LOK JAGRUTI UNIVERSITY (LJU)

L J INSTITUTE OF ENGINEERING AND TECHNOLOGY

Department of Computer Engineering

Master of Engineering (M. E)

Semester: I

Branch: Computer Engineering (Software Engineering)

Course Code:	20-CE-PG-049010105
Course Name:	Data Mining and Analysis
Category of Course:	Core
Prerequisite Course:	Databases, Probability

Teaching Scheme				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Total Hours
3	0	2	4	40

Course Objectives	
1	To understand the principles of Data warehousing and Data Mining
2	To be familiar with the Data warehouse architecture and its Implementation
3	To know the Architecture of a Data Mining system.
4	To understand the various Data Pre-processing Methods.
5	To understand the various classification and clustering techniques.
6	To get an introduction to spatial, multimedia and text mining.

Syllabus			
Unit No.	Topic	Prerequisite Topic	Teaching Hours
01	**Introduction to Data Mining** 1.1 Motivation and importance of Data Mining 1.2 Data Mining-Definition and Functionalities 1.3 KDD Process 1.4 Classification of DM Systems 1.5 Major issues in data mining	---	5 (10%)
02	**Data Preprocessing** 2.1 Data Summarization 2.2 Data Cleaning 2.3 Data Integration and Transformation 2.4 Data Reduction 2.5 Discretization and Concept Hierarchy Generation 2.6 Languages and System Architectures	---	6 (10%)
03	**Data warehouse and OLAP technology** 3.1 Data warehouse definition 3.2 Multidimensional data model(s) 3.3 Data warehouse architecture 3.4 OLAP server types	---	3 (5%)
04	**Data cube computation and Data Generalization** 4.1 Efficient methods for data cube computation 4.2 Discovery driven exploration of data cubes 4.3 Complex Aggregation 4.4 Attribute Oriented Induction for data generalization	---	4 (10%)
05	**Association Rule Mining** 5.1 Market-Basket Analysis 5.2 Frequent Itemsets 5.3 Apriori Algorithm 5.4 Improved Apriori algorithm	---	3 (15%)
06	**Advanced Association Rule Mining** 6.1 Generalized Association Rules 6.2 Multilevel Association Rules 6.3 Multidimensional Association Rules	---	4 (10%)
07	**Classification and Prediction** 7.1 Definition, Basic issues regarding classification and predication 7.2 Classification by Decision Tree Induction 7.3 Bayesian Classification 7.4 Rule Based Classification 7.5 Associative Classification 7.6 Prediction	---	5 (15%)
08	**Cluster Analysis** 8.1 What is Cluster Analysis? 8.2 Categorization of Major Clustering Methods	---	5 (10%)

	8.3 Partitioning Methods		
	8.4 Hierarchical methods		
	8.5 Density-Based Methods		
	8.6 Grid-Based Methods		
	8.7 Model Based Clustering Methods		
	8.8 Outlier Analysis.		
09	**Advanced Mining Techniques**	---	**5 (10%)**
	9.1 Spatial Data Mining		
	9.2 Multimedia Data Mining		
	9.3 Text Mining		
	9.4 Web Mining		
	9.5 Recommender systems		
	9.6 Large scale data mining		
10	**APPLICATIONS AND TRENDS IN DATA MINING**	---	**3 (5%)**
	10.1 Data Mining Applications		
	10.2 Data Mining System Products and Research Prototypes		
	10.3 Social Impacts of Data Mining		

Course Outcome

1	Study of different sequential pattern algorithms
2	Study the technique to extract patterns from time series data and it application in real world.
3	Can extend the Graph mining algorithms to Web mining
4	Help in identifying the computing framework for Big Data

Suggested Reference Books

1	Data Mining concepts and Techniques, Jiawei Han, Micheline Kamber, Elsevier
2	Data Mining, Arun K. Pujari, University Press
3	Mordern Data Warehousing, Data Mining and Visualization, George M.Marakas, Pearson
4	Data Mining, Vikram Puri And P. Radha Krishana, Oxfrod Press
5	Data Warehousing, Reema Theraja, Oxford Press
6	Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Addison Wesley,2006
7	Sequence Data Mining, G Dong and J Pei, Springer, 2007.

Proposed Evaluation Scheme by Academicians (Percentage of Weightage out of 100%)

Theory Descriptive Test		MCQ Test		Hands on Project	
Formulas and Derivation Test		**Numerical Test**		**Seminar**	

Practical Project/Hands On Project

Sr. No.	List of Practical Projects	Linked with Unit
1	Overview of SQL Server 2008 Databases and analysis services	Unit 1
2	Create your own data set (like customer, weather, agriculture etc.), load it and apply any pre-processing technique and clean the data, show results. a. Clean missing values b. Remove Data c. Data Smoothing using Binning.	Unit 2
3	Implement Naive Bayesian algorithm taking any dataset of your choice and predict the result.	Unit 7
4	Implement Association mining algorithm by taking appropriate data set and find support and confidence. Also show confusion matrix.	Unit 4
5	Implement Decision Tree algorithm by taking appropriate data set and predict the result. Calculate entropy and information gain.	Unit 7
6	To implement Bayesian Classification algorithms in C programming language	Unit 3
7	Implement K-means clustering algorithm by taking appropriate data set and predict the result.	Unit 8
8	Implement K-medoids clustering algorithm by taking appropriate data set and predict the result	Unit 8
9	Implement CART algorithm by taking appropriate data set and predict the result	Unit 7
10	Implement DBSCAN clustering algorithm by taking appropriate data set and predict the result.	Unit 8

List of Recommended MOOC Courses:

1) https://www.coursera.org/learn/cluster-analysis
2) https://www.edx.org/course/cluster-analysis
3) https://www.coursera.org/learn/text-mining?specialization=data-mining
4) https://www.udemy.com/course/data-mining-with-r-go-from-beginner-to-advanced/