

Sl. No.

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 हैं तथा भाग दो "व्यक्तिपरक" प्रकार का है और इसके कुल अंक 60 हैं।	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 the following is the use of id() function in python ?
- (A) Id returns the identity of the object
 - (B) Every object doesn't have a unique id
 - (C) Both of the above
 - (D) None of the above
- 1.2 Using _____ on categorical data will produce similar output to a Series or DataFrame of type string.
- (A) .desc()
 - (B) .describe()
 - (C) .rank()
 - (D) .dataframe()
- 1.3 Assume, you are given two lists : a = [1, 2, 3, 4, 5], b = [6, 7, 8, 9]. The task is to create a list which has all the elements of a and b in one dimension. Output : a = [1, 2, 3, 4, 5, 6, 7, 8, 9]
- Which of the following option would you choose ?
- (A) a.append(b)
 - (B) a.extend(b)
 - (C) Any of the above
 - (D) None of the above

- 1.4 You have built a machine learning model that you wish to freeze now and use later. Which of the following command can perform this task for you ? Note : Pickle library has been imported as pkl.
- (A) push(model, "file")
 - (B) save(model, "file")
 - (C) dump(model, "file")
 - (D) freeze(model, "file")
- 1.5 Set makes use of _____
Dictionary makes use of _____
- (A) keys, keys
 - (B) key values, keys
 - (C) keys, key values
 - (D) key values, key values
- 1.6 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}
- 1.7 Which of the following operation works with the same syntax as the analogous dict operations ?
- (A) Getting columns
 - (B) Setting columns
 - (C) Deleting columns
 - (D) All of the above

<p>1.8 Point out the wrong statement.</p> <p>(A) A DataFrame is like a fixed-size dict in that you can get and set values by index label</p> <p>(B) Series can be passed into most NumPy methods expecting an NDArray</p> <p>(C) A key difference between Series and NDArray is that operations between Series automatically align the data based on label</p> <p>(D) None of the above</p>	<p>2. Each statement below is either TRUE or FALSE. Choose the most appropriate one and ENTER in the “OMR” answer sheet supplied with the question paper, following instructions therein. (1x10)</p> <p>2.1 A lambda form in python does not have statements as it is used to make new function object and then return them at runtime.</p> <p>2.2 A Pearson correlation between two variables is zero but, still, their values can be related to each other.</p> <p>2.3 NDArray is also known as the alias array.</p> <p>2.4 Just like lists in python, NumPy arrays can also be sliced.</p> <p>2.5 stringlength() is a predefined Python function is used to find the length of string.</p> <p>2.6 Pandas consist of an integrated group by the engine for transforming and aggregating data sets.</p> <p>2.7 The panel is a container for Series, and DataFrame is a container for DataFrame objects.</p>
<p>1.9 What is the correct syntax of destroy in tkinter ?</p> <p>(A) destroy(object)</p> <p>(B) object.destroy()</p> <p>(C) object(destroy)</p> <p>(D) delete(object)</p>	<p>2.8 mainloop() is an infinite loop used to run the application, wait for an event to occur, and process the event till the window is not closed.</p> <p>2.9 Tkinter provides various methods to perform multithreading in python.</p>
<p>1.10 The correct way to draw a line in canvas tkinter ?</p> <p>(A) line()</p> <p>(B) canvas.create_line()</p> <p>(C) create_line(canvas)</p> <p>(D) None of the above</p>	<p>2.10 draw_shell(G, keywords) : This gives circular layout of the graph G.</p>

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

	X		Y
3.1	Unsupervised machine learning algorithms	A	axisartist
3.2	A collection of functions and objects for creating or placing inset axes.	B	bind
3.3	Includes a derived Axes implementation	C	constant
3.4	Space ticks logarithmically from min to max.	D	LogLocator
3.5	Containers that hold information which cannot be changed later	E	K-means
3.6	Core algorithm for building decision tree	F	axes_grid1.inset_locator
3.7	Numpy in the Python provides the	G	wm
3.8	Communicate with window manager	H	column_stack
3.9	associates Tcl scripts with X events	I	grid
3.10	Geometry manager that arranges widgets	J	List
		K	Apriori
		L	ID3
		M	Array

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.	Principal Component Analysis (PCA)	B.	Spinbox	C.	Entry Widget
D.	scatter(),plot()	E.	column_stack	F.	size, value
G.	arange	H.	plot(),scatter()	I.	Union
J.	plt.cla()	K.	ix_	L.	value,size
M.	Alpha()				

- 4.1 To create sequences of numbers, NumPy provides a function _____ analogous to range that returns arrays instead of lists.
- 4.2 _____ function can be used to combine different vectors so as to obtain the result for each n-uplet.
- 4.3 _____ function is used to stacks 1D arrays as columns into a 2D array.
- 4.4 Data structures in Pandas can be mutated in terms of _____ but not of _____.
- 4.5 _____ will happen to the indexes when any operations are done to unaligned series.
- 4.6 The Axes is the area on which the data is plotted with functions such as _____ and _____ .
- 4.7 _____ function is used to clear an axis.
- 4.8 _____ is used to create input fields in the GUI to take in textual input.
- 4.9 _____ is like an entry field, except it lets us select from a fixed number of values.
- 4.10 _____ is an unsupervised, non-parametric statistical technique primarily used for dimensionality reduction in machine learning.

PART TWO

(Answer any FOUR questions)

5. (a) Write the difference between data science and machine learning in terms of definition, role, and scope.

(b) What is logistic regression ?

(c) What advantages do Numpy Arrays offer over (nested) Python Lists ?

(5+5+5)

6. (a) Write a NumPy program to compute the mean, standard deviation, and variance of a given array along the second axis.

Sample output:

Original array: [0 1 2 3 4 5]

Mean: 2.5

std: 1

variance: 2.9166666666666665

(b) What is the difference between tuples and lists in Python ?

(10+5)

7. (a) Name a few libraries in Python used for Data Analysis and Scientific Computations.

(b) What is Python Pandas ?

(c) What is the application of Pandas ?

(d) What are different types of Data structures in Pandas ?

(e) What are the different ways in which a DataFrame can be created in Pandas ?

(3+3+3+3+3)

8. (a) The sales of a company (in million dollars) for each year are shown in the table below.

x (year)	2005	2006	2007	2008	2009
y (sales)	12	19	29	37	45

(i) Find the least square regression line $y = ax + b$.

(ii) Use the least squares regression line as a model to estimate the sales of the company in 2012.

- (b) Write the difference between Supervised and Unsupervised Learning. (10+5)

9. (a) Write a program to plot sin and cos graph using numpy and matplotlib, as shown in figure 1.

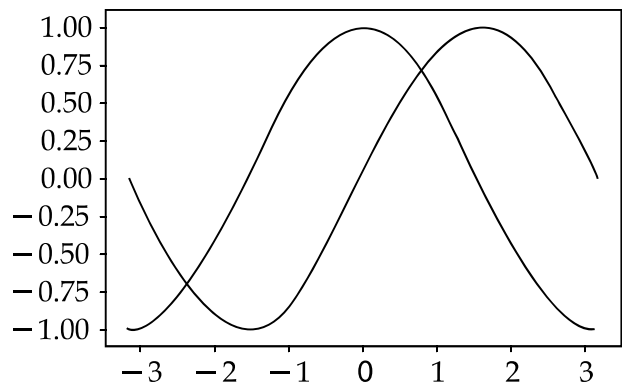


Figure-1

- (b) How can you check if a data set or time series is Random ? (10+5)

- o o o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK