

NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	Integrated B.Tech.(CSE)-MBA
Course Code:	CSI0803
Course Title:	Data Mining
Course Type:	Core
Year of Introduction:	2021-22

Credit Scheme

L	T	Practical Component				C
		LPW	PW	W	S	
3	0	2	-	-	-	4

Course Learning Outcomes (CLO):

At the end of the course, students will be able to –

1. interpret data types and understand the need of data mining
2. infer various data pre-processing techniques for the mining process
3. apply basic principles and algorithms used in practical data mining
4. evaluate and compare data mining algorithms based on different metrics and parameters for real life applications.

Syllabus:

Total Teaching hours: 30

Unit	Syllabus	Teaching hours
Unit-I	Introduction: Motivation and importance, different kinds of data, data mining functionalities, classification of data mining systems, major issues in data mining	02
Unit-II	Data Pre-processing: Data summarization, data cleaning, data integration and transformation, data reduction, data discretization and concept hierarchy generation, feature extraction, feature transformation, feature selection, introduction to Dimensionality Reduction, CUR decomposition	08
Unit-III	Mining Frequent Patterns, Associations and Correlations: Basic concept, efficient and scalable frequent item-set mining methods, mining various kind of association rules, from association mining to correlation analysis, Advanced Association Rule Techniques, Measuring the Quality of Rules.	06
Unit-IV	Classification and Prediction: Neural Network-Based Algorithms, Rule-Based Algorithms, Combining Techniques, accuracy and error measures, evaluation of the accuracy of a classifier or predictor.	04
Unit-V	Advanced Cluster Analysis: Types of data in cluster analysis, overview of major clustering methods, learning, Hierarchical Agglomerative Clustering, k-means Algorithm, Self-Organizing Maps, probabilistic model-based clustering, clustering high dimensional data	07
Unit-VI	Case studies: Applications of Distributed and parallel Data Mining. Advanced Techniques: Web Mining, Spatial Database Mining, Temporal Mining, And Multimedia Mining.	03

Self-Study: The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents

- Suggested Readings:
1. Jiawei Han and Micheline Kamber, Data mining: Concepts and Techniques, Morgan Kaufmann Publishers.
 2. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann
 3. Hand, Mannila, and Smyth., Principles of Data Mining, MIT Press
 4. Berry and Linoff, Mastering Data Mining, Wiley
 5. Delmater and Hancock, Data Mining Explained, Digital Press

Suggested List of Experiments:	Sr. No.	Title	Hours
	1	Implement Min-max normalization and z-score normalization for the given dataset.	02
	2	Apply the methods to find the missing values in the given dataset.	02
	3	Implement the suitable method with help of Quartile for outlier detection.	02
	4	Find frequent itemsets using an iterative level-wise approach based on candidate generation.	02
	5	Implement AND and OR gate using Perceptron Learning (self-implementation)	02
	6	Implement Ex-OR gate using backpropagation Neural Network. (self-implementation)	02
	7	Implement K-means clustering algorithm.	02
	8	Implement probability model based clustering algorithm.	02
	9	Getting acquainted with WEKA and R-programming.	02
	10	Use WEKA for classification.	02

Suggested Case List: -NA-