

NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	Master of Computer Application (2-Years Programme)
Course Code:	3MCAD360
Course Title:	Data Mining and Visualization
Course Type:	Departmental Elective
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. summarize the process of associations, classification and regression
2. analyse the process of wrangling, exploring and analysing data
3. apply visualization techniques for visualizing streaming data of various domains
4. evaluate visualization tools to do data analysis.

Syllabus:

Total Teaching hours: 45

Unit	Syllabus	Teaching hours
Unit-I	Data Objects and attributes: Types of Data Objects and Attribute, Nominal Attributes, Binary Attributes, Ordinal Attributes, Numeric Attributes, Discrete versus Continuous Attributes	05
Unit-II	Introduction: Introduction to Data Mining, Concept Description: Characterization and Discrimination, Mining Frequent Patterns, Associations, and Correlations, Classification and Regression for Predictive Analysis, Cluster Analysis, Outlier Analysis, Database Systems, and Data Warehouses, Information Retrieval, Major Issues in Data Mining	06
Unit-III	Introduction to data visualization: Pixel-Oriented Visualization Techniques, Geometric Projection Visualization Techniques, Icon-Based Visualization Techniques, Hierarchical Visualization Techniques, Visualizing Complex Data, and Relations	07
Unit-IV	Introduction to Data Analysis: Data wrangling, exploring, analyzing, and communicating data, work with data in Python using libraries like NumPy and Pandas, explore weather trends, investigate a dataset, Data Wrangling: Data wrangling process of gathering, assessing, and cleaning data, use Python to wrangle data programmatically and prepare it for analysis	08
Unit-V	Data Visualization with Python: Introduction to and comparison of Python for data visualization, visualization using Python libraries and techniques, standard plotting libraries, Matplotlib and Seaborn, Data management in Python and connecting to MySQL, demos of visualizations in Python	08

Unit-VI	Streaming Visualizations: Real-time data visualizations of devices connected to IoT/cloud, Introduction to streaming data, solving problems with visualizations tools: cloud/IOT applications	06
Unit-VII	Case Study: Understanding the business problem, gathering the data, processing the data, analyzing the data, visualizing the data, presenting the visuals.	05

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/
References:

1. Jiawei Han, Micheline Kamber, Jian Pei, Data Mining Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems
2. McKinney, Wes. Python for data analysis: Data wrangling with Pandas, NumPy, and Python. " OReilly Media, Inc."
3. Tim Grobmann, Mario Dobler. Data Visualization with Python, OReilly Media, Inc.
4. Anthony Aragues. Visualizing Streaming Data, OReilly Media, Inc.
5. Ellis, Byron. Real-time analytics: Techniques to analyze and visualize streaming data. John, Wiley & Sons.
6. Guido Van Smit. Python for Data Analysis: A Complete Crash Course on Python for Data Science to Learn Essential Tools and Python Libraries, NumPy, Pandas, Jupyter Notebook, Analysis and Visualization

Suggested List of Experiments:	Sr. No.	Title	Hours
	1.	To study and implement the concept of association rule mining	02
	2.	To study and implement the concept of classification and regression	04
	3.	To study and perform cluster analysis and outlier analysis	04
	4.	To study and perform data preparation and data visualization	02
	5.	To study and perform data wrangling (gathering, assessing and cleaning) on twitter dataset	04
	6.	To study the use of Matplot to create plots to present data, and format the plots. Also add information to plots such as labels, titles, legends, etc	02
	7.	To study to get acquainted with the Seaborn plotting library. Plot data using Seaborn in a variety of different plots	02
	8.	To study and visualize the Graphs and Networks	02
	9.	To Perform Real-time data visualizations of devices connected to IoT/Cloud	04
	10.	To perform Sale Trend Visualization using Tableau - A tool for data visualization	04

Suggested Case List: -NA-