|  |  |
| --- | --- |
| MONO | **DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**  **NATIONAL INSTITUTE OF TECHNOLOGY PATNA**  Ashok Raj Path, PATNA 800 005 (Bihar), India |
| Phone No.: 0612 – 2372715, 2370419, 2370843, 2371929, 2371930, 2371715 Fax – 0612- 2670631 Website: [www.nitp.ac.in](http://www.nitp.ac.in/) |

***CS6403 Data Mining & Warehousing***

**L-T-P-Cr: 3-1-0-4**

**Pre-requisites:** Fundamentals of Algorithms

**Objectives/Overview:**

* 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** |
|  | To learn data mining tasks and pre-processing activities. | PO4, PO2 |
|  | Understandability of data warehouse, architecture, schema designs, OLAP operations and servers. | PO2, PO3 |
|  | Learning market basket analysis and association rules, understanding multilevel and multi-dimensional rules and its generation techniques. | PO3 |
|  | Understanding various data classification and prediction techniques. | PO3 |
|  | Learning various clustering techniques that are used in different types of data. | PO3 |

**UNIT I: Lectures: 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 II: Lectures: 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 III: Lectures: 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 IV: Lectures: 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 V: Lectures: 8**

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/ Reference Books:**

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