

DSPC2001 DATA MINING AND DATA WAREHOUSING (3-0-0)

1. Be familiar with mathematical foundations of data mining tools..
2. Understand and implement classical models and algorithms in data warehouses and data mining
3. Characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering.
4. Master data mining techniques in various applications like social, scientific and environmental context.
5. Develop skill in selecting the appropriate data mining algorithm for solving practical problems.

Module - I

Data Mining overview, Data Warehouse and OLAP Technology Data Warehouse Architecture, Steps for the Design and Construction of Data Warehouses, A Three-Tier Data Warehouse Architecture, OLAP , OLAP Queries, Metadata Repository, Data Preprocessing – Data Integration and Transformation, Data Reduction, Data Mining Primitives, System Architectures – Data Mining Primitives: What Defines a Data Mining Task? Task-Relevant Data, The Kind of Knowledge to be Mined, KDD

Module – II

Mining Association Rules in Large Databases, Association Rule Mining, Market Basket Analysis: Association Rule Mining, Basic Concepts, Association Rule Mining A Road Map, Mining Association Rules from Frequent Itemsets, Mining Multilevel Association Rules from Transaction Databases, Multilevel Association Rules, Approaches to Mining Multilevel Association Rules, Mining Distance-Based Association Rules, From Association Mining to Correlation Analysis

Module - III

Classification and Prediction – What is Classification? What Is Prediction? Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Bayes Theorem, Classification by Back propagation, A Multilayer Feed-Forward Neural Network, MLP, RBFN, Defining a Network Topology, Classification Based of Concepts from Association Rule Mining, Other Classification Methods, k-Nearest Neighbor Classifiers, Genetic Algorithms, Fuzzy Set Approaches, Prediction, Linear and Multiple Regression, Nonlinear Regression, Other Regression Models, Classifier Accuracy

Module – IV

Cluster Analysis – What Is Cluster Analysis, Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods, Classical Partitioning Methods: k-Means and k Medoids, Partitioning Methods in Large Databases: k-Medoids, Hierarchical Methods, Agglomerative and Divisive Hierarchical Clustering, Clustering Using Wavelet Transformation, Clustering High-Dimensional Space, Model-Based Clustering Methods, Statistical Approach, Neural Network Approach, LVQ, SOM, Mining Time-Series and Sequence Data, Mining Text Databases, Mining the World Wide Web. Applications and Trends in Data Mining – Data Mining Applications, Data Mining System Products.

Course Outcome:

CO1: Learn Data Mining overview, Data Warehouse and OLAP Technology

- CO2: Gain knowledge in Data Mining Primitives, System Architectures, Mining Association Rules in Large Databases
- CO3: Learn Classification and Prediction, Classification by Back propagation, Categorization of Major Clustering Methods
- CO4: Be familiar with Applications and Trends in Data Mining.

Textbooks:

1. Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber, Morgan Kaufmann Publisher (Elseviers)
2. Data Mining Concepts, Models, Methods and Algorithms By Mehmed Kantardzic Wiley Interscience, IEEE Press.

References:

1. Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
2. Data Ware Housing Fundamentals, Pualraj Ponnaiah, Wiley Student Edition.
3. The Data Ware House Life Cycle Toolkit- Ralph Kimball, Wiley Student Edition.
4. Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University.