Sl. No.

A10.5-R5 : MACHINE LEARNING USING PYTHON

अवधि : 03 घंटे **अधिकतम अंक : 100**
DURATION : 03 Hours **MAXIMUM MARKS : 100**

ओएमआर शीट सं. : OMR Sheet No. :					

रोल नं. : **Roll No. :**

उत्तर-पुस्तिका सं. : **Answer Sheet No. :**

परीक्षार्थी का नाम : **परीक्षार्थी के हस्ताक्षर :**
Name of Candidate : _____ **Signature of Candidate :** _____

परीक्षार्थियों के लिए निर्देश :	**Instructions for Candidate :**
कृपया प्रश्न-पुस्तिका, ओएमआर शीट एवं उत्तर-पुस्तिका में दिये गए निर्देशों को ध्यानपूर्वक पढ़ें।	Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.
प्रश्न-पुस्तिका की भाषा अंग्रेजी है। परीक्षार्थी केवल अंग्रेजी भाषा में ही उत्तर दे सकता है।	Question Paper is in English language. Candidate can answer in English language only.
इस मॉड्यूल/पेपर के **दो भाग** हैं। **भाग एक** में **चार** प्रश्न और **भाग दो** में **पाँच** प्रश्न हैं।	There are **TWO PARTS** in this Module/Paper. **PART ONE** contains **FOUR** questions and **PART TWO** contains **FIVE** questions.
भाग एक ''वैकल्पिक'' प्रकार का है जिसके कुल अंक 40 है तथा **भाग दो** ''व्यक्तिपरक'' प्रकार का है और इसके कुल अंक 60 है।	**PART ONE** is Objective type and carries 40 Marks. **PART TWO** is Subjective type and carries 60 Marks.
भाग एक के उत्तर, **ओएमआर उत्तर-पुस्तिका** पर ही दिये जाने हैं। **भाग दो** की उत्तर-पुस्तिका में **भाग एक** के उत्तर **नहीं** दिये जाने चाहिए।	**PART ONE** is to be answered in the **OMR ANSWER SHEET** only. **PART ONE** is **NOT** to be answered in the answer book for **PART TWO**.
भाग एक के लिए अधिकतम समय सीमा **एक घण्टा** निर्धारित की गई है। **भाग दो** की उत्तर-पुस्तिका, **भाग एक** की उत्तर-पुस्तिका जमा कराने के पश्चात् दी जाएगी। तथापि, निर्धारित एक घंटे से पहले **भाग एक** पूरा करने वाले परीक्षार्थी **भाग एक** की उत्तर-पुस्तिका निरीक्षक को सौंपने के तुरंत बाद, **भाग दो** की उत्तर-पुस्तिका ले सकते हैं।	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.
परीक्षार्थी, उपस्थिति-पत्रिका पर हस्ताक्षर किए बिना और अपनी उत्तर-पुस्तिका, निरीक्षक को सौंपे बिना, परीक्षा हॉल/कमरा नहीं छोड़ सकते हैं। ऐसा नहीं करने पर, परीक्षार्थी को इस मॉड्यूल/पेपर में अयोग्य घोषित कर दिया जाएगा।	**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.**
प्रश्न-पुस्तिका को खोलने के निर्देश मिलने के पश्चात् एवं उत्तर लिखना आरम्भ करने से पहले उम्मीदवार जाँच कर यह सुनिश्चित कर लें कि प्रश्न-पुस्तिका प्रत्येक दृष्टि से संपूर्ण है।	After receiving the instruction to open the booklet and before starting to answer the questions, the candidate should ensure that the Question Booklet is complete in all respect.

जब तक आपसे कहा न जाए, तब तक प्रश्न-पुस्तिका न खोलें।
DO NOT OPEN THE QUESTION BOOKLET UNTIL YOU ARE TOLD TO DO SO.

PART ONE

(Answer all the questions)

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 instructions therein. **(1x10)**

1.1 Consider a List l1 is [1, 2, 3, 4, 5, 6, 7, 8, 9]. Which of the following is correct syntax for slicing operation ?

(A) print(l1[0])

(B) print(l1[:2])

(C) print(l1[:-2])

(D) all of the above

1.2 Which of the following will be true about k in k-NN in terms of variance ?

(A) When you increase the k the variance will increases

(B) When you decrease the k the variance will increases

(C) Can't say

(D) None of the above

1.3 One of the main challenge of NLP is :

(A) Handling Ambiguity of Sentences

(B) Handling Tokenization

(C) Handling POS-Tagging

(D) All of the above

1.4 Higher value of which of the hyper parameters are better for decision tree algorithm ?

1. Number of samples used for split
2. Depth of tree
3. Samples for leaf

(A) 1 and 2

(B) 2 and 3

(C) 1 and 3

(D) Can't say

1.5 The Bayes rule can be used for :

(A) Solving queries

(B) Increasing complexity

(C) Decreasing complexity

(D) Answering probabilistic query

1.6 A perceptron is :

(A) a single layer feed-forward neural network with pre-processing

(B) an auto-associative neural network

(C) a double layer auto-associative neural network

(D) a neural network that contains feedback

1.7 OpenCV stores RGB pixels in what order ?

(A) GBR

(B) RGB

(C) BRG

(D) BGR

1.8 Which of the following is **not** the promise of artificial neural network ?

(A) It can survive the failure of some nodes

(B) It can explain result

(C) It has inherent parallelism

(D) It can handle noise

1.9 Which of the following will be Euclidean Distance between the two data point A(1, 3) and B(2, 3) ?

(A) 1

(B) 2

(C) 4

(D) 8

1.10 N-grams are defined as the combination of N keywords together. How many bi-grams can be generated from given sentence :

"Machine Learning Using Python is a good course "

(A) 7

(B) 8

(C) 9

(D) 10

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 instructions therein. **(1x10)**

2.1 A perceptron adds up all the weighted inputs it receives, and if it exceeds a certain value, it outputs a 1, otherwise it just outputs a 0.

2.2 Decision Tree is a structure in which internal node represents test on an attribute, each branch represents outcome of test and each leaf node represents class label.

2.3 The output of a neural network model is a categorical attribute value.

2.4 The Python dictionary has all the keys of same type.

2.5 PoS tagging refers to the process of tagging words within sentences into their respective parts of speech and then finally labelling them.

2.6 The statement "a[2:3] = []" removes the middle element 3 from list a = [1, 2, 3, 4, 5] so that it becomes [1, 2, 4, 5].

2.7 Photometric factor variation is one of the main challenge in computer vision problems.

2.8 K-Means technique performs classification by finding the hyperplane that maximizes the margin between the two classes.

2.9 SciPy library is one of widely used library to solve computer vision problems.

2.10 K-Means technique is used in Deep Learning Networks.

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 instructions therein.** **(1x10)**

X		Y	
3.1	It is the one of classification accuracy metric when the true class label is 'No' and Predicted class label is 'Yes'	**A**	Reinforcement learning
3.2	It is the problem of getting an agent to act in the world so as to maximize its rewards	**B**	False Positive
3.3	It is a general-purpose array-processing package.	**C**	PoS tagging
3.4	It is process of removing the suffix from a word and reduce it to its root word	**D**	Manhattan distance
3.5	It is a library designed to solve computer vision problems	**E**	Tuple
3.6	It is one of the widely used activation function in deep learning	**F**	K-Means
3.7	It is the process of converting a sequence of characters into a sequence of tokens	**G**	True Positive
3.8	It refers to the process of tagging words within sentences into their respective parts of speech and then finally labelling them.	**H**	Numpy
3.9	It is one of the metric for distance calculation	**I**	False Negative
3.10	It is the one of classification accuracy metric when the true class label is 'Yes' and Predicted class label is 'Yes'	**J**	Lexical Analysis
		K	ReLU
		L	OpenCV
		M	Stemming

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 attached to the question paper, following instructions therein. **(1x10)**

A	Support Vector Machine	**B**	Confusion Matrix	**C**	OpenCV	**D**	Seaborn
E	Tensorflow	**F**	Ensemble	**G**	Dictionary	**H**	K-Means
I	Gradient Descent	**J**	Pandas Library	**K**	Reinforcement learning	**L**	ReLU
M	Tuple						

4.1 _____ is an example of clustering technique.

4.2 _____ is iterative optimization algorithm to find the minimum value (local optima) of a function.

4.3 _____ method uses combination of models to increase accuracy.

4.4 _____ are accessed via keys and not via their position in Python.

4.5 _____ is a Python data visualization library which provides a high-level interface for drawing attractive and informative statistical graphics.

4.6 _____ an open source library which provides high-performance, easy-to-use data structures and data analysis tools.

4.7 _____ library is used to solve computer vision problems.

4.8 _____ is an open source software library used to design, build, and train deep learning models.

4.9 _____ technique performs classification by finding the hyperplane that maximizes the margin between the two classes.

4.10 _____ contains information about actual and predicted classifications done by a classification system.

PART TWO

(Answer any FOUR questions)

5. (a) Write a NumPy program to convert a list consisting of 10 elements into array. Then reverse the elements of the array.

(b) Write a NumPy program to sum and compute the product of a NumPy array elements. **(8+7)**

6. (a) Differentiate between Supervised Learning and Unsupervised learning.

(b) Write a python program for reading the weather data from a file a named "weather.csv" for 50 cities of a state in India. The file has following fields :

City, Wind, Precipitation, Highest_ Temperature, Lowest_Temperature, Humidity and Atmospheric_ Pressure

Your program should do the following :

(i) Displays the 5 entries of "City"

(ii) Displays the total null entries in "Precipitation" and "Humidity"

(iii) Displays all the rows by removing an instance of missing values

(iv) Displays all the rows by removing all the rows with an instance of missing values by using pad method **(5+10)**

7. (a) Describe the following evaluation metrics of classification based on True Positive (TP),True Negative (TN),False Positive (FP),False Negative (FN) in machine learning :
(i) Accuracy
(ii) Error rate
(iii) Precision
(iv) Recall
(v) F-score
(vi) F1-measure

(b) Explain the working of k-NN algorithm.

(c) Consider the following one-dimensional data set :

x	0.5	3.0	4.5	4.6	4.9	5.2	5.3	5.5	7.0	9.5
y	C1	C1	C2	C2	C2	C1	C1	C2	C1	C1

where, y is the class label whose value can be either C1 or C2.
Classify the data point x=5.0 according to different k values using K-nearest neighbors (KNN)
(i) k=1 (ii) k=3
(iii) k=5 (iv) k=9
Show all steps. **(6+5+4)**

8. (a) Write the following NLP processing steps in Python for a input sentence "I am learning machine learning using Python"
(i) Import the required libraries
(ii) Remove the stops words
(iii) Apply the tokenization process on the input sentence
(iv) Display the different tokens
(v) Apply the PoS tagging

(b) Explain the working of Ensemble method for classification. **(7+8)**

9. Briefly explain the following (**any three**) :
(a) Perceptron Learning Algorithm
(b) Face Recognition and Detection with OpenCV python library
(c) Decision Trees
(d) Missing data handling using panda library
(e) Clustering **(5+5+5)**

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK