

# Syllabus

Subject: Data Mining  
Credit: 3

Course Code: COMP 482  
F.M: 100 (50 Internal + 50 Final)

## ***Objectives***

- To introduce students to the basic concepts and techniques of Data Mining.
- To explain data mining methodology.
- To develop skills of using recent data mining software for solving practical problems.
- To use visual techniques to describe data.
- To gain experience of doing independent study and research.

### **1. Introduction to Data Mining**

- 1.1. What is / Why data mining?
- 1.2. What kinds of data can be mined?
  - 1.2.1. Database data, data warehouses, transactional data
- 1.3. Which technologies are used?
  - 1.3.1. Machine Learning, DBMS, OLAP, Statistics, Information Retrieval
- 1.4. Applications of data mining
  - 1.4.1. Business intelligence and web search engines
- 1.5. Data Mining Goals
- 1.6. Stages of the Data Mining Process
- 1.7. Data Mining Techniques
- 1.8. Knowledge Representation Methods
- 1.9. Example: weather data

### **2. Data Warehouse and OLAP**

- 2.1 Data Warehouse and DBMS
- 2.2 Multidimensional data model
- 2.3 OLAP operations
- 2.4 Example: loan data set

### **3. Data Preprocessing**

- 3.1. Data cleaning
  - 3.1.1. Missing values, noisy data, inconsistent data
- 3.2. Data integration and transformation

- 3.3. Data reduction
- 3.4. Discretization and generating concept hierarchies
- 3.5. Installing Weka 3 Data Mining System
- 3.6. Experiments with Weka - filters, discretization

#### **4. Data Mining Knowledge Representation**

- 4.1. Task relevant data
- 4.2. Background knowledge
- 4.3. Interestingness measures
- 4.4. Representing input data and output knowledge
- 4.5. Visualization techniques
- 4.6. Experiments with Weka - visualization

#### **5. Attribute-Oriented Analysis**

- 5.1. Attribute generalization
- 5.2. Attribute relevance
- 5.3. Class comparison
- 5.4. Statistical measures
- 5.5. Experiments with Weka - using filters and statistics

#### **6. Data Mining Algorithms: Association Rules**

- 6.1. Motivation and terminology
- 6.2. Example: mining weather data
- 6.3. Basic idea: item sets
- 6.4. Generating item sets and rules efficiently
- 6.5. Correlation analysis
- 6.6. Experiments with Weka - mining association rules

#### **7. Data Mining Algorithms: Classification**

- 7.1. Basic learning/mining tasks
- 7.2. Inferring rudimentary rules: 1R algorithm
- 7.3. Decision trees
- 7.4. Covering rules
- 7.5. Experiments with Weka - decision trees, rules

#### **8. Data Mining Algorithms: Prediction**

- 8.1. The prediction task
- 8.2. Statistical (Bayesian) classification
- 8.3. Bayesian networks

- 8.4. Instance-based methods (nearest neighbor)
- 8.5. Linear models
- 8.6. Experiments with Weka - Prediction

## **9. Data Mining Algorithms: Clustering**

- 9.1. Basic issues in clustering
- 9.2. Partitioning methods: k-means, expectation maximization (EM)
- 9.3. Hierarchical methods: distance-based agglomerative and divisible clustering
- 9.4. Conceptual clustering: Cobweb
- 9.5. Experiments with Weka - k-means, EM, Cobweb

## **10. Data Mining Algorithms: Outlier Detection**

- 10.1 What are outliers?
- 10.2 Types of outliers
- 10.3 Challenges of outliers' detections
- 10.4 Outlier detection method
  - 10.4.1 Supervised, Semi-Supervised, and Unsupervised Methods
  - 10.4.2 Statistical Methods, Proximity-Based Methods
  - 10.4.3 Clustering-Based Methods
  - 10.4.4 Classification Based Methods
- 10.5 Experiment with Weka – Interquartile Range

## **References**

1. Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques* (3rd edition). Morgan Kaufmann, 2012. ISBN 978-0-12-381479-1.
2. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining* (1<sup>st</sup> edition). Pearson, 2016. ISBN 978-93-3257-140-2.
3. Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques* (3rd edition). Morgan Kaufmann, 2011. ISBN 978-0-12-374856-0. (Available as an e-book through the Athabasca University Library)
4. Udit Agarwal. *Data Mining and Data Warehousing* (2<sup>nd</sup> edition). S.K. Kataria & Sons, 2016. ISBN 978-93-5014-490-9.

**Lecture type**

Lectures will be delivered through slides presentation. All the lectures will be highly interactive with active participation of students and demonstration of real life examples.

**Note:** *Reading materials will be provided throughout the semester for further readings. It includes research papers, case studies, reports and articles.*

**Grading Policy:**

S.N	Item	Grade
1.	Internal (Subjective and MCQ)	20
2.	Mini Project + Term Paper	20
3.	Assignments	10