

DATA SCIENCE LABORATORY

I Semester: CSE								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
BCSB10	CORE	L	T	P	C	CIA	SEE	Total
		-	-	3	2	30	70	100
Contact Classes: 36	Tutorial Classes: Nil	Practical Classes: Nil			Total Classes: 36			

COURSE OBJECTIVES:
The course should enable the students to:

- I. Illustrate R objects.
- II. Make use of different types of datasets for analysis in R
- III. Define relations among variables using correlation and covariance analysis.
- IV. Analyze and differentiate the data models for predictions using R.

COURSE OUTCOMES (COs):

CO 1: Explore methods that implements neural network techniques.

CO 2: Practice the fuzzy set relations using different operations.

CO 3: Design Regression techniques for a set of data points.

CO 4: Capture an appropriate classification model for analytical tasks.

CO 5: Explore methods that implements neural network techniques.

COURSE LEARNING OUTCOMES(CLOs):

CLO 1: Understand and develop relevant programming abilities

CLO 2: Understand and intuition of the whole process line of extracting knowledge from data

CLO 3: Equip with the fundamental knowledge on basics of data science and R programming

CLO 4: Critically analyze and evaluate variety of NoSQL databases.

CLO 5: Develop the ability to build and assess Data-based models.

CLO 6: Analyze data analysis and make models using regression analysis

CLO 7: Familiarize with variety of machine learning tasks: clustering, dimensionality reduction, regression and classification

CLO 8: Understand how to formalize practical problems using methods of machine learning

CLO 9: Understand neural networks techniques solve real time problems

CLO 10: Understand the different learning algorithms

CLO 11: Chose a appropriate learning Algorithms to solve particular problems

CLO 12: Based on delivering results make a documentation for various results sets

CLO 13: Understand how to plot graphs for multivariate and matrix data

WEEK -1	R AS CALCULATOR APPLICATION
a) Using with and without R objects on console b) Using mathematical functions on console c) Write an R script, to create R objects for calculator application and save in a specified location in disk.	
WEEK -2	DESCRIPTIVE STATISTICS IN R
a) Write an R script to find basic descriptive statistics using summary, str, quartile function on mtcars& cars datasets. b) Write an R script to find subset of dataset by using subset (), aggregate () functions on iris dataset.	
WEEK -3	READING AND WRITING DIFFERENT TYPES OF DATASETS
a) Reading different types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location. b) Reading Excel data sheet in R. c) Reading XML dataset in R.	
WEEK -4	VISUALIZATIONS
a. Find the data distributions using box and scatter plot. b. Find the outliers using plot. c. Plot the histogram, bar chart and pie chart on sample data.	
WEEK -5	CORRELATION AND COVARIANCE
a. Find the correlation matrix. b. Plot the correlation plot on dataset and visualize giving an overview of relationships among data on iris data. c. Analysis of covariance: variance (ANOVA), if data have categorical variables on iris data.	
WEEK -6	REGRESSION MODEL
Import a data from web storage. Name the dataset and now do Logistic Regression to find out relation between variables that are affecting the admission of a student in a institute based on his or her GRE score, GPA obtained and rank of the student. Also check the model is fit or not. require (foreign), require(MASS).	
WEEK-7	MULTIPLE REGRESSION MODEL
Apply multiple regressions, if data have a continuous independent variable. Apply on above dataset.	
WEEK-8	REGRESSION MODEL FOR PREDICTION
Apply regression Model techniques to predict the data on above dataset.	
WEEK-9	CLASSIFICATION MODEL
a. Install relevant package for classification. b. Choose classifier for classification problem. c. Evaluate the performance of classifier.	

WEEK-10	CLUSTERING MODEL
a) Clustering algorithms for unsupervised classification. b) Plot the cluster data using R visualizations.	
Reference Books:	
Yanchang Zhao, “R and Data Mining: Examples and Case Studies”, Elsevier, 1 st Edition, 2012.	
Web References:	
1. http://www.r-bloggers.com/how-to-perform-a-logistic-regression-in-r/ 2. http://www.ats.ucla.edu/stat/r/dae/rreg.htm 3. http://www.coastal.edu/kingw/statistics/R-tutorials/logistic.html 4. http://www.ats.ucla.edu/stat/r/data/binary.csv	
SOFTWARE AND HARDWARE REQUIREMENTS FOR A BATCH OF 18 STUDENTS:	
HARDWARE: 18 numbers of Intel Desktop Computers with 2 GB RAM.	
SOFTWARE: Turbo C/ J2SE	