

Course Code: COMP 4331
Course Title: Data Mining

Course Description

This course will provide an introduction to concepts and techniques in the field of data mining. Materials include an introduction to data warehousing and OLAP, data preprocessing and the techniques used to explore the large quantities of data for the discovery of predictive models and knowledge. The course will include techniques such as nearest neighbor, decision trees, neural networks, Bayesian networks and Naive Bayes, rule-based methods, association analysis and clustering, as well as social networks and data mining applications in business and finance applications, and other emerging data mining subareas. Students learn the materials by attending lectures and implementing and applying different data analysis and mining techniques to large datasets throughout the semester.

List of Topics

- Introduction
- Data Preprocessing
 - data summarization, cleaning, integration, transformation, reduction and discretization
- Classification and Regression
 - decision trees, neural networks, Bayesian classification, rule-based classification, and various regression methods
- Cluster Analysis
 - partitioning, hierarchical and density-based clustering algorithms, as well as outlier detection techniques
- Association Analysis
 - popular frequent itemset mining algorithms
- Miscellaneous topics

Textbooks

Data Mining: Concepts and Techniques by Jiawei Han Jian Pei and Hanghang Tong

Grading Scheme

3 assignments (30%)

Midterm (25%)

Final examination (45%)