

B.E Semester: 8 Mechanical Engineering
Subject Name: Data Mining & Analysis (MA803-N-C)
[Dept. Elect.-5]

A. Course Objective:

- Learn data mining concepts understand association rules mining. Discuss classification algorithms learn how data is grouped using clustering techniques.
- To develop the abilities of critical analysis to data mining systems and applications.
- To implement practical and theoretical understanding of the technologies for data mining
- To understand the strengths and limitations of various data mining models

B. Teaching / Examination Scheme:

Teaching Scheme				Total Credit	Evaluation Scheme					
L	T	P	Total		Theory		Mid Sem Exam	CIA	Pract.	Total
Hrs	Hrs	Hrs	Hrs		Hrs	Marks	Marks	Marks	Marks	Marks
3	0	2	5	4	3	70	30	20	30	150

C. Detailed Syllabus:

Unit No.	Details
1	Introduction to Data Mining: Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing, Data Cleaning, Missing data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binaryzation, Data Transformation; Measures of Similarity and Dissimilarity- Basics.
2	Association Rules: Problem Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation; APRIORI Algorithm, The Partition Algorithms, FP-Growth Algorithms, Compact Representation of Frequent Item Set- Maximal Frequent Item Set, Closed Frequent Item Set.
3	Classification: Problem Definition, General Approaches to solving a classification problem , Evaluation of Classifiers , Classification techniques, Decision Trees-Decision tree Construction , Methods for Expressing attribute test conditions, Measures for Selecting the Best Split, Algorithm for Decision tree Induction ; Naive-Bayes Classifier, Bayesian Belief Networks; K- Nearest neighbor classification-Algorithm and Characteristics.
4	Clustering: Problem Definition, Clustering Overview, Evaluation of Clustering Algorithms, Partitioning Clustering-K-Means Algorithm, K-Means Additional issues, PAM Algorithm; Hierarchical Clustering-Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering Algorithm, Specific techniques, Key Issues in Hierarchical Clustering, Strengths and Weakness; Outlier Detection.

5	Web and Text Mining: Introduction, web mining, web content mining, web structure mining, we usage mining, Text mining –unstructured text, episode rule discovery for texts, hierarchy of categories, text clustering.
---	---

Total hours (Theory):48
Total hours (Practical):32
Total hours:80

D. Lesson Planning:

Sr. No.	Date/Week	Unit	Weight age	Topic No
1	1 st ,2 nd ,3 rd	Unit 1	20%	1
2	4 th .5 th ,6 th	Unit 2	20%	2
3	7 th , 8 th ,9 th	Unit 3	20%	3
4	10 th .11 th . 12 th	Unit 4	20%	4
5	13 th , 14 th ,15 th ,16 th	Unit 5	20%	5

E. Instructional Method & Pedagogy

1	At the start of course, the course delivery pattern , prerequisite of the subject will be discussed
2	Lecture may be conducted with the aid of multi-media projector, black board, OHP etc. & equal Weight age should be given to all topics while teaching and conduction of all examinations.
3	Attendance is compulsory in lectures and laboratory, which may carries five marks in overall evaluation.
4	One/Two internal exams may be conducted and total/average/best of the same may be converted to equivalent of 30 marks as a part of internal theory evaluation.
5	Assignment based on course content will be given to the student for each unit/topic and will be evaluated at regular interval. It may carry an importance of ten marks in the overall internal evaluation.
6	Surprise tests/Quizzes/Seminar/Tutorial may be conducted and having share of five marks in the overall internal evaluation.
7	The course includes a laboratory, where students have an opportunity to build an appreciation for the concept being taught in lectures. Suggested list of experiment is given below

F. List of Experiments :

1	Software Forensic tools such as XML Miner & WeKA
2	Implemenation of data mining alg such as Clustering, Classification and Associciation Rules
3	OLAP in standard SQL Server
4	MS Excell & Datamining plugin

G. Students Learning Outcomes:

1	Ability to perform the pre-processing of data and apply mining techniques on it.
2	Ability to identify the association rules, classification and clusters in large data sets.
3	Ability to solve real world problems in business and scientific information using data mining
4	Ability to classify web pages, extracting knowledge from the web

H. Text Books & Reference Books:

1	Data Mining- Concepts and Techniques- Jiawei Han, Micheline Kamber, Morgan Kaufmann Publishers, Elsevier 2 Edition, 2006.
2	Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbach, Pearson Education.
3	Data mining Techniques and Applications, Hongbo Du Cengage India Publishing
4	Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
5	Data Mining Principles & Applications – T.V Sveresh Kumar, B.Esware Reddy, Jagadish S Kalimani, Elsevier.
6	Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University Press