General Instructions:

1. Please read the instructions carefully.
2. This Question Paper consists of **21** questions in **two** sections: **Section A & Section B.**
3. **Section A** has Objective type questions whereas **Section B** contains Subjective type questions.
4. Out of the given **(5 + 16 =) 21** questions, a candidate has to answer **(5 + 10 =) 15** questions in the allotted (maximum) time of **2** hours.
5. All questions of a particular section must be attempted in the correct order.
6. **SECTION A - OBJECTIVE TYPE QUESTIONS (24 MARKS):**
   i. This section has **05** questions.
   ii. Marks allotted are mentioned against each question/part.
   iii. There is no negative marking.
   iv. Do as per the instructions given.
7. **SECTION B – SUBJECTIVE TYPE QUESTIONS (26 MARKS):**
   i. This section has **16** questions.
   ii. A candidate has to do **10** questions.
   iii. Do as per the instructions given.
   iv. Marks allotted are mentioned against each question/part.

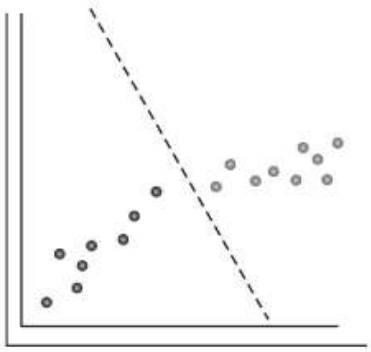
		SECTION A **(Objective Type Questions)**	
1.		**Answer any 4 out of the given 6 questions on employability skills :**	4 x 1 = 4
	(i)	Which of the following does not help in stress management ? (A) Healthy food (B) Sound sleep (C) Yoga asanas (D) Negative thoughts	1
	Ans	(D) Negative thoughts ***(1 Mark for the correct answer)***	
	(ii)	Spam refers to (A) Unnecessary images (B) Temporary files (C) Junk mails (D) Music files	1
	Ans	(C) Junk mails ***(1 Mark for the correct answer)***	

	(iii)	**Assertion (A)** : Sustainable agriculture is environment friendly. **Reason (R)** : It prevents use of chemical fertilizers to protect soil. (A) Both (A) and (R) are correct and (R) is the correct explanation of (A). (B) Both (A) and (R) are correct but (R) is not the correct explanation of (A). (C) (A) is correct but (R) is not correct. (D) (A) is not correct but (R) is correct.	1
	Ans	(A) Both (A) and (R) are correct and (R) is the correct explanation of (A) ***(1 Mark for the correct answer)***	
	(iv)	Which of the following is not a step to build self-motivation ? (A) Focusing on your goal. (B) Planning to achieve your goal. (C) Being indisciplined. (D) Finding out your strength.	1
	Ans	(C) Being indisciplined. ***(1 Mark for the correct answer)***	
	(v)	Which of the following types of communication takes place when the number of people is small enough to communicate with each other effectively ? (A) Interpersonal communication (B) Public communication (C) Intrapersonal communication (D) Small group communication	1
	Ans	(D) Small group communication ***(1 Mark for the correct answer)***	
	(vi)	Reema has started her own restaurant. She keeps on trying new ideas to make different dishes for her customers. As an entrepreneur, Reema is : (A) Impatient (B) Creative (C) Under-confident (D) Lazy	1
	Ans	(B) Creative ***(1 Mark for the correct answer)***	
2.		Answer any **5** out of the given **6** questions :	5 X 1 = 5
	(i)	**Assertion (A) :** When a machine is able to mimic human traits, it is said to be artificially intelligent. **Reason (R) :** A fully automatic washing machine is artificially intelligent. (A) Both (A) and (R) are correct and (R) is the correct explanation of (A) (B) Both (A) and (R) are correct and (R) is not the correct explanation of (A). (C) (A) is correct but (R) is not correct. (D) (A) is not correct but (R) is correct.	1

	Ans	(C) (A) is correct but (R) is not correct. ***(1 Mark for the correct answer)***	
	(ii)	Platforms such as Spotify, Facebook, Instagram, Amazon, Netflix etc. shows recommendation on the basis of what you like. Which is the technology behind this ? (A) Human Intelligence (B) Platform Intelligence (C) Artificial Intelligence (D) Application Intelligence	1
	Ans	(C) Artificial Intelligence ***(1 Mark for the correct answer)***	
	(iii)	**Statement 1** : In "When" block of 4Ws canvas we find the stakeholders. **Statement 2** : Stakeholders are the people who face a particular problem and would be benefited with the solution. (A) Both Statement 1 and Statement 2 are correct. (B) Both Statement 1 and Statement 2 are incorrect. (C) Statement 1 is correct but Statement 2 is incorrect. (D) Statement 2 is correct but Statement 1 is incorrect.	1
	Ans	(D) Statement 2 is correct but Statement 1 is incorrect. ***(1 Mark for the correct answer)***	
	(iv)	Whenever we want an AI project to be able to predict an output, we need to _____. (A) first test it using the data. (B) first train it using the data. (C) Both (A) and (B) (D) Neither (A) nor (B)	1
	Ans	(B) first train it using the data. ***(1 Mark for the correct answer)***	
	(v)	What does the term "image processing" refer to in Computer Vision ? (A) Editing videos (B) Extracting meaningful information from images (C) Playing audio files (D) Compiling codes	1
	Ans	(B) Extracting meaningful information from images ***(1 Mark for the correct answer)***	
	(vi)	A corpus contains 4 documents in which the words such as 'an, is, the' were appearing frequently. Identify the term that is used for such words. (A) Stop word (B) Rare word (C) Missing word (D) Removable word	1
	Ans	(A) Stop word ***(1 Mark for the correct answer)***	

3.		Answer any **5** out of the given **6** questions :	5 X 1 = 5
	(i)	Identify the logo of an application of AI given below. It helps us to navigate to places. 	1
	Ans	Google Maps OR Maps ***(1 Mark for the correct answer)***	
	(ii)	Which of the following data science application is not associated with genetics and genomics ? (A) To understand the impact of DNA on health. (B) To analyze reactions to drugs and disease. (C) To find individual biological connection. (D) To search the house address of a relative on the Internet.	1
	Ans	(D) To search the house address of a relative on the Internet. ***(1 Mark for the correct answer)***	
	(iii)	In Computer Vision, which of the following tasks is used for single object ? (A) Object Detection (B) Classification + Localization (C) Instance Segmentation (D) Non-Localization	1
	Ans	(B) Classification + Localization ***(1 Mark for the correct answer)***	
	(iv)	It is a domain-specific language that is designed for managing data held in different kinds of DBMS (Database Management System). It is particularly useful in handling structured data. Which computer language is this ? (A) SQL (B) CSV (C) Spreadsheet (D) TXT	1
	Ans	(A) SQL ***(1 Mark for the correct answer)***	
	(v)	Which application of NLP helps to provide an overview of a news item or blog post ? It also avoids redundancy from multiple sources and maximises the diversity of content obtained. (A) Virtual Assistants (B) Sentiment Analysis (C) Text Classification (D) Automatic Summarization	1
	Ans	(D) Automatic Summarization ***(1 Mark for the correct answer)***	

	(vi)	Which condition of evaluation does the following diagram indicate ?  (A) False Positive (B) False Negative (C) True Positive (D) True Negative	1
	Ans	(B) False Negative ***(1 Mark for the correct answer)***	
4.		Answer any **5** out of the given **6** questions :	5 x 1 = 5
	(i)	Which of the following is the correct expansion of CSV ? (A) Comma and Space Value (B) Comma Separated Values (C) Control Separated Values (D) Character Separated Values	1
	Ans	(B) Comma Separated Values ***(1 Mark for the correct answer)***	
	(ii)	**Statement 1** : Overfitting is not recommended for evaluation of a model. **Statement 2** : This is because the model will simply remember the whole training set, and will therefore always predict the correct label for any point in the training set. (A) Both Statement 1 and Statement 2 are correct. (B) Both Statement 1 and Statement 2 are incorrect. (C) Statement 1 is correct but Statement 2 is incorrect. (D) Statement 2 is correct but Statement 1 is incorrect.	1
	Ans	(A) Both Statement 1 and Statement 2 are correct. ***(1 Mark for the correct answer)***	
	(iii)	It is one of the parameters for evaluating a model's performance and is defined as the percentage of true positive cases versus all the cases where the prediction is true. Which of the following evaluation parameter is this ? (A) Precision (B) Recall (C) F1 score (D) Accuracy	1
	Ans	(A) Precision ***(1 Mark for the correct answer)***	

	(iv)	Which form of learning-based approach does the following diagram indicate ?  (A) Clustering (B) Classification (C) Regression (D) Dimensionally Reduction	1
	Ans	(B) Classification ***(1 Mark for the correct answer)***	
	(v)	Which application of NLP helps to provide an overview of a news item or blog post ? It also avoids redundancy from multiple sources and maximises the diversity of content obtained. (A) Virtual Assistants (B) Sentiment Analysis (C) Text Classification (D) Automatic Summarization	1
	Ans	(C) Text Classification ***(1 Mark for the correct answer)***	
	(vi)	Raghav can turn on and off any appliance remotely using an internet enabled device. This is an example of _____. (A) Artificial Intelligence (AI) (B) Internet of Things (IoT) (C) Computer Vision (CV) (D) Deep Learning (DL)	1
	Ans	(B) Internet of Things (IoT) ***(1 Mark for the correct answer)***	
5.		Answer any **5** out of the given **6** questions :	5 x 1=5
	(i)	Musical intelligence is a concept that (A) assesses one's ability to regulate, measure and understand numerical symbols, abstraction and logic. (B) measures the language processing skills both in terms of understanding or implementation in writing or verbally. (C) evaluates the ability to process information on the environment around us. (D) describes a person's ability to recognize and create sounds, rhythms and sound patterns.	1

	Ans	(D) describes a person's ability to recognize and create sounds, rhythms and sound patterns. ***(1 Mark for the correct answer)***	
	(ii)	With respect to evaluation, for which of the following does the prediction and reality match ? (A) True positive and False positive (B) True positive and True negative (C) False positive and False negative (D) True positive and False negative	1
	Ans	(B) True positive and True negative ***(1 Mark for the correct answer)***	
	(iii)	**Statement 1** : Various search engines and e-commerce portals now have a new feature called image-based search using computer vision. **Statement 2** : Image-based search helps in finding items, people and places by giving their sounds to the system. (A) Both Statement 1 and Statement 2 are correct. (B) Both Statement 1 and Statement 2 are incorrect. (C) Statement 1 is correct but Statement 2 is incorrect. (D) Statement 2 is correct but Statement 1 is incorrect.	1
	Ans	(C) Statement 1 is correct but Statement 2 is incorrect. ***(1 Mark for the correct answer)***	
	(iv)	In the context of NLP, which of the following words represents a stem resulting from stemming for "Studies" ? (A) Study (B) Stud (C) Studi (D) Studied	1
	Ans	(C) Studi ***(1 Mark for the correct answer)***	
	(v)	Which of the following scenarios might have a high False Negative (FN) cost ? (A) Viral Disease Outbreak (B) Spam (C) Mining (D) Image Search	1
	Ans	(A) Viral Disease Outbreak ***(1 Mark for the correct answer)***	
	(vi)	Which type of chat-bot has a wide functionality, is flexible and powerful, and works on bigger databases directly ?	1
	Ans	Smart Bot **OR** Artificial Intelligent chat-bot **OR** AI-powered chat-bot **OR**	

		Learning based chat-bot ***(1 Mark for the correct answer)*** **OR** ***(1/2 Mark for* any AI based smart bot name example like Gemini, ChatGPT)**	

SECTION - B

(SUBJECTIVE TYPE QUESTIONS)

Answer any **3** out of the given **5** questions on Employability Skills (2 x 3 = 6 marks)
Answer each question in **20 - 30** words.

6.	Explain the importance of following a healthy lifestyle in effectively dealing with stress. Write any one common factor that causes stress among the children nowadays.	2
Ans	Maintaining a healthy lifestyle will help you manage stress. Eat a healthy diet, exercise regularly and get enough sleep. Make a conscious effort to spend less time in front of a screen – television, tablet, computer and phone – and more time relaxing. Avoid using alcohol or drugs to manage stress. Stress won't disappear from your life. And stress management needs to be ongoing. Factors that causes stress among the children are ● Peer pressure ● Lack of knowledge ● Comparison ● Competition ● Academic Pressure ● Social Media and Cyberbullying ● Family Issues ● Peer Pressure ● Lack of Sleep ● Uncertainty and lack of resources/guidance about the Future ● Teenage issues like more independent, looks and hormonal changes ● Lack of healthy diet/nutrition ● Lack of outdoor activities/ physical exercise ● Health Problems ***(1 Mark each for the correct answer or any other relevant answer)***	
7.	If you are a team leader of a team of 20 people in an organization, mention any two methods that you will use for effective communication with your team members.	2

Ans	There are different methods of communication: non-verbal, verbal and visual. However, all these methods can only be effective if we follow the basic principles of professional communication skills. These can be abbreviated as 7 Cs i.e., Clear, Concise, Concrete, Correct, Coherent, Complete and Courteous. The **examples** of various methods of communication are: 1. Face-to-face informal communication 2. Notices/Posters 3. E-mail 4. Business Meetings 5. Social networks 6. Message 7. Phone call for communication 8. Newsletter 9. Blog ***(1 mark each for any correct method)*** **OR** ***(1/2 mark for each correct/relevant example)***	
8.	Write any two tasks that entrepreneurs do when they run their business.	2
Ans	The two tasks that entrepreneurs do when they run their business are : 1. Decision Making 2. Managing the business 3. Divide Income 4. Taking risk 5. Create a new method or idea or product 6. Developing a Business Plan 7. Fund management 8. Product/Service Development 9. Managing Finances 10. Hiring and Managing Employees ***(1 mark for each correct/relevant task)***	
9.	Enlist any two measures that an individual should follow to take care of his/her digital devices.	2

Ans	The measures that an individual should follow to take care of his/her digital devices are : 1. Use of updated Antivirus/Firewall 2. Removal of unnecessary files/Spam 3. Use Defragmenter/Disk cleaner (Utility software) 4. Keep the devices cool 5. Regular software updates and security maintenance 6. Proper cleaning and physical care 7. Avoid overcharging and exposure to extreme temperatures 8. Do not run too many programs at a time 9. Always plug in devices carefully 10. Use strong passwords and keep changing 11. Always logout from your device after finishing your work 12. Shutdown the devices properly 13. Get your devices insured ***(1 mark for each correct/relevant answer)***	
10.	Discuss the following problems related to sustainable development : (i) Water (ii) Fuel	2
	Water - ● Water scarcity due to wastage and overuse of water. ● Water Pollution as people are dumping garbage in the water bodies. ● Inefficient water management, due to lack of awareness regarding water harvesting / saving ● Climate Change as extreme weather such as cold or heat are seen in many places, which affect the people living there. ● Wastage of water like car cleaning with a pipe/ keeping the tap open while brushing ● Unplanned city and sewage system ● Lack of awareness/ resources for water recycling **Fuel:** ● We are using a lot of wood from trees as fuels and for construction of homes and furniture. ● As more and more trees are being cut, it is affecting the climate of the place. ● Extreme weather conditions, such as floods, extreme cold or heat, are seen in many places, which affect the people living there. ● Dependency on Fossil Fuels, Climate Change and Greenhouse Gas Emissions ● Limited Access to Clean Energy ● Over use of fuel resources ● Lack of awareness of other methods like solar energy, etc ***(1 mark for each correct/relevant problem)***	

Answer any **4** out of the given **6** questions in **20-30** words each. (4 x2 = 8 marks)

<table>
<tr><td>11.</td><td>Differentiate between Computer Vision (CV) and Natural Language Processing (NLP).</td><td>2</td></tr>
<tr><td>Ans</td><td>
<table>
<tr><th></th><th>Computer Vision</th><th>Natural Language Processing</th></tr>
<tr><td>1</td><td>Abbreviated as CV, is a domain of AI that depicts the capability of a machine to get and analyse visual information and afterwards predict some decisions about it.</td><td>Abbreviated as NLP, is a branch of artificial intelligence that deals with the spoken and written interaction between computers and humans using the natural language.</td></tr>
<tr><td>2</td><td>The entire process involves images/videos acquiring, screening, analysing, identifying and extracting information using algorithms.</td><td>Natural language refers to language that is spoken and written by people, and it attempts to extract information from the spoken and written word using algorithms.</td></tr>
<tr><td>3</td><td>This extensive processing helps computers to understand any visual content and act on it accordingly.</td><td>The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a mainly that is valuable.</td></tr>
<tr><td>4</td><td>In computer vision, Input to machines can be photographs, videos and pictures from thermal or infrared sensors, indicators and different sources.</td><td>In NLP, Input to machines can be in the form of spoken and written text using chatbot, mics, social media or phone messages etc.</td></tr>
<tr><td>5</td><td>It reads pixels from 0-255 in RGB format.</td><td>It understands the human language, its context and semantics.</td></tr>
<tr><td>6</td><td>Examples – Self driving cars and face recognition.</td><td>Examples – Voice assistants, Chatbot</td></tr>
</table>
(1 mark each for any correct/relevant comparison, maximum up to 2 marks)
</td><td></td></tr>
</table>

12.	Define the following with respect to AI project cycle : (i) Data Exploration (ii) Data Features	2
Ans	(i) Data Exploration: It is the 3rd step of the AI Project Cycle to analyse the data. It is a way to discover hidden patterns, interesting insights and useful information from the collected data. You need to visualise it in some user-friendly format so that you can: • Quickly get a sense of the trends, relationships and patterns contained within the data. • Define strategy for which model to use at a later stage. • Communicate the same to others effectively. • To visualise data, we can use various types of visual representations. • Validating or verification of the collected data to analyze that the data is according to the specifications decided, free from errors and data is meeting our needs. (ii) Data Feature Data features refer to the type of data you want to collect. It should be relevant and correct. Data features are individual measurable properties or characteristics of a dataset that are used to analyze and build models. These features, also called variables or attributes, provide context and information about the data For example, data features would be salary amount, increment percentage, increment period, bonus, etc. ***(1 mark each for any correct/relevant answer)*** ***OR*** ***(1/2 mark each for any correct example, if only example is given)***	
13.	One of the applications of Data Science is Airline Route Planning. List any two tasks that airline companies can do using Data Science.	2
Ans	Using Data Science, the airline companies can: • Predict flight delay • Decide which class of airplanes to buy • Whether to directly land at the destination or take a halt in between (For example, A flight can have a direct route from New Delhi to New York. Alternatively, it can also choose to halt in any country.) • Effectively drive customer loyalty programs ***(1 mark each for any correct/relevant answer)***	

14.	Give any two key impacts of Computer Vision on medical imaging.	2
Ans	Application of computer vision. Computer supported medical imaging applications have been a trustworthy help for physicians. It doesn't only create and analyse images but also becomes an assistant and helps doctors with their interpretation. The application is used to read and convert 2D scan images into interactive 3D models that enable medical professionals to gain a detailed understanding of a patient's health condition. Computer Vision has significantly transformed the field of medical imaging, leading to substantial advancements in diagnosis, treatment, planning and patient care. Here are the key impacts : 1. Improved diagnostic accuracy 2. Enhanced image analysis 3. Workflow efficiency 4. Personalized treatment plans 5. Advanced research and development ***(1 mark each for any correct/relevant answer)***	
15.	What is the primary difference between Human Language and Computer Language ?	2

Ans

	HUMAN LANGUAGE	COMPUTER LANGUAGE
1.	It is communication between humans.	It is instructions to the computer.
2.	Humans understand and communicate in many languages.	The computer understands only Binary Language, that is the language of 1s and 0s. Everything that is sent to the machine has to be converted to numbers.
3.	It is flexible, allows variations and can reflect emotions	It is strict, must be followed exactly and lacks emotional touch.
4.	Communication mode is Speech, writing, gestures, images, videos and using sensory organs.	Communication mode is Text-based (code)
5.	Human language is also cultural specific, with different tone, languages and dialects used in different regions and communities	It is a set of instructions that are used to communicate with computers and other digital devices. Computer language is designed to be clear and precise, allowing computers to understand and carry out specific tasks.
6.	Example English, Hindi, Punjabi, French etc.	Python, C++, Basic etc.

(1 mark each for any correct/relevant comparison, maximum upto 2 marks)
OR
(1/2 mark each if only an example is given.)

16. Suppose you are developing an AI model to detect fraudulent financial transaction risk. Describe False Positives and False Negatives in this Context. — 2

Ans

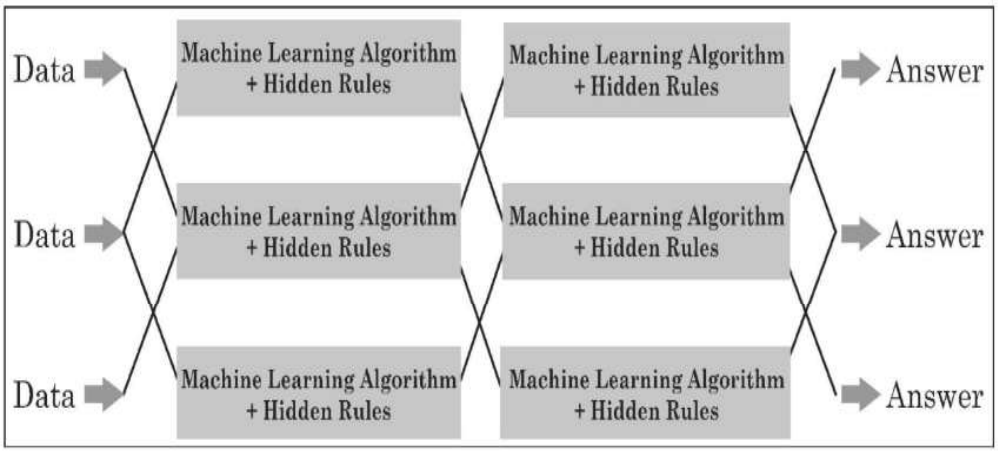
In AI fraud detection model -

1. **False Positives** are legitimate transactions that are incorrectly classified as fraudulent. This can lead to customer dissatisfaction and loss of trust/business/customers/reputation, if legitimate transactions are frequently blocked or flagged for review. A False Positive where Reality is NO but Prediction is YES, like if there is not a fraud done by a customer but the machine has predicted it wrongly as fraud.

	2. **False Negatives** are fraudulent transactions that are not detected by the model and are allowed to proceed. This can result in financial loss and damage to the institution's reputation. A False Negative where Reality is YES but Prediction is NO ***(1 mark for each correct answer)***	

Answer any 3 out of the given 5 questions in 50- 80 words each (4 x 3 = 12 marks)

17.	What do you understand by AI Bias and AI Access ? Give one example of each to support your answer.	4
Ans	**AI Bias** refers to the presence of discrimination in AI systems that leads to favor certain groups over others, due to the data used to train the model. It occurs when an AI system exhibits prejudiced behavior or produces inaccurate predictions due to flawed training data, biased algorithms, or incomplete representations of reality. Sometimes, it is required to have a bias to control a situation and keep things working. We cannot expect a machine to have any biases of its own. Any bias can transfer from the developer to the machine while the algorithm is being developed. Examples: 1. 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. 2. If you search on any search engines 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. **AI Access** refers to the availability to use AI technologies, tools, and resources. It ensures that AI systems are accessible to different communities and organizations, particularly marginalized or underserved groups, to avoid widening the digital divide. AI access can be both **technical access** (access to AI tools, platforms, and infrastructure) and **social access** (ensuring that AI is used equitably and benefits all sectors of society). The people who can afford AI enabled devices make the most of it while others who cannot are left behind. Because of this, a gap has emerged between these two classes of people and it gets widened with the rapid advancement of technology. Examples: **Duolingo**: An AI-driven language-learning app that uses machine learning algorithms to adapt lessons based on the user's skill level. This makes learning a new language more accessible to people worldwide, regardless of location or socioeconomic status. **AI-powered mobile health apps**: apps like **Babylon Health** and **Ada**	

	Health use AI to provide diagnostic advice based on symptoms. These apps allow people in low-resource areas to access healthcare consultations and diagnostics remotely, reducing the need to travel long distances to see a doctor. ***(1 ½ mark for each correct answer and ½ marks for example)*** ***OR*** ***(½ mark each if only example is written)*** ***OR*** ***(2 marks each for correct explanation with an example)***	
18.	What is the use of a problem statement template with respect to 4Ws of problem scoping ? Draw a problem statement template depicting all key elements.	4
Ans	The Problem Statement Template helps us to summarise all the key points with respect to 4Ws (Who, What, Where, Why) into one single Template so that in future, whenever there is need to look back at the basis of the problem, we can take a look at the Problem Statement Template and understand the key elements of it. ***(1 mark for correct use of problem statement template)*** ***(3 marks for correct problem statement template/ 1 mark for each column)***	
19.	Consider the following diagram. It explains how a system of organized machine learning algorithms perform certain tasks. Identify the concept and explain its working.  	4

Our	Stakeholders	Who
Has/have a problem	[issue, problem, need]	What
When/ While	[context, situation]	Where
An ideal solution would	[benefit solution for them]	Why

Ans	Neural networks are modelled after how neurons in the human brain behave. The key advantage of neural networks is that they are able to extract data features automatically without needing the input of the programmer. A neural network is essentially a system of organizing machine learning algorithms to perform certain tasks. This is a representation of the working of neural networks. A Neural Network is divided into 3 layers and each layer is further divided into several blocks called nodes. Each node has its own machine learning algorithm to accomplish, which is then passed to the next layer. ● The first layer of a Neural Network is known as the input layer. The job of an input layer is to acquire data and feed it to the Neural Network. No processing occurs at the input layer. ● Next to it, are the hidden layers. Hidden layers are the layers in which the whole processing occurs. Their name essentially means that these layers are hidden and are not visible to the user. Each node of these hidden layers has its own machine learning algorithm which it executes on the data received from the input layer. The processed output is then fed to the subsequent hidden layer of the network. There can be multiple hidden layers in a neural network system and their number depends upon the complexity of the function for which the network has been configured. Also, the number of nodes in each layer can vary accordingly. The last hidden layer passes the final processed data to the output layer which then gives it to the user as the final output. ● Similar to the input layer, the output layer too does not process the data which it acquires. It is meant for user-interface ***(1/2 mark for the correct identification , 1/2 mark for the definition)*** ***(3 marks for all the layers/ 1 mark for each layer)***	
20.	Consider the following documents : Document 1 : NLP is a domain of AI. Document 2 : NLP stands for Natural Language Processing. Implement all the four steps of Bag of Words (BoW) model to create a document vector table.	4
Ans	**Step 1 – Text Normalisation** Document 1: [nlp is a domain of AI] Document 2: [nlp stands for natural language processing] **Step 2 : Creating Dictionary** nlp, is, a, domain, of, ai, stands, for, natural, language, processing **Step 3 : Create Document vectors** nlp is a domain of ai stands for natural language processing	

	nlp	is	a	domain	of	AI	stands	for	natural	language	processing
Document 1	1	1	1	1	1	1	0	0	0	0	0
Document 2	1	0	0	0	0	0	1	1	1	1	1

Step 4 : Create Document vectors for all the documents/Term Frequency

	nlp	is	a	domain	of	AI	stands	for	natural	language	processing
Document 1	1	1	1	1	1	1	0	0	0	0	0
Document 2	1	0	0	0	0	0	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1

(½ mark each for naming the step and ½ mark each for making correct table)

21. An AI model has been developed to test specimens of blood/urine/cough etc. to diagnose ailments (diabetes/liver infection etc.). The model was tested on a data-set of about 630 tests and the resulting confusion matrix is as follows : — 4

Confusion Matrix		Reality	
		Yes	No
Prediction	Yes	110	60
	No	50	410

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

Ans

CONFUSION MATRIX		REALITY	
		YES	NO
PREDICTION	YES	110 (TP)	60 (FP)
	NO	50 (FN)	410(TN)

True Negative cases - 410

Confusion Matrix:

Precision= $\frac{TP}{TP + FP}$

= $\frac{110}{110+60}$ = $\frac{110}{170}$ = 0.647

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

= $\frac{110}{110+50}$ = $\frac{110}{160}$ = 0.688

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

= $\frac{2 * 0.647 * 0.688}{0.647+.688}$ = $\frac{0.890}{1.335}$ = 0.667

(1 mark for correct total number of true Negative value)

(½ mark for each correct formula)

(½ mark each for value substitution)

NOTE- Mathematical calculations can be ignored