Generic Elective – IV

GE33241 Data mining and Warehousing

Course Objective:

1. Understand data mining principles and techniques.
2. Building basic terminology.
3. Learning how to gather and analyze large sets of data to gain useful business understanding.
4. Learning how to produce a quantitative analysis report/memo with the necessary information to make decisions.
5. Describing and demonstrating basic data mining algorithms, methods, and tools.
6. Identifying business applications of data mining.
7. Overview of the developing areas - web mining, text mining, and ethical aspects of data mining.
8. Develop and apply critical thinking, problem-solving, and decision-making skills.

Learning Outcome:

At the end of the course, the student should be able to:

1. Learn the concepts of database technology evolutionary path which has led to the need for data mining and its applications.
2. Examine the types of the data to be mined and present a general classification of tasks and primitives to integrate a data mining system.
3. Apply preprocessing statistical methods for any given raw data.
4. Explore DWH and OLAP, and devise efficient & cost effective methods for maintaining DWHs.
5. Discover interesting patterns from large amounts of data to analyze and extract patterns to solve problems, make predictions of outcomes.
6. Comprehend the roles that data mining plays in various fields and manipulate different data mining techniques.
7. Select and apply proper data mining algorithms to build analytical applications.

Course Contents:

Module	Course Topics	Total Hours	Credits
I	**Data Mining** Overview , Motivation(for Data Mining),Data Mining, Definition & Functionalities, Data Processing, Form of Data Preprocessing ,Data Cleaning: Missing Values ,Noisy Data, Inconsistent Data, Data Integration and Transformation. , Data Reduction, Data Cube Aggregation, Dimensionality reduction, Data Compression, Numerosity Reduction, Clustering, Discretization and Concept Hierarchy generation.	30 Hours	1

II	**Data Mining Statistics and Association rule** Concept Description, Definition, Data Generalization, Analytical Characterization, Analysis of attribute relevance, Mining Class comparisons ,Statistical measures in large Databases, Measuring Central Tendency, Measuring Dispersion of Data ,Range ,Quartiles, Outliers, Box plots, Variance, Standard Deviation, Graph Displays of Basic Statistical class Description, Mining Association Rules in Large Databases, Association rule mining, Mining Single-Dimensional Boolean Association rules from transactional Databases– A priori Algorithm, Mining Multilevel Association rules from Transaction Databases, Mining Multi-Dimensional Association rules from Relational Databases.	30 Hours	1
III	**Classification and Predictions** What is Classification & Prediction ,Issues regarding Classification and prediction, Decision tree, Bayesian Classification ,Classification by Back propagation, Multilayer feed-forward Neural Network, Back propagation Algorithm, Classification methods ,K nearest neighbor classifiers, Genetic Algorithm, Cluster Analysis ,Data types in cluster analysis, Categories of clustering methods, Partitioning methods, Hierarchical Clustering-,CURE and Chameleon, Density Based Methods-DBSCAN, OPTICS, Grid Based Methods-STING, CLIQUE ,Model Based Method –Statistical Approach, Neural Network approach, Outlier Analysis.	30 Hours	1
IV	**Data Warehousing and OLAP** Overview ,Definition, Delivery Process ,Difference between Database System and Data Warehouse, Multi-Dimensional Data Model, Data Cubes ,Stars ,Snow Flakes ,Fact Constellations ,Concept hierarchy, ProcessArchitecture,ThreeTierArchitecture,DataMarting, Aggregation,Historical information ,Query Facility, OLAP function and Tools, OLAP Servers, ROLAP, MOLAP, HOLAP, Data Mining interface, Security, Backup and Recovery, Tuning Data Warehouse, Testing Data Warehouse.	30 Hours	1

Text/Reference Books:

1. M.H. Dunham, “Data Mining: Introductory and Advanced Topics” Pearson Education
2. Jiawei Han, Micheline Kamber, “Data Mining Concepts & Techniques”, Elsevier
3. Sam Anahory, Dennis Murray, “Data Warehousing in the Real World : A Practical Guide for Building Decision Support Systems, 1/e”, Pearson Education
4. Mallach, “Data Warehousing System”, McGraw –Hill