GUJARAT TECHNOLOGICAL UNIVERSITY
INFORMATION TECHNOLOGY
B. E. SEMESTER: VII

Subject Name: **Data Warehousing and Data Mining**
Subject Code: **171601**

Teaching Scheme				Evaluation Scheme			
Theory	**Tutorial**	**Practical**	**Total**	**University Exam (E)**		**Mid Sem Exam (Theory) (M)**	**Practical (Internal)**
				Theory	**Practical**		
3	0	2	5	70	30	30	20

Sr. No.	Course Contents	Total Hrs
1.	**Introduction to Data Warehousing** Why reporting and Analyzing data, Raw data to valuable information-Lifecycle of Data - What is data warehousing - The building Blocks: Defining Features - Data warehouses and data marts - Overview of the components - Metadata in the data warehouse - Need for data warehousing - Basic elements of data warehousing - trends in data warehousing.	10
2.	**Introduction to Data Mining** Motivation for Data Mining - Data Mining: On What kind of Data? -Definition and Functionalities: What kind of patterns can be mined? - Classification of DM Systems – Integration of a Data Mining system with a Database or a Data Warehouse - Issues in DM – KDD Process	10
3.	**Data Preprocessing and Data Mining Primitives** Why Preprocess the Data? – Data Cleaning – Data Integration and Transformation – Data Reduction – Discretization and Concept Hierarchy Generation – Data Mining Primitives: What Defines a Data Mining Task?	08
4.	**Concept Description and Association Rule Mining** What is concept description? - Data Generalization and summarization-based characterization - Attribute relevance - class comparisons Association Rule Mining: Market basket analysis - basic concepts - Finding frequent item sets: Apriori algorithm - generating rules – Improved Apriori algorithm – Incremental ARM – Associative Classification – Rule Mining	10

5.	**Classification and Clustering** What is classification and prediction? – Issues regarding Classification and prediction: Classification methods: Decision tree, Bayesian Classification, Rule based, CART, Neural Network, CBR, Rough set Approach, Fuzzy Logic, Genetic Algorithms – Prediction methods: Linear and non linear regression, Logistic Regression – What is Cluster Analysis? – Types of Data in Cluster Analysis – A Categorization of Major Clustering Methods, Types of Clustering Algorithms.	12
6.	**Advance Topics of Data Mining and its Applications** Mining Time-Series and Sequence Data – Mining Text Databases – Mining the World Wide Web – Data Mining Application	04

Text Books:

1. J. Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann
2. Ian H. Witten & Eibe Frank, "Data Mining practical machine learning tools & technology", 2nd edition, Elsevier.Vikram Pudi, "Data Mining", Oxford.

Reference Books:

1. Reema Thareja, "Data Warehousing", Oxford.
2. Paulraj Ponnian, "Data Warehousing Fundamentals", John Willey.
3. Data Mining Techniques, second edition, Arun K pujari, Universities Press
4. Data Warehouse Fundamentals, Pulraj Ponniah
5. Introduction to Data Mining, Ping-Ning Tan, Vipin kumar, Steinbach, pearson
6. Introduction to Data Mining with Case Studies,G.K.Gupta, PHI
7. Data Mining: Concepts and Techniques, 2nd Edition, Han, Elsevier