DECISION SUPPORT SYSTEMS
OPIM 410/672: Spring 2010

Professor	**Shawndra Hill**
Course Webpage	Accessible from webcafe: https://webcafe.wharton.upenn.edu/eRoom/opim/410-sp10-1 Links on course wiki site http://shawndra.pbworks.com/SP09OPIM410672
Classroom	JMHH F92
First/Last Class	Jan13, 2010/April 26, 2010
Class times	MW 410: 10:30 – 12. MW 672: 1:30pm-3pm
Course Assistant	
Office Hours	Fridays 2-5, or by appt
Email	shawndra@wharton.upenn.edu **subject: [DSS 672 class] … ← note!**
Telephone	Office 215-573-5677

1. Course Overview

This course will change the way you think about data and its role in business.

Businesses, governments, and society leave behind massive trails of data as a by-product of their activity. Increasingly, decision-makers rely on intelligent systems to analyze these data systematically and assist them in their decision-making. In many cases automating the decision-making process is necessary because of the speed with which new data are generated. This course connects real-world data to decision-making. Real-world examples from Finance, Marketing, and Operations are used to illustrate applications of a number of data mining methods. The use of real-world examples and cases places these techniques in context and teaches students how to avoid the common pitfalls of data mining, emphasizing that proper application of data mining techniques is as much an art as it a science. In addition to cases, the course features hands-on exercises with data mining software. The course is suitable for those interested in working with and getting the most out of data as well as those interested in understanding data mining from a strategic business perspective. It will change the way you think about data in organizations.

The goal of this course is three-fold. After taking this course you should:

1. ***Approach business problems data-analytically***. Think carefully & systematically about whether & how data can improve business performance.

2. ***Be able to interact competently on the topic of data mining for business intelligence.*** Know the basics of data mining processes, algorithms, & systems well enough to interact with CTOs, expert data miners, and business analysts. Be able to envision data-mining opportunities.

3. ***Have had hands-on experience mining data.*** Be prepared to follow up on ideas or opportunities that present themselves, e.g., by performing pilot studies

2. Instruction Method

This is primarily a lecture-based course, but student participation is an essential part of the learning process in the form of active technical and project discussion. The course will explain with detailed real-world examples the inner workings and uses of various data mining techniques. The primary emphasis is on understanding the various types of data mining techniques, and when and how to use them, and secondarily on the mechanics of how they work. Slides will be posted by 9am on the day of class.

Assignment Questions

Each class session has materials you must read prior to class. There will be a total of eight question assignments, each comprising a (multi-part) question. In addition, the assignments may involve a hands-on work that will be completed in Weka—based on a data set that we will provide. You must turn in *all* question answers on the dates they are due. They will be graded and returned promptly.

Late assignments

Turn in your assignment early if there is any uncertainty about your ability to turn it in on the due date. Assignments up to 1 day late will have their grade reduced by 50%. After one day, late assignments will receive no credit (no exceptions).

Data Mining Competition (optional for extra credit)

You will mine large data sets to get hands-on experience in formulating problems and using the various techniques discussed in class. You will use these data to build and evaluate predictive models. The assignment will include "competition" data sets: parts of the data will be held back to evaluate the models you mine. You may work in teams of 2-3 students. The best-scoring team will get 5 extra-credit points added to their overall course grade.

For the hands-on portion of assignments you will use the data mining toolkit Weka.

http://www.cs.waikato.ac.nz/ml/weka/

A tutorial on Weka will be given in class. Weka should be installed in the computer Lab. In order to use Weka at home, you must have access to a computer on which you can install software. If you do not have such a computer, please see the TA immediately so we can make alternative arrangements. The first hands-on assignment will be very easy, ensuring that you can install the software and get it running, before moving on to more challenging assignments.

Data Mining Project

You are required to complete a novel research report (~15 pages)

You are required to format the document as if to submit the report for publication to a well known decision support systems or data mining journal. However, your grade will not depend on the success of your submission, only on its quality as assessed by the instructor. Your report should include Introduction, Related Work, Methodology, Results, Comparison / Analysis, Conclusion and Future Work, and References. Your research report can be a *description of a new data mining method or algorithm you have developed*, or it can be an *analysis of a data set*.

Your instructor will help you with some project ideas and provide you with examples from prior years, though you are encouraged to choose your own, which will need approval.

3. Requirements and Grading

This is a lecture-style course, however student participation is important. Students are required to be prepared and read the material before class. Students are required to attend all sessions and discuss with the instructor any absence from class. As discussed above, you will hand-in 8 (individual) write-ups to questions that will be assigned in class and will be posted on the class webCafe site. Answers should be well thought out and concise. Points will be deducted for sloppy language and irrelevant discussion.

There will be one team project (teams of 2-3 students) in which students will address business problems with data mining techniques. Students will hand in a report (accounts for 80% of project grade) and prepare a short class presentation of their work (20% of project grade). A class discussion will follow the presentations. Details of the requirements for the project will be discussed the second week of class. There will be many examples posted on the webcafe site.

There will not be a final exam at the end of the semester.

The grade breakdown is as follows:

1. Assignment Questions (7-8 Write-ups): 35 points overall
2. Data Mining Project: 50 points
3. Participation and Class Contribution: 15 points
4. Data Mining Competition: **optional for extra credit**

4. Teaching Materials

The following are reading materials for this course:

1.   Textbook: *available at the bookstore*
Data Mining Techniques, Second Edition
by Michael Berry and Gordon Linoff
Wiley, 2004
ISBN: 0-471-47064-3

2.   Two Crows Corporation. Introduction to Data Mining: Third Edition.
Available at: http://www.twocrows.com/intro-dm.pdf

3.   Supplemental readings will be provided as the class progresses.

4.   Website (Blackboard) for this course containing lecture materials and late breaking news, accessible through the webcafe home page
(https://webcafe.wharton.upenn.edu/eRoom/opim/672-sp08-1)

5. Course Schedule (Tentative Dates)

Week	Lecture 1	Lecture 2	Monday	Wednesday
1		Introduction to the Course Readings: Chapter 1 and 2		Jan. 13
2	**MLK**	Introduction to Data Mining/Classification Readings: Chapter 1 and 2	**NO CLASS**	Jan. 20
3	Classification: Recursive partitioning and Decision Trees	Classification: Recursive partitioning and Decision Trees	Jan. 25	Jan. 27
4	Naïve Bayes	Association Rules/k-nearest Neighbor/Clustering	Feb. 1	Feb. 3
5	Classification Model Evaluation	Cost Sensitive Learning	Feb. 8	Feb. 10
6	Exploratory Data Analysis	Practical Weka Demo (Features you may want to use)	Feb. 15	Feb. 17
7	Weka Lab/ Work on DM Project in class	Case Discussion	Feb. 22	Feb. 24
8	Related Technologies (OLAP, RDB, etc)	**Group Presentations (Progress Report)**	Mar. 1	Mar. 3
	Spring Break	**Spring Break**	Mar. 8	Mar. 10
9	Genetic Algorithms	Neural Networks	Mar. 15	Mar. 17
10	Relational Learning	Relational Learning	Mar. 22	Mar. 24
11	Recommendation Systems/Collaborative Filtering	Weka Lab/ Work on DM Competition in class	Mar. 29	Mar. 31
12	Guest Speaker	Guest Speaker	Apr. 5	Apr. 7
13	Guest Speaker	Guest Speaker	Apr. 12	Apr. 14
14	**Group Presentations 1**	**Group Presentations 2**	Apr. 19	Apr. 21
15	**Group Presentations 3/ Data mining competition winner / results**		Apr. 26	