

Subject Code: 1ET1100403	Subject Title: Data Mining & Warehouse
Pre-requisite	Fundamentals of DBMS and Computer Programming

Course Objective:

- To understand data warehouse concepts, architecture, business analysis and tools.
- To understand data pre-processing and data visualization techniques.
- To study algorithms for finding hidden and interesting patterns in data.
- To understand and apply various classification and clustering techniques using tools.

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture	Tutorial	Practical	Credit	Theory		Practical		Total
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
3	1	2	5	70	30	30	20	150

Subject Contents			
Sr. No	Topic	Total Hours	Weight (%)
1	Data Mining – Introduction: Introduction to Data Mining Systems – Knowledge Discovery Process, Data Mining Techniques-Issues, applications- Data Objects and attribute types, Statistical description of data.	4	10
2	Data Preprocessing: Cleaning, Integration, Reduction, Transformation and discretization, Data Visualization, Data similarity and dissimilarity measures.	4	15
3	Data Mining - Frequent Pattern Analysis: Mining Frequent Patterns, Associations and Correlations, Mining Methods, Pattern Evaluation Method, Pattern Mining in Multilevel, Classification using Frequent Patterns.	5	15
4	Classification: 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.	6	20
5	Clustering: 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.	6	20
6	Data Warehousing, Business Analysis and On-line Analytical Processing(OLAP): 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.	9	20

Course Outcome:

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

- Design a Data warehouse system and perform business analysis with OLAP tools.
- Apply suitable pre-processing and visualization techniques for data analysis.
- Apply frequent pattern and association rule mining techniques for data analysis.
- Apply appropriate classification and clustering techniques for data analysis.

Course Learning Outcomes (CLO):

CLO	Description	Bloom's Taxonomy Level
CLO1	Design a Data warehouse system and perform business analysis with OLAP tools.	Understanding, Applying
CLO2	Apply suitable pre-processing and visualization techniques for data analysis.	Understanding, Applying
CLO3	Apply frequent pattern and association rule mining techniques for data analysis.	Applying
CLO4	Apply appropriate classification and clustering techniques for data analysis.	Applying, Evaluating

Mapping of CLOs with POs & PSOs:

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CLO1	M	H	M	H	H	M	L	M	M	M	M	M	M	H	L	L
CLO2	M	M	H	M	M	M	L	M	M	M	L	L	H	M	M	M
CLO3	H	M	H	M	H	H	M	L	M	L	M	M	H	M	M	M
CLO4	H	H	M	H	H	H	L	L	M	M	M	L	H	M	M	M

List of References:

1. Alex Berson and Stephen J.Smith, —Data Warehousing, Data Mining & OLAP||, Tata McGraw – Hill Edition, 35th Reprint 2016.
2. K.P. Soman, Shyam Diwakar and V. Ajay, —Insight into Data Mining Theory and Practice||, Eastern Economy Edition, Prentice Hall of India, 2006.
3. Ian H.Witten and Eibe Frank, —Data Mining: Practical Machine Learning Tools and Techniques||, Elsevier, Second Edition.

Text book:

1. Jiawei Han and Micheline Kamber, —Data Mining Concepts and Techniques||, Third Edition, Elsevier, 2012.

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.

1. Listing applications for mining
2. File format for data mining
3. Conversion of various data files
4. Training the given dataset for an application
5. Testing the given dataset for an application
6. Generating accurate models
7. Data pre-processing – data filters
8. Feature selection
9. Web mining
10. Text mining
11. Design of fact & dimension tables
12. Generating graphs for star schema