No. of Printed Pages : **8**

A10.5-R5 : Machine Learning Using Python

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 instructions therein. **(1x10)**

1.1. Consider a dictionary d = {'first': 1, 'second': 2, 'third': 3}. What is the result of the statement d[' second ':' third ']

(A) [2,3]

(B) (2,3)

(C) 2 3

(D) Exception is raised

1.2. Consider a List a = [1, 2, 3, 4, 5]. Select the statements that remove the middle element 3 from list a so that it equals [1, 2, 4, 5]

(A) a[2:3] = []

(B) a[2] = []

(C) a[2:2] = []

(D) a[3:4]=[]

1.3. Machine learning is :

(A) The autonomous acquisition of knowledge through the use of computer programs

(B) The autonomous acquisition of knowledge through the use of manual programs

(C) The selective acquisition of knowledge through the use of computer programs

(D) The selective acquisition of knowledge through the use of manual programs

1.4. Which of the following statements is true for k-NN classifiers ?

(A) The classification accuracy is better with larger values of k

(B) The decision boundary is smoother with smaller values of k

(C) k-NN does not require an explicit training step

(D) The decision boundary is linear

1.5. Which of the following options is/are true for K-fold cross-validation ?

1. Increase in K will result in higher time required to cross validate the result.
2. Higher values of K will result in higher confidence on the cross-validation result as compared to lower value of K.
3. If K=N, then it is called Leave one out cross validation, where N is the number of observations.

(A) 1 and 2

(B) 2 and 3

(C) 1 and 3

(D) 1, 2 and 3

1.6. How the Bayesian network can be used to answer any query ?

(A) Full distribution

(B) Joint distribution

(C) Partial distribution

(D) All of the mentioned

1.7. One of the main challenges of NLP is :

(A) Handling Ambiguity of Sentences

(B) Handling Tokenization

(C) Handling POS-Tagging

(D) All of the mentioned

1.8. Which of the following is a challenge when dealing with computer vision problems ?

(A) Variations due to geometric changes

(B) Variations due to photometric factors

(C) Image occlusion

(D) All of the above

1.9. What are the advantages of neural networks over conventional computers ?

(i) They have the ability to learn by example

(ii) They are more fault tolerant

(iii) They are more suited for real time operation due to their high computational rates

(A) (i) and (ii) are true

(B) (i) and (iii) are true

(C) Only (i)

(D) All of the mentioned

1.10. Which of the following is/are one of the important step(s) to pre-process the text in NLP based projects ?

1. Stemming

2. Stop word removal

3. Object Standardization

(A) 1 and 2

(B) 1 and 3

(C) 2 and 3

(D) 1, 2 and 3

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 All the keys in Python dictionary must be of the same type.

2.2 Python is case sensitive when dealing with identifiers.

2.3 On average, Neural networks have higher computational rates than conventional computers.

2.4. K-Means clustering algorithm terminates when mean values computed for the current iteration of the algorithm are identical to the computed mean values for the previous iteration.

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

2.6. maxsplit specifies the maximum number of splits to make on the input string.

2.7. Lower values of K in K-fold cross-validation will result in higher confidence on the cross-validation result as compared to higher value of K.

2.8. Bayes classifier technique can process both numeric and categorical input attributes.

2.9. Convolutional Neural Networks can perform various types of transformation (rotations or scaling) in an input.

2.10. "import cv2" statement Imports our OpenCV Python bindings.

3. Match words and phrases in column X with the closest related meaning / words(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 contains information about actual and predicted classifications done by a classification system	**A**	K-Means
3.2	This method uses combination of models to increase accuracy	**B**	OpenCV
3.3	This technique performs classification by finding the hyperplane that maximizes the margin between the two classes.	**C**	Gradient Descent
3.4	It is an open source software library used to design, build, and train deep learning models	**D**	Dictionary
3.5	It is a Python data visualization library which provides a high-level interface for drawing attractive and informative statistical graphics.	**E**	Ensemble
3.6	It is an open source library which provides high-performance, easy-to-use data structures and data analysis tools	**F**	Tuple
3.7	This library is used to solve computer vision problems	**G**	Computer Vision
3.8	These are accessed via keys and not via their position in Python	**H**	Tensorflow
3.9	It is an example of clustering technique	**I**	Confusion Matrix
3.10	It is iterative optimization algorithm to find the minimum value (local optima) of a function	**J**	Seaborn
		K	Support Vector Machine
		L	Pandas Library
		M	NumPy Library

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

A	False Positive	**B**	True Positive	**C**	Stemming	**D**	Reinforcement learning
E	Decision Tree	**F**	Manhattan distance	**G**	ReLU	**H**	Lexical analysis
I	PoS tagging	**J**	Sigmoid	**K**	True Negative	**L**	NumPy
M	OpenCV						

4.1 _____ is one of the metric for distance calculation.

4.2 _____ is one of the widely used activation function in deep learning.

4.3 _____ is the process of converting a sequence of characters into a sequence of tokens.

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

4.5 When the true class label is 'No' and Predicted class label is 'Yes', then the accuracy metric is _____.

4.6 When the true class label is 'Yes' and Predicted class label is 'Yes', then the accuracy metric is _____.

4.7 _____ is a general-purpose array-processing package.

4.8 _____ is a process of removing the suffix from a word and reduce it to its root word

4.9 _____ is a library designed to solve computer vision problems.

4.10 _____ is the problem of getting an agent to act in the world so as to maximize its rewards.

PART TWO
(Answer any FOUR questions)

5. (a) Write a python program which takes input as a two dimensional array and do the following operations :
- (i) Displays the transpose of the array
- (ii) Displays the inverse of the array
- (iii) Displays the identity matrix of the array
- (iv) Displays the dot product of the array

(b) What are three data structures of the pandas library ? Write an example of each displaying data

(6+9)

6. (a) Write the 5 different functions for handling missing data in a csv file using panda library. Explain each function by taking an example.

(b) Write a python program for reading the obesity data from a file named "students.csv"for class 10 students in a state of India. The file has following fields :
CITY, SCHOOL_NAME, COUNT_OVERWEIGHT, PCT_OVERWEIGHT, COUNT_OBESE, PCT_OBESE
Your program should do the following :
- (i) Displays the 5 entries of "SCHOOL_NAME"
- (ii) Displays the total null entries in "COUNT_OBESE" and "COUNT_OVERWEIGHT"
- (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
- (v) Displays all the rows by removing with an instance of missing values by using mean method

(5+10)

7. (a) What is activation function in ANN? Write 6 different activation functions used in ANN. Mention the output of each activation function.

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

(c) Apply the k-means algorithm on the following one dimensional data :
Dataset : {2, 4, 10, 12, 3, 20, 30, 11, 25}.
Consider that initial means are m_1=3 and m_2=4. The value of k=2.

(6+4+5)

8. (a) Consider the following confusion matrix data :

True Class → Predicted class ↓	Positive	Negative
Positive	2	1
Negative	1	2

Calculate the values of the following :
- (i) True Positive
- (ii) False Positive
- (iii) Accuracy
- (iv) Recall

(b) Explain the working of Naive Bayes approach using a suitable example.

(c) What are "Support Vectors" in Support Vector Machine? Explain it using an example.

(4+5+6)

9. Briefly explain the following (any three) :
- (i) The architecture and working of Convolution Neural Network
- (ii) Ensemble method of classification
- (iii) K-fold cross validation
- (iv) Face detection with a python library

(5+5+5)

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK