

Subject Code: 1ET2030201	Subject Title: DATA MINING & DATA WAREHOUSING
Pre-requisite	Fundamentals of DBMS and Computer Programming

Course Objective:

This Course is intended to explore the knowledge gathering from raw data. This course will provide various methodologies and approaches for the same. This course aims to provide various business intelligence and mining tools for the knowledge gathering.

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture	Tutorial	Practical	Credit	Theory		Practical		Total
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
2	-	2	4	60	40	30	20	150

Subject Contents			
Sr. No	Topic	Total Hours	Weight (%)
1	Introduction to Data Mining: Motivation, what is Data Mining? Data Mining Tasks, Issues in Data Mining, Applications	2	8
2	Data Preprocessing: Types of data, Data Quality, Data preprocessing, Similarity and Dissimilarity Measures	2	8
3	Data Exploration: Data Set & its Statistics, Visualization, OLAP & Multidimensional Data Analysis	4	16
4	Classification: Introduction, Applications, Decision Tree based Algorithms, Model Over-fitting, Performance Evaluation of a Classifier, Comparing Classifiers	4	16
5	Classification: Alternative Techniques: Rule Based Classifier, Nearest Neighbor Classifier, Bayesian Classification, Support Vector Machine	3	13
6	Association Rule Mining: Introduction, Applications, Market-Basket Analysis, Frequent Item sets, Apriori Algorithm, Alternative Methods. Generalized Association Rules, Multilevel Association Rules, Multidimensional Association Rules, Temporal Association Rules, Infrequent Patterns, Constrained Based Association Rules	5	23
8	Clustering: Introduction, Applications, Partitioning Algorithms, Hierarchical Algorithms, Density based Algorithms, Cluster Evaluation	4	16

Review Presentation: The student is expected to spend minimum 2 hours per week (as mentioned in the teaching and evaluation scheme) to refer at least two peer reviewed journal papers related to this domain/subject. The student is expected to identify issues/challenges and emerging trends in the domain/subject. Student is supposed to explore various video lectures (E.g. NPTEL) available in the domain/subject. Student is required to make a review-presentation on the work carried out for the same.

Recommended sites for journal papers are (1) dl.acm.org (2) springer.com (3) sciencedirect.com (4) elsevier.com (5) ieeexplore.ieee.org (6) scholar.google.co.in (7) scopus.com or others of similar repute.

Course Outcome:

At the end of this course, the student would be able to:

- Learn data mining tasks like constructing decision trees, finding association rules, classification, and Clustering.
- Learn and design the data mining algorithm
- Provide a holistic view of data mining.
- Have knowledge of database, statistical, algorithmic and application perspectives of data mining.

List of References:

1. Tan, Pang-Ning and other “*Introduction to Data Mining*” Pearson Education, 2006.
2. Han J & Kamber M, “*Data Mining: Concepts and Techniques*”, Elsevier, Second Edition, 2006
3. Dunhum M.H. & Sridhar S. “*Data Mining-Introductory and Advanced Topics*”, Pearson Education, 2006.

E-Resources / Web-Links:

1. <http://www.cs.waikato.ac.nz/ml/weka/>

List of Experiments:

Note: The experiment list provided beneath is for reference only. The course teacher may change/formulate it as per his/her methodology and requirement.

Lab #	Topic	Learning Objective
1.	Data Pre-processing	How to prepare data for mining through aggregation, discretization, feature selection, transformation etc.
2.	Classification	How to build a basic classifier such as decision tree, k-nn etc.
3.	Adv. Classification	How to build other models for classification and how to compare models etc
4.	Association Rule Mining	How to generate frequent item sets? How to apply different measures for AR mining?
5.	Clustering	How to construct clusters using different algorithms?
6.	Adv. Clustering	How to evaluate clusters? How to compare different algorithms of clustering? How to detect outliers?