GUJARAT TECHNOLOGICAL UNIVERSITY
Syllabus for Integrated MSc, 5th Semester
Branch: Computer Science
Subject Name: Data Mining
Subject Code: 1350303

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks				Total Marks
				Theory Marks		Practical Marks		
L	T	P	C	ESE(E)	PA (M)	PA (I)	ESE (V)	
4	0	2	5	70	30	20	30	150

Content:

Sr. No.	Content	Teaching Hours	Module Weightage (%)
1	**Introduction to data mining (DM):** Motivation for Data Mining - Data Mining-Definition and Functionalities – Classification of DM Systems - DM task primitives - Integration of a Data Mining system with a Database or a Data Warehouse - Issues in DM – KDD Process	5	20
2	**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	7	20
3	**Concept Description, Mining Frequent Patterns, Associations and Correlations:** What is concept description? - Data Generalization and summarization-based characterization - Attribute relevance - class comparisons, 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.	10	20
4	**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	20
5	**Cluster Analysis:** Clustering: Problem Definition, Clustering Overview, Evaluation of Clustering Algorithms, Partitioning Clustering -K-Means Algorithm, K- Means Additional issues, PAM Algorithm; Hierarchical Clustering – Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering, Strengths and Weakness; Outlier Detection, Clustering high dimensional data, clustering Graph and Network data.	10	20

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", PearsonEducation

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

Sr. No.	CO statement	Marks % weightage
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