Code No : 20AD6T02 **R20**

III B. TECH II SEMESTER REGULAR EXAMINATIONS APRIL - 2023
DATA SCIENCE, PREPARATION AND ANALYSIS
(ARTIFICIAL INTELLIGENCE AND DATA SCIENCE)

Time: 3 hours Max. Marks: 70

Note: Answer **ONE** question from each unit **(5 × 14 = 70 Marks)**

UNIT-I

1. a) What is Data Science? Explain in detail about data science process. [7M]

   b) Demonstrate in detail about exploratory data analysis with examples. [7M]

(OR)

2. a) Discuss the best practices for data cleansing and data transformation. [7M]

   b) How models can be built on various data science applications. Explain in detail. [7M]

UNIT-II

3. a) Describe the different types of frequency distributions, and provide an example of how they can be used in practice. [6M]

   b) A company has collected data on the number of sales made by each of its sales representatives in a month. The data is as follows: 23, 12, 19, 16, 24, 15, 17, 21, 14, 18. Create a frequency distribution and a relative frequency distribution for the data. Calculate the mean, median, and mode and Performance. [8M]

(OR)

4. a) A survey was conducted to gather data on the heights of a group of students. The data is as follows: 68, 70, 71, 66, 72, 69, 67, 68, 70, 73. Create a histogram and a box plot to display the data. [8M]

   b) Explain the concept of range, variance, and standard deviation, and describe how to calculate them. [6M]

UNIT-III

5. a) Compare and contrast Python lists and Numpy arrays. What are the advantages of using Numpy arrays over lists for scientific computing and data analysis? [7M]

   b) Discuss the key features of Pandas and how they can be used to read, manipulate, and analyze data. [7M]

(OR)

6. a) Explain the concept of hierarchical indexing in Pandas. How does it work and why is it useful in data analysis? [7M]

   b) How can you use Pandas to combine and merge datasets? Discuss the different methods for joining, merging, and concatenating datasets in Pandas [7M]

UNIT-IV

7. a) Discuss the situations where normal distributions are commonly used and explain how they can be used to analyze and interpret data. [7M]

   b) A standardized test has a mean score of 500 and a standard deviation of 100. If a student scores a 620 on the test, what is their z-score? How does their score compare to the rest of the students who took the test? [7M]

(OR)

8. a) Discuss the formula for calculating the least squares regression line and explain how it can be used to predict values based on a relationship between variables. [6M]

   b) A company produces light bulbs with a mean life of 800 hours and a standard deviation of 40 hours. If the company wants to guarantee that at least 95% of their bulbs will last longer than the stated life, what is the minimum life they should guarantee? [8M]

UNIT-V

9. a) Discuss the key features of matplotlib and some common types of plots that can be created using this library. [7M]

   b) Discuss the importance of using color, labeling, and other design elements in creating effective visualizations. [7M]

(OR)

10. a) What is data visualization, and why is it important in data analysis? Discuss the benefits of using visualizations to present data. [7M]

    b) How can you use a scatter plot to visualize this relationship? What are some advantages and disadvantages of using a scatter plot for this type of data? [7M]

* * * * *