

SYLLABUS

Subject:	STA 5703 Data Mining Methodology I (STA5703)
Lecture:	MW 4:30-5:45 pm, Room 220, Classroom Building I
Instructor:	Dr. Xiaogang Su Room 102, CC II (407) 823-2940 [O] xiaosu@mail.ucf.edu
Office Hours:	MW 3:30-4:30 pm
Text Book:	<i>Elements of Statistical Learning</i> , 2nd Edition Hastie, T., Tibshirani, R., and Friedman, J. H. (2008), Chapman and Hall. ISBN-13: 978-0387848570
Course Web Page:	http://pegasus.cc.ucf.edu/~xsu/CLASS/STA5703
Prerequisite:	STA 5103 and STA 5206
Computing	SAS Enterprise Miner.

- **Course Summary:** Data mining is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of observational data in order to discover meaningful patterns and models to the data owner. By applying data mining techniques, data miners can fully exploit data patterns and behavior, and gain a greater understanding of the inside of the data. The goal of data mining application in business is to produce new knowledge that decision-makers can act upon. It does this by using sophisticated techniques such as logistic regression and decision trees 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 knowledge that can be used to support decision-making and to predict new business opportunities. This course will cover data mining techniques such as clustering, PCA, regularization, decision trees, bagging, boosting, and random forests. In addition, assessments of classification rules and how to use SAS Enterprise Miner will be covered.

- **Useful References**

1. Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984), *Classification and Regression Trees*, Chapman and Hall / CRC.
2. David H., Mannila, H. and Smyth (2001). *Principles of Data Mining*, Massachusetts Institute of Technology.
3. David H. (1997). *Construction and Assessment of Classification Rules*. John Wiley & Sons, Inc.
4. *Getting Started with Enterprise Miner Software*, Version 4, SAS Institute.
5. *Decision Tree Modeling Course Notes*, SAS Institute
6. *Predictive Modeling Using Logistic Regression Course Notes*, SAS Institute.
7. Berry, M. J. A. and Linoff, G. S. (1999). *Master data Mining, The Arts and Sciences of Customer Relationship Management*. John Wiley & Sons, Inc.

8. Zhang, H. and Singer, B. (1999). *Recursive Partitioning in the Health Sciences*. New York: Springer-Verlag.

• **Course Objectives**

- Use R, SAS Enterprise Miner and SAS/STAT software effectively and efficiently
- Capable to use clustering techniques to perform task such as customer segmentation
- Capable to use CART and CHAID like tree techniques to build rule-based prediction models
- Capable of performing exploratory data analysis with trees
- Able to apply boosting, bagging, and random forests typed techniques in predictive modeling
- Apply decision theory to compare different data mining models

- **Grading:** There will be a number of computer assignments and a final project. The assignments make up 50% and the final project makes up 35% to your final score. For the final project, students are given the freedom to select data from whatever field they are interested in (and at the same time, a topic that is adequate for the course). Students should make their own plans to collect data, raise some interesting questions to study, and consult me for the adequacy of the project. Also, each student will have the opportunity to present their work in class. There will be also a couple of in-class quizzes or exams, which make up 20%. No make-up exam is given for ANY reason and no late project submission is accepted for ANY reason. If you expect to miss up to 10 hours of classes for any reason, then please do not consider taking this course.

The criterion for assigning grades is listed as follows. Incomplete grades are not an option unless in extreme instances and only with prior permission of the instructor.

Range	94+	93-90	89-87	86-83	82-80	79-77	76-73	72-70	69-67	66-63	62-60	59-0
Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F