

DATA SCIENCE LABORATORY

I Semester: CSE																							
Course Code	Category	Hours / Week			Credits	Maximum Marks																	
		L	T	P	C	CIA	SEE	Total															
BCSC12	Core	0	0	4	2	30	70	100															
Contact Classes: Nil		Total Tutorials: Nil		Total Practical Classes: 36			Total Classes: 36																
<p>I. COURSE OVERVIEW: The Course is designed in various levels of data and use of R for statistical programming, analysis and visualization of processed data. It includes the basics of mathematics, probability and statistical methods and data communication. The techniques are used in Distribution and data analysis in non- parametric statistics such as image processing, e-commerce and banking.</p> <p>II. COURSE OBJECTIVES: The students will try to learn: I. The concept of R objects. II. Make use of different types of datasets for analysis using R. III. Relations among variables using statistical analysis. IV. The different data models for predictions using R</p> <p>III. COURSE OUTCOMES: After successful completion of the course, students will be able to:</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tbody> <tr> <td style="width: 10%;">CO1</td> <td style="width: 70%;">Make Use of the container Data types for display the functional values.</td> <td style="width: 20%;">Apply</td> </tr> <tr> <td>CO2</td> <td>Demonstrate the reading and writing operations from the web, and disk.</td> <td>Apply</td> </tr> <tr> <td>CO3</td> <td>Illustrate the nature and relationships of data with effective visualizations for exploring the data.</td> <td>Apply</td> </tr> <tr> <td>CO4</td> <td>Analyze the linear data models in data exploration.</td> <td>Understand</td> </tr> <tr> <td>CO5</td> <td>Develop the classification and clustering models to label the data</td> <td>Analyze</td> </tr> </tbody> </table>									CO1	Make Use of the container Data types for display the functional values.	Apply	CO2	Demonstrate the reading and writing operations from the web, and disk.	Apply	CO3	Illustrate the nature and relationships of data with effective visualizations for exploring the data.	Apply	CO4	Analyze the linear data models in data exploration.	Understand	CO5	Develop the classification and clustering models to label the data	Analyze
CO1	Make Use of the container Data types for display the functional values.	Apply																					
CO2	Demonstrate the reading and writing operations from the web, and disk.	Apply																					
CO3	Illustrate the nature and relationships of data with effective visualizations for exploring the data.	Apply																					
CO4	Analyze the linear data models in data exploration.	Understand																					
CO5	Develop the classification and clustering models to label the data	Analyze																					
<p>IV. SYLLABUS</p> <p>Week-1: R AS CALCULATOR APPLICATION</p> <ol style="list-style-type: none"> Using with and without R objects on console Using mathematical functions on console Write an R script, to create R objects for calculator application and save in a specified location in disk <p>Week-2: DESCRIPTIVE STATISTICS IN R</p> <ol style="list-style-type: none"> Write an R script to find basic descriptive statistics using summary, str, quartile function on mtcars& cars data sets. Write an R script to find subset of dataset by using subset (), aggregate () functions on iris dataset. <p>Week-3: READING AND WRITING DIFFERENT TYPES OF DATASETS</p> <ol style="list-style-type: none"> Reading different types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location. Reading Excel data sheet in R. 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.

V. REFERENCE BOOKS:

1. Yanchang Zhao, "R and Data Mining: Examples and Case Studies", Elsevier, 1st Edition, 2012.

VI. 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>