

Course Code	Foundations of Data Analytics	L	T	P	J	C
CSE3505		2	0	2	4	4
Pre-requisite	NIL	Syllabus version				
		v.1.0				
Course Objectives:						
<ol style="list-style-type: none"> 1. To establish clearly the objectives and scope of the predictive analysis 2. Use R programming language to identify suitable data sources to agree the methodological approach 3. Validate and review data accurately and identify anomalies 4. To appreciate the current trends in data analysis procedure 5. Carry out rule-based analysis of the data in line with the analysis plan 6. Apply statistical models to perform Regression Analysis, Clustering and Classification 7. Present the results and inferences from your analysis using R tool 8. To improve document management and team work 						
Expected Course Outcome:						
Students will be able to:						
<ol style="list-style-type: none"> 1. Understand R with Business Intelligence, Business Analytics, Data and Information 2. Contextually integrate and correlate information automatically to gain faster insights 3. Implement statistical analysis techniques for solving practical problems. 4. Graphically interpret data and Find a meaningful pattern in data 5. Perform statistical analysis on variety of data. 						
Module:1	Introduction to Analytics	4 hours				
Analytics life cycle - Business analytics - lending analytics- recommendation analytics- Healthcare Analytics- financial analytics - sports analytics						
Module:2	R programming Basics	5 hours				
Introduction to R, R Studio (GUI): R Windows Environment, introduction to various data types, Numeric, Character, date, data frame, array, matrix etc.,						
Module:3	Working with datasets and files:	6 hours				
Reading Datasets, Working with different file types .txt,.csv , R studio, Files, Datasets, Extracting Datasets, Preparing datasets. Data Cleaning, Data imputation, Data conversion Analysis						
Module:4	Introduction to statistical learning and R-Programming	6 hours				
Basic statistics: mean, median, standard deviation, variance, correlation, covariance - Outliers, Combining Datasets in R, Functions and loops. Summary Statistics - Summarizing data with R - Correlation and Regression						
Module:5	Document Creation and Knowledge Sharing:	3 hours				
Access existing documents, language standards, templates and documentation tools from their organization's knowledge base. Confirm the content and structure of the documents with appropriate people, Create documents using standard templates and agreed language standards. Review documents with appropriate people and incorporate their inputs						
Module:6	Self and work Management:	3 hours				
Establish and agree their work requirements with appropriate people - Keep their immediate work area clean and tidy - utilize their time effectively - Use resources correctly and efficiently - Treat confidential information correctly - Work in line with organization's policies and						

procedures - Work within the limits of their job role		
Module:7	Team Work and Communication	3 hours
Communicate with colleagues clearly, concisely and accurately - Work with colleagues to integrate their work effectively with them - Pass on essential information to colleagues in line with organizational requirements - Work in ways that show respect for colleagues - carry out commitments they have made to colleagues - Let colleagues know in good time if they cannot carry out their commitments, explaining the reasons - Identify any problems they have working with colleagues and take the initiative to solve these problems		
Total Lecture hours		30 hours
Text Book(s)		
1.	Trevor Hastie and Rob Tibshirani, “An Introduction to Statistical Learning with Applications in R”, Springer, 2017.	
2.	Mark van der Loo, Edwin de Jonge, “Learning R Studio for R Statistical Computing”, Packt Publishing, 2012.	
3.	Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. “Mining of Massive Datasets”. Cambridge University Press. 2014.	
Reference Books		
1.	Hadley Wickham and Garrett Grolemund, “R for Data Science: Import, Tidy, Transform, Visualize, and Model Data”, O’Reilly, 2017.	
2.	Grolemund, Garrett. “Hands-on programming with R”, O’ Reilly Media, Inc., 2014.	
3.	Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze, “Introduction to Information Retrieval”, Cambridge University Press, First South Asian Edition, 2008.	
4.	Trevor Hastie, Robert Tibshirani, Jerome Friedman, “The Elements of Statistical Learning”, Springer, Second Edition, 2011.	
5.	https://www.sscnasscom.com/qualification-pack/SSC/Q2101/	
Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar		
List of Challenging Experiments (Indicative)		
1.	Understanding of R System and installation and configuration of R-Environment and R-Studio, Understanding R Packages, their installation and management	3 hours
2.	Understanding of nuts and bolts of R: a. R program Structure b. R Data Type, Command Syntax and Control Structures c. File Operations in R	3 hours
3.	Dataframes and lists	3 hours
4.	Excel and R integration with R connector.	3 hours
5.	Preparing Data in R a. Data Cleaning b. Data imputation c. Data conversion	3 hours
6.	Manipulating Matrices in R	3 hours
7.	Outliers detection using R	3 hours
8.	Correlation and N-Fold cross validation in R	3 hours
9.	Debugging and Program Efficiency in R	3 hours
10.	Visualizing data using R with different type of graphs and charts	3 hours
Total Laboratory Hours		30 hours

Mode of assessment: Assessment Examination, FAT Lab Examination			
Recommended by Board of Studies	08-02-2020		
Approved by Academic Council	No. 58	Date	26-02-2020