JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF INFORMATION TECHNOLOGY

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

Course Objectives:
The main objective of the course is to
- Introduce basic concepts and techniques of data warehousing and data mining
- Examine the types of the data to be mined and apply pre-processing methods on raw data
- Discover interesting patterns, analyze supervised and unsupervised models and estimate the accuracy of the algorithms.

Course Outcomes:
By the end of the course student will be able to
- Illustrate the importance of Data Warehousing, Data Mining and its functionalities and Design schema for real time data warehousing applications.
- Demonstrate on various Data Preprocessing Techniques viz. data cleaning, data integration, data transformation and data reduction and Process raw data to make it suitable for various data mining algorithms.
- Choose appropriate classification technique to perform classification, model building and evaluation.
- Make use of association rule mining techniques viz. Apriori and FP Growth algorithms and analyze on frequent itemsets generation.
- Identify and apply various clustering algorithm (with open source tools), interpret, evaluate and report the result.

UNIT I:
Data Warehousing and Online Analytical Processing: Data Warehouse: Basic concepts, Data Warehouse Modelling: Data Cube and OLAP, Data Warehouse Design and Usage, Data Warehouse Implementation, Introduction: Why and What is data mining, What kinds of data need to be mined and patterns can be mined, Which technologies are used, Which kinds of applications are targeted.

UNIT II:
Data Pre-processing: An Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization.

UNIT III:
Classification: Basic Concepts, General Approach to solving a classification problem, Decision Tree Induction: Attribute Selection Measures, Tree Pruning, Scalability and Decision Tree Induction, Visual Mining for Decision Tree Induction.

UNIT IV:
Association Analysis: Problem Definition, Frequent Item set Generation, Rule Generation: Confident Based Pruning, Rule Generation in Apriori Algorithm, Compact Representation of frequent item sets, FP-Growth Algorithm.

UNIT V:
Cluster Analysis: Overview, Basics and Importance of Cluster Analysis, Clustering techniques, Different Types of Clusters; K-means: The Basic K-means Algorithm, K-means Additional Issues, Bi-secting K Means,

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF INFORMATION TECHNOLOGY

Text Books:
1. Data Mining concepts and Techniques, 3/e, Jiawei Han, Michel Kamber, Elsevier,2011.
2. Introduction to Data Mining: Pang-Ning Tan & Michael Steinbach, Vipin Kumar, Pearson,2012.

Reference Books:
1. Data Mining Techniques and Applications: An Introduction, Hongbo Du, Cengage Learning.
2. Data Mining: VikramPudi and P. Radha Krishna, Oxford Publisher.
3. Data Mining and Analysis - Fundamental Concepts and Algorithms; Mohammed J. Zaki, Wagner Meira, Jr, Oxford
4. Data Warehousing Data Mining & OLAP, Alex Berson, Stephen Smith, TMH.
   http://onlinecourses.nptel.ac.in/noc18_cs14/preview
5. (NPTEL course by Prof.PabitraMitra)
   http://onlinecourses.nptel.ac.in/noc17_mg24/preview
6. (NPTEL course by Dr. NandanSudarshanam& Dr. BalaramanRavindran)
   http://www.saedsayad.com/data_mining_map.htm