CSE4011/CSE6031/CSE7011: Data Mining

Programme: M. Tech. (CSE)	Year: 1	Semester: 1
Course: Program Elective	Credits: 3	Hours: 40

Course Context and Overview:
The quantity of data, of varying types, is increasing at a phenomenal rate and it is necessary to identify patterns that are interesting out of this data. Generally, Data Mining is the process of analyzing data from different perspectives and summarizing it into useful information. It covers algorithms and computational paradigms to find patterns, regularities (irregularities) in data, perform prediction, and forecasting. This course will cover these concepts and techniques and will illustrate the whole process by taking different case studies.

Prerequisite Courses: NIL

Course Outcomes (COs):

On completion of this course, the students will have the ability to:	
CO1:	Assess raw input data and process it to provide suitable input for a range of data mining algorithms.
CO2:	Apply, evaluate and analyze data mining algorithms and report the output appropriately.
CO3:	Evaluate and implement a wide range of emerging and newly adopted methodologies and technologies to facilitate knowledge discovery.

Course Topics:

Contents	Lecture Hours
UNIT 1 **Introduction**	
Introduction to Data Mining, Introduction to Major Building Blocks – Association Pattern Mining, Classification, Clustering, Outlier Detection;	2
UNIT 2 **Data Preparation**	
Types of Attributes, Feature Extraction, Creation, and Selection, Data Cleaning, Dimensionality Reduction	3

UNIT 3 **Classification**	
Decision Trees (ID3 or C4.5 or J48): Attribute test Conditions, Best split, handling continuous attributes, Training and Testing Error, MDL, Cost-sensitive learning, Evaluating Performance of classifier.	4
Rule-based Classifiers (RIPPER): Rule ordering, Direct and Indirect Method of rule Extraction, Rule Evaluation.	2
Statistical Classifier (Naïve Bayes): Estimating conditional probability and M-estimate of conditional Probability, Bayes error rate.	4
Method of Comparing classifiers, Ensemble methods (Bagging, Boosting etc),	2
Multi-class Problem, One-class Classification	1
UNIT 4 **Association Pattern Mining**	
Market basket Analysis	0.5
Frequent Itemset Mining Algorithms (Apriori) Maximal and Closed Frequent Itemsets, Pattern Querying, Rule Generation, Evaluation of Association Pattern	4
Skewed Support Distribution	0.5
Continuous and categorical Attributes	1
Sequential Pattern Discovery	2.5
Subgraph Pattern Mining	2.5
Infrequent Patterns, Negative Patterns	2
UNIT 5 **Clustering**	
Different types of Clustering and Clusters, Representative-based Clustering, Density-based Clustering	3
Cluster Evaluation	1
Graph-based Clustering, Clustering using mixture Model, Clustering Categorical Data	3
Scalable Data Clustering	1
UNIT 6 **Outlier Analysis**	
Statistical Approaches, Clustering-based Methods	1

Textbook references:
Text Books:

1. P. Tan, M. Steinbach, V. Kumar: *"Introduction to Data Mining*," Pearson Education, 2006

Reference books:

1. Charu C. Aggarwal, *"Data Mining,"* Springer, 2015
2. K. Pujari: *"Data Mining Techniques,"* Universities Press, 3rd Edition, 2013
3. J. Han, M. Kamber, *"Data Mining: Concepts and Techniques"*, Morgan Kaufmann Publishers, Second Edition, 2006
4. M.J. Zaki & W. Meira Jr.: *"Data Mining and Analysis – Fundamental Concepts and Algorithms,"* Cambridge University Press, 2014

Prepared By: Preety Singh, Bharavi Mishra
Last Modification: 7th July 2020