INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

DATA SCIENCE							
I Semester: CSE							
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**	
BCSD03	**Elective**	**L**	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
		3	0	0	3	40	60	100
Contact Classes: 45	**Total Tutorials: Nil**	**Total Practical Classes: Nil**				**Total Classes: 45**		

I. COURSE OVERVIEW:
Data Science is a field of study that deals with the collection, analysis, and processing various data or information to extract solutions. It deals with the understanding of many structured and unstructured data with specialized knowledge to get the required insights. It is that part of science which also requires knowledge about business or commerce related fields.

II. COURSE OBJECTIVES:
The students will try to learn:
I.   The fundamental knowledge on basics of data science and R programming.
II. The programs in R language for understanding and visualization of data using statistical functions and plots.
III. How to apply hypotheses and data into actionable predictions.
IV. A range of machine learning algorithms along with their strengths and weaknesses.
V.   The document and transfer the results and effectively communicate the findings using visualization techniques.

III. COURSE OUTCOMES:
After successful completion of the course, students should be able to

CO1	**Explore** the fundamental concepts of data science	Apply
CO2	**Make use of** various data description functions in R programming for exhibiting various stages of the data science process.	Apply
CO 3	**Identify** interfacing packages for handling SQL and NoSQL databases for performing data analysis.	Analyze
CO 4	**Evaluate** data models using clustering and classification techniques.	Apply
CO 5	**Solve** various real-time problems on various hypothesis conditions by using artificial neural networks.	Apply
CO 6	**Illustrate** delivering results through documentation and visualization techniques	Understand

IV. COURSE CONTENT:
MODULE-I: INTRODUCTION (10)
Data science process, roles, stages in data science project, working with data from files, working with relational databases, exploring data, managing data, data cleaning and sampling for modeling. Introduction to R: Introduction to various data types, reading and writingdatasets, working with different file types .txt, .csv, outliers, R functions and loops; Summary statistics: Summary, str, aggregate, subset, head, tail; Probability distribution.

MODULE-II: SQL, NOSQL AND DATA ANALYSIS (10)
SQL using R, excel and R, introduction to No SQL, connecting R to No SQL databases, R with XML,

JSON; Correlation analysis; Covariance analysis, ANOVA, forecasting, heteroscedasticity, autocorrelation; Regression analysis: Regression modeling, multiple regression.

MODULE-III: DATA ANALYSIS (08)
Data analysis: Introduction, Terminology and concepts, Central tendencies and distributions, Variance, Basic machine learning algorithms, Linear regression, SVM Cluster analysis: K-means algorithm, Naive Bayes memorization methods ,evaluating clustering models and validating models.

MODULE-IV: ARTIFICIAL NEURAL NETWORKS (09)
Artificial neural networks: Introduction, neural network representation, appropriate problems for neural network learning, perceptions, multilayer networks and the back propagation algorithm Evaluation hypotheses: Motivation, estimation hypothesis accuracy, basics of sampling theory, a general approach for deriving confidence intervals, difference in error of two hypothesis,comparing learning algorithms.

MODULE-V: DELIVERING RESULTS (08)
Documentation and deployment, producing effective presentations, introduction to graphical analysis, plot() function, displaying multivariate data, matrix plots, multiple plots in one window, exporting graph, using graphics parameters, case studies.

V. TEXT BOOKS:
1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 1stEdition, 2014.
2. William N. Venables, David M. Smith, "An Introduction to R", Network Theory imited, 2ndEdition, 2009.
3. Stephen Marsland, "Machine Learning: An Algorithmic Perspective", Taylor & Francis CRCPress, 2nd Edition, 2011.

VI. REFERENCE BOOKS:
1. G. Jay Kerns, "Introduction to Probability and Statistics Using R", Youngstown State University,USA, 1st Edition, 2011.
2. William W Hsieh, "Machine Learning Methods in the Environmental Sciences", Neural Networks,Cambridge University Press, 1st Edition, 2009.
3. Chris Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, 1st Edition,1995.
4. Peter Flach, "Machine Learning", Cambridge University Press, 1st Edition, 2012.

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

VIII. E-TEXT BOOKS:
2.   https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
3.   https://www.cs.bris.ac.uk/~flach/mlbook/.
4.   http://mylovelibrabry.com/emylibraryus/free.php?asin=1466583282.