JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY: KAKINADA

KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

III Year – I Semester		L	T	P	C
		3	0	0	3
DATA WAREHOUSING AND DATA MINING					

Course Objectives:
- 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

Course Outcomes:
At the end of the course, the students will 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 techniques for data analysis
- Apply appropriate clustering techniques for data analysis

UNIT I
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.

UNIT II
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, Data Preprocessing – Cleaning, Integration, Reduction, Transformation and discretization, Data Visualization, Data similarity and dissimilarity measures.

UNIT III
Data Mining - Frequent Pattern Analysis: 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

UNIT IV
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

UNIT V
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.

Text Books:
1) Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2012.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY: KAKINADA

KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2) Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Introduction to Data Mining, Pearson,2016.

Reference Books:

1) Alex Berson and Stephen J.Smith, ―Data Warehousing, Data Mining & OLAP‖, Tata McGraw – Hill Edition, 35th Reprint 2016.
2) K.P. Soman, ShyamDiwakar 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.

e-Resources:

1) https://www.saedsayad.com/data_mining_map.htm
2) https://nptel.ac.in/courses/106/105/106105174/
3) (NPTEL course by Prof.Pabitra Mitra) http://onlinecourses.nptel.ac.in/noc17_mg24/preview
4) (NPTEL course by Dr. Nandan Sudarshanam & Dr. Balaraman Ravindran) http://www.saedsayad.com/data_mining_map.htm