

DATA SCIENCE LABORATORY

I Semester: CSE

Course Code	Category	Hours / Week			Credits	Maximum Marks		
BCSB10	Core	L	T	P	C	CIA	SEE	Total
		0	0	4	2	30	70	100
Contact Classes: Nil	Total Tutorials: Nil	Total Practical Classes: 36				Total Classes: 36		

I. COURSE OVERVIEW:

The data science laboratory provides hands-on experience with various concepts and techniques in the field of data science. Students will learn to utilize R as a powerful calculator application, enabling them to perform calculations, data manipulations, and statistical analyses efficiently. The laboratory also focuses on reading and writing different types of datasets, allowing students to work with diverse data sources and formats.

II. COURSE OBJECTIVES:

The students will try to learn:

- 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.

III. COURSE OUTCOMES:

After successful completion of the course, students should be able to:

CO 1	Make use of the container Data types for displaying the functional values.	Remember
CO 2	Demonstrate the reading and writing operations from the web, and disk.	Understand
CO 3	Illustrate the nature and relationships of data with effective visualizations for exploring the data.	Analyze
CO 4	Analyze the linear data models in data exploration.	Analyze
CO 5	Develop the classification and clustering models to label the data.	Create

IV. SYLLABUS:

LIST OF EXPERIMENTS

Week-1	R AS CALCULATOR APPLICATION
<ul style="list-style-type: none"> 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
<ul style="list-style-type: none"> 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
	<ul style="list-style-type: none"> 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
	<ul style="list-style-type: none"> 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
	<ul style="list-style-type: none"> 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
	<ul style="list-style-type: none"> a. Install relevant package for classification. b. Choose classifier for classification problem. c. Evaluate the performance of classifier.
Week-10	CLUSTERING MODEL
	<ul style="list-style-type: none"> 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 18 STUDENTS:

SOFTWARE: R Software , R Studio Software

HARDWARE: 18 numbers of Intel Desktop Computers with 4 GB RAM