



**AKKINENI NAGESWARA RAO COLLEGE:: GUDIVADA**  
(An Autonomous college under the jurisdiction of Krishna University, Machilipatnam)

**B.Sc. Honours in Data Science MINOR**

(For the batch of students admitted from 2023-24)

**I YEAR II SEMESTER**

**INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING**

Course Code: 23DS2	Admitted Batch: 2023-24	No. of Teaching Hours/week : 3	No. of Credits : 3
Year of Introduction: 2023-24	Year of offering: 2023-24	Year of Revision: 2023-24	% of Revision: 100%
Course Delivery Method: Class Room/Blended Mode/Both	C.I.A: 30 Marks	S.E.E: 70 Marks	Total: 100 Marks

**Course Objective:**

Data Science is a fast-growing interdisciplinary field, focusing on the analysis of data to extract knowledge and insight. This course will introduce students to the collection, preparation, analysis, modelling and visualization of data, covering both conceptual and practical issues. Examples and case studies from diverse fields will be presented, and hands-on use of statistical and data manipulation software will be included.

**Course Learning Outcomes:**

- Recognize the various discipline that contribute to a successful data science effort.
- Understand the processes of data science identifying the problem to be solved, data collection, preparation, modeling, evaluation and visualization.
- Be aware of the challenges that arise in Data Sciences.
- Be able to identify the application of the type of algorithm based on the type of the problem.
- Be comfortable using commercial and open source tools such as the R/Python language and its associated libraries for data analytics and Visualization.

**SYLLABUS**

**UNIT I:**

Defining Data Science and Big data, Benefits and Uses, facets of Data, Data Science Process. History and Overview of R, Getting Started with R, R Nuts and Bolts

**UNIT II:**

The Data Science Process: Overview of the Data Science Process-Setting the research goal, Retrieving Data, Data Preparation, Exploration, Modeling, data Presentation and Automation. Getting Data in and out of R, Using reader package, Interfaces to the outside world.

### UNIT III:

Machine Learning: Understanding why data scientists use machine learning-What is machine learning and why we should care about, Applications of machine learning in data science, Where it is used in data science, The modeling process, Types of Machine Learning-Supervised and Unsupervised.

### UNIT IV:

Handling large Data on a Single Computer: The problems we face when handling large data, General Techniques for handling large volumes of data, Generating programming tips for dealing with large datasets.

### UNIT V:

Sub setting R objects, Vectorised Operations, Managing Data Frames with the dplyr, Control structures, functions, Scoping rules of R, Coding Standards in R, Loop Functions, Debugging, Simulation. Case studies on preliminary data analysis.

### TEXT BOOKS:

1. DavyCielen, Arno. D.B.Maysman, Mohamed Ali, "Introducing DataScience" Manning Publications, 2016.
2. Roger D. Peng, "R Programming for Data Science" Lean Publishing, 2015.

### REFERENCE BOOKS:

1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.
2. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta, "Practical DataScience Cookbook", Packt Publishing Ltd., 2014.

Web References for case studies:

1. <https://www.kaggle.com/datasets>
2. <https://github.com/>

**Recommended Co-curricular activities:** (Co-curricular Activities should not promote copying from text book or from others' work and shall encourage self/independent and group learning)

#### A. Measurable:

1. Assignments on
2. Student seminars (Individual presentation of papers) on topics relating to
3. Quiz Programmes on:
4. Individual Field Studies/projects:
5. Group discussion on:
6. Group/Team Projects on:

#### B. General

1. Collection of news reports and maintaining a record of paper-cuttings relating to topics covered in syllabus.
2. Group Discussions on:
3. Watching TV discussions and preparing summary points recording personal observations etc., under guidance from the Lecturers
4. Any similar activities with imaginative thinking.
5. Recommended Continuous Assessment methods



**AKKINENI NAGESWARA RAO COLLEGE:: GUDIVADA**  
(An Autonomous college under the jurisdiction of Krishna University, Machilipatnam)

**B.Sc. Honours in Data Science MINOR**

(For the batch of students admitted from 2023-24)

**I YEAR II SEMESTER**

**INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING**

**Course Code: 23DS2**

**BLUE PRINT**

**Section A (5 x 4 = 20 marks)**

- 5 questions to be answered out of 8 questions
- At least 1 question must be given from each unit.

**Section B (5 x 10 = 50 marks)**

- Answer ALL questions must be given from each unit.

	<b>Section A</b>	<b>Section B</b>	<b>Total Marks</b>
Unit I	2	2	28
Unit II	1	2	24
Unit III	1	2	24
Unit IV	2	2	28
Unit V	2	2	28
	<b>8</b>	<b>10</b>	<b>132</b>



**AKKINENI NAGESWARA RAO COLLEGE:: GUDIVADA**  
(An Autonomous college under the jurisdiction of Krishna University, Machilipatnam)

**B.Sc. Honours in Data Science MINOR**

(For the batch of students admitted from 2023-24)

**I YEAR II SEMESTER**

**INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING**

**Course Code: 23DS2**

**Model Paper**

**Time: 3 hrs**

**Total: 70 M**

**Section – A**

Answer any **FIVE** of the following questions

**5 X 4 = 20 M**

1. What is data science, and Big data, What is the application of data science.
2. Explain Read R package
3. What are the applications of machine learning in data science.
4. What are the different challenges that we face when handling large data.
5. What is meant by data frame in 'R'. Explain dplyr package.
6. What are the different types of big data.
7. What are the four steps in modeling process in machine learning.
8. What is meant by debugging.

**Section – B**

Answer the following questions

**5 X 10 = 50 M**

9. a) Explain different phases of facets of data.  
(or)  
b) What is R. Describe basic commands in R with Examples ( Vectors, matrices, lists, data frames etc.)
10. a) Explaining detail the steps involved in data science process.  
(or)  
b) What are the different ways of loading data into R. with examples.
11. a) What are the different types of machine learning processes. Explain detail.  
(or)  
b) List out the importance of machine learning and gives examples in our day to day life.
12. a) What are the different techniques for handling large volumes of data.  
(or)  
b) Explain any case study that deals with large data sets.
13. a) Explain Vectorised operations, control structures, functions and loop functions in R.  
(or)  
b) Explain and give examples of exploring data using single variable and two variables.



**AKKINENI NAGESWARA RAO COLLEGE:: GUDIVADA**  
(An Autonomous college under the jurisdiction of Krishna University, Machilipatnam)

**B.Sc. Honours in Data Science MINOR**

(For the batch of students admitted from 2023-24)

**I YEAR II SEMESTER**

**INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING**

**Credits: 1**

**Course Code: 23DS2 (P)**

**LAB LIST**

**List of Experiments**

1. Installing R and R studio, with proper notes on version management, cosmetic settings and different libraries.
2. Basic operations in r with arithmetic and statistics.
3. Getting data into R, Basic data manipulation, Loading Data into R
4. Basic plotting
5. Loops and functions
6. Create Vectors, Lists, Arrays, Matrices, Data frames and operations on them.
7. Demonstrate the visualization and graphics using visualization packages like ggplot2.
8. Implement Loop functions with lapply(), sapply(), tapply(), apply(), mapply().
9. Explore data using Single Variables: Unimodal, Bimodal, Histograms, Density Plots, Bar charts
10. Explore data using two Variables: Line plots, Scatter Plots, smoothing curves, Bar charts
11. Explore and implement commands using dplyr package
12. Download a dataset and work on basic data manipulation followed by inferential statistics.