Roll No. .....

Total Page No. : 4

31N1506

B.TECH. III SEM MAIN/BACK (NEW SCHEME)
ACADEMIC SESSION 2023-24

Artificial Intelligence & Data Science-III and Other Branches

3AD4–06 - Introduction to Data Science

Common to DS, CD, AD

Time : 3 Hours]

[Max. Marks : 70

[Min. Passing Marks :

Instructions to Candidates :

Part–A : Short Answer Type Questions (up to **25** words) 10 × 2 = 20 marks. All 10 questions are compulsory.

Part–B : Analytical/Problem Solving questions 5 × 4 = 20 marks. Candidates have to answer 5 questions out of 7.

Part–C : Descriptive/Analytical/Problem Solving questions 3 × 10 marks = 30 marks. Candidates have to answer 3 questions out of 5.

Schematic diagrams must be shown wherever necessary. Any data you feel missing may suitably be assumed and stated clearly. Units of quantities used/ calculated must be stated clearly.

Use of the following supporting materials is permitted during examination. (Mentioned in form no. 205).

1_____ 2_____

Part–A

10×2=20

1. Name the smallest and largest unit of data.
2. What type of data is involved in weather forecasting and air traffic control in an Airport ?
3. Give the representation on N-gram with the help of an example.
4. Name and explain the three V's of Big Data.
5. Suppose the quantitative data is 1, 4, 12, 4, 2, 5, 7, 13, 5, 17, 8, 9, 3. Compute the value of the First Quartile (Q1), Second Quartile (Q2), and Third Quartile(Q3) values from the given data.
6. Differentiate between Data profiling and data analytics.
7. Cosine similarity is superior to Euclidean distance; why ?
8. What are the challenges faced while handling big data ? Suggest two techniques to deal with big data.
9. Write a short note on REST API. How does it help in data collection ?
10. What is Cross-Validation and its types ?

Part–B

5×4=20

1. How do you handle missing data? What imputation techniques do you recommend ? Discuss with an example.
2. Calculate the regression coefficient of X on Y for the following data :

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

3. What do you understand about Data Analytics ? Write various data analytics capabilities. Explain the pyramid architecture of data analytics.

4. Why do we use SVMs for classification, regression, and outlier detection ? Discuss its polynomial kernel function.

5. What gradient is in machine learning, and why is it essential?

6. Explain True Positive, True Negative, False Positive, and False Negative in Confusion Matrix with an example.

7. How is machine learning used in day-to-day life ?

Part–C

3×10=30

1. Suppose we want to investigate whether students will pass or fail in a subject based on the information like they studied, cheated, and slept well before the exam. You are given the following data for five students. There are three features, "Studied," "Slept," and "Cheated." The column "Result" shows the label we want to predict.

	Studied	Slept	Cheated	Result
Student 1	Yes	No	No	Passed
Student 2	Yes	No	Yes	Failed
Student 3	No	Yes	No	Failed
Student 4	Yes	Yes	Yes	Failed
Student 5	Yes	Yes	No	Passed

(a) What is the entropy H (Result) at the root node of the following given data with the decision tree ?

(b) Draw the decision tree where every split maximizes the information gain.

2. Let us assume two points, as (x1, yl) and (x2, y2) in the two-dimensional coordinate plane if we want to compute the distance between two points P(0, 4) and Q(6, 2) using Euclidean distance, then find the distance between P and Q.

3. Consider the following data given in the following table. Compute range in terms of central tendency.

S. No.	Country	Population
1	Switzerland	6,32, 4014
2	Italy	7,32, 1211
3	Canada	12,23, 1221
4	USA	9,21, 6409
5	Macedonia	5,21, 4021

4. Write a short note on the following :

(a) What is the main reason to use REST over other APIs like SOAP ?

(b) Explain JSON objects and their structure.

5. Write a short note on the following :

(a) Explain about mapping problems to machine learning.

(b) Explain about Poisson and normal distribution.

****