

# GUJARAT UNIVERSITY

K. S. SCHOOL OF BUSINESS MANAGEMENT  
[Five Years' (Full-time) Integrated Degree Course]

## Semester-8 [M.Sc. (CA & IT)]

**Subject Code: - KS\_C\_CC-485**

**Subject Name: - Data Mining and Data Analytics (Theory)**

**Course Credit: - 3**

### Objective:

To understand the need of Data Warehouses over Databases, differentiate between RDBMS schemas & Data Warehouse Schemas, to understand the concept of Analytical Processing (OLAP) and its similarities & differences with respect to Transaction Processing (OLTP), to understand the need for Data Mining, Data Analytics and advantages to the business world, to get a clear idea of various classes of Data Analytics techniques, their need, scenarios (situations) and scope of their applicability.

Unit No.	Course Content	Weight-age (%)
1	<b>Data Warehousing and OLAP:</b> - Basic Concepts of Data Warehouse, Data Warehouse Modeling: Data Cube and OLAP, Data Warehouse Design and Usage, Data Warehouse Implementation, Data Generalization by Attribute-Oriented Induction	(20%)
2	<b>Data Preprocessing and Data Mining:</b> - An Overview of Data Preprocessing, Data Cleaning, Data Integration, Data Reduction, Introduction to Data Mining, Types of Data and Patterns that can be mined, Data Mining Applications	(20%)
3	<b>Mining Association Rules:</b> - Market Basket Analysis, Frequent and closed item sets, Association rules, Apriori Algorithm, Generating Association rules from frequent item sets, Pattern Evaluation Method	(20%)
4	<b>Classification and Regression:</b> - <b>Linear Correlation, Linear regression</b> , An Overview of Classification, Supervised vs. Unsupervised classification, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, <b>Support vector machine, Nearest Neighbor Classifier</b> , multiple linear regression	(20%)
5	<b>Cluster Analysis and Outlier Detection:</b> - An Overview, Typical Requirement of Clustering in Data Mining, K-means clustering, Types of Data in Cluster Analysis: Interval-scaled Variables, Binary Variables, Nominal, Ordinal & Ratio-scale variables, Overview of Major Clustering Methods, An Overview of Outlier and Outlier Detection Methods.	(20%)

**Recommended Lecture Scheme:** Approximately 40 to 45 hours in a semester

**Recommended Practical Scheme:** Not Applicable

**Assignment:** Five assignments should be given.

**Main Reference Books:**

1. Data Mining: Concepts & Techniques ,Third Edition, Morgan Kaufmann Publishers  
By Jiawei Han & Micheline Kamber

**Reference Books:**

1. Introduction to Data Mining with Case Studies, PHI  
By G.K. Gupta
2. Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals, Wiley-India By Paulraj Ponniah
3. Data Mining Methods & Models, Wiley-India  
By Daniel T. Larose
4. Data Mining, Oxford University Press  
By Vikram Pudi & P. Radhakrishnan
5. Data Warehousing, Data Mining & OLAP, TataMcGraw-Hill Michael  
By Alex Berson & Stephen J. Smith
6. Data Mining Techniques, Wiley-India  
By J. A. Berry & Gordon S. Linoff