IT401: DATA MINING AND DATA WAREHOUSING
CREDITS = 6 (L=4, T=0, P=2)

Course Objective:

To extract knowledge from data repository for data analysis, frequent pattern, classification and prediction.

Teaching and Assessment Scheme:

Teaching Scheme			Credits	Assessment Scheme				Total Marks
				Theory		Practical		
L	T	P	C	ESE	CE	ESE	CE	
4	0	2	6	70	30	30	20	150

Course Contents:

Unit No.	Topics	Teaching Hours
1	**Data Warehousing:** Introduction to data warehousing- Data warehousing components, Building a data warehouse, Difference between database system and data warehouse, Data warehouse architecture- 3 Tier architecture, Warehouse schema design, Data extraction, Cleanup & transformation tools, Multi-dimensional data model, Data cubes- Stars, Snowflakes, Fact constellations, Concept hierarchy, Online analytical processing- Typical OLAP operations.	10
2	**Data Mining:** Introduction of data mining - Definition and functionalities, Classification of DM systems, DM task primitives, Integration of a data mining system with a database and data warehouse - Issues in DM, KDD process.	06
3	**Data Preprocessing:** Data Pre-processing, Data cleaning, Data integration and transformation, Data reduction, Discretization and concept hierarchy generation, Data mining primitives, Languages and system architectures, Concept description: Characterization and comparison, Analytical characterization, Mining class comparison.	10
4	**Association Rule Mining:** Association rule mining, Mining of single dimensional Boolean association rules, Multilevel association rules and multidimensional association rules, Correlation analysis, Constraint based association mining.	09

5	**Classification:** Basic issues regarding classification and predication, Classification by decision Tree, Bayesian classification, Classification by back propagation, Associative classification, Prediction, Classifier accuracy.	10
6	**Clustering:** Cluster analysis, Basic issues, Clustering using partitioning methods, Hierarchical methods, Density based methods, Grid based methods and model based methods, Algorithms for outlier analysis.	10
7	**Advance Mining:** Introduction to spatial mining, Text mining and web mining with related algorithms, Introduction to Big data analytics.	05
	TOTAL	**60**

List of References:

1. Jiawei Han, Micheline Kamber, *"Data Mining concepts and Techniques",* Elsevier.
2. Arun K. Pujari, *"Data Mining"*, University Press.
3. Paulraj Ponnian, *"Data Warehousing Fundamentals"*, John Willey.

Course Outcomes (COs):
At the end of this course students will be able to …

1. Understand warehousing architectures and tools for systematically organizing and use their data to make strategic decisions.
2. Understand KDD process for finding interesting pattern from warehouse.
3. Remove redundancy and incomplete data from the dataset using data preprocessing methods.
4. Characterize the kinds of patterns that can be discovered by association rule mining.
5. Discover interesting patterns from large amounts of data to analyze for predictions and classification.
6. Develop a data mining application for data analysis using various tools.