Question Paper

Exam Date & Time: 12-Dec-2023 (09:30 AM - 12:30 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

THIRD SEMESTER B.TECH (AI & ML) END SEMESTER EXAMINATIONS, DEC 2023

INTRODUCTION TO DATA ANALYTICS [CSE 2126]

Marks: 50 **Duration: 180 mins.**

A

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1) Develop the python script for the following data processing: (5)

A)

a. Reading a table without header from CSV_01.csv file. Name the columns as RollNo, Name and CGPA.

b. Reading the second sheet of the excel file Excel_02.xls

c. Reading a table where data/column headings were separated by one or more whitespace characters from myData_03.txt file. Use Regular Expressions for parsing text file. Skip comments at lines 0, 2 and 5.

d. Convert a 3X3 table with values from 0 to 8 to .html format and print it.

e. Read Json data in tablular form from Books.json file.

B) Develop a python script to create and display an array of size 3X3 with numbers from 1 to 9 using np.arange() function. Find the maximum, minimum, sum and standard deviation of each of the rows and columns of the array. Also extract and display the middle row and column of the array. (3)

C) Consider FIVE students having roll no, name and marks in three different subjects, Maths, Physics and Chemistry. Design a python script to create a data frame of student's data with roll no as the index. Also, list the student names, having null values in Maths. (2)

2) a) Design a Python script to display the output string from a given input string, where all occurrences of input string's first character are changed to '$', except the first character itself. [sample Input: monument ; output: monu$ent] (5)

A)

b) Develop a Python function that takes a list of words as input and returns the longest word and the length of the longest one as output. [sample Input: 'a','ab','abc'; Output: 'abc', 3].

B) Design a Python script to print the no. of even numbers in a list. Consider only one occurrence of any number and use the functions, remove_duplicate(), which removes duplicates from a list, even_no() to check whether a given no. is even or not and main() function. [Sample Input: [1,2,3,4,2,4,6]; Output: No. of Even No's: 3] (3)

C) Consider the Fruits stock at two stores, which are stored as dictionaries: Store1: banana(100), mango(50) and Store2: apple:(60), mango(30). Convert this stock data into Pandas series with index, banana, apple, mango and pineapple and add them to get the total stock in two stores. Fill the NaN values with Zero before adding the stocks. (2)

3) Develop a Pandas script to perform Left join, right join and outer join operations on the two datasets (5)

A) using keys as given below. Also, compute the results of left join, right join and outer join operations on the datasets and tabulate the same.

Dataset1:

	Key1	Key2	P	Q
0	K0	K0	P0	Q0
1	K0	K1	P1	Q1
2	K1	K0	P2	Q2
3	K2	K1	P3	Q3

Dataset2:

	Key1	Key2	R	S
0	K0	K0	R0	S0
1	K1	K0	R1	S1
2	K1	K0	R2	S2
3	K2	K0	R3	S3

B) Develop a python script, which computes frequency of each character in a string. Also, it replaces blank spaces in a string with the least frequent character.[Sample input: 'All Apple'; output: 'AlleApple'] (3)

C) How will you convert the first character of each string element in a series to uppercase? Write appropriate python script to achieve the same. [Sample input:'aba', 'good'; output: 'Aba', 'Good'] (2)

4) In a binary classification using KNN classifier, 25 out of 40 samples with class A are correctly classified, while 7 out of 10 samples with class B are correctly classified. Calculate the following by mentioning the suitable formulas: (5)

A)

a. TP, FP, TN and FN.

b. Confusion matrix and tabulate the same.

c. Accuracy, precision and recall.

d. Specificity, Fallout and differences between the same.

B) Develop a Python function that takes two lists and returns True if they have at least one common member. [sample input: [1,2,3,4,5] , [5,6,7,8,9]; output: True] (3)

C) How to transform given data into discrete categories using pandas functions, illustrate with an example. (2)

5) Apply k-means clustering technique (with k=3) the following two dimensional data set : A(1,2), B(2,3), C(3,5),D(4,3), E(5,6),F(6,4), G(7,8), H(8,5). Use Manhattan distance (d = |x1 - x2| + |y1 - y2|) with max. 3 iterations. Consider A (1,2), B(2,3) and C(3,5) as initial centroids. (4)

A)

B) Design a Python script that has the dictionary of your friends' names as keys and phone numbers as its values. Print the dictionary in a sorted order. Prompt the user to enter the name and check if it is present in the dictionary. If the name is not present, then enter the details in the dictionary. [Sample input: {"Ram": '1234-567"}; output: 'The Phone number of Ram is '1234-567'] (4)

C) Illustrate, how to evaluate the performance of a classification model using the ROC curve with a suitable diagram. (2)

-----End-----