



No:-

Date:

CS64120: Data Mining

L-T-P-Cr: 3-0-2-4

Objectives:

- Explaining the importance data warehousing and data mining.
- Learning the knowledge discovery process.
- Learning data mining tasks and study their well-known techniques.

Course Outcomes:

At the end of the course, a student should have:

Sl. No.	Outcome	Mapping to POs
1.	To learn data mining tasks and pre-processing activities.	PO4, PO2
2.	Understandability of data warehouse, architecture, schema designs, OLAP operations and servers.	PO2, PO3
3.	Learning market basket analysis and association rules, understanding multilevel and multi-dimensional rules and its generation techniques.	PO3
4.	Understanding various data classification and prediction techniques.	PO3
5.	Learning various clustering techniques that are used in different types of data.	PO3

Unit 1

Lecture 7

Data Mining, Data Mining task primitives, Integration of Data Mining system with the database, Major issues in Data Mining, Data Pre-processing, Descriptive data summarization, Data cleaning, Data integration and transformation, Data reduction, Data Discretization.

Unit 2

Lecture 8

Data Warehouse, Multidimensional data model, Data Warehouse architecture, Three tier Data Warehouse architecture, Metadata repository, Types of OLAP servers, Data Warehousing to Data Mining.

Unit 3

Lecture 10

Frequent patterns, Market basket analysis, Association Rule, Support and Confidence, overview of multilevel association rule, multidimensional association rule, closed itemset, maximal itemset, Apriori algorithm, Generating association rule from frequent itemset, Mining frequent itemsets without candidate generation (FP- growth), Mining multilevel association rules, Mining multidimensional association rules, Mining quantitative association rules, Association analysis to correlation analysis.

Unit 4

Lecture 9

Classification and Prediction: Classification by Decision Tree Induction, Attribute selection measures, Bayes Theorem, Predicting a class label using Bayesian classification, A multilayer feed forward neural network, Classification by Backpropagation, Prediction: Linear Regression, Nonlinear Regression.

Unit 5

Lecture 9

Cluster Analysis: Types of Data in Cluster Analysis, Categorization of the major clustering methods, Partitioning methods: k-Means and k-Medoids, Hierarchical Methods: Agglomerative and Divisive Hierarchical Clustering, Balanced Iterative Reducing and clustering using hierarchies, Density Based Methods: DBSCAN, Grid Based Methods: STING, Model based Clustering Methods: Expectation-Maximization.

Text Books:

1. Data Mining Concepts and Techniques by Jiawei Han, Micheline Kamber, Elsevier.

Reference books:

2. Data Mining. A tutorial-based Primer by Roiger, Michael W. Geatz and Pearson Education.
3. Data Mining- Introductory and Advanced Topics by Margaret H. Dunham, Pearson