GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Diploma Engineering
Level: Diploma
Branch: Information Technology
Subject Code: DI04016081
Subject Name: Data Mining and Warehousing

w. e. f. Academic Year:	2025-26
Semester:	4th
Category of the Course:	Professional Elective - II

Prerequisite:	Students should know the basics of databases and SQL queries. They should also be familiar with simple programming concepts and data structures. This background will help them to easily understand data warehousing and data mining concepts.
Rationale:	In today's world, a huge amount of data is generated every day. To make useful decisions, this data needs to be stored properly and analyzed effectively. This course helps students learn how to organize large data using data warehousing and how to find useful patterns and knowledge using data mining. By learning these skills, students will be able to understand real-life applications like business analysis, market trends, and customer behavior, which are very important in the IT industry.

Course Outcome:
After Completion of the Course, Student will able to:

No	Course Outcomes	RBT Level
01	Understand the data mining process and its practical applications.	U
02	Remember the basic concepts of data warehousing and differences between OLTP and OLAP.	R
03	Apply basic techniques like classification, clustering, and association rule mining.	A
04	Understand schema models, fact/dimension tables, and OLAP operations.	U
05	Apply simple data mining tools to analyze datasets and observe results.	A

**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(M)	PA(I)	ESE(V)	
3	0	2	4	70	30	20	30	150

Course Content:

Unit No.	Content	No. of Hours	% of Weightage
1.	**Fundamental Concepts of Data Mining** 1.1 Data Mining 1.1.1 Define data mining 1.1.2 Difference between data mining and database 1.1.3 Common terms used: patterns, trends, & predictions 1.2 Steps in Knowledge Discovery Process (KDD) 1.2.1 Data selection, Data preprocessing, Data mining 1.2.2 Interpretation and evaluation of results 1.3 Importance of Data Mining 1.4 Applications of Data Mining in real life 1.4.1 Education: analyzing student performance, Business: customer purchase analysis, Healthcare: disease prediction, Banking: fraud detection	08	18
2.	**Basics of Data Warehousing** 2.1 Meaning of data, information, and knowledge 2.1.1 Definition and examples of data 2.1.2 How data becomes information 2.1.3 What is knowledge 2.1.4 Real-life examples 2.2 Introduction to Data Warehouse 2.2.1 Meaning and purpose of a data warehouse 2.2.2 Key features 2.2.3 Difference between database and data warehouse 2.2.4 Importance in decision-making 2.3 Difference between OLTP and OLAP 2.3.1 Meaning of OLTP and OLAP 2.3.2 Major differences and Examples 2.4 Data Marts 2.4.1 Meaning of a data mart 2.4.2 Difference between data warehouse and data mart 2.4.3 Types of data marts 2.4.4 Uses of data marts in organizations	07	15

3.	**Data Mining Techniques** 3.1 Classification 3.1.1 Meaning of classification 3.1.2 Working of Decision Tree and Naïve Bayes 3.1.3 Real-life examples 3.2 Clustering 3.2.1 Meaning of clustering 3.2.2 Working of K-means 3.2.3 Applications of clustering 3.3 Association Rule Mining 3.3.1 Meaning of association rule mining 3.3.2 Basics of Apriori algorithm 3.3.3 Market Basket Analysis	12	27
4.	**Data Warehouse Design** 4.1 Star Schema and Snowflake Schema 4.1.1 Meaning of Star and Snowflake Schema 4.1.2 Key differences between Star and Snowflake, Examples 4.2 Fact Table and Dimension Table 4.2.1 Definition of Fact Table and Dimension Table 4.2.2 Relationship between fact and dimension tables, Examples 4.3 ETL (Extract, Transform, Load) process 4.4 Basic OLAP operations 4.4.1 Roll-up, Drill-down, Slice, Dice	10	22
5.	**Tools and Applications** 5.1 Introduction to data mining tools 5.1.1 Orange, RapidMiner, KNIME 5.2 Case studies of Data Mining 5.2.1 Business: Market Basket Analysis, Education: Student performance analysis, Healthcare: Disease prediction and patient data analysis 5.2.2 Results and insights from each case study 5.3 Future scope of Data Mining and Warehousing 5.3.1 Emerging applications 5.3.2 Role in decision-making and business intelligence 5.3.3 Ethical considerations	08	18
	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**
15	**25**	**30**	**-**	**-**	**-**

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:

Sr. No.	Title of Book	Author	Publication with place, year and ISBN
1	Data Mining: Concepts and Techniques	Jiawei Han, Micheline Kamber	Morgan Kaufmann / Elsevier, 978-0123814791
2	Data Warehousing: Concepts, Techniques, Products & Applications	C. S. R. Prabhu,	PHI Learning, India, 978-8120336278
3	Data Mining & Data Warehousing	K. A. Balasubramaniam, P. Arul Prabu, D. Prem Raja, L. Anitha,	Dr. BGR Publications, India, 2025, 978-81-991359-7-0
4	Data Warehousing & Data Mining	N. P. Gopalan & B. Sivaseelvan,	PHI Learning, India, 978-8120338128
5	Introduction to Data Mining with Case Studies	G. K. Gupta,	PHI Learning, India, 978-8120350021

(b) Open source software and website:

Open Source Software's

1. Orange
   - Open source machine learning and data visualization.
   - Great for clustering, classification, and visualization.
   - Website: https://orangedatamining.com/

2. RapidMiner (Community Edition)
   - Visual drag-and-drop data mining tool.
   - Supports machine learning, text mining, and predictive analytics.
   - Website: https://rapidminer.com/
3. KNIME Analytics Platform
   - Open-source platform for data analysis and reporting.
   - Good for hands-on ETL (Extract, Transform, Load) and data workflows.
   - Website: https://www.knime.com/
4. MySQL / MariaDB
   - Open-source database software.
   - Useful for creating a Data Warehouse schema (star, snowflake).
   - Website: https://www.mysql.com/

Useful Websites

1. UCI Machine Learning Repository
   - Collection of free datasets for practice.
   - Website: https://archive.ics.uci.edu/ml/index.php
2. Kaggle Datasets
   - Free datasets and beginner projects.
   - Website: https://www.kaggle.com/datasets
3. GitHub Repositories
   - Many free data mining sample projects and codes.
   - Website: https://github.com/
4. Google Dataset Search
   - Easy way to find open datasets online.
   - Website: https://datasetsearch.research.google.com/
5. Towards Data Science (Blog on Medium)
   - Beginner-friendly tutorials on data mining and data warehousing.
   - Website: https://towardsdatascience.com/

Suggested Course Practical List:

Sr. No.	Practical Outcomes (PrOs)	Unit No.	Approx. Hrs. required
1	Study of KDD (Knowledge Discovery in Databases) process with examples.	1	2

2	Identify and discuss real-life applications of data mining (group discussion / mini case study).	1	2
3	Study of Data Warehousing concepts through case studies/examples.	2	2
4	Compare OLTP and OLAP databases using simple SQL queries.	2	2
5	Perform Classification using Decision Tree in data mining tools.	3	2
6	Perform Classification using Naïve Bayes in data mining tools.	3	2
7	Apply Clustering using K-Means in data mining tools.	3	2
8	Apply Association Rule Mining using Apriori algorithm in data mining tools.	3	2
9	Simple project: Apply two techniques (Classification + Clustering) on a sample dataset and compare results.	3	2
10	Design a simple Star Schema for a sales database.	4	2
11	Design a Snowflake Schema for a student performance database.	4	2
12	Perform basic OLAP operations (Roll-up, Drill-down, Slice, Dice) using sample data in SQL.	4	2
13	Introduction and hands-on with Data Mining Tool.	5	2
14	Perform a case study analysis (e.g., Market Basket, Student Result Analysis, or Healthcare data).	5	2
15	Mini Project: Apply OLAP + one data mining technique on a dataset and present findings.	5	2
	Total		**30**

List of Laboratory/Learning Resources Required:

Sr. No.	Laboratory/Learning Resources/Equipment Name with Broad Specifications	PrO. No.
1	Computer system with operating system: Windows 7 or higher version, macOS, and Linux, with 4GB or higher RAM	All

Suggested Project List:

1. Student Performance Analysis
   - Collect marks data of students and apply classification to predict pass/fail.
2. Market Basket Analysis
   - Use Apriori algorithm to find association rules in shopping datasets (e.g., "bread → butter").
3. Library Book Borrowing Analysis
   - Create a small dataset of book transactions and apply clustering to group students by reading habits.
4. Online Shopping Trends
   - Analyze product purchase data using OLAP operations (roll-up, drill-down, slice, dice).
5. College Attendance Monitoring
   - Design a star schema for student attendance and analyze attendance trends.
6. Healthcare Data Mining
   - Use classification (Decision Tree/Naïve Bayes) to predict diseases from patient symptoms dataset.
7. Movie Rating Prediction
   - Use clustering to group movies by rating and user preferences.
8. Sales Data Warehouse
   - Create a small sales data warehouse (Star Schema) and perform OLAP queries for decision-making.
9. Social Media Post Analysis
   - Collect simple social media data (likes, shares) and analyze trends using clustering.
10. Mini Data Mining Tool Demo
    - Use data mining tool to compare two techniques (e.g., Decision Tree vs. Naïve Bayes) on the same dataset and present findings.

Suggested Activities for Students:

1. Prepare charts/posters showing differences between OLTP and OLAP systems.
2. Draw schemas (Star and Snowflake) for small sample data like sales or student records.
3. Collect real-life examples where data mining is used (e.g., online shopping, banking, healthcare).
4. Use data mining tool to perform simple classification and clustering on a given dataset.
5. Write a short report on how OLAP operations (Roll-up, Drill-down, Slice, Dice) can help in decision-making.
6. Mini-project in groups: Analyze a small dataset (like student marks or product sales) using one data mining technique and present results.
7. Case study discussion: Market Basket Analysis or Student Performance Analysis.
8. Prepare a presentation on future applications of Data Mining and Warehousing in daily life.

* * * * * * *