**Reg. No. :** | | | | | | | | | | | | |

**Question Paper Code: U5316**

M.E. DEGREE EXAMINATION, NOV 2024

Professional Elective

Computer Science and Engineering

21PCS516 – ANALYTICAL DATA SCIENCE

(Regulations 2021)

Duration: Three hours | Maximum: 100 Marks

Answer ALL Questions

PART A - (5 x 20 = 100 Marks)

| | | | | |
|---|---|---|---|---|
| 1. | (a) | Explain the role of matrix operations in machine learning algorithms. How does linear algebra facilitate tasks such as feature extraction and dimensionality reduction? | CO1-U | (20) |
| | | Or | | |
| | (b) | Explain the significance of problem formulation in data science. How do defining clear objectives and success criteria contribute to structured problem-solving? | CO1-U | (20) |
| 2. | (a) | Develop a plan to use a histogram to analyze the distribution of exam scores in a classroom. Include steps for data preprocessing, binning, and interpretation of results. | CO2- App | (20) |
| | | Or | | |
| | (b) | Apply the concept of PCA to reduce the dimensionality of a dataset containing multiple features. Demonstrate how PCA helps in preserving important information while reducing noise and redundancy. | CO2- App | (20) |
| 3. | (a) | The features in a data set are as given below. Find the dimensionality of the data set. $\begin{bmatrix} 2 & 1 & -1 \\ 4 & -1 & 5 \\ -2 & 3 & 4 \end{bmatrix}$ | CO2-App | (20) |
| | | Or | | |
| | (b) | Find all the distance measures for the points (1,2,3,4,5) and (5,6,7,8,9). | CO2-App | (20) |

4. (a) (i) How do we overcome data discovery challenges? CO2- App (20)
(ii) Discuss data discovery use cases.

Or

(b) (i) Sort a list of numbers in ascending order using Python CO2- App (20)
(ii) Write a python program to convert marks obtained into grade.

5. (a)

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **INT** | 15 | 23 | 18 | 23 | 24 | 22 | 22 | 19 | 19 | 16 | 24 | 11 | 24 | 16 | 23 |
| **EXT** | 49 | 63 | 58 | 60 | 58 | 61 | 60 | 63 | 60 | 52 | 62 | 30 | 59 | 49 | 68 |

CO2- App (20)

Apply k means clustering to the above data set

Or

(b) Explain in detail about data pre-processing and preparation with python code. CO2- App (20)