ST 532	Advanced Data Mining	Spring 2007

Instructor: J. Michael Hardin, Ph.D.
Office: 149 Bidgood
E-mail: mhardin@cba.ua.edu
Website www.cba.ua.edu/~mhardin
Phone: 348-8901

Office Hours: Mon and Wed 11:00 a.m. – 12:00 noon; 1:30 – 3:00 p.m.
Other times by appointment

Class time: 3:30 – 4:45 p.m.

Classroom: BD 357

Textbook: (Required)
1. Applied Data Mining: Statistical Methods for Business and Industry, John Wiley & Sons, 2003.
2. Instructors notes (to be provided – see web site)

Prerequisite: ST 531 (Data Mining), ST 522 (Statistical Data Management), or Permission of instructor.

Course Description: This course examines in detail a variety of the data mining methodologies introduced in the first course and provides students with insight and understanding into the advanced aspects of the algorithms and techniques employed in data mining. Emphasis will be placed on a detailed study of the major techniques of data mining including logistic regression, neural networks, decision trees, general classifier theory, and unsupervised learning methods. Mathematical details of these techniques as well as the computer techniques for their implementation will be examined. Students are shown how these details relate to various options in the corresponding Enterprise Miner node.

The SAS programming language and SAS's Enterprise Miner will be used to accomplish these tasks. Other software packages such as CART, MARS, and SPLUS also may be used.

ST 532	Advanced Data Mining	Spring 2007

Course Learning Objectives:

1. The student will be able to discuss the various data mining algorithms and provide both an intuitive and detailed presentation of their assumptions and how they are structured.
2. The student will be able to correctly specific options within SAS Enterprise Miner nodes to implement the major data mining methodologies.
3. The student will be able to prepare oral and written presentations discussing the results and conclusions from a data mining/ business intelligence project. Further, they will be able to field questions and explain the their work to audiences with varying mathematical and statistical backgrounds such as commonly found in business and industry.
4. The student will be able to implement new data mining techniques using SAS programming or other appropriate statistical programming languages such as R, SPLUS, or MatLab.
5. The student will be able to design classifiers/ predictive models to achieve a given cost benefit through their understanding of statistical decision theory. This objective includes the identification of appropriate factors in constructing a profit matrix and in determining appropriate prior vectors.

Grading Scheme: The final grades will be based on the following distribution

Homework (Project-oriented) (Close to every two weeks)	15%
Pop Quizzes	10%
Class Participation	10%
Exam 1	20%
Exam 2	20%
Final Project (due 5/10/2007)	25%

Students with Special Needs:

If any student has a disability or special need that may prevent him or her from participating fully in class activities, please contact me as soon as possible. Every reasonable accommodation will be made to make sure that all students have an opportunity to learn and benefit from the course.

Policies:

1. Homework projects will be assigned on approximately every two weeks.

2. No late assignments will be accepted.

3. Assignments must be submitted at the beginning of the class session on the date due. Papers or assignments submitted after the deadline will be penalized one letter grade for each day late.

4. Attendance and class participation are an expected standard. Students are expected to attend all class sessions and actively participate in class discussions.

5. Equivalence or make-up examinations will be administered only at the instructor's discretion.

6. All papers, projects, and reports are expected to be presented in professional format (i.e., wordprocessed, correct usage, professional appearing format).

7. Grades of "I" for the course are only given in unusual and limited situations, and at the instructor's discretion. For any case in which the Instructor and the student have agreed for the student to receive a course grade of "I" for the term, it is the student's responsibility to

ensure that all work is completed and received by the instructor. If the work is submitted, then the instructor will submit a change of grade to be effective by the end of the term in which the work is accepted. The exact time for submitting the change of grade, however, will be at the instructor's discretion and convenience.

ST 532	Advanced Data Mining	Spring 2007

Course Topics

1. Decision Theory for Classification

2. Introduction to Classification

3. Logistic Regression

4. Introduction to Generalized Linear Models

5. Decision Trees

6. Linear Discriminant Methods

7. Neural Networks

8. Cluster Analysis

9. Miscellaneous Topics
   may include naïve bayes, Bayesian networks, genetic algorithms, and text mining as time permits

FINAL EXAM Date

Date- Thursday, May 10, 2007