GUJARAT TECHNOLOGICAL UNIVERSITY
Syllabus for Integrated MSc, 5th Semester
Branch: Information Technology
Subject Name: Data Warehouse and Mining
Subject Code: 1350503

Teaching and Examination Scheme

Teaching Scheme			Credits	Examination Marks				Total Marks
				Theory Marks		Practical Marks		
L	T	P	C	ESE(E)	PA (M)	PA (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Content:

Sr. No.	Content	Teaching Hours	Module Weightage (%)
1.	**Data Warehouse:** What is it, Who Need It, and Why?, Things to Consider, Managing the Data Warehouse, Data Warehouse Design Methodology, Data Marts and Start Schema Design, Fundamentals of ETL Architecture, Partitioning Data, Indexing Data.	5	10
2.	**Data mining –** Introduction, Data mining on what kind of data , Data mining functionalities classification of Data mining systems, Major issues in Data mining **Mining Association rules in large databases** - Association rule mining, Mining single-Dimensional Boolean association rules from Transactional databases, Mining multi-Dimensional Association rules from relational Databases and Data Warehouses.	13	30
3.	**Classification and Prediction:** Classification vs. prediction, Issues regarding classification and prediction, Statistical-Based Algorithms, Distance-Based Algorithms, Decision Tree Based Algorithms, Neural Network-Based Algorithms, Rule-Based Algorithms, Combining Techniques, accuracy and error measures, evaluation of the accuracy of a classifier or predictor. Neural Network Prediction methods: Linear and nonlinear regression, Logistic Regression Introduction of tools such as DB Miner / WEKA / DTREG DM Tools	10	25
4.	**Cluster analysis –** Introduction types of data in cluster analysis a categorization of major clustering methods portioning methods, hierarchical methods, Density based methods,: DBSCAN, Grid-based method : STRING , Model based clustering method: Statistical Approach, outlier analysis..	10	25
5.	**Web Mining:** Introduction to Web Mining, Web content mining, Web usage mining, Web Structure mining, Web log structure.	2	10

Suggested Specification table with Marks (Theory):70

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
10	30	10	10	5	5

R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Reference Books:

1. J. Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann
2. M. Kantardzic, "Data mining: Concepts, models, methods and algorithms, John Wiley &Sons Inc.
3. M. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education.
4. Ning Tan, Vipin Kumar, Michael Steinbanch Pang, "Introduction to Data Mining", Pearson Education

Course Outcome:

After learning the course, the students should be able to:

CO-1	Perform the preprocessing of data and apply mining techniques on it.	20
CO-2	Identify the association rules, classification, and clusters in large data sets	30
CO-3	Solve real world problems in business and scientific information using data mining.	20
CO-4	Use data analysis tools for scientific applications.	15
CO-5	Implement various supervised machine learning algorithms.	15