A9.5-R5 - ARTIFICIAL INTELLIGENCE CONCEPTS AND R PROGRAMMING

DURATION : 03 Hours   **MAXIMUM MARKS : 100**

OMR Sheet No. : ☐☐☐☐☐

Roll No. : ☐☐☐☐☐☐   **Answer Sheet No. :** ☐☐☐☐☐

Name of Candidate : _____ ; **Signature of Candidate :** _____

INSTRUCTIONS FOR CANDIDATES :

• Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.

• Question Paper is in English language. Candidate has to answer in English language only.

• There are **TWO PARTS** in this Module/Paper. **PART ONE** contains **FOUR** questions and **PART TWO** contains **FIVE** questions.

• **PART ONE** is Objective type and carries **40** Marks. **PART TWO** is Subjective type and carries **60** Marks.

• **PART ONE** is to be answered in the **OMR ANSWER SHEET** only, supplied with the question paper, as per the instructions contained therein. **PART ONE** is **NOT** to be answered in the answer book for **PART TWO**.

• Maximum time allotted for **PART ONE** is **ONE HOUR**. Answer book for **PART TWO** will be supplied at the table when the Answer Sheet for **PART ONE** is returned. However, Candidates who complete **PART ONE** earlier than one hour, can collect the answer book for **PART TWO** immediately after handing over the Answer Sheet for **PART ONE** to the Invigilator.

• **Candidate cannot leave the examination hall/room without signing on the attendance sheet and handing over his/her Answer Sheet to the invigilator. Failing in doing so, will amount to disqualification of Candidate in this Module/Paper.**

• After receiving the instruction to open the booklet and before answering the questions, the candidate should ensure that the Question Booklet is complete in all respects.

DO NOT OPEN THE QUESTION BOOKLET UNTIL YOU ARE TOLD TO DO SO.

PART ONE

(Answer all the questions. Each question carries ONE mark)

1. **Each question below gives a multiple choice of answers. Choose the most appropriate one and enter in the "OMR" answer sheet attached to the question paper, following instructions therein.** **(1x10)**

1.1 Which algorithm is used in the Game tree to make decisions of Win/Lose ?

(A) Heuristic Search Algorithm

(B) DFS/BFS algorithm

(C) Greedy Search Algorithm

(D) Min/Max algorithm

1.2 Which AI technique enables the computers to understand the associations and relationships between objects and events ?

(A) Pattern Matching

(B) Cognitive Science

(C) Relative Symbolism

(D) Heuristic Processing

1.3 Which of the following are informed search methods ?

(A) Best First Search

(B) A* search

(C) Memory bound Heuristic Search

(D) All of the Above

1.4 A feature F1 can take certain value : A, B, C, D, E, & F and represent grade of students from a college. Which feature type is used ?

(A) Nominal

(B) Categorical

(C) Boolean

(D) Ordinal

1.5 Which of the following techniques would perform better for reducing dimensions of a data set ?

(A) removing columns which have high variance in data

(B) removing columns which have too many missing values

(C) removing columns with dissimilar data trends

(D) None of the above

1.6 Database query is used to uncover this type of knowledge :

(A) hidden

(B) shallow

(C) deep

(D) multidimensional

1.7 Number of main statistical methods are used in data analysis :

(A) 3

(B) 2

(C) 4

(D) 5

1.8 Out of the following values, which one is not possible in probability ?

(A) $P(x) = 1$

(B) $\Sigma x \, P(x) = 5$

(C) $P(x) = 0.2$

(D) $P(x) = -0.3$

1.9 Which plot shows to display the relationship between two numerical variables ?

(A) Pie chart

(B) Scatter plot

(C) Bar chart

(D) Area chart

1.10 What will be the output of following code ?

thislist <- list("Mumbai", "Delhi", "Gujarat")

thislist[2]

(A) Mumbai

(B) Delhi

(C) Gujarat

(D) MumbaiDelhi

2. **Each statement below is either TRUE or FALSE. Choose the most appropriate one and ENTER in the "OMR" sheet attached to the question paper, following instructions therein.** **(1x10)**

2.1 Rational agent always does the right things.

2.2 Google ML Kit is an AI tool.

2.3 The choice phase of decision making considers ways to solve problems, fill needs or take advantage of opportunities.

2.4 Artificial Intelligence (AI) is the science of making machines imitate computer thinking and behavior.

2.5 Arithmetic mean of two regression coefficients is greater than the correlation coefficient.

2.6 Relational database is unstructured data.

2.7 K-means is a clustering algorithm that is guaranteed to converge.

2.8 Sampling from a Bayes net using likelihood weighting will systematically overestimate the posterior of a variable conditioned on one of its descendants.

2.9 There are three number types in R.

2.10 In R, maximum() functions can be used to find the highest number in a set.

3. **Match words and phrases in column X with the closest related meaning/word(s)/phrases in column Y. Enter your selection in the "OMR" answer sheet attached to the question paper, following instructions therein.**

(1x10)

	X		Y
	X		**Y**
3.1	Poisson distribution	**A.**	R-Programming
3.2	Clustering	**B.**	Natural Language processing
3.3	rm(<variable name>)	**C.**	Semi-supervised learning
3.4	Chatbot	**D.**	The number of times a coin lands on heads when you toss it five times
3.5	To detect fraudulent usage of credit cards	**E.**	Unsupervised Learning
3.6	Dynamically typed language	**F.**	F1 Score and Accuracy
3.7	To predict numerical values that represent the probability of dogs becoming ill based on their age and body fat percentage	**G.**	Outlier Analysis
3.8	Evaluation metrics	**H.**	The amount of time cars wait at a red light
3.9	Binomial distribution	**I.**	Deep Learning
3.10	K-Nearest Neighbor (KNN) Algorithm	**J.**	Euclidean distance
		K.	Multiple Linear Regression
		L.	To delete the variable in R-Programming
		M.	Supervised Learning

4. **Each statement below has a blank space to fit one of the word(s) of phrases in the list below. Enter your choice in the "OMR" answer sheet attached to the question paper, following instructions therein.**

(1x10)

A.	NLP	**B.**	Reinforcement learning	**C.**	Crossover
D.	Data frame	**E.**	Turing Test	**F.**	[]
G.	Zero	**H.**	()	**I.**	Heuristic
J.	One	**K.**	Independent	**L.**	PCA
M.	Information				

4.1 A technique that was developed to determine whether a machine could or could not demonstrate the artificial intelligence known as the _____.

4.2 The basis of decision making depends upon the availability of _____ and how we experience and understand it.

4.3 Amazon Alexa is a voice-controlled digital or virtual assistant software that takes voice commands to make to-do lists, place online orders, schedule reminders and answer queries through internet searches. Alexa uses _____.

4.4 _____ is an example of a deterministic algorithm.

4.5 As the learning agent interacts with the environment while moving from one state to another, it is rewarded for success but penalized for failure. The Learning model is _____.

4.6 _____ is the most significant phase in genetic algorithm.

4.7 In correlation coefficient formula a result of _____ indicates no relation at all.

4.8 If any regression coefficient's value is zero, the two variables are _____.

4.9 _____ brackets are used for indexing in R-Programming.

4.10 A _____ is a table or a two-dimensional array-like structure in R-Programming.

PART TWO

(Answer any FOUR Questions)

5. (a) Compare goal based agents with utility based agents in Artificial Intelligence.

   (b) Explain the Role of Artificial Intelligence in Healthcare.

   (c) Differentiate Supervised Learning and Unsupervised Learning. Give one example for each of them. **(5+5+5)**

6. (a) Calculate Precision, F1 Score and Accuracy from the following Confusion Matrix.

N=660	True Positive	True Negative
Predicted Positive	500	10
Predicted Negative	50	100

   (b) What is Probability distribution ? Explain discrete probability distribution with example.

   (c) Use the K-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters :

   $A1 = (2, 10)$, $A2 = (2, 5)$, $A3 = (8, 4)$, $A4 = (5, 8)$, $A5 = (7, 5)$, $A6 = (6, 4)$, $A7 = (1, 2)$, $A8 = (4, 9)$.

   (i) The new clusters (i.e. the examples belonging to each cluster).

   (ii) The centroid of the new clusters. **(3+6+6)**

7. (a) Explain association rule mining with example.

   (b) Differentiate between structured and unstructured data type.

   (c) Find the mean, median and mode for following data : 11, 2, 3, 10, 7, 2, 15, 23, 27 **(6+6+3)**

8. (a) How to plot line graph in R ? What is the syntax ? Explain with parameters and example.

   (b) Suppose that the Attendance data warehouse consists of the following dimensions : Sales Fact, Date Dimension, Customer Dimension, Product Dimension and Store Dimension are its dimension tables; Quarter Dimension, Month Dimension, Week Dimension, State Dimension, Brand Dimension and City Dimension are the sub-dimensions of those five Dimensions. Draw the snowflake schema.

   (c) Which are the types of operators in R programming ? Explain in detail. **(5+5+5)**

9. (a) List the extraction techniques used for dimensionality reduction and explain any one with example.

   (b) How Data and AI is transforming the Media and Entertainment Industry ?

   (c) Write a R script to draw a bar chart for the following data. Mention the name of x-axis, y-axis and chart as Month, Revenue and Revenue Chart.

   $X = \{8, 15, 24, 6, 18\}$

   $Y = \{"Mar","Apr","May","Jun","Jul"\}$ **(5+5+5)**

- o O o -

SPACE FOR ROUGH WORK