

23CS31E4 - DATA MINING & DATA WAREHOUSING

Course Category:	Professional Elective - I	Credits:	3
Course Type:	Theory	Lecture-Tutorial-Practical:	3-0-0
Prerequisite:	<ul style="list-style-type: none"> Knowledge in Data Warehousing & Data Mining. 	Sessional Evaluation: Univ. Exam Evaluation: Total Marks:	30 70 100
Course Objectives:	Students undergoing this course are expected:		
	<ul style="list-style-type: none"> Familiarize with mathematical foundations of data mining tools. Introduce classical models and algorithms in data warehouses and data mining. Investigate the kinds of patterns that can be discovered by association rule mining, classification and clustering. Explore data mining techniques in various applications like social, scientific and environmental context 		

Course Outcomes:	Upon successful completion of the course, the students will be able to:				
	CO1	Design a Data warehouse system and perform business analysis with OLAP tools.			
	CO2	Apply suitable pre-processing and visualization techniques for data analysis.			
	CO3	Apply frequent pattern and association rule mining techniques for data analysis.			
	CO4	Design appropriate classification and clustering techniques for data analysis.			
	CO5	Infer knowledge from raw data.			
Course Content:	UNIT-I				
	Basic Concepts – Data Warehousing Components – Building a Data Warehouse – Database Architectures for Parallel Processing – Parallel DBMS Vendors – Multidimensional Data Model – Data Warehouse Schemas for Decision Support, Concept Hierarchies -Characteristics of OLAP Systems – Typical OLAP Operations, OLAP and OLTP.				
	UNIT-II				
	Introduction to Data Mining Systems – Knowledge Discovery Process – Data Mining Techniques – Issues – applications- Data Objects and attribute types, Statistical description of data, Data Preprocessing – Cleaning, Integration, Reduction, Transformation and discretization, Data Visualization, Data similarity and dissimilarity measures				
	UNIT-III				
	Mining Frequent Patterns, Associations and Correlations – Mining Methods- Pattern Evaluation Method – Pattern Mining in Multilevel, Multi Dimensional Space – Constraint Based Frequent Pattern Mining, Classification using Frequent Patterns.				

	<p style="text-align: center;"><u>UNIT-IV</u></p> <p>Decision Tree Induction – Bayesian Classification – Rule Based Classification – Classification by Back Propagation – Support Vector Machines — Lazy Learners – Model Evaluation and Selection-Techniques to improve Classification Accuracy. Clustering Techniques – Cluster analysis-Partitioning Methods – Hierarchical Methods – Density Based Methods – Grid Based Methods – Evaluation of clustering – Clustering high dimensional data- Clustering with constraints, Outlier analysis-outlier detection methods..</p> <p style="text-align: center;">.</p> <p style="text-align: center;"><u>UNIT-V</u></p> <p>WEKA TOOL: Datasets – Introduction, Iris plants database, Breast cancer database, Auto imports database – Introduction to WEKA, The Explorer – Getting started, Exploring the explorer, Learning algorithms, Clustering algorithms, Association–rule learners.</p>
Text Books & References Books:	<p>TEXTBOOKS:</p> <p>1.Jiawei Han and Micheline Kamber, —Data Mining Concepts and Techniques, Third Edition, Elsevier, 2012</p> <p>REFERENCE BOOKS:</p> <p>1. Alex Berson and Stephen J.Smith, —Data Warehousing, Data Mining & OLAP®, Tata McGraw – Hill Edition, 35th Reprint 2016.</p> <p>2. K.P. Soman, Shyam Diwakar and V. Ajay, —Insight into Data Mining Theory and Practice, Eastern Economy Edition, Prentice Hall of India, 2006.</p> <p>3. Ian H.Witten and Eibe Frank, —Data Mining: Practical Machine Learning Tools and Techniques, Elsevier, Second Edition.</p>
E-Resources:	NPTEL