INSTITUTE	**FACULTY OF TECHNOLOGY**
PROGRAM	**BACHELOR OF TECHNOLOGY (COMPUTER ENGINEERING)**
SEMESTER	**5**
COURSE TITLE	**DATA WAREHOUSING & DATA MINING**
COURSE CODE	**01CE0527**
COURSE CREDITS	**3**

Objective:

1 The goal of the course is to understand and apply different data warehousing and data mining techniques where data warehousing techniques focus on storing historical data along with data analysis and data mining techniques focus on data preprocessing data and applying mining techniques for discovering the knowledge from data

Course Outcomes: After completion of this course, student will be able to:

1 Understand the principles and architecture of Data Warehousing

2 Understand the principles, architecture and process of Data Mining.

3 Apply data pre-processing techniques for data cleansing

4 Analyze and evaluate the algorithms used for identifying the frequent items and Association Rules

5 Apply Classification, Prediction and Clustering algorithms.

Pre-requisite of course: Understanding of Database Management Systems, Data Structures

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
2	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Data Warehousing** Introduction to Data Warehouse, Data warehouse vs Operational DBMS,, Multidimensional data model: Data cubes, Conceptual Modelling of Data warehouse: schemas, Multidimensional Data, OLAP operations, Comparison of OLAP with OLTP, Data Warehouse architecture,, OLAP Server Architecture: ROLAP, MOLAP	6
2	**Data Mining** Motivation and importance of data mining, Different kinds of data, Data mining functionalities, Classification of data mining systems, Data mining task primitives, Integration of a data mining system with a Database or a Data Warehouse, KDD Process	5

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
3	**Data Pre-processing** Data pre-processing overview, Major tasks in data pre-processing, Introduction to Data cleaning, Data cleaning process for incomplete and noisy data, Introduction to Data integration, Correlation Analysis for nominal and numeric data,, Covariance analysis for numeric data, Introduction to data transformation, Data normalization techniques, Box and whisker plot using five number summary, Quartile for outlier detection	8
4	**Mining Frequent Patterns, Associations and Correlations:** Basic concept of frequent pattern analysis, Market Basket Analysis, Apriori algorithm, Frequent item-set mining methods, Mining various kind of association rules, Measuring the Quality of Rules, Applications of association rules	5
5	**Data Mining Applications** Data mining applications in Healthcare, Finance, Fraud Detection, Enterprise Operations & Process Mining, E-commerce & Retail, Entertainment & Media, Professional Sports, Agriculture & Farming, Smart Cities & Urban Planning, Telecommunications, Airlines & Travel	4
Total Hours		**28**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Practical 1** Analysis and comparison of data mining and data warehousing tools.	2
2	**Practical 2** Analysis Weka modules.	2
3	**Practical 3** Demonstration of types datasets.	2
4	**Practical 4** Demonstration on pre-processing techniques.	2
5	**Practical 5** Apply pre-processing techniques on dataset.	2
6	**Practical 6** Demonstration on APRIORI algorithm.	2
7	**Practical 7** Apply APRIORI algorithm on dataset.	2
8	**Practical 8** Demonstration of dataset loading and exploration using Kaggle Notebooks.	2
9	**Practical 9** Demonstration of dataset loading and exploration using Google Colab.	2

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
10	**Practical 10** Apply and analyse data cleaning using Google Colab.	2
11	**Practical 11** Apply data mining techniques on dataset using Google Colab.	2
12	**Practical 12** Project on real life Data Mining applications.	2
Total Hours		**24**

Textbook :

1 Data mining concepts and techniques, J. Han ,M Kamber, Morgan Kaufmann 4th edition, 2022

References:

1 Data Mining Introductory and Advanced Topics, Data Mining Introductory and Advanced Topics, M.H. Dunham, Pearson Education., 2006

2 Introduction to Data Mining",, Introduction to Data Mining",, Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Pearson Education., 2006

3 Data warehousing- concepts, Techniques, Products and Applications, Data warehousing- concepts, Techniques, Products and Applications, Prabhu, Hall of India., -

4 Data Warehousing, Data Mining and OLAP, Data Warehousing, Data Mining and OLAP, Alex Berson and Stephen J. Smith,, Tata McGraw Hill Edition, Tenth Reprint, 1997

5 "Introduction to Data Mining, "Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Pearson Education, 2006

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
0.00	10.00	40.00	30.00	20.00	0.00

Instructional Method:

1 The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.

2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.

3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Instructional Method:

4 Students will use supplementary resources such as online videos, NPTEL videos, e-courses, Virtual Laboratory.

Supplementary Resources:

1 https://www.cs.waikato.ac.nz/ml/weka/
2 https://altair.com/altair-rapidminer
3 https://orangedatamining.com/
4 https://www.sas.com/en_us/insights/analytics/data-mining.html
5 https://onlinecourses.nptel.ac.in/noc21_cs06/preview
6 https://www.kaggle.com/docs/notebooks
7 https://colab.research.google.com/notebook