

Kadi Sarva Vishwavidyalaya
Faculty of Engineering and Technology
First Year Master of Engineering (Computer Engineering)
In Effect from Academic Year 2017-18

Subject Code: MECE103-N	Subject Title: DATA WAREHOUSING AND DATA MINING
--------------------------------	--

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
04	00	02	06	05	03	70	30	20	30	150

LEARNING OBJECTIVES:

The educational Objectives of this Course are:

- To Introduce various Data Mining Applications in real world scenario
- To be learning more about various mining tools for analysis and decision making
- Applying efficient mining methods to solve engineering problems
- Learning concepts of Business Intelligence in solutions, organizational changes, products, technologies and methods to organize key data to improve performance and profit.

OUTLINE OF THE COURSE:

Sr. No	Title of the Unit	Minimum Hours
1	Data Warehousing Fundamentals	8
2	Data Pre-processing	7
3	Data Extraction, Transformation and Loading (ETL)	8
4	Introduction to Business Intelligence	10
5	Mining Frequent Patterns, Associations, and Correlations	7
6	Classification and Prediction	8
7	Clustering	8
8	Current and Future Research Trends	8

Total Hours (Theory):64

Total Hours (Lab):32

Total Hours: 96

Kadi Sarva Vishwavidyalaya
Faculty of Engineering and Technology
First Year Master of Engineering (Computer Engineering)
In Effect from Academic Year 2017-18

DETAILED SYLLABUS:

Sr. No	Topic	Lecture Hours	Weight age (%)
1	Data Warehousing fundamentals <ul style="list-style-type: none"> • Introduction • A Multi-Dimensional Data Model • Data Warehouse Architecture • Data Warehouse Implementation • From Data Warehouse to Data Mining to Business Intelligence 	8	13
2	Data Pre-processing <ul style="list-style-type: none"> • Data Cleaning • Data Integration and Transformation • Data Reduction • Data Discretization and Concept Hierarchy Generation 	7	10
3	Data Extraction, Transformation and Loading (ETL) <ul style="list-style-type: none"> • Extracting the Data • Transforming the Data • Loading the Data into a DW System • ETL Using Export Import • Challenges for ETL • ETL Tools • Difference between ETL and BI Tools 	8	13
4	Introduction to Business Intelligence <ul style="list-style-type: none"> • Introduction • A Data Framework For BI Structured Vs. Semi-Structured Data Framework Architecture For Structured Data Architecture For Semi-Structured Data • BI as a Product, Process, Solution and Tools • Factor driving Business Intelligence • Role of Data, Information and Knowledge in Data Warehouse, Data Mining and Decision Support System • Difference between BI and other related technologies. • Utilization and benefits of BI in Organization. • Obstacles to BI • Business Intelligence User Tools • Research issues in BI 	10	15

Kadi Sarva Vishwavidyalaya
Faculty of Engineering and Technology
First Year Master of Engineering (Computer Engineering)
In Effect from Academic Year 2017-18

5	Mining Frequent Patterns, Associations, and Correlations <ul style="list-style-type: none"> • Basic Concepts and a Road Map • Efficient and Scalable Frequent Item set Mining Methods • Mining Various Kinds of Association Rules • Constraint based Association Mining 	7	10
6	Classification and Prediction <ul style="list-style-type: none"> • The fundamentals of classification systems • Issues regarding Classification and prediction • Differences between classification, recommendation, and clustering • Applications of classification • Classification methods: Decision tree, Bayesian Classification, Rule based, CART Neural Network CBR Rough set Approach Fuzzy Logic Genetic Algorithms • Prediction methods: • The fundamentals of Prediction • Linear and nonlinear regression • Accuracy and Error Measures • Accuracy of Classifier 	8	13

Kadi Sarva Vishwavidyalaya
Faculty of Engineering and Technology
First Year Master of Engineering (Computer Engineering)
In Effect from Academic Year 2017-18

7	Clustering <ul style="list-style-type: none"> • Introduction to clustering • Types of Data in Cluster Analysis Interval-Scaled Variables, Binary Variables Categorical, Ordinal, and Ratio-Scaled Variables Variables of Mixed Types and Vector Objects • Measuring the similarity of items • Exploring distance measures Euclidean, Squared Euclidean, Manhattan distance, Cosine, Tanimoto, Weighted distance • Categorization of Major Clustering Methods K-means clustering Fuzzy k-means clustering Hierarchical Methods CLARANS, BIRCH, ROCK, Chameleon • Model-based clustering Deficiencies of k-means Dirichlet clustering Latent Dirichlet allocation (LDA) 	8	13
8	Current and Future Research Trends <ul style="list-style-type: none"> • Multirelational Data Mining, Spatial mining, web mining, text mining, Ensemble Classifier (Multiple Classifier, Bagging, Boosting, Stacking), Incremental learning 	8	13
	Total	64	100

GUIDELINES:

- Theory hours are to be utilized to teach the concepts
- Practical hours are to be utilized for making exhaustive programming practice for real life examples.
- The theory examination should contain approximate 2/3 weight age for logical examples.

INSTRUCTIONAL METHOD AND PEDAGOGY:

- At the start of course, the course delivery pattern, prerequisite of the subject will be discussed.
- Lectures will be conducted with the aid of multi-media projector, black board, OHP etc.
- Attendance is compulsory in lecture and laboratory which carries 10 marks in overall evaluation.
- One internal exam will be conducted as a part of internal theory evaluation.
- Assignments based on the course content will be given to the students for each unit and will be evaluated at regular interval evaluation.
- Surprise tests/Quizzes/Seminar/tutorial will be conducted having a share of five marks in the overall internal evaluation.
- The course includes a laboratory, where students have an opportunity to build an appreciation for the concepts being taught in lectures.

Kadi Sarva Vishwavidyalaya
Faculty of Engineering and Technology
First Year Master of Engineering (Computer Engineering)
In Effect from Academic Year 2017-18

- Experiments shall be performed in the laboratory related to course contents.

STUDENTS LEARNING OUTCOME:

On successful completion of the course, the student will:

- Be able to Compare various mining methods
- Be able to understand and identify the analytical characteristics of mining algorithms.
- Employ algorithm to model engineering problems, when appropriate.

RECOMMENDED STUDY MATERIAL:

REFERENCE BOOKS:

1. Data Mining concepts and Techniques by Jiawei Han, Micheline Kamber – Elsevier.
2. M. Kantardzic, “Data mining: Concepts, models, methods and algorithms, John Wiley & Sons Inc.
3. Business Intelligence by Rajiv Sabherwal, Irma Becerra-Fernandez, Wiley Publications, John Wiley & Sons, Inc.

Web Links:

1. <http://www.kdd.org/>
2. <http://www.statsoft.com/Textbook/Elementary-Statistics-Concepts>
3. <http://www.kdnuggets.com/websites/blogs.html>
4. <http://facweb.cs.depaul.edu/mobasher/classes/ect584/lecture.html>

LIST OF EXPERIMENTS:

Sr. No.	Practical Aim
1	Data Preprocessing Techniques in Standard Tool like Excel Miner/Mat Lab
2	Perform ETL on any standard dataset (Export – Import, Data Pump etc.)
3	Generating different types of graphs on different types of data.
4	Implement and simulate different classification algorithm on standard dataset
5	Implement and simulate different clustering algorithm on standard dataset
6	Future prediction on Data mining Tool
7	Computing association rule with TANAGRA and WEKA
8	Building decision tree with TANAGRA and WEKA. Error rate estimation using a Cross validation.
9	Generate intelligent report for enterprise data using BI tools
10	A Survey paper on latest research in Data Mining and Business Intelligence