

**Department Of Computer Science and Engineering**  
**Kathmandu University**  
**Dhulikhel, Kavre**



**Subject: Data Mining**  
**Level: B.E./B.Sc 4<sup>th</sup> Year**

**Course Code: COMP 482**  
**Credit Hours: 3**

**Course 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

**CHAPTERS**

**Chapter 1: Introduction to Data Mining [3 hrs]**

- 1.1. Data mining
- 1.2. Types of data and data sources
- 1.3. Technologies used
- 1.4. Applications of data mining
- 1.5. Data Mining Goals
- 1.6. Data Mining Methodologies
  - a. CRISP-DM, SEMMA, KDD
- 1.7. Representing input data and output knowledge
- 1.8. Common Tools for Data Mining
  - a. Weka, Orange, R etc.
- 1.9. Data Mining and ethics

**Chapter 2: Data Exploration, Preprocessing and Visualization [8 hrs]**

- 2.1. Non-graphical and graphical methods for exploring data
- 2.2. Data cleaning: Missing values, noisy data, inconsistent data
- 2.3. Data integration
- 2.4. Data transformation
- 2.5. Data reduction
- 2.6. Discretization and generating concept hierarchies
- 2.7. Data visualization tools and techniques

## **Chapter 3: Data Warehouse and OLAP [3 hrs]**

- 3.1. Data Warehouse and DBMS
- 3.2. ETL (Extract, Transform, Load) vs ELT (Extract, Load, Transform)
- 3.3. Data Warehouse vs Data Mart vs Data Lake
- 3.4. Multidimensional data model
- 3.5. OLTP (Online Transaction Processing) vs OLAP (Online Analytical Processing)
- 3.6. OLAP operations: Roll-up, drill-down, pivoting, slice and dice, select
- 3.7. Relational OLAP: Database schema types: Star schema, Snowflake schema, Star Constellation schema

## **Chapter 4: Data Mining Algorithms: Association Rules [4 hrs]**

- 4.1. Motivation and terminology
- 4.2. Basic idea: itemsets, frequent itemsets, rules, support, confidence
- 4.3. Generating itemsets and rules efficiently: Apriori Algorithm
- 4.4. Evaluation of association rules

## **Chapter 5: Data Mining Algorithms: Classification and Prediction (Supervised Machine learning) [14 hrs]**

- 5.1. Basic concepts
- 5.2. Decision trees
- 5.3. Bayesian Classifiers
  - Naive Bayes Classifier
  - Bayesian networks
- 5.4. Nearest neighbor classifiers
- 5.5. Linear Regression, Logistic Regression
- 5.6. Artificial Neural Networks (ANN)
- 5.7. Support Vector Machine (SVM)
- 5.8. Classifier Evaluation: Accuracy, Precision, Recall

## **Chapter 6: Data Mining Algorithms: Clustering (Unsupervised Machine Learning) [5 hrs]**

- 6.1. Cluster analysis
- 6.2. Different types of clusters
- 6.3. Partitioning methods: k-means, expectation maximization (EM)
- 6.4. Hierarchical methods: distance-based agglomerative and divisive clustering
- 6.5. Cluster validation: Internal validation criteria, external validation criteria

## **Chapter 7: Evaluation and Credibility [3 hrs]**

- 7.1. Training and Testing
- 7.2. Holdout, Cross-validation, Leave-One-Out Cross-validation, Bootstrap
- 7.3. ROC curves, Recall-Precision Curves
- 7.4. Loss function

## **Chapter 8: Data Mining Algorithms: Outlier Detection [5 hrs]**

- 8.1. Outliers
- 8.2. Types of outliers
- 8.3. Challenges of outliers detections
- 8.4. Outlier detection method
  - Supervised, Semi-Supervised, and Unsupervised Methods
  - Statistical Methods, Proximity-Based Methods
  - Clustering-Based Methods
  - Classification-Based Methods

### **Books**

1. Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques* (3<sup>rd</sup> 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* (3<sup>rd</sup> edition). Morgan Kaufmann, 2011. ISBN 978-0-12-374856-0.
4. Udit Agarwal. *Data Mining and Data Warehousing* (2<sup>nd</sup> edition). S.K. Kataria & Sons, 2016.