Total No. of Questions : 09

Total No. of Pages : 02

B.Tech. (CSE) (Sem.–6)

DATA SCIENCE

Subject Code : BTCS616-18

M.Code : 79256

Date of Examination : 17-05-2023

Time : 3 Hrs.

Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :

1. SECTION-A is COMPULSORY consisting of TEN questions carrying TWO marks each.
2. SECTION-B contains FIVE questions carrying FIVE marks each and students have to attempt any FOUR questions.
3. SECTION-C contains THREE questions carrying TEN marks each and students have to attempt any TWO questions.

SECTION-A

1. Write briefly :

   a) Differentiate the terms to clarify their meaning : data-warehousing, data-mining, data-processing, data-cleaning and data-extraction.

   b) What do you mean by 'data-science' ? Also list the processes involved in it.

   c) What is data pre-processing? How it is used to transform data and why?

   d) Discuss and exemplify 'structured' and 'unstructured' data.

   e) What are the error measuring and expectation minimization algorithms?

   f) What are data visualization tools?

   g) Define Mean Square Error (MSE).

   h) Differentiate between model 'underfitting' and 'overfitting'.

   i) What is 'machine' in machine learning programming paradigm?

   j) Give co-variance relationship measurement of two variable X and Y.

SECTION-B

2. Data of five students with study hours and marks obtained by each is given in the following table. Using K-means algorithm classify students into two groups.

No	Student ID	Study Hours	Marks Obtained
1	CS-1234	34	60
2	CS-2574	50	85
3	CS-3456	35	22
4	CS-1255	45	80
5	CS-2590	28	55

3. Discuss application of data science techniques for forecasting in different areas.

4. A given dataset's distribution Karl Pearson's coefficient of skewness is 0.65. standard deviation is 13 and mean is 59.2. Find its mode and median.

5. Find the median value for the result marks of 20 students and compare it with the mode: 2, 3, 7, 8, 8, 8, 9, 10, 10, 11, 12, 12, 14, 14, 16, 17, 17, 19, 19, 20

6. For the given variables X and Y values in the table, calculate the regression coefficients and obtain the lines of regression:

X	1	2	3	4	5	6	7
Y	8	9	10	13	12	11	14

SECTION-C

7. For two series of index numbers, P for price index and S for stock of the commodities. the mean and standard deviation is 100 and 8 for P and 103 and 4 for S. respectively. The correlation coefficient between P and S is 0.4. Find the regression lines with these data value P on S and S on P respectively.

8. What is the role accuracy score, precision score, recall and area under ROC curve to evaluate the performance of a data science model.

9. Write short notes on :

   a) Residual Plot

   b) Residual Value

   c) Actual Value

   d) Predicted Value.

NOTE : Disclosure of Identity by writing Mobile No. or Making of passing request on any page of Answer Sheet will lead to UMC against the Student.