Course Code	CSE506
Course Name	Data Mining
Credits	4
Course Offered to	UG/PG
Course Description	Data mining aims at finding the useful patterns in large data sets. Interest in the field is motivated by the growth of computerized data collection due to ubiquity of Internet enabled devices. This course will cover a set of techniques designed to be used for finding interesting patterns from the data. The techniques include classification, clustering, association rule minin and sequence mining. Students will learn and use the open source R statistical software, see http://www.r-project.org, and machine learning packages such as Weka in this course.

Pre-requisites

Pre-requisite (Mandatory)	Pre-requisite (Desirable)	Pre-requisite(other)
CSE202 Fundamentals of Database Systems	Programming in Java	
CSE101 Intro to Programming		
MTH100 Maths I		
MTH201 Probability & Statistics		

Post Conditions*(For suggestions on verbs please refer the second sheet)

CO1	CO2	CO3	CO4
Students are able to discuss basic applications, concepts and techniques of data mining such as association rules mining, classification, clustering and sequence mining.	Students are able to use data mining software (Weka, R etc.) to solve practical problemsrelated to association rules mining, classification, clustering and sequence mining.	Students are able to analyze the performance of different data mining techniques.	

Weekly Lecture Plan

Week Number	Lecture Topic	COs Met	Assignment/Labs/Tutorial
1-3	Exploratory data analysis on numeric and categorical attributes, frequent pattern mining (introduction), partitional clustering (K-Means, EM), Classification (Decision Trees)	CO1, C02	One programming assignment to introduce the students to basics of pre-processing and analyzing a small dataset using R. One assignment containing practice questions on frequent pattern mining .

4-6	Exploratory data analysis on high dimensional data, PCA, SVD, LDA, Frequent pattern mining (Itemset summaries), sequence pattern mining	CO2	A programming assignment to apply PCA/SVD techniques on real dataset.
7-9	Probabilistic classification, support vector machines, kernel svm, classifier evaluation	CO3	A programming assignment to apply different classifiers as well as evaluate their performance.
10-13	Hierachical/Density based clustering, Spectral clustering, Graph Mining	C02, C03	A programming assignment on hierarchical/density based clustering.

*Please insert more rows if required

Weekly Lab Plan

Week Number	Laboratory Exercise	COs Met	Platform (Hardware/Software)

*Please insert more rows if required

Assessment Plan

Type of Evaluation	% Contribution in Grade
End-sem	25
Mid-sem	25
Quiz	10
Project	10
Homework/Programming Assignments	30

Resource Material

Type	Title
Reference	Database Mining and Analysis Mohammed J. Zaki, Wagner Meira JR.

Reference	Introduction to Data Mining Michael Steinbach, Parag-Nin Tan, Vipin Kumar
Reference	Anand Rajaraman and Jeffrey David Ullman: Mining of Massive Datasets
Other Materials	