

# CSCI 568: Data Mining

## Fall/Winter 2011, M/W/F 1-2PM, 212 Coolbaugh Hall

Yong Joseph Bakos  
231 Chauvenet Hall (subject to change)  
(303) 653-3017  
ybakos@mines.edu

**Office hours** are M/W from 3 – 5PM, or by appointment.

**Course Web home:** <http://mines.humanoriented.com/568/>

**Prerequisites:** CSCI403 Database Management or approval of instructor, and a passion for programming.

**Texts:** Programming Collective Intelligence. Segaran. O'Reilly Media. 2007. ISBN 0596529325  
Introduction to Data Mining. Tan, Steinbach, Kumar. Addison-Wesley. 2006. ISBN 0321321367  
Beautiful Data. Segaran, Hammerbacher. O'Reilly Media. 2009. ISBN 0596157118

### Course Objectives

The goal of this course is to understand and *to be able apply* the basic concepts of data mining. Topics include:

- Data mining theory and algorithms
- Mathematical foundations of data mining tools
- Programming data mining algorithms
- Acquiring, parsing, filtering, mining, representing, refining and interacting with data
- Data visualization

At the end of this course you should have a complete data mining portfolio containing implementations in the programming language of your choice.

### Grading

- Attendance & Participation 20%
- Projects & Portfolio 60%
- Midterm 10%
- Quizzes 10%

### Attendance & Participation

You are expected to be present for class (of course!) and participate in our data mining discussions. This is a professional-level class that demands your participation in order to gain a comprehensive understanding of modern data mining. Besides, it's fun discussing and debating topics with you.

### Project

Most of your labor in this class will involve continuous work on a variety of small data mining projects. You will be involved in all steps of the data mining process (collecting and preparing data, mining, refining, visualization, discussion and reporting). We will essentially be mining data while playing with modern data mining tools, and in many ways the projects are exploratory, experimental and fun.

## Portfolio

Over the course of the semester you will receive homework assignments that in the end will be used to assemble a data mining portfolio that demonstrates your knowledge of the data mining process and specific strategies used in data mining. Each assignment will have a specified due date and your entire portfolio will be submitted at the end of the semester. The algorithms in each assignment will be constantly discussed in class, and it is critical to keep up.

## Exam

One midterm exam will be conducted the week of October 6 2008.

There is no final exam (yay!).

A makeup examination can be arranged only when a student has an emergency (eg, medical emergency or urgent family matter). The student may be asked to provide the instructor with an appropriate document, such as a doctor's note.

## Accommodation

If you need certain accommodation based on disability, talk to the instructor in person so that appropriate arrangements can be made.

## Course Schedule

This schedule is not fixed in stone and is subject to change according to the actual progress of the course.

<b>Week</b>	<b>Lecture</b>	<b>Reading</b>
1	Introduction	DM ch 1, CI ch 1, BD ch1
2	Filtering, Similarity	DM ch 2, CI ch 2, BD ch2
3	Clustering	CI ch 3, DM ch 2, BD ch 2, 3
4	Matching and Ranking, Neural Nets	CI ch 4, DM ch 8, BD ch 3, 4
5	Stochastic Optimization, Genetic Algorithms	CI ch 5, BD ch 5, 6
6	Classification	CI ch 6, BD ch 7, 8
7	Decision Trees	CI ch 7, DM ch 4, BD ch 9, 10
8	<b>Midterm</b>	
9	Graphs, Neighbors	CI ch 8, DM ch 9, BD ch 11, 12
10	Advanced Classification, Support-Vector Machines	CI ch 9, DM ch 5
11	Features, Non-Negative Matrix Factorization	CI ch 10
12	Genetic Programming	CI ch 11
13	Case Studies & Project	
14	Visualization	DM ch 3, processing
15	Visualization	Processing doco
16	Algorithm Review	CI ch 12

## On Collaboration & Academic Integrity

Students are encouraged to discuss and collaborate as much as possible. However, it is obviously not acceptable to copy another student's solution. Your work must be your own. In addition, simply copying solutions found online is not acceptable. Be aware that homework assignments, project and midterm will not just focus on producing correct code, but explaining how things work.

Please see the Student Handbook for details on academic dishonesty. No exceptions will be made for students found simply giving away or taking another's solutions.

### Academic Integrity Pledge

I pledge to uphold the high standards of academic ethics and integrity expressed by the Colorado School of Mines Student Honor Code by which I am bound. In particular, I will not misrepresent the work of others as my own, nor will I give or receive unauthorized assistance in the performance of academic coursework. I understand that my instructor will report any infraction of academic integrity to the Department Head and that any such matter will be investigated and prosecuted fully.

Name (print): \_\_\_\_\_

Signature: \_\_\_\_\_