

Course Code	PIT21C202J	Course Name	DATA MINING AND DATA WAREHOUSING	Course Category	C	Professional Core	L	T	P	C
							3	0	4	5

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):		Learning			Program Learning Outcomes (PLO)														
		1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
The purpose of learning this course is to:																			
CLR-1 :	Understand and implement classical models and algorithms in data warehouses and data mining																		
CLR-2 :	Be familiar with mathematical foundations of data mining tools																		
CLR-3 :	Characterize the kinds of patterns that can be discovered by association rule mining																		
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 :	Conduct investigations of complex problems: Use research based knowledge and research methods including design of experiments																		
Course Learning Outcomes (CLO):		Learning			Program Learning Outcomes (PLO)														
		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
CLO-1 :	Understand the functionality of the various data mining and data warehousing component	2	80	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-2 :	Appreciate the strengths and limitations of various data mining and data warehousing models	3	85	75	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-3 :	Explain the analyzing techniques of various data	3	75	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-4 :	Describe different methodologies used in data mining and data warehousing.	3	85	80	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-5 :	Compare different approaches of data warehousing and data mining with various technologies	3	85	75	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-6 :	Foundations of Software development: The ability to grasp the software development lifecycle and methodologies of software systems.	3	80	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M

Duration (Hour)		21	21	21	21	21
S-1	SLO-1	Introduction to Data Mining	Association Analysis- Market basket analysis	Classification Techniques-introduction	CLUSTERING Analysis - Introduction	Introduction to Data Warehousing
	SLO-2	Data mining as the Evolution of information technology	Frequent Itemsets, closed itemsets, Association rules	-Supervised Vs Unsupervised classifications	Overview of clustering methods	Operational database systems versus Data warehouses
S-2	SLO-1	Data and large datasets	Methods of Frequent itemset mining	Decision tree induction Attribute selection measures Tree pruning	K- means method k-medoids method	Why have a separate data warehouse
	SLO-2	Kinds of Pattern in data mining	Apriori Algorithm	Scalability and decision tree induction	Hierarchical method	Data warehousing-Multi-tiered architecture- Data warehouse models
S-3	SLO-1	Technology used in data mining	Generating Association Rules from frequent Itemset	Bayes' Classification method-Bayes theorem	Agglomerative vs Divisive Hierarchical clustering	Extraction, Transformation,
	SLO-2	Machine learning, and	Pattern Growth Approach	Naive Bayesian Classification	Hierarchical clustering	Loading- Meta data repository Architecture of DW
S4-7	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.
	SLO-2					
S8	SLO-1	Database systems	Vertical Data Format	Rule Bases Classification	BIRCH	Multi dimensional data model-Data cube
	SLO-2	Data warehouse	Mining and closed and Max Patterns	IF-Then Rules for classification	Chameleon method	Schemas for multidimensional data models
S9	SLO-1	Kinds of applications- Business Intelligence	Pattern Evaluation methods	Rule Extraction	Probabilistic Hierarchical clustering	Role of concept hierarchies
	SLO-2	Web search Engines	Pattern mining -	Rule Extraction from decision tree	Density based method-	OLAP
S10	SLO-1	DM versus Knowledge Discovery in Databases	Mining Multi level associations	Rule Induction	DBSCAN	OLAP operations

	SLO-2	Data Mining Issues in Mining methodology	Mining multidimensional associations	Metrics for evaluating classifier performance	OPTICS	Querying multidimensional databases
S11-14	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, Kappa statistic.	Laboratory 14: create a query based on multidimensional databases
	SLO-2					
S15	SLO-1	User interaction, Efficiency and scalability Diversity of data types	Mining quantitative association rules	Cross validation	DENCLUE	Data warehouse design and uses
	SLO-2	Data preprocessing	Mining rare patterns and negative patterns	Holdout method	Grid based clustering methods	DW design process
S16	SLO-1	Data Mining Metric -	Constraint based frequent pattern mining	Bootstrap	STING	DW usages for Information processing
	SLO-2	Social Implications of Data Mining	Meta Rule	classification	CLIQUE	From OLAP to Multidimensional data mining
S17	SLO-1	Overview of Applications of Data Mining	Constraints based pattern generation	Classification by Back propagation	Evaluation of clustering methods	Data warehouse Implementations
	SLO-2	Data Objects and Attributes types	Graph Mining- Frequent sub-graph mining	Support vector machine	Measuring cluster quality	OLAP Server Architectures
S18-21	SLO-1					

	SLO-2	Laboratory 3: Perform data preprocessing tasks and Demonstrate performing association rule mining on data sets	Laboratory 6: 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 1d3, 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.
--	-------	--	--	--	---	--

Learning Assessment											
Bloom's Level of Thinking		Continous 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, 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, IT Analyst, Tata Consultancy Services	Dr. Neelanarayanan,, Professor, School of Computer Science and Engineering, VIT Chennai	Dr. S. Kanchana
		Mrs. S. Parimala