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

A. Course Objective:

• Students will be able to use mining tool.

• To Students are able to perform various data warehouse related exercise.

B. Teaching / Examination Scheme:

Teaching Scheme				Evaluation Scheme						
L	Т	Р	Total	Total Credit	The	eory	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
	Introduction to Data Mining (DM)
1	Motivation for Data Mining - Data Mining-Definition and Functionalities – Classification of DM
1	Systems - DM task primitives - Integration of a Data Mining system with a Database or a Data
	Warehouse - Issues in DM – KDD Process
	Data Pre-processing
	Why to pre-process data? - Data cleaning: Missing Values, Noisy Data - Data Integration and
2	transformation - Data Reduction: Data cube aggregation, Dimensionality reduction - Data
2	Compression - Numerosity Reduction - Data Mining Primitives - Languages and System
	Architectures: Task relevant data - Kind of Knowledge to be mined - Discretization and Concept
	Hierarchy.
	Concept Description and Association Rule Mining
	What is concept description? - Data Generalization and summarization-based characterization -
3	Attribute relevance - class comparisons Association Rule Mining: Market basket analysis - basic
	concepts - Finding frequent item sets: Apriori algorithm - generating rules - Improved Apriori
	algorithm – Incremental ARM – Associative Classification – Rule Mining
	Classification and Prediction
	What is classification and prediction? – Issues regarding Classification and prediction:
4	Classification methods: Decision tree, Bayesian Classification, Rule based, CART, Neural
	Network Prediction methods: Linear and nonlinear regression, Logistic Regression Introduction
	of tools such as DB Miner /WEKA/DTREG DM Tools
	Data Mining for Business Intelligence Applications
5	Data mining for business Applications like Balanced Scorecard, Fraud Detection, Clickstream
	Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance
	and CRM etc., Data Analytics Life Cycle: Introduction to Big data Business Analytics - State of
	the practice in analytics role of data scientists Key roles for successful analytic project - Main
	phases of life cycle - Developing

	core deliverables for stakeholders.
6	Advance TopicsIntroduction and basic concepts of following topics. Clustering, Spatial mining, web mining, textmining, Big Data: Introduction to big data: distributed file system – Big Data and its importance,Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using mapreduce, Matrix-Vector Multiplication by Map Reduce. Introduction to Hadoop architecture:Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands , Anatomy ofFile Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduceparadigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & HadoopConfiguration – HDFS Administering – Monitoring & Maintenance.

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
2	$4^{th} .5^{th} , 6^{th}$	Unit 2	20%	3
3	7^{th} , 8^{th} , 9^{th}	Unit 3	20%	4
4	$10^{\text{th}} . 11^{\text{th}} . 12^{\text{th}}$	Unit 4	20%	5
5	13 th , 14 th ,15 ^{th,} ,16 th	Unit 5	20%	6

E. Instructional Method & Pedagogy

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

F. List of Practicles:

1	Learning of data collection techniques.
2	Data storage practice for big data.
3	Store the basic information about students such as roll no, name, date of birth, and address of student using various collection types such as List, Set and Map
4	Retrieve various types of documents from students collection

5	To find documents from Students collection
6	Develop Map Reduce Work Application
7	Creating the HDFS tables and loading them in Hive and learn joining of tables in Hive
8	Data analysis using –Excel.
9	Data separation and plotting using- scilab
10	Data organisation using -excel or Scilab.

G. Students Learning Outcomes:

1	The student can identify	different areas and applications	of Data Mining and Analysis.
-	5	11	0 5

H. Text Books & Reference Books:

1	J. Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann
2	Jo M. Kantardzic, "Data mining: Concepts, models, methods and algorithms, John Wiley & Sons
	Inc.
3	Paulraj Ponnian, "Data Warehousing Fundamentals", John Willey
4	M. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education
5	G. Shmueli, N.R. Patel, P.C. Bruce, "Data Mining for Business Intelligence: Concepts,
5	Techniques, and Applications in Microsoft Office Excel with XLMiner", Wiley India.