

**NATIONAL INSTITUTE OF TECHNOLOGY, JAMSHEDPUR**  
**Department of Computer Applications**  
FOURTH SEMESTER 2020-2021  
**Course Handout**

Dated: 20-01-2021

**Course Code** : CA3405  
**Course Title** : Data Warehousing and Data Mining (Elective – III)  
**Course Structure** : 3-1-0  
**Course Instructor** : Dr. D. K. Shaw

**Syllabus**

**Unit – I**

**Data Warehousing:** Data Warehousing Components, Building a Data Warehouse, Warehouse Database, Mapping the Data Warehouse to a Multiprocessor Architecture, DBMS Schemas for Decision Support, Data Extraction, Cleanup and Transformation Tools, Metadata.

**Unit – II**

**Business Analysis:** Reporting and Query Tools and Applications, Tool Categories, The Need for Applications, On line Analytical Processing (OLAP), Multidimensional Data Model, Multidimensional Vs Multi-relational OLAP, OLAP Tools and the Internet.

**Unit – III**

**Data Mining:** Introduction, Data and their Types, Data Mining Functionalities, Mining Frequent Patterns, Classification of Data Mining, Data Mining Task Primitives, Integration of Data Mining System with Data Warehouse, Data Mining Issues, Data Preprocessing.

**Unit – IV**

**Association Rule Mining and Classification:** Associations and Correlations, Mining Methods, Mining Various Kinds of Association Rules, Correlation Analysis, Constraint Based Association Mining, Classification and Prediction- Basic Concepts, Decision Tree Induction, Bayesian Classification, Rule Based Classification, Classification by Back Propagation, Support Vector Machine (SVM), Methods of Prediction.

**Unit – V**

**Clustering and Trends in Data Mining:** Cluster Analysis, Categorization of Major Clustering Methods- Partitioning Methods, Hierarchical Methods, Density Based Methods, Grid Based Methods, Evaluation of Clustering, Advanced Cluster Analysis- Probabilistic Model Based Clustering, Clustering High Dimensional Data, Clustering Graph and Network Data, Outlier Analysis, Applications of Data Mining, Trends in Data Mining.

## Books

### Text Books

1. Jiawei Han and Micheline Kamber: Data mining: Concepts and Techniques, 3<sup>rd</sup> Edition, Morgan Kaufmann.
2. Alex Berson and Stephen J. Smith: Data Warehousing, Data Mining & OLAP, Tata McGraw Hill.

### Reference Book

1. Marakas: Modern Data Warehousing, Mining and Visualization, Pearson Education Pvt. Ltd.
2. Varsha Bhosale and Deepali Vora: Data Warehousing & Data Mining, Technical Publications
3. D. Hand, H. Mannila and P. Smyth: Principles of Data Mining, PHI
4. S. Prabhu: Data Warehousing: Concepts, Techniques, Products and Applications, 2/e, New Age International
5. Rajeev Parida: Principles and Implementation of Data Warehousing, Laxmi Publication

### Evaluation Scheme

Sl. No.	Component	Duration (minutes)	% Weightage	Marks	Date & Time
1	Mid Term Exam.	120	30	30	Academic Calendar
2	End Semester Exam.	180	50	50	Academic Calendar
3	Assignment Submission	Before End Term Exam.	10	10	Academic Calendar
4	Class Test	50	5	5	To be announced
5	Attendance for Theory & Tutorial		5	5	

**Course Instructor**  
**CA3405**