

DATA MINING

Course Code: 22CD1103

L	T	P	C
3	0	0	3

COURSE OUTCOMES: At the end of the Course the student shall be able to

- CO1: Interpret the given data statistically. (L2)
- CO2: Classify data using various Classification techniques. (L3)
- CO3: Apply Association rule mining to find frequent patterns (L3)
- CO4: Build models using unsupervised learning techniques. (L4)
- CO5: Apply different techniques for Anomaly Detection. (L3)

UNIT-I

10 Lectures

INTRODUCTION:

Data Mining, Data Mining tasks, Types of Data, Data Quality, Data Processing, Measures of Similarity and Dissimilarity, Exploring Data: Data Set, Summary Statistics, Visualization, OLAP and multidimensional data Analysis.

Learning Outcomes: At the end of the module the student will be able to

1. Identify the appropriate preprocessing technique to apply on the data. (L1)
2. Describe different multi-dimensional analysis techniques. (L2)
3. Interpret summary statistics of the data. (L2)

UNIT-II

10 Lectures

CLASSIFICATION:

Basic Concepts, Decision Trees, and model evaluation: General approach for solving a classification problem, Decision Tree induction, Model overfitting: Due to presence of noise, due to lack of representation samples, Evaluating the performance of classifier, Rule based classifier, Nearest Neighborhood classifier, Bayesian Classifier.

Learning Outcomes: At the end of the module the student will be able to

1. Model the classifier appropriate to the problem. (L3)
2. Predict the class labels using appropriate classification algorithm. (L3)
3. Build models for classification and evaluate its performance. (L3)

UNIT-III

10 Lectures

ASSOCIATION ANALYSIS:

Problem Definition, Frequent Item-set generation, Rule generation, compact representation of frequent item sets, FP-Growth Algorithm, Handling categorical and continuous attributes, concept hierarchy, sequential patterns: Problem formulation, sequential pattern discovery, Apriori-like method, Candidate generation, Candidate pruning, support counting.

Learning Outcomes: At the end of the module the student will be able to

1. Use algorithms to mine frequent patterns. (L3)
2. Demonstrate methods of sequential pattern discovery. (L3)
3. Identify frequent subgraph patterns. (L1)

UNIT-IV

10

Lectures

CLUSTERING:

Overview, Types of Clustering, Types of clusters, K-means, Basic Agglomerative Hierarchical clustering algorithm, DBSCAN, Cluster evaluation: Supervised and unsupervised cluster evaluation.

Learning Outcomes: At the end of the module the student will be able to

1. Compare various clustering methods. (L2)
2. Examine given data to identify relevant clusters. (L3)
3. Use cluster evaluation methods to identify the quality clusters. (L3)

UNIT-V

10 Lectures

ANOMALY DETECTION:

Preliminaries: Causes of Anomalies, Approaches of Anomaly Detection, Issues.

Statistical approaches: Detecting outliers in a univariate Normal Distribution, Outliers in a Multivariate Normal Distribution, A mixture model approach for anomaly Detection, Proximity-based outlier detection, Density-based outlier detection, Clustering-based outlier detection

Learning Outcomes: At the end of the module the student will be able to

1. identify the causes of anomalies in data. (L1)
2. summarize outlier detection methods. (L2)
3. apply appropriate statistical approaches to detect outliers. (L3)

TEXT BOOKS:

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, *Introduction to DataMining*, 2nd edition, Pearson education, 2018.

REFERENCE BOOKS:

1. Jiawei Han & Micheline Kamber, *Data Mining, "Concepts and Techniques"*, 3rd edition, Morgan Kaufmann Publishers, 2012.
2. Margaret H Dunham, *Data Mining Introductory and advanced topics*, 6th edition, Pearson Education, 2009
3. Arun K Pujari, *Data Mining Techniques*, 1st edition, University Press, 2005.
4. GK Gupta, *Introduction to Data Mining with Case Studies*, 3rd edition, Prentice Hall, 2014.
5. [K. P. Soman](#), [Shyam Diwakar](#), [V. Ajay](#), *Data Mining Theory and Practice*, 1st edition, PHI, 2006.

WEB REFERENCES:

1. https://onlinecourses.nptel.ac.in/noc21_cs06/preview
2. <https://www.coursera.org/specializations/data-mining>