1. Course Information

About the Course

As an introductory course on data mining, this course introduces the concepts, algorithms, techniques, and systems of data warehousing and data mining, including (1) what is data mining? (2) get to know your data, (3) data preprocessing, integration and transformation, (4) design and implementation of data warehouse and OLAP systems, (5) data cube technology, (6&7) mining frequent patterns and association: basic concepts and advanced methods, (8&9) classification: basic concepts and advanced techniques, and (10) cluster analysis: basic concepts. The course will serve both senior-level computer science undergraduate students and the first-year graduate students interested in the field. Also, the course may attract students from other disciplines who need to understand, develop, and use data warehouse and data mining systems to analyze large amounts of data.

Prerequisites

- Background: "Data Structure and Software Principles" or consent of instructor (good statistics and machine learning knowledge will help better understanding the course materials).
- Programming: We will give one or two programming assignments. You will need to be familiar with at least one programming language, such as C++, or Java. We will not cover programming-specific issues in this course.

Textbook

- Jiawei Han, Micheline Kamber and Jian Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011. See the book’s home page for course slides and other reference materials.

Reference

The following texts are recommended but not required, for reference, and are also on reserve at Grainger Engineering Library. There are numerous other books or online resources on data mining available.


Lecture slides contain most technical briefing and reference materials. Please study the materials in class preparation and class review.

There are many research papers that will help understand the course contents. Please check the references of this course to obtain further information.

Course Format, Activities, and Evaluation

This course will draw materials mainly from the textbook (including the 3rd edition Textbook Draft), the course slides are important references. Students will study the materials and complete all the course requirements.

Reading: Before and After Classes

We encourage students to read ahead, before lectures for the materials to be discussed. Please check the schedule page to see what will be covered in each lecture before the class begins.

Homeworks and programming assignments

There will be five assignments, spaced out over the course of the semester. Among these assignments, two will be a programming assignments.
Examinations

There will be two exams. The midterm exam will be 90 minutes in length, and the final will be 3 hours in length. We will not normally give make-ups for missed exams; please see the policies.

Extra Quarter-Unit Work: Course project

This course is designed for three-hour credit. However, graduate students (including online students) and some undergraduate students may take this course for one extra unit if you wish to do more work and show your in-depth understanding of course materials. Those taking the class for 4 credits are required to finish a course project. Please refer to project description for more details.

Evaluation

We plan to determine final grades of the course in the following way:

- Written Assignments: 15% (3 homework assignments expected)
- Programming assignments: 20% (two programming assignments expected)
- Midterm exam: 30%
- Final exam: 35%
- Project only for one extra unit students: 25%. The overall scores will be scaled proportionally.