NIRMA UNIVERSITY

Institute:	Institute of Technology, School of Technology
Name of Programme:	MTech CSE, MTech CSE (Cyber Security)
Course Code:	6CS365ME25
Course Title:	Data Mining and Visualization
Course Type:	Department Elective-II
Year of Introduction:	2025-26

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

Course Learning Outcomes (CLO):
At the end of the course, students will be able to –

1. demonstrate data domains and corresponding analysis tasks, including multivariate data, networks, text and cartography (BL2)
2. apply data mining techniques to solve problems in other disciplines in a mathematical way (BL3)
3. analyse the data using various statistical tools (BL4) (BL4)
4. build a data visualisation system for interdisciplinary problems. (BL6)

Unit	Contents	Teaching Hours (Total 30)
Unit-I	**Data Understanding:** types of data, information and uncertainty, classes and attributes, interactions among attributes, relative distributions, and summary statistics. **Data Visualization:** using different tools - refine data and create, edit, alter, and display their visualizations (x-y graph, bar chart, pie chart, cube, etc.) **Data Quality:** inaccurate data, sparse data, missing data, insufficient data, imbalanced data **Social Challenges:** data ownership, data security, ethics and privacy	08
Unit-II	**Data Reduction and Feature Enhancement:** standardizing data, sampling data, using principal components to eliminate attributes, limitations and pitfalls of principal component analysis (PCA), the curse of dimensionality **Association Analysis:** association rule learning, the Apriori algorithm, FP-Growth, market basket analysis	07
Unit-III	**Showing Complex Data:** Organizational Models, Preattentive Variables, Sorting and Rearranging, Searching and Filtering, Datatips, Data Spotlight, Dynamic Queries, Data Brushing, Local Zooming, Sortable Table, Radial Table, Muti-Y Graphs, Treemap, Small Multiples	05
Unit-IV	**Model Selection and Validation:** training error and optimism, the Bayes error rate, inductive bias, the bias-variance trade-off, overfitting, **Ensemble Learning:** bootstrap aggregating (bagging), boosting, stacking/blending, random subspaces, random forests	05

Unit-V	**Recent Trends in Data Mining:** Recommender Systems, Reinforcement Learning, Active Learning, Semi-supervised Learning, Transfer Learning, Deep Learning, Data Stream Mining.	05

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

Suggested Readings/ References:

1. Jack G. Zheng, Data Visualization for Business Intelligence, Taylor & Francis
2. Jiawei Han & Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann
3. Jennifer Tidwell, Designing Interfaces, O'Reilly
4. Edward Tufte, The Visual Display of Quantitative Information, Graphics Press LLC
5. Ben Fry, Visualizing Data, O'Reilly
6. Noah Iliinsky, Julie Steele, Designing Data Visualizations, O'Reilly
7. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson.
8. Wes McKinney, Python for Data Analysis, O'Reilly
9. S. Nagabhushana, Data Warehousing OLAP and Data Mining, New Age Publishers.

Suggested List of Experiments:

Sr. No.	Name of Experiments/Exercises	Hours
1	(a): Aim: Data Domain selection and Identification of Characteristics of selected Dataset of different formats. 1. What data domain you have selected? 2. What is the information dataset contains? 3. Identify the characteristics of various fields of the dataset. (The distribution, inference etc.) 4. What is the insight (knowledge) we can generate for the selected dataset? 5. What are the patterns available in the dataset? (b): For selected dataset generate Five Number Summary using Python. Also generate mode and midrange, outlier detection using the concept of Quartile method and other. Compare the results	06
2	Case Study for Data visualization using Tableau. Use the dataset selected in practical -1 and design an interactive Dashboard for analyzing data for selected KPI	04
3	Data Preprocessing (Data Quality): a) Aim: Implement data smoothing and data normalization methods. Redundancy analysis using Pearson correlation and Chi- Square. Discretization by Intuitive Partitioning	04
4	Data Reduction and Feature Enhancement: Aim: Implement Dimensionality reduction and Feature selection technique with selected dataset	04
5	Data Analysis and Visualization Aim: Apply Classification Techniques for selected datasets to visualize the results	04

6	Data Analysis and Visualization Aim: Apply Clustering Techniques for selected datasets and visualize the results	04
7	Pattern Analysis and Visualization Aim: Implementation of various classification and regression techniques and visualize the result with selected visualization tools like Tableau.	04