INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad -500 043

COMPUTER SCIENCE AND ENGINEERING

COURSE DESCRIPTOR

Course Title	**DATA SCIENCE LABORATORY**				
Course Code	BCSB10				
Programme	M.Tech				
Semester	I	CSE			
Course Type	LAB				
Regulation	IARE - R18				
Course Structure	**Theory**			**Practical**	
	Lectures	**Tutorials**	**Credits**	**Laboratory**	**Credits**
	-	-	-	3	2
Course Faculty	Dr. R Obulakonda Reddy, Associate Professor CSE				

I. COURSE OVERVIEW:

The course introduces the concepts of R Programming Language. Moreover the course pays a special attention to solve typical uncertainty problems which are primarily explored by R Programming concepts.

II. COURSE PRE-REQUISITES:

Level	Course Code	Semester	Prerequisites	Credits
PG	-	-	-	-

III. MARKS DISTRIBUTION:

Subject	SEE Examination	CIA Examination	Total Marks
Dist Science	70 Marks	30 Marks	100

IV. DELIVERY / INSTRUCTIONAL METHODOLOGIES:

✔	LCD / PPT	✔	Student viva	✔	Mini Project	✘	Videos
✔	Open Ended Experiments						

V. EVALUATION METHODOLOGY:

Each laboratory will be evaluated for a total of 100 marks consisting of 30 marks for internal assessment and 70 marks for semester end lab examination. Out of 30 marks of internal assessment, continuous lab assessment will be done for 20 marks for the day to day performance and 10 marks for the final internal lab assessment.

Semester End Examination (SEE): The semester end lab examination for 70 marks shall be conducted by two examiners, one of them being Internal Examiner and the other being External Examiner, both nominated by the Principal from the panel of experts recommended by Chairman, BOS. The emphasis on the experiments is broadly based on the following criteria:

20 %	To test the preparedness for the experiment.
20 %	To test the performance in the laboratory.
20 %	To test the calculations and graphs related to the concern experiment.
20 %	To test the results and the error analysis of the experiment.
20 %	To test the subject knowledge through viva – voce.

Continuous Internal Assessment (CIA):
CIA is conducted for a total of 30 marks (Table 1), with 20 marks for continuous lab assessment during day to day performance, 10 marks for final internal lab assessment.

Table 1: Assessment pattern for CIA

Component	**Laboratory**		**Total Marks**
Type of Assessment	Day to Day Performance	Final internal lab assessment	
CIA Marks	20	10	30

Continuous Internal Examination (CIE):
One CIE exams shall be conducted at the end of the 16th week of the semester. The CIE exam is conducted for 10 marks of 3 hours duration.

Preparation	**Performance**	**Calculations and Graph**	**Results and Error Analysis**	**Viva**	**Total**
2	2	2	2	2	10

VI. HOW PROGRAM OUTCOMES ARE ASSESSED:

Program Outcomes (POs)		**Strength**	**Proficiency assessed by**
PO 1	An ability to analyze a problem, and to identify and define the computing requirements appropriate to its solutions.	3	Laboratory practices, student viva
PO 2	Solve complex heterogeneous data intensive analytical based problems of real time scenario using state of the art hardware/software tools	3	Laboratory practices, student viva
PO 3	Demonstrate a degree of mastery in emerging areas of CSE/IT like IoT, AI, Data Analytics, Machine Learning, cyber security, etc.	3	Laboratory practices, Mini project

3 = High; 2 = Medium; 1 = Low

VII. COURSE OBJECTIVES(COs):

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.

VIII. COURSE OUTCOMES(COs):

CO Code	COs	At the end of the course, the student will have the ability to	PO's Mapped	Strength of Mapping
BCS001.01	CO 1	Understand the process and different stages of data science and relevant data descriptions in R.	PO 1,PO2	2
BCS001.02	CO 2	Illustrate various SQL, NOSQL databases connecting with R and perform correlation and regression analysis	PO 1, PO 2	2
BCS001.03	CO 3	Evaluate different data models and perform clustering analysis.	PO 1, PO 2	2
BCS001.04	CO 4	Solve various real time problems using artificial neural networks techniques and comparing different learning algorithms.	PO2, PO3	2
BCS001.05	CO 5	Explore on various ways to deliver results through and plots of multivariate data and matrix data	PO 2,PO3	2

IX. MAPPING COURSE OUTCOMES LEADING TO THE ACHIEVEMENT OF PROGRAM OUTCOMES

Course Outcomes (COs)	Program Outcomes						
	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO 1	2	2					
CO 2	2	2	1				
CO 3			2		1		
CO 4			1		1	1	
CO 5			1		1	1	

3 = High; 2 = Medium; 1 = Low

X. ASSESSMENT METHODOLOGIES –DIRECT

CIE Exams	PO 1, PO 2, PO 3	SEE Exams	PO 1, PO 2, PO 3	Laboratory Practices	PO 1, PO 2, PO 3	Student Viva	PO 1, PO2, PO5
Mini Project	PO7						

XI. ASSESSMENT METHODOLOGIES -INDIRECT

✔	Early Semester Feedback	✔	End Semester OBE Feedback
✘	Assessment of Mini Projects by Experts		

XII. SYLLABUS

LIST OF EXPERIMENTS	
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, 1st Edition, 2012.	
Web References:	
1. www.cs.put.poznan.pl/pawelw/sus/dcs07.doc 2. https://developer.apple.com/library/mac/documentation	
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	

XIII. COURSE PLAN:

The course plan is meant as a guideline. Probably there may be changes.

Week No.	Topics to be covered	Learning Outcomes
1	R as Calculator Application	CO1
2	Descriptive Statistics in R	CO1
3	Reading and Writing Different Types of Datasets	CO3
4	Visualizations	CO1
5	Correlation and Covariance	CO2
6	Regression Model	CO2
7	Multiple Regresstion Model	CO3
8	Regression Model for Prediction	CO1
9	Classification Model	CO3
10	Clustering Model	CO1

Prepared by:
Dr. R Obulakonda Reddy, CSE **HOD, CSE**