Introduction to Data Mining
Fall Quarter, 2004

Description: Data mining is concerned with the extraction of novel knowledge from large amounts of data. This course introduces and studies the concepts, issues, and techniques of data mining. Topics include data preparation and feature selection, association rules, classification, clustering, evaluation and validation, scalability, spatial and sequence mining, and data mining applications. This course is designed for senior year undergraduate students and graduate students. 3 hours lecture, 2 hours lab.

Prerequisites: CS 400 (Introduction to Database Systems), or CS 406 (Introduction to AI), or equivalents, or with consent of the instructor. Implicitly, CS 400 (Data Structures) is also required.

Instructor: Dr. Guozhu Dong. 430 RC.

Phone & Email: (937)-775-5066, gdong@cs.wright.edu

Class details: TTh 2:15pm - 3:30pm, 105 Biological Sciences Building
Office hours: 2:30-5:30pm, T and Th; 4:30-5:20 W. Use e-mail for short questions.


Resources: Class directory: Uptodate slides will be provided in class directory /afs/ecsna1/users/cs/gdong/699/public on gamma. Slides at the course website may not be up to date.
Course website: http://www.cs.wright.edu/~gdong/mining03/WSUCS499DataMining.html
Students may find this webpage useful: http://www.kdnuggets.com/, especially its pointers to datasets.
There are many Java programs for data mining at www.cs.waikato.ac.nz/ml/weka.

Topics:
• Introduction to Data Mining
• Data preparation/Preprocessing
• Classification Methods
• Performance Evaluation
• Clustering Methods
• Association Rules
• Spatial Mining
• Sequence Mining
• Applications

Grading: Homeworks: 10%; Midterm: 25%; Final: 35%; Projects 30%.
Final grade: A=[90,100], B=[80,90), C=[70,80), D=[60,70), F=[0,60).

Handouts: Handouts, and other course material will be distributed in class. It is the students' responsibility to collect them. The instructor plans to make these, plus slides masters, available in the course website.

Important dates:
10/5, in class midterm.
3:15-5:15pm, Thursday, 11/18/04, Final.