PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)

Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (Data Science)

III B. Tech – I Semester CSE (Data Science)

DATA WAREHOUSING AND DATA MINING

Course Code	20DS4501C	Year	III	Semester	I
Course Category	PEC	Branch	CSE(Data Science)	Course Type	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Data Structures DBMS
Continuous Internal Evaluation	30	Semester End Examination	70	Total Marks	100

Course Outcomes	
Upon successful completion of the course, the student will be able to	
CO1	Describe the fundamental concepts, principles, and techniques in Data Mining and Warehousing. **L2**
CO2	Apply supervised learning algorithms to build predictive models for classification problems. **L3**
CO3	Utilize unsupervised learning techniques to discover meaningful patterns and groupings within unlabeled data. **L3**
CO4	Analyze data mining problems, choose suitable algorithms, and critically assess their performance and limitations. **L4**

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of correlations (3:High, 2: Moderate, 1:Low)														
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	2													
CO2												2	3	
CO3												2	3	
CO4		3										2		

Syllabus		
Unit No.	**Contents**	**Mapped CO**
I	**Data Mining:** Introduction, Need of Data Mining, Definition, KDD process, Kinds of Data, Kinds of Patterns, Applications, Major Issues in Data Mining. **Data:** Introduction**,** Data Objects**,** Attribute Types**-**Nominal, Binary, Interval-Scaled, Ratio-Scaled, Discrete versus Continuous Attributes,Similarity and distance measures	CO1
II	**Data Warehousing:** Introduction, Definition, three-tier data warehousing architecture, OLTP vs OLAP Systems, Data Warehouse Models(Enterprise Warehouse, Data Mart, and Virtual), Warehouse Schemas for Multidimensional Data Models (Stars, Snowflakes, and Fact Constellations), Typical OLAP Operations.	CO1
III	**Classification: Introduction**, General Frame Work to Classification, Applications. **Decision Tree Induction**: Introduction, Decision Tree Representation, Attribute Selection Measures, Decision Tree Learning Algorithm, Tree Pruning, Issues in Decision Tree, Metrics for Evaluating Classifier Performance.	CO1, CO2 CO4
IV	**Mining Frequent Patterns, Associations**: Introduction, Frequent Itemsets, Association Rules, **Frequent Itemset Mining Methods:** Apriori Algorithm: Finding Frequent Itemsets by Confined Candidate Generation, Example, Pattern-Growth Approach for Mining Frequent Itemsets, Example.	CO1, CO3, CO4
V	**Cluster Analysis**: Introduction to Cluster Analysis, Basic Clustering Methods, Measures of Similarity and Dissimilarity, Metrics for Evaluating Clustering Performance, Applications. **Partitioning Methods**: K-Means and K-Medoids algorithms, Applications. **Hierarchical Methods**: Agglomerative and Divisive Approaches, Linkage Criteria.	CO1, CO3, CO4

Learning Resources
Text Books
1. Data Mining Concepts and Techniques: Jiawei Han, Micheline Kamber, Jian Pei, Third Edition, 2011, Morgan Kaufmann
Reference Books
1. Introduction to Data Mining, PANG-NING TAN, MICHAEL STEINBACH, ANUJ KARPATNE, VIPIN KUMAR, Second Edition, 2021, Pearson Education
e- Resources & other digital material
1.https://nptel.ac.in/courses/106105174