

<b>Subject Code : 1CS2010402</b>	<b>Subject Title: DATA WAREHOUSE &amp; DATA MINING</b>
<b>Pre-requisite :</b>	Knowledge of RDBMS, OLTP and OLAP

**Course Objective:**

The objectives of the course are to:

- To understand the need of Data Warehouses over Databases, and the difference between usage of operational and historical data repositories.
- To understand the concept of Analytical Processing (OLAP) and its similarities & differences with respect to Transaction Processing (OLTP).
- To understand the need for pre-processing
- To learn the algorithms used for various type of Data Mining problems

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	--	4	60	40	--	--	100

Subject Contents			
Sr. No	Topic	Total Hours	Weight (%)
<b>1</b>	<b>Data Warehouse and OLAP Technology:</b> Introduction to Data Warehouse, Multi-dimensional Data Model, Data Warehouse Architecture, OLAP Servers.	5	20
<b>2</b>	<b>Data Mining Introduction:</b> What is Data Mining – on What Kind of Data, Data Mining Functionalities – What Kind of Patterns Can be Mined, Are All of the Patterns Interesting, Classification of Data Mining Systems, Major Issues in Data Mining.	9	15
<b>3</b>	<b>Data Preprocessing:</b> Need for Pre-processing, Data Cleaning, Data Integration & Transformation, Data Reduction, Concept Hierarchy Generation. Thrashing, Locality.	7	15
<b>4</b>	<b>Mining Frequent Patterns, Associations, and Correlations:</b> Basic Concepts: Market Basket Analysis, Efficient and Scalable frequent Itemset Mining Methods, From Association Mining to Correlation Analysis	8	15
<b>5</b>	<b>Classification:</b> Introduction to Classification and Prediction, Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification.	8	15
<b>6</b>	<b>Cluster Analysis:</b> Introduction to Cluster Analysis, Types of Data in Cluster Analysis, A categorization of Major Clustering Methods, Partitioning Methods.	8	15
<b>7</b>	<b>Application and Trends in Data Mining:</b> Data Mining Applications, Data Mining System Products and Research Prototypes.	3	5

**Course Outcome:**

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

- Ability to apply pre-processing on existing operational & historical data for creation of Data Warehouse
- Ability to apply Apriori algorithm for Association Mining
- Ability to apply Decision Tree and Bayesian algorithms for Classification
- Ability to mine Statistical Measures in large databases

**List of References**

1. Jiawei Han & Micheline Kamber, "Data Mining: Concepts & Techniques", Morgan Kaufmann Publishers (2002)
2. W. H. Inmon, "Building the Data Warehouse", Wiley Dreamtech India Pvt. Ltd.
3. Mohanty, Soumendra, "Data Warehousing: Design, Development and Best Practices", Tata McGraw Hill (2006)