



# MARRI LAXMAN REDDY INSTITUTE OF TECHNOLOGY AND MANAGEMENT

(AN AUTONOMOUS INSTITUTION)

(Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad)

Accredited by NAAC with 'A' Grade & Recognized Under Section 2(f) & 12(B) of the UGC act, 1956

## COURSE CONTENT

DATA SCIENCE								
IV-I Semester: CE / CSD / CSE / CSM / ECE / EEE / ME								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
2470529	Foundation	L	T	P	C	CIA	SEE	Total
		3	0	0	3	40	60	100
Contact Classes: 48	Tutorial Classes: Nil	Practical Classes: Nil			Total Classes: 48			
Prerequisites: There are no prerequisites to take this course.								

### Course Overview:

- Data Science is an interdisciplinary course that integrates statistics, programming, and data visualization to extract meaningful insights from data.
- This course introduces students to the core concepts of data science, including data types, statistical inference, data modeling, and visualization techniques, with hands-on exposure to R programming.
- Emphasis is placed on understanding data, summarizing it statistically, visualizing patterns, and building regression models for prediction and interpretation.

### Course Objectives:

- To Learn concepts, techniques and tools they need to deal with various facets of data science practice, including data collection and integration
- Understand the basic types of data and basic statistics
- Identify the importance of data reduction and data visualization techniques
- Introduce the fundamental concepts of Data Science, including big data, datafication, and statistical inference.
- Develop practical skills in R programming for handling vectors, matrices, factors, data frames, and lists.

### Course Outcomes: After Completion of the Course, Students should be able to

- After completion of the course, student should be able to
- Understand basic terms of statistical modeling and data science
- Implement of R programming concepts
- Utilize R elements for data visualization and prediction
- Differentiate between various data types and attributes, and
- summarize data using statistical measures and graphical representations.
- Construct data visualizations (charts, plots, histograms, scatterplots)
- and perform regression analysis (linear and multiple regression) for interpreting and modelling of data.

**UNIT - I:** Definition of Data Science- Big Data and Data Science hype – and getting past the hype - Datafication- Current landscape of perspectives - Statistical Inference - Populations and samples – Statistical modeling, probability distributions, fitting a model – Over fitting. Basics of R: Introduction, R-Environment Setup, Programming with R, Basic Data Types.

### UNIT - II:

Data Types & Statistical Description: Types of Data: Attributes and

Measurement, Attribute, The Type of an Attribute, The Different Types of Attributes, Describing Attributes by the Number of Values, Asymmetric Attributes, Binary Attribute, Nominal Attributes, Ordinal Attributes, Numeric Attributes, Discrete versus Continuous Attributes. Basic

Statistical Descriptions of Data: Measuring the Central Tendency: Mean, Median, and Mode, Measuring the Dispersion of Data: Range, Quartiles, Variance, Standard Deviation, and Interquartile Range, Graphic Displays of Basic Statistical Descriptions of Data.

**UNIT - III:** Vectors: Creating and Naming Vectors, Vector Arithmetic, Vector sub setting, Matrices: Creating and Naming Matrices, Matrix Sub setting, Arrays, Class. Factors and Data Frames: Introduction to Factors: Factor Levels, Summarizing a Factor, Ordered Factors, Comparing Ordered Factors, Introduction to Data Frame, subsetting of Data Frames, Extending Data Frames, Sorting Data Frames. Lists: Introduction, creating a List: Creating a Named List, Accessing List Elements, Manipulating List Elements, Merging Lists, Converting Lists to Vectors

**UNIT - IV:**

Conditionals and Control Flow: Relational Operators, Relational Operators and Vectors, Logical Operators, Logical Operators and Vectors, Conditional Statements. Iterative Programming in R: Introduction, While Loop, For Loop, Looping Over List. Functions in R: Introduction, writing a Function in R, Nested Functions, Function Scoping, Recursion, Loading an R Package, Mathematical Functions in R.

**UNIT - V:** Charts and Graphs: Introduction, Pie Chart: Chart Legend, Bar Chart, custom charts, cyclical data and circular area charts, Dual Axis charts, Box Plot, Histogram, Line Graph: Multiple Lines in Line Graph, Scatter Plot.

Regression: Linear Regression Analysis, Multiple Linear regression

**TEXT BOOKS:**

1. Doing Data Science, Straight Talk from The Frontline. Cathy O’Neil and Rachel Schutt, O’Reilly,2014.
2. K G Srinivas, G M Siddesh, “Statistical programming in R”, Oxford Publications.

**REFERENCE BOOKS:**

1. Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, 3rd ed.The Morgan Kaufmann Series in Data Management Systems

**ELECTRONIC RESOURCES:**

1. <https://www.geeksforgeeks.org/r-language/r-tutorial>
2. <https://www.programiz.com/r>
3. <https://www.w3schools.com/r>
4. <https://www.programiz.com/statistics>
5. <https://www.khanacademy.org/math/statistics-probability>

## **MATERIALS ONLINE:**

1. Course template
2. Tutorial question bank
3. Tech talk and Concept Video topics
4. Open-ended experiments
5. Definitions and terminology
6. Assignments
7. Model question paper – I
8. Model question paper – II
9. Lecture notes
10. E-Learning Readiness Videos (ELRV)