

Course Code	UCA20D05J	Course Name	DATA ANALYSIS USING R	Course Category	D	Discipline Specific Elective Course	L	T	P	C
							4	0	4	6

Pre-requisite Courses	Nil	Co-requisite Courses	Nil	Progressive Courses	Nil
Course Offering Department	Computer Applications	Data Book / Codes/Standards	Nil		

Course Learning Rationale (CLR):	The purpose of learning this course is to,	Learning	Program Learning Outcomes (PLO)
----------------------------------	--	----------	---------------------------------

CLR-1 :	Learn in-depth concepts, methods and applications of data mining	1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CLR-2 :	Learn how to start looking at data from the perspective of the data scientist																		
CLR-3 :	Experimenting with different data mining techniques for knowledge discovery																		
CLR-4 :	Use R software for data import and export, data exploration and visualization, and for data analysis tasks																		
CLR-5 :	Demonstration on how to perform classification and clustering data mining tasks on real time datasets																		
CLR-6 :	Build an effective model and perform model evaluation based on performance metrics																		

Course Learning Outcomes (CLO):	At the end of this course, learners will be able to:	Level of Thinking (Bloom)	Expected Proficiency (%)	Expected Attainment (%)	Fundamental Knowledge	Application of Concepts	Link with Related Disciplines	Procedural Knowledge	Skills in Specialization	Ability to Utilize Knowledge	Skills in Modeling	Analyze, Interpret Data	Investigative Skills	Problem Solving Skills	Communication Skills	Analytical Skills	ICT Skills	Professional Behavior	Life Long Learning
CLO-1 :	Understand Data Mining and its various tasks	2	85	80	L	H	H	H	H	-	-	M	M	L	-	H	-	-	-
CLO-2 :	Perform Linear Regression and Logistic Regression on a dataset	3	85	80	L	H	H	H	H	-	-	M	M	L	-	H	-	-	-
CLO-3 :	Extract interesting rules from dataset using Association Rule Mining	3	85	80	L	H	H	H	H	-	-	M	M	L	-	H	-	-	-
CLO-4 :	Apply various Classification Algorithms in data mining	3	85	80	L	H	H	H	H	-	-	M	M	L	-	H	-	-	-
CLO-5 :	Perform unsupervised learning using various Clustering Techniques	3	85	80	L	H	H	H	H	-	-	M	M	L	-	H	-	-	-
CLO-6 :	Effectively use R programming constructs and packages to perform mining on different datasets	3	85	80	L	H	H	H	H	-	-	M	M	L	-	H	-	-	-

Duration (hour)	24	24	24	24	24
S-1	SLO-1 Why Data Mining? Evolution of Information Technology	Simple Linear Regression	Introduction to Association Rules	Basic Concepts-What is Classification?	What is Cluster Analysis?
	SLO-2 What is data mining?	Estimating the Coefficients	What is an Association Rule?	General Approach to Classification	Overview of Basic Clustering Approaches
S-2	SLO-1 What is Data Warehouse?	Assessing the accuracy of the Coefficient estimate	Methods to discover Association Rules	Decision Tree Induction	Requirements for Cluster Analysis
	SLO-2 DW Architecture	Assessing the accuracy of the Model	Market Basket Analysis: A motivational Example	Attribute Selection Methods	Clustering Paradigms
S-3	SLO-1 OLAP	Multiple Linear regression	Basic Concepts: Frequent Item Sets, Closed Item Sets	Decision Tree Classifier- Using Entropy	Similarity and Distance Metrics
	SLO-2 Warehouse Schema	Multiple Linear Regression-Continued	Frequent Item Set Mining- Apriori Algorithm	Decision Tree Classifier- Using Gini Index	Characteristics of Clustering Algorithms

S-4	SLO-1	Data Mining Goals	Estimating the Regression Coefficients	Candidate Generation	Tree Pruning Techniques	Partitioning Algorithms
	SLO-2	KDD Vs Data Mining	Other Considerations in Regression Model	Generating Frequent Item Sets-Example	Rule Based Classifier	k-Means Clustering- Centroid Based Technique
S 5-8	SLO-1	Lab1: Installing R and RStudio, RStudio	Lab: 4 Package RODBC to ODBC connect databases	Lab 7: Association rule mining using APRIORI algorithm for titanic dataset	Lab 10: Build a decision tree for the iris data with package party, rpart	Lab 13:k-means clustering of iris data
	SLO-2	Using Functions e	Analyze correlation between variables in Australian CPI dataset.			
S-9	SLO-1	Steps in KDD Process	Multiple Linear regression	Apriori Algorithm-Examples	Using IF-THEN rules for Classification	Partitioning Method: k-Means Clustering-Example
	SLO-2	What kind of data can be mined?	Estimating the Regression Coefficients	Improving the Efficiency of Apriori	Rule Extraction from Decision Tree	k-Medoids Algorithms
S-10	SLO-1	Data Mining Techniques	K-Nearest Neighbour	A Pattern Growth Approach for Mining Frequent Item Sets	Bayes Classification Methods	Example: k-Medoids
	SLO-2	Data Mining Techniques - Continued	K-NN Demonstration with example	FP-Growth Algorithm – Basic Idea	Bayes Theorem	Hierarchical Methods
S-11	SLO-1	What kind of patterns can be mined?	Compare LR with k-NN	FP-Growth Algorithm with example	Naïve Bayes Classifier	Agglomerative and Divisive Methods
	SLO-2	Are all patterns interesting?	Evaluation for regression	FP-Growth Algorithm –Examples	Naïve Bayes –Example	Distance Measures in Algorithmic Methods
S-12	SLO-1	Data Objects and Attribute Types	Model selection and over-fitting	Mining Closed and Max Patterns	Model Evaluation and Selection	Dendrogram
	SLO-2	Nominal, Binary, Ordinal, Numeric, Discrete data types	Logistic regression	Mining Frequent Item Sets Using Vertical Data format	Techniques to improve Classification Accuracy	Density Based Methods: DBSCAN
S 13-16	SLO-1	Lab2: Overview -Working in the Console	Lab 5: Prediction with Simple Linear Regression	Lab 8: Inspection of Association rules	Lab 11: Classification using Naïve Bayes classifier	Lab 14: Implementation of DBSCAN algorithm using iris data
	SLO-2	• Arithmetic Operators • Logical Operations				
S-17	SLO-1	What technologies are used?	The Logistic Model	Which Patterns are Interesting?	Ensemble Techniques-Bagging	DBScan Algorithm
	SLO-2	Data exploration and visualization	Estimating the Regression Coefficients	Strong Rules are not Necessarily Interesting	Boosting	Evaluation of Clustering
S-18	SLO-1	Data Visualization Techniques	Making Predictions	Interestingness Measure-Support, Confidence	Artificial Neural Network-Introduction	Measuring Clustering Quality
	SLO-2	Visualizing Complex data and relations	Multiple Logistic Regression	Generalization of Association Patterns	Defining the Network Topology	Determining the number of Clusters
S-19	SLO-1	Issues and Challenges in Data mining	Linear Discriminant Analysis	Pattern Evaluation Methods	Multi Layer Feed Forward Neural Network	Scalable Clustering Algorithms
	SLO-2	Applications of Data Mining	LDA Continued	From Association analysis to correlation analysis	Back Propagation	Applications of Data Mining in Finance
S-20	SLO-1	Case Study: Weather Data	LDA for p=1	Comparison of Pattern Evaluation Measures	Inside the Black Box- Back Propagation	Applications of Data Mining in Business
	SLO-2	Case Study – Discussion	LDA for p>1	Discussion on Different Algorithms	Neural Network –Perceptron Example	Applications of Data Mining in Social Networks
S 21-	SLO-1	Lab:3	Lab 6: Predict the probability of occurrence with logistic regression.	Lab 9: Visualize association rules, including scatter plot, balloon plot,	Lab 12: Classification using Perceptron	Lab 15: Demonstrate hierarchical clustering on iris data
	SLO-2	Getting Help in R and Quitting				

24	RStudio	graph	Stock Market Prediction
----	---------	-------	-------------------------

Learning Resources	1. Arun K Pujari, "Data Mining Techniques", Univeristy Press 2. J Han and M Kamber, "Data Mining : Concepts and Techniques", Third Edition, Morgan Kaufmann Publishers 3. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani,(2013), "An Introduction to Statistical Learning with Applications in R", Springer 4. Yanchang Zhao, "R and Data Mining: Examples and Case Studies" yanchang@rdatamining.com , http://www.RDataMining.com
--------------------	---

Learning Assessment											
Level	Bloom's Level of Thinking	Continuous Learning Assessment (50% weightage)								Final Examination (50% weightage)	
		CLA – 1 (10%)		CLA – 2 (10%)		CLA – 3 (20%)		CLA – 4 (10%)#			
		Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice
Level 1	Remember	20%	20%	15%	15%	15%	15%	15%	15%	15%	15%
	Understand										
Level 2	Apply	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
	Analyze										
Level 3	Evaluate	10%	10%	15%	15%	15%	15%	15%	15%	15%	15%
	Create										
	Total	100 %		100 %		100 %		100 %		100 %	

CLA – 4 can be from any combination of these: Assignments, Seminars, Tech Talks, Mini-Projects, Case-Studies, Self-Study, MOOCs, Certifications, Conf. Paper etc.,

Course Designers		
Experts from Industry	Experts from Higher Technical Institutions	Internal Experts
Mr.G.Muruganandam, Group Project Manager, HCL Technologies, Chennai	Dr.S.Gopinathan, Professor, University of Madras, Chennai	1. Mrs.Anita Jasmine, SRM IST
Mr.M. Hemachandar, Tech Lead, Wipro Limited, Chennai		2. Mrs. Ramla, SRM IST