

DATA WAREHOUSING AND DATA MINING

(Common to CSE & IT)

Course Code :13CT1122

| L | T | P | C |
|---|---|---|---|
| 4 | 1 | 0 | 3 |

Course Educational Objectives:

To introduce the student to various data warehousing and data mining techniques. The course will cover all the issues of KDD process and will illustrate the whole process by examples of practical applications.

- ❖ To make the student capable of applying data mining techniques in real time applications.
- ❖ To make the student capable to compare and contrast different conceptions of data mining as evidenced in both research and application.
- ❖ Explain the role of finding associations in commercial market basket data.
- ❖ Identify and characterize sources of noise, redundancy, and outliers in presented data.
- ❖ To get an idea about the data that how it is going to be classified into clusters

Course Outcomes:

At the end of the course the student will be able to

- ❖ Understand the application of data mining techniques in real time applications
- ❖ Understand the Comparing and contrast different conceptions of data mining
- ❖ Understand the finding associations in commercial market basket data

- ❖ Understand the identifying and characterizing the noise, redundancy and outliers in presented data
- ❖ Understand about the clusters.

UNIT-I

(12 Lectures)

INTRODUCTION:

Data mining-On what kinds of Data, what kinds of patterns can be mined, which technologies are used, which kinds of applications are targeted, major issues in Data Mining.

DATA PREPROCESSING: An Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data discretization.

UNIT-II

(12 Lectures)

DATA WAREHOUSE AND OLAP TECHNOLOGY:

Data Warehouse: Basic concepts, Data Warehouse Modeling: Data Cube and OLAP, Data Warehouse Implementation, Data cube computation method- Multi way array aggregation for full cube computation.

DATA GENERALISATION: Data generalization by Attribute-Oriented Induction.

UNIT-III

(12 Lectures)

MINING FREQUENT PATTERNS, ASSOCIATION AND CORRELATIONS:

Basic Concepts, Efficient and Scalable Frequent Item set Mining Methods, Which patterns Are interesting?-Pattern Evaluation methods.

UNIT-IV

(12 Lectures)

CLASSIFICATION: BASIC CONCEPTS:

Classification by Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Classification by Back propagation.

UNIT-V

(12 Lectures)

CLUSTER ANALYSIS: BASIC CONCEPTS:

Cluster analysis, Partitioning Methods (k- Means, k- Medoids) Hierarchical Methods: Agglomerative Vs Divisive, (BIRCH), Density-Based Methods: DBSCAN, Grid-Based Methods (STING).

TEXT BOOKS:

- 1 Jlawei Han & Micheline Kamber, “*Data Mining – Concepts and Techniques*”, 3rd Edition, Morgan Kaufmann Publishers, 2011.
- 2 Margaret H Dunham, “*Data Mining Introductory and advanced topics*”, 6th Edition, Pearson Education, 2009.

REFERENCES:

1. Arun K Pujari, “*Data Mining Techniques*”, 1st Edition, University Press, 2005.
2. Pang- Ning Tan, Michael Steinbach, Vipin Kumar, “*Introduction to Data Mining*”, 1st Edition, Pearson Education, 2012.
3. Sam Aanhory & Dennis Murray, “*Data Warehousing in the Real World*”, 1st Edition, Pearson Edn Asia, 2008.
4. Paulraj Ponnaiah, “*Data Warehousing Fundamentals*”, 1st Edition, Wiley student Edition, 2007.
5. Ralph Kimball, “*The Data Warehouse Life Cycle Tool Kit*”, 2nd Edition, Wiley student Edition, 2005.

