**MODEL QUESTION PAPER**

|  |  |
| --- | --- |
| **Code No: CT3545** | **SRGEC-R20** |
| **III B.Tech II Semester Regular Examinations** | |
| **DATA SCIENCE**  **(Artificial Intelligence and Data Science)** | |
| **Time: 3 Hours Max. Marks: 70** | |

**Note:** Answer all questions. All Questions carry Equal Marks

**5 × 14 = 70M**

**Unit - I**

|  |  |  |
| --- | --- | --- |
|  |  | BL |
| 1. | a) Define statistical learning. How do you estimate function f(x) using x? (7M)  b) Describe KNN regression? Compare linear regression with KNN. (7M) | L1 |
|  | **(OR)** |  |
| 2. | a) Develop Python program for calculating Multiple linear regression. (8M)  b) Summarize Simple linear regression. (6M) | L2 |
| **Unit – II** | | |
| 3. | a) What is ‘training Set’ and ‘test Set’ in a Machine Learning Model? How Much Data Will You Allocate for Your Training, Validation, and Test Sets? (7M)  b) What Is ‘naive’ in the Naive Bayes Classifier? Explain about Naive Bayes Classifier. (7M) | L1 |
|  | **(OR)** |  |
| 4. | a) What is Bias and Variance in a Machine Learning Model? (7M)  b) Explain the Confusion Matrix with Respect to Machine Learning Algorithms. (7M) | L4 |
| **Unit – III** | | |
| 5. | a) What’s the best way to draw multiple lines in a single figure with neat example? (7M)  b) How to Plot Histogram in Python using Matplotlib? (7M) | L4 |
|  | **(OR)** |  |
| 6. | a) What is Matplotlib? How do you install Matplotlib for python? What are some of the features provided by Matplotlib? (7M)  b) Discuss about two sample t-test with unequal variance and equal variance with example. (7M) | L2 |
| **Unit - IV** | | |
| 7. | a) Discuss Python Code to retrieve data from Java Script Object Notation format. (7M)  b) Discuss about Data imputations and Data acquisition using Python programming. (7M) | L5 |
|  | **(OR)** |  |
| 8. | a) Make use of the following data and answer the questions given   |  |  |  |  | | --- | --- | --- | --- | | **SL NO** | **Roll No** | **Marks 1** | **Marks 2** | | 1 | 3001 | 41 | NA | | 2 | 3002 | 50 | 71 | | 3 | 3003 | NA | 80 | | 4 | 3004 | 77 | 66 | | 5 | 3005 | 54 | 85 | | 6 | 3006 | 50 | 76 | | 7 | 3007 | NA | 70 | | 8 | 3008 | 85 | NA | | 9 | 3009 | 48 | 73 | | 10 | 3010 | 77 | 83 |   Develop a Python program to replace the missing values with column average. (8M)  b) Explain about Dimensionality Reduction. (6M) | L3 |
| **Unit - V** | | |
| 9. | a) Discuss about (i) Drill-down and (ii) Pivot (7M)  b) Discuss about data de-duplication and data summarization. (7M) | L1 |
|  | **(OR)** |  |
| 10. | a) Discuss Roll-up and Slice with Python. (7M)  b) Explain data visualization using CUBEs with neat example. (7M) | L1 |

BL ==> Blooms taxonomy Level (**L1**-Remembering, **L2**-Understanding, **L3**-Applying, **L4**-Analyzing, **L5**- Evaluating, **L6**- Creating)