

The Hong Kong University of Science and Technology

UG Course Syllabus

Course Title: Data Mining

Course Code: COMP4331

No. of Credits: 3-credit

Any Pre/co-requisites: ELEC 2600 OR IEDA 2520 OR IEDA 2540 OR ISOM 2500 OR LIFS 3150 OR MATH 2411 OR MATH 2421 OR MATH 2431

Name: Prof. James KWOK

Email: jamesk@cse.ust.hk

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, 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, nearest-neighbor classification, and various regression methods
- Cluster Analysis
 - partitioning, hierarchical and density-based clustering algorithms
- Association Analysis
 - popular frequent itemset mining algorithms
- Miscellaneous topics

Textbooks

Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber

Grading Scheme

3 assignments (30%)

Midterm (20%)

Final examination (50%)