DATA WAREHOUSING AND DATA MINING

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

Course Outcomes:
At the end of the course, a student will be able to:
CO1: Interpret the problem in real time applications, and apply the appropriate data mining technique for effective results.
CO2: Analyze the different conceptions of data mining and compute OLAP operations.
CO3: Infer associations and correlations in the given Information domain.
CO4: Apply basic techniques of classification for various applications like banks, health organizations.
CO5: Analyze the real time data using various clustering techniques.

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.