SUPERNOVA-LEARNPYTHON
CLASS X : SAMPLE PAPER - 3
ARTIFICIAL INTELLIGENCE (SUBJECT CODE - 417)
MARKING SCHEME

SECTION A: OBJECTIVE TYPE QUESTIONS

1. Answer any 4 out of the given 6 questions on Employability Skills.

(i) (a) Operating System
(ii) (b) Emotional barrier
(iii) (c) Balancing the well-being of ourselves and future generations
(iv) (d) Time-management
(v) (c) Patience
(vi) Firewall

2. Answer any 5 out of the given 6 questions.

(i) (a) A machine that gets trained first on the training data and then optimizes itself according to its own experiences
(ii) (a) Who, What, Where, Why
(iii) (b) Analyzing and interpreting the collected data
(iv) (b) Recall
(v) Resolution
(vi) (c) Confusion Matrix

3. Answer any 5 out of the given 6 questions.

(i) (d) Copying text
(ii) (b) AI models can be broadly categorized into four domains.
(iii) (a) matplotlib, pandas
(iv) (c) II
(v) True
(vi) (a) Bag of Words

4. Answer any 5 out of the given 6 questions.

(i) (c) Data Acquisition
(ii) (a) Genetics & Genomics
(iii) (b) Airline Route Plan
(iv) (d) F1 Score
(v) corpus
(vi) (a) Overfitting

5. Answer any 5 out of the 6 given questions.

(i) (a) imshow()
(ii) (d) mean(), mode(), median()
(iii) Token
(iv) Named Entity Recognition (NER)
(v) Precision
(vi) (c) Spatial Visual Intelligence

SECTION B: SUBJECTIVE TYPE QUESTIONS

Answer any 3 out of the given 5 questions on Employability Skills. (2 x 3 = 6 marks)

6. (a) Always use strong passwords to login to your computer, which means a combination of uppercase letters, lowercase letters, numbers and special characters. Use passwords that are difficult to guess to prevent unauthorized people from using the computer.

   (b) Always install antivirus software and enable a firewall to prevent unauthorized access and viruses from entering the computer system.

   (c) Always encrypt data to prevent unauthorized usage of the computer system. This practice is usually done by banks and companies where customers' data is stored.

   (d) Always provide confidential details to secure sites by checking the address bar of the browser. If the site address starts with https:// and has a lock symbol, then it is safe to use.

7. Visual communication is a method of conveying information, ideas or messages through visual elements such as images, graphics, charts, diagrams and symbols. It is highly effective because it makes complex information more understandable or engaging.

8. The four positive effects of entrepreneurship on society are:

   (i) Fosters creativity

   (ii) Accentuates economic growth

   (iii) Encourages welfare of the society

   (iv) Creates jobs and employment opportunities

9. (a) **Refuse:** We must refuse to use products that harm the environment so that we can create a sustainable environment.

   (b) **Upcycle:** It is the process which involves creativity and innovation, giving a new look to the old product and making it look desirable.

10.

Eustress		Distress	
(i)	It is a positive stress	(i)	It is a negative stress
(ii)	It gives the feeling of excitement and joy	(ii)	It gives the feelings of anxiety and unease
(iii)	*Examples:* Starting a new job, planning a trip, etc.	(iii)	*Examples:* Chronic illness, failure, etc.

Answer any 4 out of the given 6 questions in 20-30 words each. (2 x 4 = 8 marks)

11. Data visualization is the representation of data with the use of common graphics, such as charts, plots, graphs, diagrams and animations. It is an important part of data science. It helps us to understand and communicate in a visual and easy manner.

    The two Python libraries which are used for visualizing data are matplotlib and seaborn.

12. TF-IDF stands for Term Frequency–Inverse Document Frequency. It is a technique used in Natural Language Processing to measure the importance of words in a collection of documents. It is 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 which is a metric that shows how many documents in the corpus contain a particular term.

    Its formula is:

$$TF = \frac{\text{number of times the term appears in a document}}{\text{Total number of words in the document}}$$

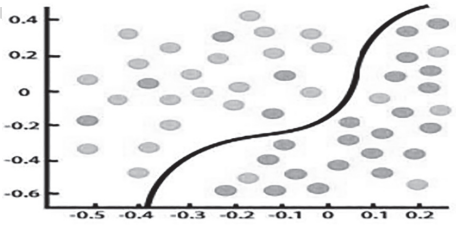
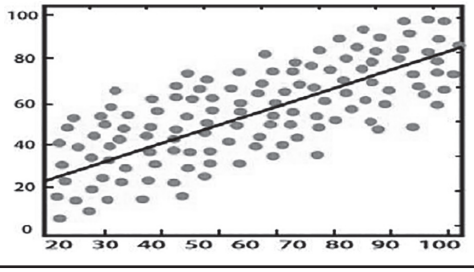
$$IDF = \frac{\log (\text{total number of documents in the corpus}}{(\text{number of documents in the corpus containing the term})}$$

$$TF–IDF = TF * IDF$$

13. Evaluation is important because:
    (i) It ensures that the model is operating correctly and optimally.
    (ii) It is an initiative to understand how well it achieves its goals.
    (iii) It helps to determine what works well and what could be improved in a program.
14. The four applications of AI in computer vision are:
    (i) Self-driving Vehicles
    (ii) Facial Recognition
    (iii) Augmented Reality
    (iv) Medical Imaging and Diagnostics
15. The two different approaches for AI modelling are:
    (i) Rule-based Approach
    (ii) Learning-based Approach
    (i) **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.
    (ii) **Learning-based Approach** is where the relationship or patterns in data are not defined by the developer. In this approach, random data is fed into the machine and it is left to the machine to figure out patterns and trends out of it.
16. (i) **Job Displacement** – AI can lead to job displacement in various industries. It may lead to financial instability and challenges in finding new employment.
    (ii) **Data Privacy** – AI systems need a secure storage and processing of data to prevent unauthorized access. Security breaches can lead to the exposure of sensitive personal information, leading to privacy violations.

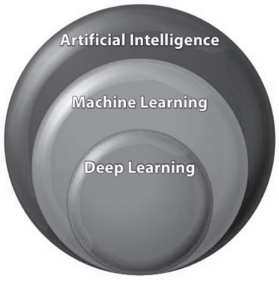
Answer any 3 out of the given 5 questions in 50-80 words each. **(4 x 3 = 12 marks)**

17.

	Classification		**Regression**
(i)	In Classification algorithm, the output variable is always categorical type.	(i)	In Regression algorithm, the output variable is continuous type.
(ii)	Classification finds the decision boundary which divides the datasets into different classes.	(ii)	It finds the best fit line which predicts the output more accurately.
(iii)	Classification algorithm solves classification problems like identifying spam emails.	(iii)	Regression algorithm solves regression problems like predicting house price.
(iv)	Graph: 	(iv)	Graph: 

18. Yes. Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL) and Data Science (DS) are related to each other as they work symbiotically but these fields cannot be synonymous.

    Artificial Intelligence (AI) is a broader field in which Machine Learning (ML) and Deep Learning (DL) are subsets of AI.



    Whereas, Data Science is a field where AI tools, such as Machine Learning and Deep Learning, are used to extract meaningful insights and to analyze large volumes of data. Data Science includes data mining, data cleaning, data exploration and data visualization.

19. NLTK, which stands for Natural Language Toolkit, is a popular Python library for NLP that provides various tools and functionalities to assist in data processing. It simplifies complex tasks.

    Its features are as follows: -

    (a) It provides functions to break down text into individual words or sentences, which is essential for many natural language processing (NLP) tasks.

    (b) It provides Named Entity Recognition (NER) tool for identifying and classifying named entities in text, such as names of people, organizations, locations, etc.

    (c) It includes tools for sentiment analysis, allowing you to determine the sentiment (positive, negative, neutral) expressed in a piece of text.

20. Calculation:

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

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

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

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

    Accuracy = (50+20) / (50+20+20+10)

    = (70/100)

    = 0.7

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

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

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

    = (50 / (50 + 20))

    = (50/70)

    = 0.714

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

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

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

= 50 / (50 + 60)

= 50 / 110

= 0.45

F1 Score: It is defined as the measure of balance between precision and recall.

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

= 2 * (0.714 *0.5) / (0.714 + 0.5)

= 2 * (0.357 / 1.214)

= 2* (0.29406)

= 0.588

Therefore,

Accuracy = 0.7

Precision = 0.714

Recall = 0.45

F1 Score = 0.588

Within the test, there is a trade-off. But Recall is not a good evaluation metric. Recall metric needs to improve more because of the following reasons:

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

False Negative (impacts Recall): A person is predicted as low risk but has a heart attack.

Therefore, since False Negatives miss actual heart patients, recall metric needs more improvement. False Negatives are more dangerous than False Positives.

21. The AI Project Cycle can be described as a systematic and sequential process from the initial planning phase till the completion and review phase.

There are 5 stages of the AI Project Cycle.

(i) **Problem Scoping** – Problem scoping is the process of defining and understanding the specific boundaries and details of a problem before starting an AI project. It helps us clarify what needs to be solved and what is the best approach to solve our problem.

(ii) **Data Acquisition** – This stage involves gathering and collecting the relevant data for the AI project. The sources of data are databases, web scraping, social media, sensors, surveys and APIs.

(iii) **Data Exploration** – Data exploration is a way to discover hidden patterns, interesting insights and useful information from the collected data. Data exploration helps us make important decisions and find insights that can be used to improve things.

(iv) **Modelling** – AI modelling refers to the process of creating a mathematical or statistical representation of a problem. The process of AI modelling has three essential components: Data, algorithms and Training.

(v) **Evaluation** – Evaluation helps us understand how well the model is performing and whether it meets the desired objectives with the help of metrics and evaluation techniques.