

635 - Data Analytics

Course Objectives: To Learn about the most effective data analytics methods to solve problems and achieve insight.

Course Contents:

Data: types of data, data quality, data preprocessing, measures of similarity and dissimilarity, Exploring data: Iris dataset, summary statistics, visualization.

Data Warehousing and OALP; Aggregation Operations, models for data Warehousing, star schema, fact and dimension tables Life cycle of data warehouse development, data warehouse architecture, efficient methods of data cube computation, relationship between data warehouse and data mining.

Association Analysis: Frequent itemset generation, rule generation, compact representation of frequent itemsets, FP-growth algorithm, evaluation of association patterns.

Association Analysis: advanced topics: handling categorical and continuous attributes, Handling concept hierarchy, sequential patterns, subgraph patterns, infrequent patterns.

Classification: general approach, decision tree induction, model overfitting, evaluating performance of a classifier, methods of comparing classifiers.

Classification: alternative techniques: rule based classifier, nearest neighbor classifier, Bayesian classifier, ANN, SVM, ensemble methods.

Cluster Analysis: K-means, agglomerative hierarchical clustering, DBSCAN, cluster evaluation, density based clustering, BIRCH, CURE.

Anomaly detection: statistical approaches, proxy-based outlier detection, density-based outlier detection, cluster based techniques.

Main Reading:

1. Peng-Ning Tan, Michael Steinbach and Vipin Kumar, Introduction to Data Mining, Pearson Education.
2. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques," 1st Edition Indian Reprint 2001, Harcourt India Private Limited, ISBN 1-55860-489-8.
3. Arun K Pujari, "Data Mining Techniques". Universities Press.