SIXTH SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
CE603T(i)	Data Warehousing and Mining (PE-II)	3	-	-	3	30	70	100

Course Objectives	Course Outcomes
This course is intended to ● Provide the concepts of data warehouse and data mining, ● Use tools and techniques for Knowledge Discovery in Databases preprocessing and classification ● Gain knowledge of how data is grouped using clustering techniques.	Student will be able to: ● Understand the basics of data warehousing, including how it differs from operational databases. ● Analyze the link between data warehousing and mining, grasp fundamentals, recognize patterns, identify technologies, and evaluate major mining issues. ● Analyze data processing by examining data quality, addressing missing values and noisy data through cleaning processes, exploring integration, reduction, and transformation strategies, ● Perform classification and clustering of data using suitable classification and clustering algorithms. ● Compare and contrast different classifiers, and different clustering methods.

Unit I [8 Hrs]
Data warehouse: Introduction to Data warehouse, Difference between operational database systems and data warehouses, Data warehouse Characteristics, Data warehouse Architecture and its Components, Extraction-Transformation-Loading, Logical(Multi-Dimensional), Data Modeling, Schema Design, Star and Snow-Flake Schema, Fact Constellation, Fact Table, Fully Addictive, Semi-Addictive, Non Addictive Measures; Fact-Less-Facts, Dimension Table Characteristics; OLAP Cube, OLAP Operations, OLAP Server Architecture-ROLAP, MOLAP and HOLAP.

Unit II [7 Hrs]
Data Mining: Fundamentals of data mining, Relationship of data warehousing and data mining, what is data mining, kind of data mined, kind of patterns mined, Technologies used for mining, kind of application targeted by mining, Major issues in data mining, Data mining applications

Unit III [7 Hrs]
Data Preprocessing: Data Processing An overview: Data Quality, Major Task in data Prepossessing. Data Cleaning-Missing Values, Noisy Data, Data cleaning as a Process. Data Integration, Data Reduction, Data Transformation Strategies Overview

Unit IV [8 Hrs]
Classification: Basic Concepts, General approaches to solving a classification problem, Decision Tree Induction-Attributes Selection Measure, Tree Pruning.Bayes Classification Methods: Bayes Theorem, Naive Bayesian Classification, Rule Based Classification: using If- then rules for classification, Rule Extraction from decision Tree. Classification by Back propagation -A multi-layer Feed forward Neural Network, Lazy Learners - K- Nearest Neighbor Classifiers

Unit V [7 Hrs]
Cluster Analysis: - what is cluster analysis, Requirements of Cluster Analysis, Overview of Categorization Basic Clustering Methods, Partitioning Methods, Hierarchical Clustering - Agglomerative Methods and divisive methods, key issues in hierarchical clustering. Outlier Detection: Types of outliers, Outliers Detection methods

Text Books

S.N	Title	Authors	Edition	Publisher
1	Data Mining Concepts and Techniques	Jiawei Han, Micheline Kamber and Jian Pei	3rd Edition	Elsevier

Reference Books

S.N	Title	Authors	Edition	Publisher
1	Iintroduction to Data Mining	Pang-Ning Tan, Michael S, Vipin Kumar	-	Pearson Education
2	Introduction to Data Mining with Case Studies	G. K. Gupta	Easter Economy Edition	PHI

		July 2023	1.1	Applicable for 2023-24
Chairman - BoS	Dean – Academics	Date of Release	Version	