

<b>Course Code</b>	21CSE355T	<b>Course Name</b>	DATA MINING AND ANALYTICS	<b>Course Category</b>	E	PROFESSIONAL ELECTIVE	L	T	P	C
							2	1	0	3

<b>Pre-requisite Courses</b>	Nil	<b>Co-requisite Courses</b>	Nil	<b>Progressive Courses</b>	Nil
<b>Course Offering Department</b>	School of Computing	<b>Data Book / Codes / Standards</b>	Nil		

Course Learning Rationale (CLR):		The purpose of learning this course is to:											
CLR-1:	introduce the basic concepts of pattern discovery and data preparation												
CLR-2:	understand the importance of Association and Correlation Algorithms												
CLR-3:	comprehend and apply various Classifiers												
CLR-4:	work with the foundation for Clustering												
CLR-5:	perform Outlier Analysis and Explore a data mining tool												
Course Outcomes (CO):		At the end of this course, learners will be able to:											
CO-1:	do the preprocessing of data before mining of data for patterns												
CO-2:	make use of Association and Correlations Algorithms for framing association rules												
CO-3:	apply as well as Compare the performance of various classifiers												
CO-4:	utilize different Clustering algorithms for generalization												
CO-5:	identify Outliers in the data given												

Program Outcomes (PO)														Program Specific Outcomes		
1	2	3	4	5	6	7	8	9	10	11	12					
Engineering Knowledge	Problem Analysis	Design/development of solutions	Conduct investigations of complex problems	Modern Tool Usage	The engineer and society	Environment & Sustainability	Ethics	Individual & Team Work	Communication	Project Mgt. & Finance	Life Long Learning	PSO-1	PSO-2	PSO-3		
1	2	-	-	-	-	-	-	-	-	-	-	2	-	-		
1	2	-	-	3	-	-	-	-	-	-	-	2	-	-		
1	2	-	-	3	-	-	-	-	-	-	-	2	-	-		
1	-	-	-	3	-	-	-	-	-	-	-	2	-	-		
1	2	-	-	3	-	-	-	-	-	-	-	2	-	-		

<b>Unit-1 - Data Mining Introduction</b>	<b>9 Hour</b>
Introduction: Kinds of Data- Kinds of Patterns-Data Objects and Attribute Type- Data Visualization -Data Preprocessing: Data cleaning, Data Integration, Data Transformation, Data Discretization and Data Reduction: Attribute Subset Selection-Histograms, Clustering, Sampling	
<b>Unit-2 - Associations and Correlations</b>	<b>9 Hour</b>
Market Basket Analysis – Apriori Algorithm – Mining Frequent Itemsets without Candidate Generation – Mining Frequent Itemsets Using Vertical Data Format – Mining Closed Frequent Itemsets – Mining Multilevel Association Rules – Mining Multidimensional Association Rules – Correlation Analysis – Constraint-Based Association Mining	
<b>Unit-3 - Classification and Prediction</b>	<b>9 Hour</b>
Basic Concepts- Decision Tree Induction-Attribute selection Measures-ID3 and CART algorithms, Tree Pruning-Bayes Classification Methods: Bayes’ Theorem, Naive Bayesian Classification - Classification by Backpropagation- Support Vector Machines-Lazy learners: KNN-Metrics for evaluating classifier performance-Techniques to improve classification accuracy-Prediction: Regression Analysis	
<b>Unit-4 - Cluster Analysis</b>	<b>9 Hour</b>
Cluster Analysis: Partitioning Methods- Hierarchical Methods: Agglomerative versus Divisive Hierarchical Clustering-Probabilistic Model based Clustering - BIRCH, DBSCAN, STING, CLIQUE Techniques- Evaluation of clustering Techniques	
<b>Unit-5 - Outliers and Statistical Approaches in Data Mining</b>	<b>9 Hour</b>
Introduction to outliers, Challenges in detecting Outliers, Outlier Detection Methods - Supervised, Semisupervised, Unsupervised- Statistical Data Mining approaches - Data mining in Recommender Systems, Data mining for Intrusion Detection, Data Mining for Financial Analysis	

<b>Learning Resources</b>	1. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2012 2. Ian H. Witten, Eibe Frank and Mark A. Hall "Data Mining: Practical Machine Learning Tools and Techniques", Fourth Edition, Elsevier, 2017.
---------------------------	---

Learning Assessment							
	Bloom's Level of Thinking	Continuous Learning Assessment (CLA)				Summative Final Examination (40% weightage)	
		Formative CLA-1 Average of unit test (50%)		Life-Long Learning CLA-2 (10%)			
		Theory	Practice	Theory	Practice	Theory	Practice
Level 1	Remember	15%	-	15%	-	15%	-
Level 2	Understand	25%	-	20%	-	25%	-
Level 3	Apply	30%	-	25%	-	30%	-
Level 4	Analyze	30%	-	25%	-	30%	-
Level 5	Evaluate	-	-	10%	-	-	-
Level 6	Create	-	-	5%	-	-	-
	Total	100 %		100 %		100 %	

<b>Course Designers</b>		
Experts from Industry	Experts from Higher Technical Institutions	Internal Experts
Expert Member from TCS		
1. Dr. T. Ruso, Senior Project Lead, HCL Technologies, Chennai	1. Khanna Nehemiah, Associate Professor, Anna University Chennai	