IV Year B.Tech. ECM I - Semester

L	T	P	To	C
4	-	-	4	4

CS425 DATAWAREHOUSING & DATAMINING (ELECTIVE - IV)

Course Description and Objectives:

This course is about knowing of how to make use of historical data so that high end business decision can be taken for the growth of an organization. The main objective of this course is to designing the intelligent machines which can take risk business decisions behalf of humans using the datamining techniques like classification, clustering, outlier detection, association rule mining.

Course Outcomes:

Students are able to

· *Learn the basic concepts of Database Technology Evaluation steps and also understood the need of data mining and its functionalities*
· *Explore the efficient and effective maintenance of Data Warehouses.*
· *Apply the data mining functionalities like Clustering, Classification, Association Analysis to real world data.*
· *Discover interesting patterns and association rules from huge volume of data used to do classifications and predictions.*
· *Gain knowledge on developing areas like Web Mining, Text Mining, and Spatial Mining.*

UNIT I - Introduction & Data Warehousing and Online Analytical Processing

Why Data Mining, What is Data Mining, Kinds of Data, Kinds of Patterns, and Technologies used, Kinds of applications adopted, Major issues in Data Mining. Basic Concepts, Data Warehouse Modeling, Data Warehouse Design and Usage, Data Warehouse Implementation, Data Generalization by Attribute-Oriented Induction

UNIT II - About Data & Data Preprocessing :

Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization, Measuring Data Similarity and Dissimilarity. An Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization

UNIT III - Data Cube Technology, Mining Frequent Patterns, Associations, and Correlations & Advanced Pattern Mining :

Preliminary Concepts, Data Cube Computation Methods, Processing Advanced Kinds of Queries by Exploring Cube Technology, Multidimensional Data Analysis in Cube Space

Basic Concepts and Methods: Basic Concepts, Frequent Itemset Mining Methods, Which Patterns Are Interesting?—Pattern Evaluation Methods

Pattern Mining in Multilevel, Multidimensional Space, Constraint-Based Frequent Pattern Mining

UNIT IV - Classification & Advanced Classification :

Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Model Evaluation and Selection, Techniques to Improve Classification Accuracy

Bayesian Belief Networks, Classification by Back propagation, Support Vector Machines, Classification Using Frequent Patterns, Lazy Learners, Other Classification Methods

UNIT V - Cluster Analysis & Advanced Cluster Analysis :

Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Evaluation of Clustering

Probabilistic Model-Based Clustering, Clustering High-Dimensional Data

TEXT BOOKS:

1. Jiawei Han Micheline Kamber – "Data Mining Concepts & Techniques", Third Edition, Morgan Kaufmann Publishers, 2012.

REFERENCE BOOKS :

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining", First Edition,2012.
2. Ralph Kimball, Margy Ross, "The Data Warehouse Toolkit", first edition John Wiley and Sons Inc., 2002.
3. Alex Berson, Stephen Smith, Kurt Thearling, "Building Data Mining Applications for CRM", first edition, Tata McGraw Hill, 2000.
4. Margaret Dunham, "Data Mining: Introductory and Advanced Topics", first edition, Prentice Hall, 2002.
5. Paulraj Ponnaiah, "Data Warehousing Fundamentals", first edition, Wiley Publishers, 2001.