Semester 1

Course No.:	Name: Data Warehouse And Data Mining	Credits: 3-0-0-6	Prerequisites: NIL

Course Objectives:

1. To understand the fundamental concepts of data warehousing, including architecture, extraction, transformation, and loading.
2. To explore various techniques of data preprocessing and cleaning.
3. To learn various data mining techniques such as classification, clustering, and association rule mining.
4. To gain knowledge of advanced data mining topics such as text mining, multimedia mining, and web mining.

Course Outcomes:

1. Students will be able to design, build, and maintain data warehouses and perform efficient data extraction, transformation, and loading.
2. Students will be able to analyze, preprocess, and clean data to prepare it for data mining.
3. Students will be able to apply various data mining techniques for knowledge discovery in large datasets.
4. Students will be able to identify and solve problems related to data mining and warehousing and understand its applications in various industries.

MODULE 1 - DATA WAREHOUSING AND BUSINESS INTELLIGENCE (9 hours):

- Data warehousing components
- Building a Data warehouse
- Data Warehouse Architecture
- DBMS Schemas for Decision Support
- Data Extraction, Clean-up, and Transformation Tools
- Metadata
- Reporting, Query tools and Applications
- Online Analytical Processing (OLAP)
- OLAP and Multidimensional Data Analysis.

MODULE 2 - PREPROCESSING AND DATA MINING TECHNIQUES (9 hours):

- Data Mining Functionalities
- Data Pre-processing
- Data Cleaning
- Data Integration and Transformation
- Data Reduction
- Data Discretization and Concept Hierarchy Generation
- Association Rule Mining: Efficient and Scalable Frequent Itemset Mining Methods
- Mining Various Kinds of Association Rules
- Association Mining to Correlation Analysis
- Constraint-Based Association Mining.

MODULE 3 - CLASSIFICATION AND PREDICTION (9 hours):

- Issues Regarding Classification and Prediction
- Classification by Decision Tree Introduction
- Bayesian Classification
- Rule-Based Classification
- Classification by Backpropagation
- Support Vector Machines
- Associative Classification
- Lazy Learners
- Other Classification Methods
- Prediction
- Accuracy and Error Measures
- Evaluating the Accuracy of a Classifier or Predictor
- Ensemble Methods
- Model Selection.

MODULE 4 - CLUSTER ANALYSIS (6 hours):

- Types of Data in Cluster Analysis
- A Categorization of Major Clustering Methods
- Partitioning Methods
- Hierarchical methods
- Density-Based Methods
- Grid-Based Methods
- Model-Based Clustering Methods
- Clustering High Dimensional Data
- Constraint-Based Cluster Analysis
- Outlier Analysis.

MODULE 5 - ADVANCED TOPICS (6 hours):

- Multidimensional Analysis and Descriptive Mining of Complex Data Objects
- Spatial Data Mining
- Multimedia Data Mining
- Text Mining
- Mining the World Wide Web.

ASSIGNMENTS:

1.Data Warehousing:
- Choose a real-world dataset and implement data extraction, cleaning, and transformation tools to build a data warehouse.
- Use data reporting and query tools to perform online analytical processing (OLAP) and multidimensional data analysis.

2.Data Mining:
- Choose a real-world dataset and perform data pre-processing techniques such as data cleaning, integration, transformation, and reduction.
- Implement and evaluate different association rule mining algorithms on the pre-processed dataset to find correlations between different variables.

3.Classification and Prediction:
- Choose a real-world dataset and perform data pre-processing techniques to prepare the data for classification and prediction.
- Implement and evaluate different classification algorithms such as decision trees, Bayesian classification, support vector machines, and associative classification on the preprocessed dataset.

4.Cluster Analysis:

- Choose a real-world dataset and perform data pre-processing techniques to prepare the data for clustering.
- Implement and evaluate different clustering algorithms such as partitioning methods, hierarchical methods, density-based methods, and model-based clustering methods on the pre-processed dataset.

5.Mining Object, Spatial, Multimedia, Text and Web Data:
- Choose a real-world dataset and perform multidimensional analysis and descriptive mining of complex data objects.
- Perform spatial data mining, multimedia data mining, text mining, and mining the World Wide Web on the chosen dataset.

TEXT BOOKS:
1. Jiawei Han, Micheline Kamber and Jian Pei"Data Mining Concepts and Techniques", Third Edition, Elsevier, 2011.
2. Alex Berson and Stephen J. Smith "Data Warehousing, Data Mining & OLAP", Tata McGraw – Hill Edition, Tenth Reprint 2007.
3. K.P. Soman, Shyam Diwakar and V. Ajay "Insight into Data mining Theory and Practice", Easter Economy Edition, Prentice Hall of India, 2006.
4. G. K. Gupta "Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006.
5. Pang-Ning Tan, Michael Steinbach and Vipin Kumar "Introduction to Data Mining", Pearson Education, 2007.