Course Code	PCS21S02J	Course Name	DATA MINING AND DATA WAREHOUSING		Course Category	S	Skill Enhancement Course	L	T	P	C
								1	0	2	2

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

Course Learning Rationale (CLR):	The purpose of learning this course is to:		Learning			Program Learning Outcomes (PLO)															
			1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CLR-1 :	Introduce the learners to the concept of data mining and warehousing		Level of Thinking (Bloom)	Expected Proficiency (%)	Expected Attainment (%)	Engineering Knowledge	Problem Analysis	Design & Development	Analysis, Design, Research	Modern Tool Usage	Society & Culture	Environment & Sustainability	Ethics	Individual & Team Work	Communication	Project Mgt. & Finance	Life Long Learning	PSO - 1	PSO - 2	PSO - 3
CLR-2 :	Learn the applications of Data Mining																			
CLR-3 :	Fundamentals of Classification and Clustering Techniques																			
CLR-4 :	Master data mining techniques in various applications like social, scientific and environmental context.																			
CLR-5 :	Develop skill in selecting the appropriate data mining algorithm for solving practical problems.																			
CLR-6 :	Understand the architecture of data warehouse, data marts, modeling of data in a data warehouse																			

Course Learning Outcomes (CLO):	At the end of this course, learners will be able to:																				
CLO-1 :	Understand the functionality and limitations of the various Data mining techniques		2	80	70	H	H	H	H	H										
CLO-2 :	Develop application oriented data mining models		3	85	75	H	H	H	H	H										
CLO-3 :	Have a clear understanding of the Learning methods		3	75	70	H	H	H	H	H										
CLO-4 :	Describe different methodologies used in data warehousing.		3	85	80	H	H	H	H	H										
CLO-5 :	Understand organization of data in a Data warehouse		3	85	75	H	H	H	H	H										
CLO-6 :	Build a basic data warehouse		3	80	70	H	H	H	H	H										

Duration (Hour)		9	9	9	9	9
S-1	SLO-1	Introduction to Data Mining, Data mining as the Evolution of information technology	Association Analysis-Market basket analysis, Methods of Frequent itemset mining, Apriori Algorithm	Classification Techniques-introduction, Decision tree induction Attribute selection Measures, Bayes' Classification method- Bayes theorem	CLUSTERING Analysis – Introduction, K- means method, k-mediods method	Introduction to Data Warehousing, Why have a separate data warehouse, Multi-tired architecture- Data warehouse models
	SLO-2	Data and large datasets, Kinds of Pattern in data mining, Technology used in data mining	Frequent Itemsets, closed itemsets, Association rules, Generating Association Rules from frequent Itemset, Pattern Growth Approach	-Supervised Vs Unsupervised classifications, Tree pruning, Scalability and decision tree induction	Overview of clustering methods, Hierarchical clustering, Agglomerative clustering	Operational database systems versus Data warehouses, Extraction, Transformation, Loading- Meta data repository Architecture of DW
S 2-3	SLO-1	Laboratory 1: Explore machine learning tool "WEKA" Downloading and/or installation of WEKA data mining toolkit	Laboratory 4: Perform data preprocessing tasks and Demonstrate performing association rule mining on data sets	Laboratory 7: - Demonstrate performing classification on data sets	Laboratory 10: Demonstrate performing clustering of data sets	Laboratory 13: Creation of a Data Warehouse. Kappa statistic.
	SLO-2					
S 4	SLO-1	Database systems, Machine learning, Web search Engines, Data Mining Issues in Mining methodology	Vertical Data Format, Pattern Evaluation methods, Pattern mining	Naive Bayesian Classification, Rule Extraction, Rule Induction	Probablistic Hierarchical clustering , Density based method	Multi dimensional data model-Data cube, Role of concept hierarchies

Duration (Hour)		9	9	9	9	9
	SLO-2	Data warehouse – Introduction, Kinds of applications-Business Intelligence, DM versus Knowledge Discovery in Databases	Mining of closed and Max Patterns, Mining Multi level associations, Mining multidimensional associations	Rule Based Classification : IF-Then Rules for classification, Confusion Matrix	Chameleon method, DBSCAN, OPTICS	Schemas for multidimensional data models, OLAP, operations, Querying multidimensional databases
S 5-6	SLO-1	Laboratory 2: Perform data preprocessing tasks	Laboratory5: Explore various options available in Weka for preprocessing data	Laboratory8: Explore various options available in Weka for preprocessing data	Laboratory 11: Load each dataset into Weka and run 1d3, J48 classification algorithm. Study the classifier output. Compute entropy values.	Laboratory 14: create a query based on multidimentional databases
	SLO-2					
S 7	SLO-1	User interaction, Efficiency and scalability,Diversity of data types	Mining quantitative association rules , Graph Mining-Frequent sub-graph mining	Precision , Classification by Back propagation	DENCLUE, Evaluation of clustering methods	Data warehouse design and uses, Data warehouse Implementations
	SLO-2	Data pre-processing, Overview of Applications of Data Mining, Data Objects and Attributes types	Mining rare patterns and negative patterns, Constraints based pattern generation	Recall, Support vector machine	Grid based clustering methods, Measuring cluster qualilty	DW design process, OLAP Server Architectures
S 8-9	SLO-1	Laboratory 3: Perform data preprocessing tasks and Demonstrate performing association rule mining on data sets	Laboratory6: Explore various options available in Weka for preprocessing data and apply unsupervised filters like Discretization, Resample filter, etc. on each dataset	Laboratory 9: Load each dataset into Weka and run Id3, J48 classification algorithm. Study the classifier output. Compute entropy values, Kappa statistic.	Laboratory 12: Load each dataset into Weka and run simple k-means clustering algorithm with different values of k (number of desired clusters). Study the clusters formed. Observe the sum of squared errors and centroids, and derive insights	Laboratory 15: Creation of a Data Warehouse.
	SLO-2					

Learning Resources	1. Data mining and warehousing, S. Prabhu, N.Venatesan, New Age International, 2007 2. Data Mining, Concepts and Techniques, JiaweiHan, MichelineKambar, Jian Pie, 3rd edition, 2011.	5. 3.“Introduction to data mining” by Tan, Steinbach & Kumar (2006)

Learning Assessment

	Bloom's Level of Thinking	Continuous Learning Assessment (100% weightage)							
		CLA – 1 (20%)		CLA – 2 (20%)		CLA – 3 (30%)		CLA – 4# (30%)	
		Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice
Level 1	Remember Understand	10%	10%	10%	10%	10%	10%	10%	10%
Level 2	Apply Analyze	20%	20%	20%	20%	20%	20%	20%	20%
Level 3	Evaluate Create	20%	20%	20%	20%	20%	20%	20%	20%
	Total	100%		100%		100%		100 %	

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

Course Designers		
Experts from Industry	Experts from Higher Technical Institutions	Internal Experts

Mr. S. Karthik, Assistant Consultant, Tata Consultancy Services	Dr.S.Sasikala, Associate Professor and Head, Dept. of Computer Science, University of Madras	Dr. S. Kanchana
		Dr.Sweety