No. of Printed Pages : **8**

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 supplied with the question paper, following the instructions therein. (1x10)

1.1 Which of the following is an application of Artificial Intelligence ?

(A) It helps to exploit vulnerabilities to secure the firm

(B) Text analytics and NLP

(C) Easy to create a website

(D) It helps to deploy applications on the cloud

1.2 OLAP stands for :

(A) Online analytical processing

(B) Online analysis processing

(C) Online application processing

(D) Online aggregate processing

1.3 Which of the following refers to the problem of finding abstracted patterns (or structures) in the unlabelled data ?

(A) Supervised learning

(B) Unsupervised learning

(C) Hybrid learning

(D) Reinforcement learning

1.4 Where is data warehousing used ?

(A) Transaction system

(B) Logical system

(C) Decision support system

(D) None

1.5 Normal Distribution is symmetric about _____

(A) Variance

(B) Mean

(C) Standard deviation

(D) Covariance

1.6 In a Poisson Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean value is given by :

(A) $m = np$

(B) $m = (np)^2$

(C) $m = np(1-p)$

(D) $m = p$

1.7 If you explicitly want an integer, you need to specify the _____ suffix.

(A) D

(B) R

(C) L

(D) K

SPACE FOR ROUGH WORK

1.8 What will be the output of the following R code ?

```
> x <- vector("numeric", length = 10)

> x
```

(A) 10

(B) 0 0 0 0 0 0 0 0 0 0

(C) 01

(D) 00120

1.9 Which of the following is **not** a supervised machine learning algorithm ?

(A) K-means

(B) Naïve Bayes

(C) SVM for classification problems

(D) Decision tree

1.10 Linear regression belongs to which category ?

(A) Neither supervised nor unsupervised learning

(B) Both supervised and unsupervised learning

(C) Unsupervised learning

(D) Supervised learning

2. Each statement below is either TRUE or FALSE. Choose the most appropriate one and enter your choice in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

2.1 In Linear Regression model, Error is the difference between the actual value and predicted value and the goal is to reduce this difference.

2.2 Data warehouse is generally updated in real-time.

2.3 Linear regression analysis is used to predict the value of a variable based on the value of another variable.

2.4 The correlation coefficient remains unaffected by scale changes.

2.5 In supervised machine learning, a model makes predictions or decisions based on past or labelled data.

2.6 Naïve Bayes works best when the training set is very small.

2.7 False positives are those cases that wrongly get classified as False but are True.

2.8 In an association problem, we identify patterns of associations between different variables or items.

2.9 The slope of the regression line of Y on X is also referred to as the Regression coefficient of Y on X.

2.10 Number Inf represents infinity in R .

3. Match words and phrases in column X with the closest related meaning/word(s)/ phrase (s) in column Y. Enter your selection in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

	X		Y
3.1	Artificial Intelligence	**A.**	Continuous Random Distribution
3.2	SVM (Support Vector Machine)	**B.**	Sort()
3.3	Normal Distribution	**C.**	OLAP
3.4	arrange in increasing order.	**D.**	Hyperplane
3.5	supervised machine learning	**E.**	Array
3.6	Decision Tree	**F.**	It is based on XML/RDF
3.7	root mean square error (RMSE)	**G.**	John McCarthy
3.8	Software Tool that analyse data	**H.**	standard deviation of the residuals
3.9	Semi-structured data	**I.**	Binary classification
3.10	A basic data structure of R	**J.**	Naïve Bayes
		K.	Random Forest
		L.	Guido van Rossum
		M.	Vector

4. Each statement below has a blank space to fit one of the word(s) or phrase(s) in the list below. Enter your choice in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

A.	Independent variable	**B.**	Modeling relationships	**C.**	OLAP
D.	Linear line	**E.**	Standard Deviation	**F.**	Training dataset
G.	Mean	**H.**	Variance	**I.**	Correlation
J.	List	**K.**	OLTP	**L.**	square root
M.	Data Frame				

4.1. The residuals or errors are normally distributed with a _____ of zero and a constant variance.

4.2 A table with all possible values of a random variable and its corresponding probabilities is called _____.

4.3 The learner is trying to predict housing prices based on the size of each house. The variable "size" is _____.

4.4 Regression analysis is _____ within the data.

4.5 The correlation coefficient is the _____ of the coefficient of determination.

4.6 In a Poisson Distribution, the mean and _____ are equal.

4.7 In classification, a computer is trained against a _____.

4.8 _____ is useful for administering day-to-day transactions of an organization.

4.9 A scatter plot is a chart used to plot a _____ between two or more variables at the same time.

4.10 _____ is a two-dimensional tabular structure with rows and columns as dimensions.

PART - TWO

(Answer any FOUR Questions)

5. (a) List down 5 libraries used in R Programming Language for data visualization. Explain three of them briefly.

(b) What is scatter plot ? Explain with an example of how to create one scatter plot using R- Libraries. **(8+7)**

6. (a) Discuss real-life applications of Artificial Intelligence.

(b) Differentiate between classification and regression with the help of suitable example. **(8+7)**

7. (a) What do you mean by Root Mean Square Error ? Explain with the help of proper example.

(b) How OLTP is different from OLAP ? Give one application of each. **(8+7)**

8. (a) How Data Frames in R Programming are used ? Explain with example.

(b) How can you conclude about the model's performance using the confusion matrix ? **(10+5)**

9. (a) Differentiate between Structured, Semi-Structured and Unstructured Data.

(b) Explain R Data Types - Vectors, Matrices. List in brief. **(10+5)**

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK