face, browsing history, or your geographic location, contact list etc. This data is collected with user's consent which he/she gives at the time of installing an app by clicking on "yes" or "allow" options which clearly means that we ourselves are giving permissions to the Apps.

Permissions by themselves are harmless and even useful to provide users a good mobile experience.

This data is collected to provide us with a lot of facilities and features which have made our lives easier. Another reason to collect the data is to provide us with customized recommendations and notifications according to our choices.

8

One more reason to collect the data is to make their app more accurate and efficient.

5. Should AI replace laborious jobs? Is there an alternative for major unemployment?

Yes, AI should replace laborious jobs.

- AI can replace laborious jobs like lifting of heavy items, working in mines etc.
- AI can indeed automate most repetitive and physical tasks.
- In future, AI would be a good option in the field of architecture and construction.

OR

No, AI should not replace laborious jobs completely as if it replaces laborious jobs completely, then there will be no source of income for the daily wage workers due to unemployment. So, industry owners can use some machines but more of man power. Hence the production will not get affected as humans are smarter than machines since they were the ones who invented AI.

Note: As this is an open-ended question so both the answers (yes/No) are correct but it must be with correct justification.

Is there an alternative for major unemployment?

- AI taking over laborious jobs won't create unemployment. It is just a groundless fear. The standard view of technical change is that some jobs are displaced by the substitution of machines for labour, but that the fear of total displacement is misplaced because new jobs are created, largely due to the technology-fuelled increase in productivity. Humans have always shifted away from work suitable for machines and to other jobs.
- The basic fact is that technology eliminates jobs, not work. If this level of AI revolution will happen, lots of job opportunities will be created. For example: 20-30 years ago, being an accountant was a lucrative job, but AI took over this job but this created a lot of opportunities, it raised the demand of a software engineer, data scientist, etc.
- It will open doors to skillful jobs rather than doing laborious tasks.
- Thus, we will be able to cope with the level of major unemployment, if AI took over laborious jobs.
- 6. As Artificially Intelligent machines become more and more powerful, their ability to accomplish tedious tasks is becoming better. Hence, it is now that AI machines have started replacing humans in factories. While people see it in a negative way and say AI has the power to bring mass unemployment and one day, machines would enslave humans, on the other hand, other people say that machines are meant to ease our lives. If machines over take monotonous and tedious tasks, humans should upgrade their skills to remain their masters always.

What according to you is a better approach towards this ethical concern? Justify your answer.

• AI taking over laborious jobs won't create unemployment. It is just a groundless fear. The standard view of technical change is that some jobs are displaced by the substitution of machines for labour, but that the fear of total displacement is misplaced because new jobs are created, largely due to the technology-fuelled increase in productivity. Humans have always shifted away from work suitable for machines and to other jobs.

- The basic fact is that technology eliminates jobs, not work. If this level of AI revolution will happen, lots of job opportunities will be created. For example: 20-30 years ago, being an accountant was a lucrative job, AI took over this job but this created a lot of opportunities, it raised the demand of a software engineer, data scientist, etc.
- It will open doors to skillful jobs rather than doing laborious tasks.
- Thus, we will be able to cope with the level of major unemployment, if AI took over laborious jobs.
- 7. List down various sensors that are present in a smartphone. Also list down the type of data which gets collected through them.
 - ACCELEROMETER [helps running AR applications and track steps]
 - GPS [Location Data]
 - Gyroscope [Orientation Data]
 - Magnetometer [Direction and Magnetic Field Data]
 - Biometric Sensors [Fingerprint ,Iris, Face data]

New Additions

- 1. (Case Study) AI and robotics have raised some questions regarding liability. Take for example the scenario of an 'autonomous' or AI-driven robot moving through a factory. Another robot surprisingly crosses its way and our robot draws aside to prevent collision. However, by this manoeuvre the robot injures a person.
 - a) Who can be held liable for damages caused by autonomous systems?

It is actually very difficult to blame anyone in such a scenario. Here is the situation where AI Ethics come in to the picture. Here, the choices might differ from person to person and one must understand that nobody is wrong in this case. Every person has a different perspective and hence he/she takes decisions according to their moralities. But still if someone is to be liable then it should be the programmer who has designed the algorithm of the autonomous vehicle as he/she should have considered all the exceptional conditions that could arise.

b) List two AI Ethics.

(Any two out of the following) AI Bias, AI Access, Data privacy, AI for kids.

CBSE | DEPARTMENT OF SKILL EDUCATION

ARTIFICIAL INTELLIGENCE

QUESTION BANK - CLASS 10

CHAPTER 3: AI PROJECT CYCLE

One (01) Mark Questions

1. Name all the stages of an AI Project cycle.

Problem Scoping, Data Acquisition, Data Exploration, Modeling, Evaluation

2. What are sustainable development goals?

The Sustainable Development Goals (SDGs), also known as the Global Goals, were adopted by all United Nations Member States in 2015 as a universal call to action to end poverty, protect the planet and ensure that all people enjoy peace and prosperity.

OR

The Sustainable Development Goals (SDGs) or Global Goals are a collection of 17 interlinked goals designed to be a "blueprint to achieve a better and more sustainable future for all" so that the future generations may live in peace and prosperity.

- 3. Name the 4Ws of problem canvases under the problem scoping stage of the AI Project Cycle.
 - a. Who, b. what c. where d. why

4. What is Testing Dataset?

The dataset provided to the model ML. algorithm after training the algorithm

5. Mention the types of learning approaches for AI modeling.

Supervised, unsupervised and re-enforcement

6. What is the objective of evaluation stage?

It is to evaluate whether the ML algorithm is able to predict with high accuracy or not before deployment.

7. Fill in the blank:

The analogy of an Artificial Neural Network can be made with _____? (Parallel Processing)

8. Which of the following is not an authentic source for data acquisition?

a. **Sensors b. Surveys c. Web Scraping d. System Hacking** System Hacking

- 9. Which type of graphical representation suits best for continuous type of data like monthly exam scores of a student?

Two (02) Mark Questions

1. What are the two different approaches for AI modelling? Define them.

There are two approaches for AI Modelling; Rule Based and Learning Based.

The Rule based approach generates pre-defined outputs based on certain rules programmed by humans. Whereas, machine learning approach has its own rules based on the output and data used to train the models.

OR

Rule Based Approach Refers to the AI modelling where the relationship or patterns in data are defined by the developer. The machine follows the rules or instructions mentioned by the developer, and performs its task accordingly. Whereas in Learning based approach, the relationship or patterns in data are not defined by the developer. In this approach, random data is fed to the machine and it is left to the machine to figure out patterns and trends out of it

2. What is a problem statement template and what is its significance?

The problem statement template gives a clear idea about the basic framework required to achieve the goal. It is the 4Ws canvas which segregates; what is the problem, where does it arise, who is affected, why is it a problem? It takes us straight to the goal.

3. Explain any two SDGs in detail.

1. No Poverty: This is Goal 1 and strives to End poverty in all its forms everywhere globally by 2030. The goal has a total of seven targets to be achieved.

2. Quality Education: This is Goal 4 which aspires to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. It has 10 targets to achieve.

* (Any two goals can be defined)

4. Mention the precautions to be taken while acquiring data for developing an AI Project.

It should be from an authentic source, and accurate. Look for redundant and irrelevant data parameters that does not take part in prediction.

5. What do you mean by Data Features?

The type of data to collect, It should be relevant data.

6. Write the names for missing stages in the given AI project cycle:



Problem scoping, Evaluation

7. Draw the icons of the following SDGs:

Gender Equality

Clean Water and sanitation

8. Draw the graphical representation of Classification AI model. Explain in brief.

Classification: The classification Model works on the labelled data. For example, we have 3 coins of different denomination which are labelled according to their weight then the model would look for the labelled features for predicting the output. This model works on discrete dataset which means the data need not be continuous.

OR

In classification, data is categorized under different labels according to some parameters given in input and then the labels are predicted for the data.



9. Draw the graphical representation of Regression AI model. Explain in brief.

Regression: These models work on continuous data to predict the output based on patterns. For example, if you wish to predict your next salary, then you would put in the data of your previous salary, any increments, etc., and would train the model. Here, the data which has been fed to the machine is continuous.

OR

Regression is the process of finding a model for distinguishing the data into continuous real values instead of using discrete values. It can also identify the distribution movement depending on the historical data.



10. Draw the graphical representation of Clustering AI model. Explain in brief.

Clustering: It refers to the unsupervised learning algorithm which can cluster the unknown data according to the patterns or trends identified out of it. The patterns observed might be the ones which are known to the developer or it might even come up with some unique patterns out of it.

OR

Clustering is the task of dividing the data points into a number of groups such that data points in the same groups are more similar to other data points in the same

group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.



11. Explain Data Exploration stage.

In this stage of project cycle, we try to interpret some useful information out of the data we have acquired. For this purpose, we need to explore the data and try to put it uniformly for a better understanding. This stage deals with validating or verification of the collected data and to analyze that:

- The data is according to the specifications decided.
- The data is free from errors.
- The data is meeting our needs.

12. What are the features of an Artificial Neural Network?

Any Artificial Neural Network, irrespective of the style and logic of implementation, has a few basic features as given below.

- The Artificial Neural Network systems are modelled on the human brain and nervous system.
- They are able to automatically extract features without feeding the input by programmer.
- Every node of layer in a Neural Network is compulsorily a machine learning algorithm.
- It is very useful to implement when solving problems for very huge datasets.

OR

- It can work with incomplete knowledge and may produce output even with incomplete information.
- It has fault tolerance which means that corruption of one or more cells of ANN does not prevent it from generating output.
- It has the ability to learn events and make decisions by commenting on similar events.
- It has Parallel processing capability i.e. ANN have numerical strength that can perform more than one job at the same time.

OR

- Neural Networks have the ability to learn by themselves and produce the output that is not limited to the input provided to them.
- The input is stored in its own networks instead of a database; hence the loss of data does not affect its working.

- These networks can learn from examples and apply them when a similar event arises, making them able to work through real-time events.
- Even if a neuron is not responding or a piece of information is missing, the network can detect the fault and still produce the output.
- They can perform multiple tasks in parallel without affecting the system performance

13. What is the purpose of getting AI Ready?

The world is changing with each day and we have huge data coming our way. The purpose of getting AI ready means taking steps to collect data around relevant systems, equipment, and procedures; and storing and curating that data in a way that makes it easily accessible to others for use in future AI applications.

OR

The purpose of getting AI ready specifies the responsible and optimum use of huge amount of data around us to create and implement into such systems and applications which should make life of future generations more organized and sustainable. This process may lead to better lives for mankind.

14. What are the different types of sources of data from where we can collect reliable and authentic datasets? Explain in brief.

Data can be a piece of information or facts and statistics collected together for reference or analysis. Whenever we want an AI project to be able to predict an output, we need to train it first using data There could be many ways and sources from where we can collect reliable and authentic datasets namely Surveys, Web scrapping, Sensors, Cameras, Observations, Research, Investigation, API etc.

Sometimes Internet is also used to acquire data but the most important point to keep in mind is that the data should be taken from reliable and authentic websites only. Some reliable data sources are UN, Google scholar, Finance, CIA, Data.gov etc.

Four (04) Mark Questions

1. Explain the AI Project Cycle in detail.

The steps involved in AI project cycle are as given:

- The first step is Scope the Problem by which, you set the goal for your AI project by stating the problem which you wish to solve with it. Under problem scoping, we look at various parameters which affect the problem we wish to solve so that the picture becomes clearer
- Next step is to acquire data which will become the base of your project as it will help you in understanding what the parameters that are related to problem scoping.
- Next, you go for data acquisition by collecting data from various reliable and authentic sources. Since the data you collect would be in large quantities, you can try to give it a visual image of different types of representations like graphs, databases, flow charts, maps, etc. This makes it easier for you to interpret the patterns in which your acquired data follows.
- After exploring the patterns, you can decide upon the type of model you would build to achieve the goal. For this, you can research online and select various models which give a suitable output.
- You can test the selected models and figure out which is the most efficient one.
- The most efficient model is now the base of your AI project and you can develop your algorithm around it.
- Once the modelling is complete, you now need to test your model on some newly

fetched data. The results will help you in evaluating your model and hence improving it.

Finally, after evaluation, the project cycle is now complete and what you get is your AI project.

2. Explain the relation between data size and model performance of an Artificial Neural Network.

The basis for any kind of AI development is BIG DATASET. The performance of any AI based application depends on the data supplied

ANN models are also known as Learning models and are used for prediction purposes. These are mostly developed without paying much cognizance to the size of datasets that can produce models of high accuracy and better generalization. Although, the general belief is that, large dataset is needed to construct a predictive learning model. To describe a data set as large in size, perhaps, is circumstance dependent, thus, what constitutes a dataset to be considered as being big or small is somehow vague.

In fact, the quantity of data partitioned for the purpose of training must be of good representation of the entire sets and sufficient enough to span through the input space. It must be authentic and relevant to give better model performance.

3. Draw the 4Ws problem canvas and explain each one of them briefly.

The 4Ws problem canvas is the basic template while scoping a problem and using this canvas, the picture becomes clearer while we are working to solve it.

- a) **Who:** The "Who" block helps you in analyzing the people getting affected directly or indirectly due to it? Under this, you find out who the 'stakeholders' to this problem are and what you know about them. Stakeholders are the people who face this problem and would be benefitted with the solution.
- b) **What:** Under the "What" block, you need to look into what you have on hand. At this stage, you need to determine the nature of the problem. What is the problem and how do you know that it is a problem?
- c) **Where:** In this block, you need to focus on the context/situation/location of the problem. It will help you look into the situation in which the problem arises, the context of it, and the locations where it is prominent.
- d) **Why:** in the "Why" canvas, think about the benefits which the stakeholders would get from the solution and how would it benefit them as well as the society.

4. Differentiate between rule-based and learning-based AI modelling approaches.

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

For example, suppose you have a dataset comprising of 100 images of apples and 100 images of bananas. To train your machine, you feed this data into the machine and label each image as either apple or banana. Now if you test the machine with the image of an apple, it will compare the image with the trained data and according to the labels of trained images, it will identify the test image as an apple. This is known as Rule based approach. The rules given to the machine in this example are the labels given to the machine for each image in the training dataset.

Learning Based Approach: In this approach, the machine learns by itself. It refers to the AI modelling where the relationship or patterns in data are not defined by the developer. In this approach, random data is fed to the machine and it is left on the

machine to figure out patterns and trends out of it. Generally, this approach is followed when the data is un labelled and too random for a human to make sense out of it. For example, suppose you have a dataset of 1000 images of random stray dogs of your area. You would put this into a learning approach-based AI machine and the machine would come up with various patterns it has observed in the features of these 1000 images which you might not have even thought of!

5. What is an Artificial Neural Network? Explain the layers in an artificial neural network.

Artificial Neural Network: Modeled in accordance with the human brain, a Neural Network was built to mimic the functionality of a human brain. The human brain is a neural network made up of multiple neurons, similarly, an Artificial Neural Network (ANN) is made up of multiple perceptrons.

A neural network consists of three important layers:

Input Layer: As the name suggests, this layer accepts all the inputs provided by the programmer.

Hidden Layer: Between the input and the output layer is a set of layers known as Hidden layers. In this layer, computations are performed which result in the output. There can be any number of hidden layers

Output Layer: The inputs go through a series of transformations via the hidden layer which finally results in the output that is delivered via this layer.

6. What is the need of an AI Project Cycle? Explain.

Project cycle is the process of planning, organizing, coordinating, and finally developing a project effectively throughout its phases, from planning through execution then completion and review to achieve pre-defined objectives.

Our mind makes up plans for every task which we have to accomplish which is why things become clearer in our mind. Similarly, if we have to develop an AI project, the AI Project Cycle provides us with an appropriate framework which can lead us towards the goal.

The major role of AI Project Cycle is to distribute the development of AI project in various stages so that the development becomes easier, clearly understandable and the steps / stages should become more specific to efficiently get the best possible output. It mainly has 5 ordered stages which distribute the entire development in specific and clear steps: These are Problem Scoping, Data Acquisition, Data Exploration, Modelling and Evaluation.

7. Explain the following:

a. Supervised Learning

b. Unsupervised Learning

• **Supervised learning** is an approach to creating artificial intelligence (AI), where the program is given labelled input data and the expected output results.

OR

• **Supervised learning** is a learning in which we teach or train the machine using data which is well labelled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples (data) so that supervised learning algorithm analyses the training data (set of training examples) and produces a correct outcome from labelled data.

• In a supervised learning model, the dataset which is fed to the machine is labelled. It means some data is already tagged with the correct answer. In other words, we can say that the dataset is known to the person who is training the machine only then he/she is able to label the data.

• **Unsupervised Learning**: An unsupervised learning model works on unlabeled dataset. This means that the data which is fed to the machine is random and there is a possibility that the person who is training the model does not have any information regarding it. The unsupervised learning models are used to identify relationships, patterns and trends out of the data which is fed into it. It helps the user in understanding what the data is about and what are the major features identified by the machine in it.

OR

• **Unsupervised learning** is the training of a machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns and differences without any prior training of data.

8. Differentiate between classification and clustering algorithms with the help of suitable examples.

Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters. In Classification, a computer program is trained on the training dataset and based on that training; it categorizes the data into different classes. The task of the classification algorithm is to find the mapping function to map the input(x) to the discrete output(y).

Example: The best example to understand the Classification problem is Email Spam Detection. The model is trained on the basis of millions of emails on different parameters, and whenever it receives a new email, it identifies whether the email is spam or not. If the email is spam, then it is moved to the Spam folder.

Regression is a process of finding the correlations between dependent and independent variables. It helps in predicting the continuous variables such as prediction of Market Trends, prediction of House prices, etc. The task of the Regression algorithm is to find the mapping function to map the input variable(x) to the continuous output variable(y).

Example: Suppose we want to do weather forecasting, so for this, we will use the Regression algorithm. In weather prediction, the model is trained on the past data, and once the training is completed, it can easily predict the weather for future days.

OR

Classification is the process of finding or discovering a model (function) which helps in separating the data into multiple categorical classes. In classification, the group membership of the problem is identified, which means the data is categorized under different labels according to some parameters and then the labels are predicted for the data.

Regression is the process of finding a model or function for distinguishing the data into continuous real values instead of using classes. Mathematically, with a regression problem, one is trying to find the function approximation with the minimum error deviation. In regression, the data numeric dependency is predicted to distinguish it. The Regression analysis is the statistical model which is used to predict the numeric data instead of labels. It can also identify the distribution movement depending on the available data or historic data.

CBSE Question Bank – AI – Class 10 – Chapter 3 AI Project Cycle

OR

OR

Key Differences between Classification and Regression

- The Classification process models a function through which the data is predicted in discrete class labels. On the other hand, regression is the process of creating a model which predicts continuous quantity.
- The classification algorithms involve decision tree, logistic regression, etc. In contrast, regression tree (e.g. Random forest) and linear regression are the examples of regression algorithms.
- Classification predicts unordered data while regression predicts ordered data.
- Regression can be evaluated using root mean square error. On the contrary, classification is evaluated by measuring accuracy.



- 9. Five sustainable Development Goals are mentioned below. Write 2 problems under each goal that you think should be addressed for achieving the goal.
 - a. Quality Education
 - b. Reduced Inequalities
 - c. Life on Land
 - d. No Poverty
 - e. Clean Water and Sanitation

a. Quality Education:

- i. Providing education remotely, leveraging hi-tech, low-tech and no-tech approaches;
- ii. Ensure coordinated responses and avoid overlapping efforts;
- iii. Ensuring return of students to school when they reopen to avoid an upsurge in dropout rates.

b. Reduced inequalities:

- i. Reduction of relative economic inequalities inequality in some countries having poorest and most vulnerable communities.
- ii. Improving the situations in countries with weaker health systems.

c. Life on Land:

- i. Prevention of Deforestation caused by humans and restoration of land
- ii. Preventions and cure of diseases that are transmissible between animals and humans

d. No Poverty

- i. Creation of Strong social protection systems to prevent people from falling into poverty
- ii. Reduction of social exclusion, and high vulnerability of certain populations to disasters and diseases.
- iii. Responsible distribution of resources.

e. Clean Water and Sanitation

- i. To increase access to clean drinking water and sanitation mostly in rural areas
- ii. Managing our water sustainably to manage our production of food and energy.

10. Do ethics in AI hamper data acquisition stage? Justify your answer.

Data acquisition is the most important factor or stage as the entire project development is based on the acquired data. There are several ethical issues which must always be considered when planning any type of data collection.

We need to understand that the data which is collected is ethical only if the provider agrees to provide. For example, in case of smartphone users, data is collected by clicking on allow when it asks for permission and by agreeing to all the terms and conditions. But at the same time if one does not want to share his/her data with anyone then this ethical issue hampers the acquisition process and lowers the accuracy or amount of data required for development.

Hence Regardless of the type of data collection, it is absolutely necessary to gain the approval of the community from which the data will collected otherwise.

CBSE | DEPARTMENT OF SKILL EDUCATION

ARTIFICIAL INTELLIGENCE QUESTION BANK – CLASS 10

CHAPTER 7: NATURAL LANGUAGE PROCESSING

One (01) Mark Questions

1. What is a Chabot?

A chatbot is a computer program that's designed to simulate human conversation through voice commands or text chats or both. Eg: Mitsuku Bot, Jabberwacky etc.

OR

A chatbot is a computer program that can learn over time how to best interact with humans. It can answer questions and troubleshoot customer problems, evaluate and qualify prospects, generate sales leads and increase sales on an ecommerce site.

OR

A chatbot is a computer program designed to simulate conversation with human users. A chatbot is also known as an artificial conversational entity (ACE), chat robot, talk bot, chatterbot or chatterbox.

OR

A chatbot is a software application used to conduct an on-line chat conversation via text or text-to-speech, in lieu of providing direct contact with a live human agent.

2. What is the full form of NLP? Natural Language Processing

3. While working with NLP what is the meaning of?

- **a.** Syntax
- **b.** Semantics

Syntax: Syntax refers to the grammatical structure of a sentence. **Semantics**: It refers to the meaning of the sentence.

4. What is the difference between stemming and lemmatization?

Stemming is a technique used to extract the base form of the words by removing affixes from them. It is just like cutting down the branches of a tree to its stems. For example, the stem of the words *eating*, *eats*, *eaten* is *eat*.

Lemmatization is the grouping together of different forms of the same word. In search queries, lemmatization allows end users to query any version of a base word and get relevant results.

OR

Stemming is the process in which the affixes of words are removed and the words are converted to their base form.

In lemmatization, the word we get after affix removal (also known as lemma) is a meaningful one. Lemmatization makes sure that lemma is a word with meaning and hence it takes a longer time to execute than stemming.

OR

Stemming algorithms work by cutting off the end or the beginning of the word, taking into account a list of common prefixes and suffixes that can be found in an inflected word.

Lemmatization on the other hand, takes into consideration the morphological analysis of the words. To do so, it is necessary to have detailed dictionaries which the algorithm can look through to link the form back to its lemma.

5. What is the full form of TFIDF?

Term Frequency and Inverse Document Frequency

6. What is meant by a dictionary in NLP?

Dictionary in NLP means a list of all the unique words occurring in the corpus. If some words are repeated in different documents, they are all written just once as while creating the dictionary.

7. What is term frequency?

Term frequency is the frequency of a word in one document. Term frequency can easily be found from the document vector table as in that table we mention the frequency of each word of the vocabulary in each document.

8. Which package is used for Natural Language Processing in Python programming? Natural Language Toolkit (NLTK). NLTK is one of the leading platforms for building Python programs that can work with human language data.

9. What is a document vector table?

Document Vector Table is used while implementing Bag of Words algorithm. In a document vector table, the header row contains the vocabulary of the corpus and other rows correspond to different documents.

If the document contains a particular word it is represented by 1 and absence of word is represented by 0 value.

OR

Document Vector Table is a table containing the frequency of each word of the vocabulary in each document.

10. What do you mean by corpus?

In Text Normalization, we undergo several steps to normalize the text to a lower level. That is, we will be working on text from multiple documents and the term used for the whole textual data from all the documents altogether is known as corpus.

OR

A corpus is a large and structured set of machine-readable texts that have been produced in a natural communicative setting.

OR

A corpus can be defined as a collection of text documents. It can be thought of as just a bunch of text files in a directory, often alongside many other directories of text files.

Two (02) Mark Questions

1. What are the types of data used for Natural Language Processing applications? Natural Language Processing takes in the data of Natural Languages in the form of written words and spoken words which humans use in their daily lives and operates on this.

2. Differentiate between a script-bot and a smart-bot.

(Any 2 differences)

Script-bot	Smart-bot
 A scripted chatbot doesn't carry even a glimpse of A.I Script bots are easy to make Script bot functioning is very limited as they are less powerful. Script bots work around a script which is programmed in them No or little language processing skills 	 Smart bots are built on NLP and ML. Smart -bots are comparatively difficult to make. Smart-bots are flexible and powerful. Smart bots work on bigger databases and other resources directly NLP and Machine learning skills are required. Wide functionality

3. Give an example of the following:

- Multiple meanings of a word
- Perfect syntax, no meaning

• Example of Multiple meanings of a word -

His face turns red after consuming the medicine

Meaning - Is he having an allergic reaction? Or is he not able to bear the taste of that medicine?

• Example of Perfect syntax, no meaning-

Chickens feed extravagantly while the moon drinks tea.

This statement is correct grammatically but it does not make any sense. In Human language, a perfect balance of syntax and semantics is important for better understanding.

4. What is inverse document frequency?

To understand inverse document frequency, first we need to understand document frequency.

Document Frequency is the number of documents in which the word occurs irrespective of how many times it has occurred in those documents.

In case of inverse document frequency, we need to put the document frequency in the denominator while the total number of documents is the numerator.

For example, if the document frequency of a word "AMAN" is 2 in a particular document then its inverse document frequency will be 3/2. (Here no. of documents is 3)

5. Define the following:

- Stemming
- Lemmatization

Stemming: Stemming is a rudimentary rule-based process of stripping the suffixes ("ing", "ly", "es", "s" etc) from a word.

Stemming is a process of reducing words to their word stem, base or root form (for example, books — book, looked — look).

Lemmatization: Lemmatization, on the other hand, is an organized & step by step procedure of obtaining the root form of the word, it makes use of vocabulary (dictionary importance of words) and morphological analysis (word structure and grammar relations).

The aim of lemmatization, like stemming, is to reduce inflectional forms to a common base form. As opposed to stemming, lemmatization does not simply chop off inflections. Instead it uses lexical knowledge bases to get the correct base forms of words.

OR

Stemming is a technique used to extract the base form of the words by removing affixes from them. It is just like cutting down the branches of a tree to its stems. For example, the stem of the words *eating*, *eats*, *eaten* is *eat*.

Lemmatization is the grouping together of different forms of the same word. In search queries, lemmatization allows end users to query any version of a base word and get relevant results.

OR

Stemming is the process in which the affixes of words are removed and the words are converted to their base form.

In lemmatization, the word we get after affix removal (also known as lemma) is a meaningful one. Lemmatization makes sure that lemma is a word with meaning and hence it takes a longer time to execute than stemming.

OR

Stemming algorithms work by cutting off the end or the beginning of the word, taking into account a list of common prefixes and suffixes that can be found in an inflected word.

Lemmatization on the other hand, takes into consideration the morphological analysis of the words. To do so, it is necessary to have detailed dictionaries which the algorithm can look through to link the form back to its lemma.

6. What do you mean by document vectors?

Document Vector contains the frequency of each word of the vocabulary in a particular document.

In document vector vocabulary is written in the top row. Now, for each word in the document, if it matches with the vocabulary, put a 1 under it. If the same word appears again, increment the previous value by 1. And if the word does not occur in that document, put a 0 under it.

7. What is TFIDF? Write its formula.

Term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.

The number of times a word appears in a document divided by the total number of words in the document. Every document has its own term frequency.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}}$$

8. Which words in a corpus have the highest values and which ones have the least? Stop words like - and, this, is, the, etc. have highest values in a corpus. But these words do not talk about the corpus at all. Hence, these are termed as stopwords and are mostly removed at the pre-processing stage only.

Rare or valuable words occur the least but add the most importance to the corpus. Hence, when we look at the text, we take frequent and rare words into consideration.



9. Does the vocabulary of a corpus remain the same before and after text normalization? Why?

No, the vocabulary of a corpus does not remain the same before and after text normalization. Reasons are –

- In normalization the text is normalized through various steps and is lowered to minimum vocabulary since the machine does not require grammatically correct statements but the essence of it.
- In normalization Stop words, Special Characters and Numbers are removed.
- In stemming the affixes of words are removed and the words are converted to their base form.

So, after normalization, we get the reduced vocabulary.

10. What is the significance of converting the text into a common case?

In Text Normalization, we undergo several steps to normalize the text to a lower level. After the removal of stop words, we convert the whole text into a similar case, preferably lower case. This ensures that the case-sensitivity of the machine does not consider same words as different just because of different cases.

11. Mention some applications of Natural Language Processing.

Natural Language Processing Applications-

- Sentiment Analysis.
- Chatbots & Virtual Assistants.
- Text Classification.
- Text Extraction.
- Machine Translation
- Text Summarization
- Market Intelligence
- Auto-Correct

12. What is the need of text normalization in NLP?

Since we all know that the language of computers is Numerical, the very first step that comes to our mind is to convert our language to numbers.

This conversion takes a few steps to happen. The first step to it is Text Normalization. Since human languages are complex, we need to first of all simplify them in order to make sure that the understanding becomes possible. Text Normalization helps in cleaning up the textual data in such a way that it comes down to a level where its complexity is lower than the actual data.

13. Explain the concept of Bag of Words.

Bag of Words is a Natural Language Processing model which helps in extracting features out of the text which can be helpful in machine learning algorithms. In bag of words, we get the occurrences of each word and construct the vocabulary for the corpus. Bag of Words just creates a set of vectors containing the count of word occurrences in the document (reviews). Bag of Words vectors are easy to interpret.

14. Explain the relation between occurrence and value of a word.



plot of occurrence of words versus their value

As shown in the graph, occurrence and value of a word are inversely proportional. The words which occur most (like stop words) have negligible value. As the occurrence of words drops, the value of such words rises. These words are termed as rare or valuable words. These words occur the least but add the most value to the corpus.

15. What are the applications of TFIDF?

TFIDF is commonly used in the Natural Language Processing domain. Some of its applications are:

- Document Classification Helps in classifying the type and genre of a document.
- Topic Modelling It helps in predicting the topic for a corpus.
- Information Retrieval System To extract the important information out of a corpus.
- Stop word filtering Helps in removing the unnecessary words out of a text body.

16. What are stop words? Explain with the help of examples.

"Stop words" are the most common words in a language like "the", "a", "on", "is", "all". These words do not carry important meaning and are usually removed from texts. It is possible to remove stop words using <u>Natural Language Toolkit (NLTK)</u>, a suite of libraries and programs for symbolic and statistical natural language processing.

17. Differentiate between Human Language and Computer Language.

Humans communicate through language which we process all the time. Our brain keeps on processing the sounds that it hears around itself and tries to make sense out of them all the time.

On the other hand, the computer understands the language of numbers. Everything that is sent to the machine has to be converted to numbers. And while typing, if a single mistake is made, the computer throws an error and does not process that part. The communications made by the machines are very basic and simple.

Four 04 Mark Questions

1. Create a document vector table for the given corpus: Document 1: We are going to Mumbai Document 2: Mumbai is a famous place. Document 3: We are going to a famous place. Document 4: I am famous in Mumbai.

We	Are	going	to	Mumbai	is	а	famous	place	Ι	am	in
1	1	1	1	1	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	1	0	0	0
1	1	1	1	0	0	1	1	1	0	0	0
0	0	0	0	1	0	0	1	0	1	1	1

2. Classify each of the images according to how well the model's output matches the data samples:



Here, the red dashed line is model's output while the blue crosses are actual data samples.

- The model's output does not match the true function at all. Hence the model is said to be under fitting and its accuracy is lower.
- In the second case, model performance is trying to cover all the data samples even if they are out of alignment to the true function. This model is said to be over fitting and this too has a lower accuracy
- In the third one, the model's performance matches well with the true function which states that the model has optimum accuracy and the model is called a perfect fit.

3. Explain how AI can play a role in sentiment analysis of human beings?

The goal of sentiment analysis is to identify sentiment among several posts or even in the same post where emotion is not always explicitly expressed.

Companies use Natural Language Processing applications, such as sentiment analysis, to identify opinions and sentiment online to help them understand what customers think about their products and services (i.e., "I love the new iPhone" and, a few lines later "But sometimes it doesn't work well" where the person is still talking about the iPhone) and overall *

Beyond determining simple polarity, sentiment analysis understands sentiment in context to help better understand what's behind an expressed opinion, which can be extremely relevant in understanding and driving purchasing decisions.



4. Why are human languages complicated for a computer to understand? Explain. The communications made by the machines are very basic and simple. Human communication is complex. There are multiple characteristics of the human language that might be easy for a human to understand but extremely difficult for a computer to understand.

For machines it is difficult to understand our language. Let us take a look at some of them here:

Arrangement of the words and meaning - There are rules in human language. There are nouns, verbs, adverbs, adjectives. A word can be a noun at one time and an adjective some other time. This can create difficulty while processing by computers.

Analogy with programming language- Different syntax, same semantics: 2+3 = 3+2 Here the way these statements are written is different, but their meanings are the same that is 5. Different semantics, same syntax: 2/3 (Python 2.7) $\neq 2/3$ (Python 3) Here the statements written have the same syntax but their meanings are different. In Python 2.7, this statement would result in 1 while in Python 3, it would give an output of 1.5.

Multiple Meanings of a word - In natural language, it is important to understand that a word can have multiple meanings and the meanings fit into the statement according to the context of it.

Perfect Syntax, no Meaning - Sometimes, a statement can have a perfectly correct syntax but it does not mean anything. In Human language, a perfect balance of syntax and semantics is important for better understanding.

These are some of the challenges we might have to face if we try to teach computers how to understand and interact in human language.

5. What are the steps of text Normalization? Explain them in brief.

Text Normalizationin Text Normalization, we undergo several steps to normalize the text to a lower level.

Sentence Segmentation - Under sentence segmentation, the whole corpus is divided into sentences. Each sentence is taken as a different data so now the whole corpus gets reduced to sentences.

Tokenisation- After segmenting the sentences, each sentence is then further divided into tokens. Tokens is a term used for any word or number or special character occurring in a sentence. Under tokenisation, every word, number and special character is considered separately and each of them is now a separate token.

Removing Stop words, Special Characters and Numbers - In this step, the tokens which are not necessary are removed from the token list.

Converting text to a common case -After the stop words removal, we convert the whole text into a similar case, preferably lower case. This ensures that the case-sensitivity of the machine does not consider same words as different just because of different cases.

Stemming In this step, the remaining words are reduced to their root words. In other words, stemming is the process in which the affixes of words are removed and the words are converted to their base form.

Lemmatization -in lemmatization, the word we get after affix removal (also known as lemma) is a meaningful one.

With this we have normalized our text to tokens which are the simplest form of words present in the corpus. Now it is time to convert the tokens into numbers. For this, we would use the Bag of Words algorithm

6. Through a step-by-step process, calculate TFIDF for the given corpus and mention the word(s) having highest value.

Document 1: We are going to Mumbai Document 2: Mumbai is a famous place. Document 3: We are going to a famous place. Document 4: I am famous in Mumbai.

Term Frequency

Term frequency is the frequency of a word in one document. Term frequency can easily be found from the document vector table as in that table we mention the frequency of each word of the vocabulary in each document.

We	Are	Going	to	Mumbai	is	а	famous	Place	Ι	am	in
1	1	1	1	1	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	1	0	0	0
1	1	1	1	0	0	1	1	1	0	0	0
0	0	0	0	1	0	0	1	0	1	1	1

Inverse Document Frequency

The other half of TFIDF which is Inverse Document Frequency. For this, let us first understand what does document frequency mean. Document Frequency is the number of documents in which the word occurs irrespective of how many times it has occurred in those documents. The document frequency for the exemplar vocabulary would be:

We	Are	going	to	Mumbai	is	а	Famous	place	Ι	am	in
2	2	2	2	3	1	2	3	2	1	1	1

Talking about inverse document frequency, we need to put the document frequency in the denominator while the total number of documents is the numerator. Here, the total number of documents are 3, hence inverse document frequency becomes:

We	Are	going	to	Mumbai	is	а	Famous	Place	Ι	am	in
4/2	4/2	4/2	4/2	4/3	4/1	4/2	4/3	4/2	4/1	4/1	4/1

The formula of TFIDF for any word W becomes: TFIDF(W) = TF(W) * log (IDF(W))

The words having highest value are – Mumbai, Famous

7. Normalize the given text and comment on the vocabulary before and after the normalization:

Raj and Vijay are best friends. They play together with other friends. Raj likes to play football but Vijay prefers to play online games. Raj wants to be a footballer. Vijay wants to become an online gamer.

Normalization of the given text:

Sentence Segmentation:

- 1. Raj and Vijay are best friends.
- 2. They play together with other friends.
- *3. Raj likes to play football but Vijay prefers to play online games.*
- 4. Raj wants to be a footballer.
- 5. Vijay wants to become an online gamer.

Tokenization:

Raj and Vijay are best friends.	Raj	and	Vijay	are	best	friends	
They play together with other friends	They	play	Together	with	other	friends	

Same will be done for all sentences.

Removing Stop words, Special Characters and Numbers: In this step, the tokens which are not necessary are removed from the token list. So, the words and, are, to, an, (Punctuation) will be removed.

Converting text to a common case:

After the stop words removal, we convert the whole text into a similar case, preferably lower case.

Here we don't have words in different case so this step is not required for given text. Stemming:

In this step, the remaining words are reduced to their root words. In other words, stemming is the process in which the affixes of words are removed and the words are converted to their base form.

Word	Affixes	Stem
Likes	- <i>S</i>	Like
Prefers	- <i>S</i>	Prefer
Wants	- <i>S</i>	want

In the given text Lemmatization is not required.

Given Text

Raj and Vijay are best friends. They play together with other friends. Raj likes to play football but Vijay prefers to play online games. Raj wants to be a footballer. Vijay wants to become an online gamer.

Normalized Text

Raj and Vijay best friends They play together with other friends Raj likes to play football but Vijay prefers to play online games Raj wants to be a footballer Vijay wants to become an online gamer

CBSE | DEPARTMENT OF SKILL EDUCATION

ARTIFICIAL INTELLIGENCE QUESTION BANK – CLASS 10

CHAPTER 8: EVALUATION

One (01) Mark Questions

1. Define Evaluation.

Moving towards deploying the model in the real world, we test it in as many ways aspossible. The stage of testing the models is known as <u>EVALUATION</u>.

OR

<u>Evaluation</u> is a process of understanding the reliability of any AI model, based on outputs by feeding the test dataset into the model and comparing it with actual answers.

OR

<u>Evaluation</u> is a process that critically examines a program. It involves collecting and analyzing information about a program's activities, characteristics, and outcomes. Its purpose is to make judgments about a program, to improve its effectiveness, and/or to inform programming decisions.

2. Which two parameters are considered for Evaluation of a model?

Prediction and Reality are the two parameters considered for Evaluation of a model. The "Prediction" is the output which is given by the machine and the "Reality" is the real scenario, when the prediction has been made?

3. What is True Positive?

- The predicted value matches the actual value
- The actual value was positive and the model predicted a positive value

4. What is True Negative?

- The predicted value matches the actual value
- The actual value was negative and the model predicted a negative value

5. What is False Positive?

- The predicted value was falsely predicted
- The actual value was negative but the model predicted a positive value
- Also known as the Type 1 error

6. What is False Negative?

- The predicted value was falsely predicted
- The actual value was positive but the model predicted a negative value
- Also known as the Type 2 error

Two (02) Mark Questions

1. What is meant by Overfitting of Data?

Overfitting is "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably".

(OR)

An Overfitted Model is a statistical model that contains more parameters than can be justified by the data. Here, to evaluate the AI model it is not necessary to use the data that is used to build the model. Because AI Model remembers the whole training data set, therefore it always predicts the correct label for any point in the training dataset. This is known as Overfitting

(OR)

Models that use the training dataset during testing, will always results in correct output. This is known as overfitting.

2. What is Accuracy? Mention its formula.

Accuracy is defined as the percentage of correct predictions out of all theobservations. A prediction is said to be correct if it matches reality. Here we have twoconditions in which the Prediction matches with the Reality, i.e., True Positive and True Negative. Therefore,

Formula for Accuracy is

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\%$$

Where *TP* = True Positives, *TN* = True Negatives, *FP* = False Positives, and *FN* = False Negatives.

3. What is Precision? Mention its formula.

Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true.

That is, it takes into account the True Positives and False Positives.

$$Precision = \frac{TP}{TP + FP} * 100\%$$

4. What is Recall? Mention its formula.

Recall is defined as the fraction of positive cases that are correctly Identified.

$$Recall = \frac{TP}{TP + FN}$$

5. Why is evaluation important? Explain.

Importance of Evaluation

Evaluation is a process that critically examines a program. It involves collecting and analyzing information about a program's activities, characteristics, and outcomes. Its purpose is to make judgments about a program, to improve its effectiveness, and/or to inform programming decisions.

- Evaluation is important to ensure that the model is operating correctly and optimally.
- Evaluation is an initiative to understand how well it achieves its goals.
- Evaluations help to determine what works well and what could be improved in a program

6. How do you suggest which evaluation metric is more important for any case?

F 1 Evaluation metric is more important in any case. F1 score sort maintains a balance between the precision and recall for the classifier. If the precision is low, the F1 is low and if the recall is low again F1 score is low.

The F1 score is a number between 0 and 1 and is the harmonic mean of precision and recall

$$F1 Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

When we have a value of 1 (that is 100%) for both Precision and Recall. The F1 score would also be an ideal 1 (100%). It is known as the perfect value for F1 Score. As the values of both Precision and Recall ranges from 0 to 1, the F1 score also ranges from 0 to 1.

7. Which evaluation metric would be crucial in the following cases? Justify your answer.

a. Mail Spamming b. Gold Mining c. Viral Outbreak

Here, Mail Spamming and Gold Mining are related to FALSE POSITIVE cases which are expensive at cost. But Viral Outbreak is a FALSE NEGATIVE case which infects a lot of people on health and leads to expenditure of money too for checkups.

So, False Negative case (VIRAL OUTBREAK) are more crucial and dangerous when compared to FALSE POSITIVE cases.

(OR)

- a. If the model always predicts that the mail is spam, people would not look at it and eventually might lose important information. False Positive condition would have a high cost. (predicting the mail as spam while the mail is not spam)
- b. A model saying that there exists treasure at a point and you keep on digging there but it turns out that it is a false alarm. False Positive case is very costly. (predicting there is a treasure but there is no treasure)
- c. A deadly virus has started spreading and the model which is supposed to predict a viral outbreak does not detect it. The virus might spread widely and infect a lot of people. Hence, False Negative can be dangerous

8. What are the possible reasons for an AI model not being efficient? Explain. Reasons of an AI model not being efficient:

- a. <u>Lack of Training Data</u>: If the data is not sufficient for developing an AI Model, or if the data is missed while training the model, it will not be efficient.
- <u>b.</u> <u>Unauthenticated Data / Wrong Data:</u> If the data is not authenticated and correct, then the model will not give good results.
- <u>c.</u> <u>Inefficient coding / Wrong Algorithms:</u> If the written algorithms are not correct and relevant, Model will not give desired output. <u>Not Tested:</u> If the model is not tested properly, then it will not be efficient.
- d. <u>Not Easy</u>: If it is not easy to be implemented in production or scalable.
- <u>Less Accuracy</u>: A model is not efficient if it gives less accuracy scores in production or test data or if it is not able to generalize well on unseen data.
 (Any three of the above can be selected)

9. Answer the following:

• Give an example where High Accuracy is not usable.

SCENARIO: An expensive robotic chicken crosses a very busy road a thousand times per day. An ML model evaluates traffic patterns and predicts when this chicken can safely cross the street with an accuracy of 99.99%.

Explanation: A 99.99% accuracy value on a very busy road strongly suggests that the ML model is far better than chance. In some settings, however, the cost of making even a small number of mistakes is still too high. 99.99% accuracy means that the expensive chicken will need to be replaced, on average, every 10 days. (The chicken might also cause extensive damage to cars that it hits.)

• Give an example where High Precision is not usable.

Example: "Predicting a mail as Spam or Not Spam"

False Positive: Mail is predicted as "spam" but it is "not spam".

False Negative: Mail is predicted as "not spam" but it is "spam".

Of course, too many False Negatives will make the spam filter ineffective but False Positives may cause important mails to be missed and hence Precision is not usable.

Four (04) Mark Questions

1. Deduce the formula of F1 Score? What is the need of its formulation?

The F1 Score, also called the F score or F measure, is a measure of a test's accuracy.

It is calculated from the precision and recall of the test, where the precision is the number of correctly identified positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of correctly identified positive results divided by the number of all samples that should have been identified as positive.

The F1 score is defined as the weighted harmonic mean of the test's precision and recall. This score is calculated according to the formula.

Formula:

 $F1 Score = 2 * \frac{Precision * Recall}{Precision + Recall}$

Necessary:

F-Measure provides a single score that balances both the concerns of precision and recall in one number.

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's 1, while the model is a total failure when it's 0. F1 Score is a better metric to evaluate our model on real-life classification problems and when imbalanced class distribution exists.

2. What is a confusion matrix? Explain in detail with the help of an example.

Confusion Matrix:

A Confusion Matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

(or)

A 2x2 matrix denoting the right and wrong predictions might help us analyse the rate of success. This matrix is termed the Confusion Matrix.

Evaluation of the performance of a classification model is based on the counts of test records correctly and incorrectly predicted by the model.

Therefore, Confusion Matrix provides a more insightful picture which is not only the performance of a predictive model, but also which classes are being predicted correctly and incorrectly, and what type of errors are being made.

The confusion matrix is useful for measuring Recall (also known as Sensitivity), Precision, Accuracy and F1 Score.

The following confusion matrix table illustrates how the 4-classification metrics are calculated (TP, FP, FN, TN), and how our predicted value compared to the actual value in a confusion matrix

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

Let's decipher the matrix:

The target variable has two values: Positive or Negative

The columns represent the actual values of the target variable The rows represent the predicted values of the target variable True Positive, True Negative, False Positive and False Negative in a Confusion Matrix

True Positive (TP)

The predicted value matches the actual value The actual value was positive and the model predicted a positive value True Negative (TN) The predicted value matches the actual value

The actual value was negative and the model predicted a negative value False Positive (FP) – Type 1 error The predicted value was falsely predicted The actual value was negative but the model predicted a positive value ● Also known as the Type 1 error False Negative (FN) – Type 2 error

The predicted value was falsely predicted

The actual value was positive but the model predicted a negative value also known as the Type 2 error

Example:

Case: Loan (Good loan & Bad loan)

Bad Loan =	1	Cost of FN > Cost of FP	
Good Loan	=0	Actual	Good loan predicted as a bad loan
		Bad Loan (1)	Good Loan (0)
	Bad Loan (1)	✓ []	X
Predict	Good Loan (0)	× 🖏	لي الم

The result of TP will be that bad loans are correctly predicted as bad loans.

The result of TN will be that good loans are correctly predicted as good loans.

The result of FP will be that (actual) good loans are incorrectly predicted as bad loans.

The result of FN will be that (actual) bad loans are incorrectly predicted as good loans. The banks would lose a bunch of money if the actual bad loans are predicted as good loans due to loans not being repaid. On the other hand, banks won't be able to make more revenue if the actual good loans are predicted as bad loans. Therefore, the cost of False Negatives is much higher than the cost of False Positives.

3. Calculate Accuracy, Precision, Recall and F1 Score for the following Confusion Matrix on Heart Attack Risk. Also suggest which metric would not be a good evaluation parameter here and why?

The Confusion	Reality: 1	Reality: 0
Matrix		
Prediction: 1	50	20
Prediction: 0	10	20

The Confusion	Reality: 1	Reality: 0	
Matrix			
Prediction: 1	50	20	70
Prediction: 0	10	20	30
	60	40	100

Calculation:

Accuracy:

Accuracy is defined as the percentage of correct predictions out of all the observations

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\%$$

Where True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Accuracy = (50+20) / (50+20+20+10)

Precision:

Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true.

 $Precision = \frac{True \ Positive}{All \ Predicted \ Positives} * 100\%$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

= (50 / (50 + 20)) = (50/70) = **0.714**

Recall: It is defined as the fraction of positive cases that are correctly identified.

Recall = <u>
True Positive</u> + False Negative

$$Recall = \frac{TP}{TP + FN}$$

= 50 / (50 + 60) = 50 / 110 = **0.5**

F1 Score:

F1 score is defined as the measure of balance between precision and recall.

 $F1 Score = 2 * \frac{Precision * Recall}{Precision + Recall}$ = 2 * (0.714 * 0.5) / (0.714 + 0.5)= 2 * (0.357 / 1.214)

= 2* (0.29406) = 0.58

Therefore, Accuracy= 0.7 Precision=0.714 Recall=0.5 F1 Score=0.588

Here within the test there is a tradeoff. But Recall is not a good Evaluation metric. Recall metric needs to improve more.

Because,

False Positive (impacts Precision): A person is predicted as high risk but does not have heart attack.

False Negative (impacts Recall): A person is predicted as low risk but has heart attack. Therefore, False Negatives miss actual heart patients, hence recall metric need more improvement.

False Negatives are more dangerous than False Positives.

4. Calculate Accuracy, Precision, Recall and F1 Score for the following Confusion Matrix on Water Shortage in Schools: Also suggest which metric would not be a good evaluation parameter here and why?

The Confusion Matrix (Water Shortage in	Reality: 1	Reality: 0
School)		
Prediction: 1	75	5
Prediction: 0	5	15

	Reality: 1	Reality: 0	
Prediction: 1	75	5	80
Prediction: 0	5	15	20
	80	20	100

Calculation:

Accuracy

Accuracy is defined as the percentage of correct predictions out of all the observations

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\%$$

Where True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

= (75+15) / (75+15+5+5) = (90 / 100) =0.9

Precision:

Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true.

$$Precision = \frac{True \ Positive}{All \ Predicted \ Positives} * 100\%$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

= 75 / (75+5) = 75 /80 = 0.9375

Recall:

It is defined as the fraction of positive cases that are correctly identified.

Recall = <u>True Positive</u> + False Negative

$$Recall = \frac{TP}{TP + FN}$$

= 75 / (75+5) = 75 /80 = 0.9375

F1 Score:

F1 score is defined as the measure of balance between precision and recall. = 2 * ((0.9375 * 0.9375) / (0.9375 + 0.9375))

 $F1 \ Score = 2 * rac{Precision * Recall}{Precision + Recall}$

Therefore,

Accuracy= 0.9% Precision=0.9375% Recall=0.9375% F1 Score=0.

Here precision, recall, accuracy, f1 score all are same

5. Calculate Accuracy, Precision, Recall and F1 Score for the following Confusion Matrix on SPAM FILTERING: Also suggest which metric would not be a good evaluation parameter here and why?

Confusion Matrix on SPAM FILTERING:	Reality: 1	Reality: 0
Prediction: 1	10	55
Prediction: 0	10	25

Confusion Matrix on SPAM FILTERING:	Reality: 1	Reality: 0	
Prediction:1	10	55	65
Prediction: 0	10	25	35
	20	80	100

Accuracy is defined as the percentage of correct predictions out of all the observations

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\%$$

Where True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

= (10 + 25) / (10+25+55+10) = 35 / 100 = 0.35

Precision:

Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true.

$$Precision = \frac{True \ Positive}{All \ Predicted \ Positives} * 100\%$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

= 10 / (10 +55) = 10 /65 = 0.15

Recall:

It is defined as the fraction of positive cases that are correctly identified.

 $Recall = \frac{True \ Positive}{True \ Positive + False \ Negative}$

$$Recall = \frac{TP}{TP + FN}$$

F1

Score: = 10 / (10 + 10) = 10 / 20 = 0.5

F1 score is defined as the measure of balance between precision and recall.

$$F1 \ Score = 2 * rac{Precision * Recall}{Precision + Recall}$$

Therefore, = 2 * ((0.15 * 0.5) / (0.15 + 0.5)) = 2 * (0.075 / 0.65) = 2 * 0.115 = 0.23 Accuracy= 0.35 Precision= 0.15 Recall= 0.5 F1 Score= 0.23

Here within the test there is a tradeoff. But Precision is not a good Evaluation metric. Precision metric needs to improve more. Because,

False Positive (impacts Precision): Mail is predicted as "spam" but it is not.

False Negative (impacts Recall): Mail is predicted as "not spam" but spam Of course, too many False Negatives will make the Spam Filter ineffective. But False Positives may cause important mails to be missed. Hence, Precision is more important to improve