GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering
Level: PG
Subject Code : ME02000361
Subject Name : Data Preparation and Analysis

+

w. e. f. Academic Year:	A.Y. 2024-25
Semester:	2
Category of the Course:	Professional Elective Course

Prerequisite:	Fundamentals of Data Science
Rationale:	This course is offered for understanding the principles and techniques of preparing, cleaning, transforming, and analyzing data to extract meaningful insights and support data-driven decision-making.

Course Outcome:

After Completion of the Course, Student will able to:

No	Course Outcomes	RBT Level
1	Represent data in an informative manner using measures and visualizations.	Understand
2	Pre-process data to make it suitable for more accurate analysis and model building.	Apply
3	Implement correlation and regression analysis techniques for numerical data.	Apply
4	Implement classification techniques for categorical data.	Apply
5	Implement clustering and association mining methods.	Apply

**Revised Bloom's Taxonomy (RBT)*

Teaching and Examination Scheme:

Teaching Scheme (in Hours)			Total Credits L+T+ (PR/2)	Assessment Pattern and Marks				Total Marks
				Theory		Tutorial / Practical		
L	T	PR	C	ESE (E)	PA / CA (M)	PA/CA (I)	ESE (V)	
3	0	2	4	70	30	20	30	150

Course Content:

Unit No.	Content	No. of Hours	% of Weightage
1.	**Understanding Data** Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization (Pixel-Oriented, Geometric Projection, Icon-Based, Hierarchical), Measuring Data Similarity and Dissimilarity for Different Types of Attributes	07	15
2.	**Data Preprocessing** Data Preprocessing: An Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization	07	15
3.	**Correlation and Regression Analysis** Introduction to Correlation, Correlation Coefficients (Pearson and Spearman Rank), Auto-Correlation, Introduction to SLR(Simple Linear Regression), SLR Model Building, Estimation of Parameters Using OLS, Interpretation of SLR Coefficients	11	25
4.	**Classification** Basic Concepts, Decision Tree Induction,Bayes Classification Methods, Model Evaluation and Selection, Techniques to Improve Classification Accuracy, Bayesian Belief Networks, Classification by Backpropagation, Support Vector Machines, Classification Using Frequent Patterns, Lazy Learners (or Learning from Your Neighbors)	10	25
5.	**Clustering and Association** Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods, Cluster Analysis, Partitioning Methods, Hierarchical Methods	10	20
	Total	**45**	**100**

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks (in %)					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
10	**25**	**45**	**10**	**10**	**---**

Where R: Remember; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (as per Revised Bloom's Taxonomy)

References/Suggested Learning Resources:

(a) Books:

1. Data Mining. Concepts and Techniques -Jiawei Han, Micheline Kamber, Jian Pei, 3rd Edition, Morgan Kaufmann Pub
2. Business Analytics: The Science of Data - Driven Decision Making -U. Dinesh Kumar, 2nd Edition, Wiley
3. Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining - Glenn J. Myatt, Wiley
4. Machine Learning - Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, Pearson

(b) Open source software and website:

1. Python, WEKA
2. NPTEL MOOC Course – Data Mining - https://nptel.ac.in/courses/106105174
3. NPTEL MOOC Course - Data Analytics with Python - https://nptel.ac.in/courses/106107220
4. NPTEL MOOC Course - Python for Data Science - https://nptel.ac.in/courses/106106212

1. Rahul Dubey, “An Introduction to Internet of Things: Connecting Devices, Edge Gateway, and Cloud with Applications”, Cengage India Publication
2. Raj Kamal, “Internet of Things: Architecture and Design Principles, Mc Graw Hill Education
3. Hanes et al “IoT Fundamentals”, Cisco Press
4. Vijay Madisetti and Arshdeep Bahga, “Internet of Things (A Hands-on-Approach)”, , Paperback, 2015.
5. A. McEwen, H. Cassimally, “Designing the Internet of Things”, Wiley, 2013.
6. Yashwant Kanetkar, “21 Internet of Things Experiments”, Kindle edition
7. Adeel Javed, “Building Arduino projects for Internet of Things”, Apress publication
8. Donald Noris, “The Internet of Things: Do it yourself Projects with Arduino, Raspberry PI and
9. BeagleBone Black” Mc Graw Hill Publication
10. Adrian McEwen & Hakim Cassimally, “Designing the Internet of things”, Willey publication
11. Rahul Dubey, “An Introduction to Internet of Things: Connecting Devices, Edge Gateway, and Cloud with Applications”, Cengage India Publication
12. Raj Kamal, “Internet of Things: Architecture and Design Principles, Mc Graw Hill Education
13. Hanes et al “IoT Fundamentals”, Cisco Press
14. Vijay Madisetti and Arshdeep Bahga, “Internet of Things (A Hands-on-Approach)”, , Paperback, 2015.
15. A. McEwen, H. Cassimally, “Designing the Internet of Things”, Wiley, 2013.
16. [Yashwant Kanetkar, “21 Internet of Things Experiments”, Kindle edition
17. Adeel Javed, “Building Arduino projects for Internet of Things”, Apress publication
18. Donald Noris, “The Internet of Things: Do it yourself Projects with Arduino, Raspberry PI and
19. BeagleBone Black” Mc Graw Hill Publication
20. Adrian McEwen & Hakim Cassimally, “Designing the Internet of things”, Willey publication
21. Rahul Dubey, “An Introduction to Internet of Things: Connecting Devices, Edge Gateway, and Cloud with Applications”, Cengage India Publication

Suggested Course Practical List:

The list of practicals should be prepared for the implementations of various methods of data pre-processing and analysis using Python and WEKA.

List of Laboratory/Learning Resources Required:

Programming Languages (Python), WEKA

* * * * * * *