

Code No: **R20A6701****MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY****(Autonomous Institution – UGC, Govt. of India)****III B.Tech I Semester Supplementary Examinations, June 2025****Introduction to Data Science****(B.Tech-AIML)**

Roll No									
---------	--	--	--	--	--	--	--	--	--

Time: 3 hours**Max. Marks: 70**

Note: This question paper Consists of 5 Sections. Answer **FIVE** Questions, Choosing **ONE** Question from each SECTION and each Question carries 14 marks.

SECTION-I

- | | | | BCLL | CO(s) | Marks |
|---|--|-------------------------------------------------------------------------------------------------------|-------------|--------------|--------------|
| 1 | | What are the essential Python libraries for data scientists, and how do they assist in data analysis? | L1 | CO-I | [14M] |

OR

- | | | | | | |
|---|----------|------------------------------------------------------------------|-----------|-------------|-------------|
| 2 | A | How can data be sorted and ranked in Python using pandas? | L2 | CO-I | [7M] |
| | B | Explain how to plot data in Python using Matplotlib and Seaborn. | L3 | CO-I | [7M] |

SECTION-II

- | | | | | | |
|---|----------|--------------------------------------------------------------------------------|-----------|--------------|-------------|
| 3 | A | What is Exploratory Data Analysis (EDA), and how is it used to summarize data? | L2 | CO-II | [7M] |
| | B | Describe how data distribution is measured in statistics | L2 | CO-II | [7M] |

OR

- | | | | | | |
|---|--|----------------------------------------------------------------------------------------|-----------|--------------|--------------|
| 4 | | Explain the frequency approach to statistical inference and its applications in detail | L3 | CO-II | [14M] |
|---|--|----------------------------------------------------------------------------------------|-----------|--------------|--------------|

SECTION-III

- | | | | | | |
|---|--|-------------------------------------------------------------------------------------------------------------|-----------|---------------|--------------|
| 5 | | Explain the training, validation, and test set split in supervised learning in detail with suitable example | L3 | CO-III | [14M] |
|---|--|-------------------------------------------------------------------------------------------------------------|-----------|---------------|--------------|

OR

- | | | | | | |
|---|----------|--------------------------------------------------------------------------------|-----------|---------------|-------------|
| 6 | A | Compare and contrast Support Vector Machines (SVM) and random forests | L2 | CO-III | [7M] |
| | B | What are random forests, and how do they operate in supervised learning tasks? | L3 | CO-III | [7M] |

SECTION-IV

- | | | | | | |
|---|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------------|--------------|
| 7 | | Define multiple linear regression and explain how it extends simple linear regression to model the relationship between multiple independent variables and one dependent variable. Provide an example. | L4 | CO-IV | [14M] |
|---|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------------|--------------|

OR

- | | | | | | |
|---|--|----------------------------------------------------------------------------------------------------|-----------|--------------|--------------|
| 8 | | What are quality measures of clustering, and how are they used to evaluate clustering performance? | L3 | CO-IV | [14M] |
|---|--|----------------------------------------------------------------------------------------------------|-----------|--------------|--------------|

SECTION-V

- | | | | | | |
|---|--|------------------------------------------------------------------------------------|-----------|-------------|--------------|
| 9 | | Explain the PageRank algorithm and its significance in ranking nodes in a network. | L4 | CO-V | [14M] |
|---|--|------------------------------------------------------------------------------------|-----------|-------------|--------------|

OR

- 10** Discuss the importance of visualizing networks and the role of drawing graphs in network analysis. **L3 CO-V [14M]**
