DATA WAREHOUSING AND DATA MINING

Course code: 13IT2114   **L   P   C**
                                                              4   0   3

Pre requisites: Database Management Systems.

Course Educational Objectives:
To introduce the student to various data warehouses 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.

1. To make the student capable of applying data mining techniques in real time applications.
2. To make the student capable to compare and contrast different conceptions of data mining as evidenced in both research and application.
3. Explain the role of finding associations in commercial market basket data.
4. Identify and characterize sources of noise, redundancy, and outliers in presented data.
5. 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

1. Understand the application of data mining techniques in real time applications
2. Understand the Comparing and contrast different conceptions of data mining
3. Understand the finding associations in commercial market basket data
4. Understand the identifying and characterizing the noise, redundancy and outliers in presented data
5. Understand about the clusters

UNIT- I
Introduction: Data mining-On what kinds of Data, Data Mining Functionalities, Classification of Data Mining systems, Data Mining Task Primitives, Integration of a Data Mining System with a Database or Data Warehouse System, Major issues in Data Mining.

Data Preprocessing: Descriptive data summarization, Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation.

UNIT-II
Data Warehouse and OLAP Technology: Multidimensional Data Model, Data Warehouse Architecture, Data Warehouse Implementation, From Data Warehousing to Data Mining.
Data Cube Computation and Data Generalization: Efficient methods for Data Cube Computation, Further Development of Data Cube and OLAP Technology, Attribute-Oriented Induction.

UNIT-III
Mining Frequent Patterns, Association and Correlations: Basic Concepts, Efficient and Scalable Frequent Item set Mining Methods, Mining Various kinds of Association Rules, From Association Mining to Correlation Analysis, Constraint Based Association.

UNIT-IV
Classification and Prediction-1: Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Classification by Backpropagation.
Classification and Prediction-2: Support Vector Machines, Association Classification, Other Classification Methods, Prediction, Accuracy and Error Measures, Evaluating the Accuracy of a Classifier or Predictor.

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

Text Books:

1. Jlawei Han & Micheline Kamber, *Data Mining – Concepts and Techniques*, 3rd Edition, Morgan Kaufmann Publishers, 2008.

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.