CA6ELT01- DATA MINING (Core)

Theory:4 hrs. per week

Credits:4

Unit 1: (12 hrs.)

Introduction Data Mining, Data Ware House, Transactional Databases, Data Mining Functionalities Characterization and Discrimination, Mining frequent patterns, Association and correlation, Classification and Prediction, Cluster Analysis, Classification of Data Mining Systems, Data Mining Task Primitive, Integration of Data Mining systems, Major issues in Data Mining, Data integration and transformation, Data reduction, Data discretization.

Unit 2: (12 hrs.)

Data Warehouse and OLAP technology Data Warehouse, Multidimensional data Model, Data warehouse architecture, Data Warehouse implementation, OLAP, Data Warehouse and data mining.

Unit 3: (18 hrs.)

Association Rules and Classification Concepts Efficient and Scalable Frequent item set Mining methods, Mining various kind of association rules, from association mining to Co-relation analysis, Classification and prediction, Issues, Classification by Decision tree induction, Bayesian Classification, Rule-based classification, Support Vector Machines, Learning from your neighbors, Prediction.

Unit 4: (18 hrs.)

Cluster Analysis Definition, Types of data in cluster analysis, A categorization major Clustering methods- Partitioning methods, K-means and k-medoids, from k-medoids to CLARANS, Hierarchical methods, Density based methods.

Unit 5: (12 hrs.)

Mining Complex Data Spatial Data Mining, Multimedia Data Mining, Text Mining and Mining WWW.

Book of study:

1. Jiawei Han and Micheline Kamber - Data Mining - Concepts and Techniques,Second Edition, Elsevier, 2006

Reference:

1. Witten and Frank - Data Mining Practical Machine Learning Tools and Techniques, Second Edition, Elsevier, 2005

2. Soman, Divakar and Ajay, Data Mining Theory and Practice, PHI, 2006

3. Margaret H Dunham- Data Mining –Introductory and Advanced Topics, Fourth Edition, Person

2006