ST 531 Course Syllabus
Introduction to Data Mining

Fall 2007

Instructor: J. Michael Hardin, Ph.D.
Professor
149 Bidgood

Office Phone: 348-8901
E-Mail: mhardin@cba.ua.edu

Office Hours: Mon. & Wed. 10:00am – 12:00 noon; 1:00 – 2:00PM, and by appointment

Location & Time: BD 367, 3:30-4:45pm, Monday & Wednesday

Prerequisite: undergraduate statistics course, graduate standing, and Permission of instructor

Course Description

An increasing number of organizations are collecting massive quantities of operational data. These databases, however, often have been isolated islands of technologies, not well integrated into the decision-making activities of the organizations with the result that these various data have not added value consistent with their potential. The data warehouse provides an enterprise-wide perspective. Knowledge discovery and data mining provides a set of tools with which to mine the data warehouse for previously unknown patterns that are strategically useful to the mission of the organization.

Data mining is the process of selecting, exploring, and modeling large amounts of data to uncover previously unknown patterns of data. By applying data mining techniques, data analysts can fully exploit large databases, identifying potentially useful patterns and behavior, and gain a greater understanding of the data.

The goal of data mining in business applications is to produce new knowledge that decision-makers can act upon. It does this by using sophisticated techniques such as artificial intelligence to build a model of the real world based on data collected from a variety of sources including corporate transactions, customer histories and demographics, and from external sources such as credit bureaus. This model produces patterns in the information that can support decision making and predict new business opportunities.

Course Objectives:

An overview of data mining techniques, especially predictive modeling, will be introduced in this course. Particular attention will be given to the techniques utilizing regression and logistics regression this semester. Next semester, we will further our study in predictive modeling through a study decision trees, neural networks. Unsupervised learning methods such a association analysis, and cluster analyses will also be examined next semester. A project will be required this semester as well as next semester.

Required Texts: Discovering Knowledge in Data: An Introduction to Data Mining, Larose, D.T., John Wiley & Sons, 2005.

Data Mining Methods and Models, Larose, D.T., John Wiley & Sons, 2005.

Introduction to Data Mining and Knowledge Discovery, 2nd Ed. (can be downloaded in pdf format from www.twocrows.com)

Reference Texts: Machine Learning and Statistics: The Interface, edited by G. Nakhaeizadeh & C.C. Taylor, John Wiley & Sons, 1997

Learning from Data: Artificial Intelligence and Statistics V, edited by D. Fisher & HJ. Lenz, Springer, 1996

Construction and Assessment of Classification Rules, David J. Hand, John Wiley & Sons, Inc.

Classification and Regression Trees, Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984), Chapman and Hall.

Getting Started with Enterprise Miner Software, Version 4, SAS Institute

Decision Tree Modeling Course Notes, SAS Institute

Predictive Modeling Using Logistic Regression Course Notes, SAS Institute

Master data Mining, The Arts and Sciences of Customer Relationship Management by Michael J. A. Berry and Gordon S. Linoff, John Wiley & Sons, Inc.

Machine Learning, by T.M. Mitchell, McGraw-Hill, 1997.

An Introduction to Bayesian Networks, by F.V. Jensen, Springer, 1996

Data Warehousing: Strategies, Technologies, and Techniques, by R. Mattison, McGraw-Hill, 1996

Data Mining: Concepts and Techniques by J. Han and M. Kamber
Data Mining Techniques, by M. Berry & G. Linoff, John Wiley & Sons, 1997

Solving Data Mining Problems Through Pattern Recognition, by R. Kennedy, Y. Lee, B. Roy, C. Reed, & R. Lippman, Prentice Hall, 1997

Neural Networks: A Comprehensive Foundation, 2nd Ed, by S. Haykin, Prentice Hall, 1999

Machine Learning, Neural and Statistical Classification, edited by D. Michie, D.J. Spiegelhalter & C.C. Taylor, Ellis Horwood, 1994

Neural Networks for Pattern Recognition, by C. M. Bishop, Oxford Press, 1995

The Nature of Statistical Learning Theory, by V.N. Vapnik, Springer, 1995

Recursive Partitioning in the Health Sciences, by H. Zhang & B. Singer, Springer, 1999

Learning From Data: Concepts, Theory, and Methods, by V. Cherkassky & F. Mulier, John Wiley & Sons, 1999

Data Warehousing, Data Mining, & OLAP, by A. Berson & S.J. Smith, McGraw-Hill, 1997

Course Topics:

1. Overview and Motivation for Data Mining

2. Introduction to history and concepts of data mining

3. Introduction to the use JMP as a Business Intelligence Tool

4. Review of Measurement Issues and Data Preparation

5. Visualizing and Concepts of exploratory data analysis

6. Overview of Predictive modeling

7. Predictive Data Mining Methods for regression

8. Predictive Data Mining Methods for logistic regression

9. Introduction to SAS Enterprise Miner

10. Project Preparation and Presentation

Evaluation:

Project & Presentation (Final)	25%
Homework & Assignments	10%
Attendance & Class participation	5%
Pop Quizzes	10%
Test 1 (in class, closed book)	25%
Test 2 (in class, closed book)	25%

- Generally, there will be a homework assignment every two weeks. These assignments will be selected from the material in the text or other reading for the course
- **Academic dishonesty on any exam or other course work will result in a grade of zero for that work**

Note: The project will be composed of a written portion due the day and time of the final (no late papers will be accepted—please plan ahead) and an in class presentation of the project. The presentations will occur during the week of December 3. The final time for this class is Wednesday, December 10, 2007

Make-up Exam Policy:

Make-up exams or extending assignment due dates will be allowed only in extreme instances and with **advanced permission** of the instructor. If any student feels that a sudden illness is sufficiently extreme to warrant a make-up test or to extend the assignment due date, the instructor must be provided with documentation prepared by an appropriate medical authority.

Attendance Policy:

Students are expected to attend and be prepared in all classes. Further, students are expected to participate and contribute to class discussions. Failure to attend classes and/or participate in class discussions will result in the lost of the points for class participation component of the student's grade. The lost of these points may result in an entire letter grade drop in the student's final grade.

Students with Special Needs:

Disability Access Statement: To request disability accommodations, please contact the Office of Disability Services at 348-4285. After consultation with that office, contact your professor. However, it is the student's responsibility to make arrangements for the accommodations on a timely basis. Special arrangements for exams must be made at least one week prior to the exam date or your instructor is not required to provide requested accommodations. Any request for special arrangements made less than one week prior to an exam date may not be able to be honored.

Emergency Policy:

In the event of an emergency, we will adhere to the following actions in accordance with University policies. FIRE/FIRE ALARM: Evacuate the building and stay out of the building at a safe distance until authorized to return. TORNADO WARNING: Move to the Lower Level, inside classrooms, offices or corridors. Remain until the warning has expired. Classes are cancelled until the warning expires.