राष्ट्रीय प्रौद्योगिकी संस्थान पटना / NATIONAL INSTITUE OF TECHNOLOGY PATNA
संगणक विज्ञान एंव अभियांत्रिकी विभाग / DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
अशोक राजपथ, पटना-८००००५, बिहार / ASHOK RAJPATH, PATNA-800005, BIHAR
Phone No.: 0612-2372715, 2370419, 2370843, 2371929 Ext- 200, 202 Fax-0612-2670631 Website: www.nitp.ac.in

No:- Date:

CSXX1918: Data Mining and Warehousing L-T-P-Cr: 3-0-0-3

Objectives:

- Explaining the importance data warehousing and data mining.
- Understanding data warehousing and OLAP operations
- Learning the knowledge discovery process.
- Learning market basket analysis and association rules.
- Different data classification and clustering techniques.

Course Outcomes:

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

Sl. No.	Outcome	Mapping to POs
1.	Application of pre-processing techniques on the dataset.	PO1, PO2
2.	Understanding of data warehouse and its architecture, schema designs, and OLAP operations.	PO1, PO2, PO3
3.	Understanding market basket analysis, multilevel and multi-dimensional association rules. Generation methods and application of association rules.	PO1, PO2, PO3
4.	Understanding various data classification and prediction techniques.	PO1, PO2, PO3
5.	Apply various clustering techniques that are used in different types of data.	PO1, PO2, PO3

Unit 1 Lecture 6

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, Heirarchical Methods: Agglomerative and Divisive Heirarchical 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