DATA SCIENCE

I Semester: CSE

Course Code	Category	Hours / Week		Credits	Maximum Marks			
		L	Т	Р	С	CIA	SEE	Total
BCSB06	Elective	3	0	0	3	30	70	100
Contact Classes: 45	Total Tutoria	als: Nil	I Total Practical C		lasses: Nil	Total Classes: 45		es: 45

I. COURSE OVERVIEW:

The course offers comprehensive coverage of key topics including R programming, SQL, NoSQL, and data analysis. It equips students with the necessary skills to document and transfer their research findings while effectively communicating them through visualization techniques.

II.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. To apply hypotheses and data into actionable predictions.
- IV. The 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:

CO 1	Make use of various data description functions in for exhibiting various stages of the data science process.	Understand
CO 2	Identify interfacing packages for handling SQL and NoSQL databases for performing data analysis	Remember
CO 3	Evaluate models using clustering and classification techniques.	Evaluate
CO 4	Solve various real-time problems on various hypothesis conditions by using artificial neural networks.	Apply
CO 5	CO 5 Illustrate delivering results through documentation and visualization techniques	

IV. SYLLABUS:

UNIT-I INTRODUCTION

Classes: 10

Data science process, roles, stages in data science project, working with data from files, working with relational databases, exploring data, managing data, cleaning and sampling for modeling; Introduction to R: Introduction to various data types, numeric, character, date, data frame, array, matrix etc., reading and writing datasets, working with different file types .txt, .csv, outliers, R functions and loops; Summary statistics: Summary, str, aggregate, subset, head, tail; Probability distribution.

|--|

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.

UNIT-IIIDATA MODELSClasses: 08

Choosing and evaluating models, mapping problems to machine learning, evaluating clustering models, validating models.

Cluster analysis: K-means algorithm, Naive Bayes memorization methods, unsupervised methods.

UNIT-IV	ARTIFICIAL NEURAL NETWORKS	Classes: 09
---------	----------------------------	-------------

Artificial neural networks: Introduction, neural network representation, appropriate problems for neural network learning, perceptions, multilayer networks and the back propagation algorithm, remarks on 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 hypotheses, comparing learning algorithms.

|--|

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.

Classes: 08

Text Books:

- 1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 1st Edition, 2014.
- William N. Venables, David M. Smith, "An Introduction to R", Network Theory Limited, 2nd Edition, 2009.
- 3. Stephen Marsland, "Machine Learning: An Algorithmic Perspective", Taylor & Francis CRC Press, 2nd Edition, 2011.

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.

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.

E-Text Books:

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