A10.1-R5 : DATA SCIENCE USING PYTHON

अवधि : 03 घंटे
DURATION : 03 Hours

अधिकतम अंक : 100
MAXIMUM MARKS : 100

ओएमआर शीट सं. :
OMR Sheet No. :

रोल नं. :
Roll No. :

उत्तर-पुस्तिका सं. :
Answer Sheet No. :

परीक्षार्थी का नाम :
Name of Candidate : _____

परीक्षार्थी के हस्ताक्षर :
;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. 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 Which of these measures are used to analyze the central tendency of data ?

(A) Mean and Normal Distribution

(B) Mean, Median and Mode

(C) Mode, Alpha & Range

(D) Standard Deviation, Range and Mean

1.2 Which keyword is used to define methods in Python ?

(A) Function

(B) Def

(C) Method

(D) All of these

1.3 How would you import a decision tree classifier in sklearn ?

(A) from sklearn.decision_tree import DecisionTreeClassifier

(B) from sklearn.ensemble import DecisionTreeClassifier

(C) from sklearn.tree import Decision TreeClassifier

(D) None of these

1.4 Suppose you are defining a tuple given below :

tup = (1, 2, 3, 4, 5)

to update the value of tuple at 2nd index to 10, Which of the following option will you choose ?

(A) tup(2) = 10

(B) tup[2] = 10

(C) tup{2} = 10

(D) None of the above

1.5 Suppose list1 is [2, 33, 222, 14, 25], What is list1[−1] ?

(A) Error

(B) None

(C) 25

(D) 2

1.6 What is the maximum possible length of an identifier ?

(A) 31 characters

(B) 63 characters

(C) 69 characters

(D) None of the above

1.7 Which of the following is true about Residuals ?

(A) Lower is better

(B) Higher is better

(C) (A) or (B) depend on the situation

(D) None of the above

SPACE FOR ROUGH WORK

1.8 Which of the following lines of code will result in an error ?

(A) s = {abs}

(B) s = {4, 'abc', (1,2)}

(C) s = {2, 2.2, 3, 'xyz'}

(D) s = {san}

1.9 Which of the following methods do we use to find the best fit line for data in Linear Regression ?

(A) Least Square Error

(B) Maximum Likelihood

(C) Logarithmic Loss

(D) Both (A) and (B)

1.10 What is correct syntax to copy one list into another ?

(A) listA = listB[]

(B) listA = listB[:]

(C) listA = listB[]()

(D) listA = listB

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 *NumPy, SciPy, Pandas, SciKit* are few libraries in Python used for Data Analysis and Scientific computations.

2.2 Computation of NumPy is faster than list or loops.

2.3 *Matplotlib* and *Seaborn* are the library one prefer for plotting in Python Language.

2.4 It is not a good practice to replace the NULL values present in the data frame.

2.5 Covariance measures both the strength and direction of the relationship between two variables whereas correlation indicates only the direction of the relationship between two variables.

2.6 For a given dataset, if we already know how the correct output will look then it will be a supervised machine learning.

2.7 Mean is preferred over Median to describe the characteristics of a population.

2.8 R2 score mainly gives us or tells us how perfectly the line fits for the given set of data.

2.9 Outliers are the data points which are either too small or too big from the normal data points available in the dataset.

2.10 Overfitting is when model cannot cover enough data points and is unable to predict on most data points whereas under fitting is where model tries to cover each and every data point and cannot be generalized as would fail on some random dataset.

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)

	Column X		Column Y
3.1	Tkinter	**A.**	Seaborn
3.2	aleatory	**B.**	NumPy, SciPy
3.3	Pylab	**C.**	Interpreting variances
3.4	Chebyshev	**D.**	Statistical programming method
3.5	Hat values	**E.**	Include a backwards elimination feature
3.6	NumPy similar	**F.**	Machine Learning Library
3.7	Linear regression	**G.**	Random variable
3.8	Multivariate Adaptive Regression Splines	**H.**	Graphical User Interface
3.9	Top of matplotlib	**I.**	Diagnosing data entry errors
3.10	SciKit-Learn	**J.**	Pandas
		K.	Object
		L.	Unsupervised machine learning algorithm
		M.	Global variable

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.	Version Control	B.	Pop	C.	Push	D.	Len(s)
E.	Git	F.	Pandas	G.	Predictive	H.	Boosting
I.	Data Cleansing	J.	Slicing	K.	Use Message	L.	Byte array
M.	Supervised Learning algorithm						

4.1 _____ is performed by data scientist after acquiring the data in its argument to end of array.

4.2 _____ Python library are preferred to be used for Data Mining.

4.3 _____ systems record changes to a file over time.

4.4 _____ command allows you to update the repository.

4.5 _____ create a GUI component for displaying multiple-lines of text.

4.6 s = {1, 2, 4, 3}, _____ returns 4.

4.7 _____ revision control system.

4.8 _____ uses data on some object to predict values for other object statement in some other Python source file.

4.9 _____ is a computationally fast way to methodically access a range of items from sequence types.

4.10 _____ technique comes under practical machine learning.

PART TWO
(Answer any FOUR questions.)

5. Diamonds :

This dataset contains the prices and other attributes of many diamonds.

Content

Column Name	Description
price	price in Rs. (3260--18,823)
carat	weight of the diamond (0.2--5.01)
cut	quality of the cut (Fair, Good, Very Good, Premium, Ideal)
color	diamond colour, from J (worst) to D (best)
clarity	a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
x	length in mm (0--10.74)
y	width in mm (0--58.9)
z	depth in mm (0--31.8)
depth	total depth percentage = z / mean(x, y) = 2 * z / (x + y) (43--79)
table	width of top of diamond relative to widest point (43--95)

(a) What is the main difference between a Pandas series and a single-column DataFrame in Python ?

(b) Write a Pandas program to read a dataset from diamonds DataFrame and modify the default columns values and print the first 6 rows.

(c) Write a Pandas program to remove the second column of the diamonds Dataframe. **(5+5+5)**

6. (a) Name and explain few methods/techniques used in Statistics for analyzing the data.

(b) What is correlation and covariance in statistics ?

(c) Explain different types of merges available in Pandas. **(4+4+7)**

7. (a) Write a Python program to draw a line using given axis values taken from a text file, with suitable label in the x-axis, y-axis and a title.

Test Data :

test.txt

1 2

2 4

3 1

(b) Write a Python program to create multiple types of charts (a simple curve and plot some quantities) on a single set of axes. **(7+8)**

8. (a) Write a NumPy program to convert a list of numeric value into a one-dimensional NumPy array.

(b) Write a NumPy program to reverse an array.

(c) Write a NumPy program to test whether each element of a 1-D array is also present in a second array. **(5+5+5)**

9. (a) Explain different widget classes built into tkinter.

(b) Write a Python GUI program to create a Check button widget using tkinter module.

(c) Explain the difference between supervised and unsupervised machine learning. **(6+5+4)**

- o 0 o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK