

PRINCIPLES OF DATA SCIENCE

III Semester: CSE (DS)																																									
Course Code	Category	Hours / Week			Credits	Maximum Marks																																			
ACDC01	Core	L	T	P	C	CIA	SEE	Total																																	
		3	1	0	4	30	70	100																																	
Contact Classes: 45		Tutorial Classes: 15		Practical Classes: Nil		Total Classes: 60																																			
Prerequisites: There are no prerequisites to take this course.																																									
<p>I. COURSE OVERVIEW: The course is designed to introduce to the basics of data science and use R for statistical programming, computation, graphics, and modeling. Topics covered include flavors of data, basic mathematics, probability, statistics and data visualization. The main objective of the course is to teach a range of topics and concepts related to the data science process. This course reaches to student by power point presentations, lecture notes, and lab which will give you the chance to apply knowledge of data science process.</p> <p>II. COURSE OBJECTIVES: The students will try to learn:</p> <ol style="list-style-type: none"> The fundamental knowledge on basics of data science and R programming. The programs in R language for understanding and visualization of data using statistical functions and plots. The fundamentals of how to obtain, store, explore, and model data efficiently. The fundamentals of probability and statistics for data science. <p>III. COURSE OUTCOMES: After successful completion of the course, students should be able to:</p> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <thead> <tr> <th style="width: 10%;">CO</th> <th style="width: 70%;">Outcome</th> <th style="width: 20%;">Action</th> </tr> </thead> <tbody> <tr> <td>CO 1</td> <td>Recognize the different levels of Data Science concepts for visualization of data.</td> <td>Remember</td> </tr> <tr> <td>CO 2</td> <td>Demonstrate the data visualization and statistical techniques , for describing data structure properties.</td> <td>Understand</td> </tr> <tr> <td>CO 3</td> <td>Describe the R programming for manipulation of data sets .</td> <td>Remember</td> </tr> <tr> <td>CO 4</td> <td>Apply the advanced data structure for efficient data storage.</td> <td>Apply</td> </tr> <tr> <td>CO 5</td> <td>Make Use of the container Data types for display the functional values.</td> <td>Apply</td> </tr> <tr> <td>CO 6</td> <td>Analyse the basics of probability and statistics models for data exploration .</td> <td>Analyze</td> </tr> <tr> <td>CO 7</td> <td>Apply the Bayes theorem for design new data model.</td> <td>Apply</td> </tr> <tr> <td>CO 8</td> <td>Make use of Hypothesis testing for statistical analytics fordestroying target based on the mission requirements.</td> <td>Understand</td> </tr> <tr> <td>CO 9</td> <td>Identify the effective and ineffective visualizations ,for data communication.</td> <td>Apply</td> </tr> <tr> <td>CO 10</td> <td>Outline the correlation and causation for relationship of different variables.</td> <td>Understand</td> </tr> </tbody> </table> <p>IV. SYLLABUS:</p> <p>MODULE – I: Flavors of Data (09) Structured versus unstructured data, Quantitative and qualitative data, The four levels of data: Nominal level, Ordinal level, Interval level, and Ratio level, The five steps of Data Science: Ask an interesting question, obtain the data, explore the data, model the data, communicate and visualize the results, Explore the data.</p> <p>MODULE – II: Introduction to R Programming (09) How to run R, R Sessions and Functions, Basic Math, Variables, Data Types, Vectors, Conclusion, Advanced Data Structures, Data Frames, Lists, Matrices, Arrays, Classes, R Programming Structures, Control Statements,</p>									CO	Outcome	Action	CO 1	Recognize the different levels of Data Science concepts for visualization of data.	Remember	CO 2	Demonstrate the data visualization and statistical techniques , for describing data structure properties.	Understand	CO 3	Describe the R programming for manipulation of data sets .	Remember	CO 4	Apply the advanced data structure for efficient data storage.	Apply	CO 5	Make Use of the container Data types for display the functional values.	Apply	CO 6	Analyse the basics of probability and statistics models for data exploration .	Analyze	CO 7	Apply the Bayes theorem for design new data model.	Apply	CO 8	Make use of Hypothesis testing for statistical analytics fordestroying target based on the mission requirements.	Understand	CO 9	Identify the effective and ineffective visualizations ,for data communication.	Apply	CO 10	Outline the correlation and causation for relationship of different variables.	Understand
CO	Outcome	Action																																							
CO 1	Recognize the different levels of Data Science concepts for visualization of data.	Remember																																							
CO 2	Demonstrate the data visualization and statistical techniques , for describing data structure properties.	Understand																																							
CO 3	Describe the R programming for manipulation of data sets .	Remember																																							
CO 4	Apply the advanced data structure for efficient data storage.	Apply																																							
CO 5	Make Use of the container Data types for display the functional values.	Apply																																							
CO 6	Analyse the basics of probability and statistics models for data exploration .	Analyze																																							
CO 7	Apply the Bayes theorem for design new data model.	Apply																																							
CO 8	Make use of Hypothesis testing for statistical analytics fordestroying target based on the mission requirements.	Understand																																							
CO 9	Identify the effective and ineffective visualizations ,for data communication.	Apply																																							
CO 10	Outline the correlation and causation for relationship of different variables.	Understand																																							

Loops, - Looping Over Nonvector Sets,- If-Else, Arithmetic and Boolean Operators and values, Default Values for Argument, Return Values, Functions are Objects, Recursion.

MODULE – III: Basic Mathematics and Probability for Data Science (09)

Mathematics: Vectors and matrices, Arithmetic symbols, Graphs, Logarithms/exponents, Set theory, Linear algebra.

Probability: Basic definitions, Probability, Bayesian versus Frequentist, Compound events, Conditional Probability, The rules of probability, Collectively exhaustive events, Bayes theorem, Random variables.

MODULE – IV: Statistics for Data Science (09)

Statistics: Obtaining data, Sampling data, Measuring Statistics, The Empirical rule, Point estimates, Sampling distributions, Confidence intervals, Hypothesis tests.

MODULE – V: Communicating Data (09)

Identifying effective and ineffective visualizations: Scatter plots, Line graphs, Bar charts, Histograms, Box plots. Graphs and Statistics lie: Correlation versus causation, Simpson’s paradox, Verbal Communication, The why/how/what strategy of presenting.

V. TEXT BOOKS:

1. Sinan Ozdemir, “Principles of Data Science”, Packt.
2. Norman Matloff , “The Art of R Programming”, Cengage Learning.

VI. REFERENCE BOOKS:

1. Cathy O’Neil and Rachel Schutt, “Doing Data Science, Straight Talk From The Frontline”, O’Reilly, 2014.
2. G. Jay Kerns, “Introduction to Probability and Statistics Using R”, First Edition.
3. Nina Zumel, John Mount, “Practical Data Science with R”, Manning Publications, 1st Edition, 2014.

VII. WEB REFERENCES:

1. https://en.wikipedia.org/wiki/R_programming_language
2. <http://www.r-bloggers.com/how-to-learn-r-2/#h.obx6jyuc9j7t>.
3. <http://www.tutorialspoint.com/r/>