

DATA WAREHOUSING AND DATA MINING

Course Code	23CS3501	Year	III	Semester	I
Course Category	PCC	Branch	CSE	Course Type	Practical
Credits	3	L-T-P	3-0-0	Prerequisites	Data Structures, Algorithms, Probability & Statistics, Data Base Management Systems
Continuous Internal Evaluation :	30	Semester End Evaluation:	70	Total Marks:	100

Course Outcomes

Upon successful completion of the course, the student will be able to:

CO1	Understand the basic concepts of data mining, data warehousing, and machine learning.	L2
CO2	Apply supervised learning algorithms to solve classification and regression problems.	L3
CO3	Apply unsupervised learning techniques such as clustering for pattern discovery.	L3
CO4	Analyze real-world problems and choose appropriate machine learning algorithms for effective solutions.	L4

Syllabus

Unit No.	CONTENTS	Mapped CO
I	Data Warehousing and Data Mining: Types of Data, Types of Patterns Technologies, Applications, Major issues, Data Objects & Attribute Types, Statistical Descriptions of Data. Data Warehousing: Basic concepts, Data Warehouse Modeling: Data Cube and OLAP. Data Preprocessing: An Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization (Text Book-1)	CO1
II	Introduction to Machine Learning: Evolution of Machine Learning, Paradigms for ML, Types of Data, Stages in Machine Learning, Data Sets Used, Machine learning Applications. Nearest Neighbor-Based Models: Introduction to Proximity Measures, Distance Measures, Different Classification Algorithms Based on the Distance Measures, KNN Regression, Performance Measures. (Text Book-2)	CO1, CO2, CO4

III	Classification: Basic Concepts, Decision Tree Induction-Attribute Selection Measures, Tree Pruning. Bayes Classification Methods, Bayes Theorem, Naïve Bayes Classification, Techniques to improve Classification accuracy.(Text Book-1)	CO1,CO2, CO4
IV	Linear Discriminants for Machine Learning: Perceptron Classifier, Support Vector Machines, Linearly Non-Separable Case, Non-linear Support vector machine, Logistic Regression, Linear Regression. (Text Book -2)	CO1,CO2, CO4
V	Clustering: Introduction to Clustering, Clustering of Patterns, Divisive Clustering, Agglomerative Clustering, Partitional Clustering-K-Means Clustering, Soft Clustering-Fuzzy C-Means Clustering, Expectation Maximization-Based Clustering. (Text Book -2)	CO1, CO3, CO4

Learning Resources	
Text Books	
<ol style="list-style-type: none"> 1. Data Mining concepts and Techniques, 3rd edition, Jiawei Han, Michel Kamber, Elsevier, 2011. 2. “Machine Learning Theory and Practice”, M N Murthy, V S Ananthanarayana, Universities Press (India), 2024. 3. “Machine Learning”, Tom M. Mitchell, McGraw-Hill Publication, 2017 	
Reference Books	
<ol style="list-style-type: none"> 1. “Machine Learning: An Algorithmic Perspective”, Second Edition, Stephen Marsland, CRC Press 2. “Machine Learning in Action”, Peter Harrington, DreamTech 3. “Introduction to Data Mining”, Pang-Ning Tan, Michel Stenbach, Vipin Kumar, 7th Edition, 2019. 	
E-Resources & other digital material	
<ol style="list-style-type: none"> 1. https://www.coursera.org/learn/machine-learning 2. https://nptel.ac.in/courses/106/106/106106139/ 	