CA5T2	DATA WAREHOUSING AND DATA MINING	Credits : 4

Lecture Hours : 4 periods / week

Internal assessment : 30   Marks

Semester and Examination: 70   Marks

Course Description:

       This course makes the students to understand different concepts and techniques of data mining and data warehousing, including architecture, design, implementation and applications. To make the students appreciate the Association Rules for Transactional databases. To make the learners aware of various classification and prediction methods. To make the students understand various clustering algorithms.

Course Objectives:
- Compare and contrast different conceptions of data mining as evidenced in IT Sector.
- Explain the role of finding associations in commercial market basket data.
- Characterize the kinds of patterns that can be discovered by association rule mining.
- Describe how to extend a relational system to find patterns using association rules.
- Evaluate methodological issues underlying the effective application of data mining.
- Identify and characterize sources of noise, redundancy, and outliers in presented data.
- Identify mechanisms (on-line aggregation, anytime behavior, interactive visualization) to close the loop in the data mining process.
- Describe why the various close-the-loop processes improve the effectiveness of data mining.

UNIT-I :
Introduction: Fundamentals of data mining, Data Mining Functionalities, Classification of Data Mining systems, Major issues in Data Mining, Data Warehouse and OLAP Technology for Data Mining Data Warehouse, Multidimensional Data Model, Data Warehouse Architecture, Data Warehouse Implementation, Further Development of Data Cube Technology, From Data Warehousing to Data Mining.

UNIT-II :
Data Preprocessing: Needs Preprocessing the Data, Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation, Online Data Storage.

UNIT-III :
Data Mining Primitives, Languages, and System Architectures: Data Mining Primitives, Data Mining Query Languages, Designing Graphical User Interfaces Based on a Data Mining Query Language Architectures of Data Mining Systems,

UNIT-IV :
Concepts Description: Characterization and Comparison: Data
Generalization and Summarization- Based Characterization, Analytical Characterization: Analysis of Attribute Relevance, Mining Class **Comparisons:** Discriminating between Different Classes, Mining Descriptive Statistical Measures in Large Databases.

UNIT-V :

Mining Association Rules in Large Databases: Association Rule Mining, Mining SingleDimensional Boolean Association Rules from Transactional Databases, Mining Multilevel Association Rules from Transaction Databases, Mining Multidimensional Association Rules from Relational Databases and Data Warehouses, From Association Mining to Correlation Analysis, Constraint-Based Association Mining.

UNIT-VI :

Classification and Prediction: Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Classification by Back propagation, Classification Based on Concepts from Association Rule Mining, Other Classification Methods, Prediction, Classifier Accuracy.

UNIT-VII :

Cluster Analysis Introduction : Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods, Partitioning Methods, Density-Based Methods, Grid-Based Methods, ModelBased Clustering Methods, Outlier Analysis.

UNIT-VIII :

Mining Complex Types of Data: Multidimensional Analysis and Descriptive Mining of Complex, Data Objects, Mining Spatial Databases, Mining Multimedia Databases, Mining Time-Series and Sequence Data, Mining Text Databases, Mining the World Wide Web.

Learning Resources

Text Books:
1. Data Mining – Concepts and Techniques - JIAWEI HAN & MICHELINE KAMBER Harcourt India, 2/e,2001.
2. Data Mining Techniques – ARUN K PUJARI, University Press, 2/e, 2010.
3. Building the DataWarehouse- W. H. Inmon, Wiley Dreamtech India Pvt. Ltd.,4/e, 2005.

Reference Books:
1. Data Warehousing in the Real World – SAM ANAHORY & DENNIS MURRAY, Pearson Education, 2004.
2. Data Warehousing Fundamentals – PAULRAJ PONNAIAH WILEY STUDENT EDITION, 2/e, 2001.
3. The Data Warehouse Life cycle Tool kit – RALPH KIMBALL WILEY STUDENT EDITION, 2/e, 2011.