

DATA MINING AND WAREHOUSING

Introduction to Data Mining and Warehousing

Basic Concepts: Data Mining, kinds of patterns that can be mined, Data Mining versus Database systems, Data preparation, cleaning and visualization.

Data Warehousing: Differences between database systems and Data Warehouse, Data Warehouse architecture and its components, Warehouse versus Data Mining (OLTP & OLAP), OLAP tools, Data cubes, Multidimensional Data.

Data Mining Techniques

Association Rules: Association rule, Mining association rules, frequent sets and border sets, algorithms for mining association rules – A priori algorithm, partition algorithm, Pincer-search algorithm, Border algorithm, FP-tree growth algorithm

Clustering: Hierarchical versus Partitional clustering, types of data in clustering, Partitional algorithms – k-means, k-medoids, PAM, CLARA, CLARANS. Density based clustering algorithm – DBSCAN. Hierarchical algorithms – BIRCH, CURE. Categorical clustering algorithms – ROCK, CACTUS.

Decision Trees : Introduction, tree construction principle.

Other techniques for Data Mining : Introduction to Genetic algorithms, Artificial Neural Network and Rough sets and their application in the domain of data mining. Introduction to Web Mining, Text Mining, Temporal data mining.

Recommended Readings

- K. Puzari, “Data Mining Techniques”, University Press
- Jiawei Han & Micheline Kamber, “Data Mining Concepts and Techniques”, Morgan Kaufmann & Harcourt India.
- K. Jain and R. C. Dukes, “Algorithms for Clustering Data”, Prentice-Hall