DATA WAREHOUSING AND DATA MINING

Course Code: 15CT2101

L	P	C
3	0	3

Pre requisites: Database Management Systems.

Course Outcomes:
At the end of the course, a student will be able to
CO1: Apply data pre-processing techniques.
CO2: Design data warehouse schema.
CO3: Discover associations and correlations in givendata.
CO4: Apply classification techniques.
CO5: Apply clustering techniques.

UNIT- I (10-Lectures)

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 (10-Lectures)

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 (10-Lectures)

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 (10-Lectures)

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 (10-Lectures)

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.

Web References: www.thearling.com/text/admwhite/dmwhite.html