



**B. Tech.
Semester VI**

DATA WAREHOUSING AND DATA MINING

XXXXXX

EFFECTIVE FROM July-2021

Syllabus version: 1.00

Subject Code	Subject Title	Teaching Scheme			
		Hours		Credits	
		Theory	Practical	Theory	Practical
XXXXXX	Data Warehousing and Data Mining	3	2	3	1

Subject Code	Subject Title	Theory Examination Marks		Practical Examination Marks	Total Marks
		Internal	External	CIE	
XXXXXX	Data Warehousing and Data Mining	40	60	50	150

Objectives of the course:

- To introduce students with basic concepts of data warehouse, data mining and its applications.
- To provide clarity of mathematical fundamentals for data mining algorithms.
- To familiarize with wide range of rule mining, classification and clustering algorithms.

Course outcomes:

Upon completion of the course, the student shall be able to

C01: Explore fundamentals of OLAP and cost-effective methods for maintaining data warehouse.

C02: Examine types of data to be mined and apply preprocessing statistical methods to it.

C03: Examine various association rule mining techniques to apply on real time applications.

C04: Discover interesting patterns from large amount of data using classification and prediction algorithms.

C05: Discover interesting patterns from large amount of data using clustering algorithms.

C06: Analyze various data mining applications using data mining functionalities.

Sr. No.	Topics	Hours
Unit – I		
1	Introduction to Data Warehousing: Basic concepts of data warehouse, Difference between operational database system and data warehouses, Need of data warehouse, Multidimensional data model - Data cube, Stars, Snowflakes, Fact constellations, Measures; OLAP operations in the multidimensional data model, Data warehouse architecture - Three-tier data warehouse architecture, models, ETL, Metadata repository.	6
Unit – II		
2	Introduction to Data Mining and Data Preprocessing: Basics of data mining, Knowledge Discovery from Data (KDD) process, Types of data, Data mining functionalities and issues, Need for data preprocessing, Data cleaning, Data integration, Data	11

	transformation, Data reduction, Discretization and concept hierarchy generation.	
Unit – III		
3	Mining Frequent Patterns, Associations and Correlations: Basic concepts - Market basket analysis, Frequent itemsets, Closed itemsets, Association rules; Frequent itemset mining methods - Apriori algorithm, Generating association rules from frequent itemsets, Improving the efficiency of Apriori, Pattern-Growth approach for mining frequent itemsets, Mining frequent itemsets using vertical data format, Mining closed and max patterns.	6
Unit – IV		
4	Classification: Introduction to classification and prediction, Issues regarding classification and prediction, Classification methods - Decision tree, Bayesian classification, Rule based classification; Introduction to other classification methods - k-Nearest Neighbor (KNN), Neural Network, Case-Based Reasoning (CBR), Rough set approach, Fuzzy logic, Genetic algorithms; Basics of prediction methods - Linear and nonlinear regression; Metrics for evaluating classifier performance.	10
Unit – V		
5	Clustering: Introduction to cluster analysis, Categorization of major clustering methods, Partitioning methods - k-Means, k-Medoids; Hierarchical method - Agglomerative and Divisive hierarchical clustering, Density based method - DBSCAN, Introduction to Grid-based methods.	6
Unit – VI		
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 applications.	6

Sr. No.	Data Warehousing and Data Mining (Practical)	Hours
1	Cases study of any three data mining applications and make a detailed note on them.	4
2	Write summarized description of any 10 data mining tools including Weka and Rapid Miner.	2
3	Perform following tasks as per given instruction: a. Apply Data Cleaning methods on given data set using Weka. b. To perform data cleaning by handling missing values. Preprocessing for the missing value, by replacing them with all the following for the given dataset: 1) With the global constant like Unknown or-infinity.	6

	2) Use the central tendency of attribute Mean (numerical attributes). 3) Use the class wise attribute Mean or median. 4) Apply all three normalization methods to any one numeric attribute.	
4	Perform following task as per given instructions. a. Simulation of Apriori algorithm using Weka tool. b. Implement the Apriori algorithm for frequent itemset mining for given data set. 1 2 1 3 4 2 5 7 1 2 3 2 3 5 6	4
5	Perform simulation of FP-Growth algorithm using Rapid Miner.	2
6	Implement the k-means clustering algorithm.	4
7	Create an ID3 based classification model for the given dataset.	4
8	Case study of any two advance data mining application.	4

Text book:

1. J. Han, M. Kamber - "Data Mining Concepts and Techniques", Morgan Kaufmann, 3rd Edition.

Reference books:

1. Ian H. Witten, Eibe Frank - "Data Mining practical machine learning tools & technology", Elsevier, 2nd edition.
2. Vikram Pudi - "Data Mining", Oxford.
3. Reema Thareja - "Data Warehousing", Oxford.
4. Paulraj Ponnian - "Data Warehousing Fundamentals", John Willey.
5. Ping-Ning Tan, Vipin Kumar, Steinbach – "Introduction to Data Mining", Pearson.
6. G. K. Gupta– "Introduction to Data Mining with Case Studies", PHI.

Course objectives and Course outcomes mapping:

- To introduce students with basic concepts of data warehouse, data mining and its applications: CO1, CO2, CO6
- To provide clarity of mathematical fundamentals for data mining algorithms.: CO2, CO3, CO4, and CO5
- To familiarize with wide range of rule mining, classification and clustering algorithms.: CO3, CO4, and CO5

Course units and Course outcomes mapping:

Unit No.	Unit Name	Course Outcomes					
		CO1	CO2	CO3	CO4	CO5	CO6

1	Introduction to Data Warehousing	✓					
2	Introduction to Data mining and Data Preprocessing		✓				
3	Mining Frequent Patterns, Associations and Correlations			✓			
4	Classification				✓		
5	Clustering					✓	
6	Advance Topics of Data Mining and its Applications						✓

Programme outcomes:

- PO 1: Engineering knowledge: An ability to apply knowledge of mathematics, science, and engineering.
- PO 2: Problem analysis: An ability to identify, formulates, and solves engineering problems.
- PO 3: Design/development of solutions: An ability to design a system, component, or process to meet desired needs within realistic constraints.
- PO 4: Conduct investigations of complex problems: An ability to use the techniques, skills, and modern engineering tools necessary for solving engineering problems.
- PO 5: Modern tool usage: The broad education and understanding of new engineering techniques necessary to solve engineering problems.
- PO 6: The engineer and society: Achieve professional success with an understanding and appreciation of ethical behaviour, social responsibility, and diversity, both as individuals and in team environments.
- PO 7: Environment and sustainability: Articulate a comprehensive world view that integrates diverse approaches to sustainability.
- PO 8: Ethics: Identify and demonstrate knowledge of ethical values in non-classroom activities, such as service learning, internships, and field work.
- PO 9: Individual and team work: An ability to function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
- PO 10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give/receive clear instructions.
- PO 11: Project management and finance: An ability to demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
- PO 12: Life-long learning: A recognition of the need for, and an ability to engage in life-long learning.

Programme outcomes and Course outcomes mapping:

Programme Outcomes	Course Outcomes					
	C01	C02	C03	C04	C05	C06
P01		✓	✓	✓	✓	
P02	✓					✓
P03			✓	✓	✓	
P04						✓
P05						
P06						
P07						
P08						
P09						✓
P010						
P011						
P012						