

BUAL 5610/6610/6616 PREDICTIVE MODELING II

Instructor: Pei Xu pei.xu@auburn.edu

Semester: Fall 2014

Classroom: Lowder 21

Class Time: MW 9:30 am - 10:45 am

Office: 420 Lowder bldg

Office Hours: MW 2:00 pm - 3:00 pm; others by appt.

COURSE DESCRIPTION & OBJECTIVES:

Predictive modeling data mining process involves the art of exploring and analyzing, by automatic or semiautomatic algorithms, large quantities of observational data in order to discover meaningful patterns and/or build predictive models. Data mining techniques can be used to build knowledge from information adding value to the ever-increasing electronic data. Data Mining emerges as an interdisciplinary field with joint inputs from statistics, computer science, machine learning, and artificial intelligence. This multidisciplinary field continues to deliver contemporary techniques and tools for the analysis of very large collections of data.

This course covers major concepts and algorithms of data mining. Majority of the course will be taught using the SAS Enterprise Miner software. Upon completion of the course students will be prepared to tackle the SAS Certified Predictive Modeler using SAS Enterprise Miner 7 Credential Exam. The final project will demonstrate all the data mining techniques covered in the course and furthermore expose students working with real industry data. At the end of the course students will be proficient in utilizing data mining techniques to exploit data patterns and behavior, gain insider understanding of the data, and produce new knowledge that decision-makers can act upon.

COURSE MATERIALS:

There is no required textbook for this course. All the necessary material was provided to you by the instructor. For those who would like some additional reading, the following books are recommended. Both books have electronic version available at the library of Auburn University.

1. Anasse Bari, Mohamed Chaouchi, Tommy Jung. “*Predictive Analytics for Dummies* (2014)”. ISBN-13: 978-1118728963
2. David F. Nettleton. “*Commercial Data Mining: Processing, Analysis and Modeling for Predictive Analytics Projects* (The Savvy Manager's Guides, 2014)” ISBN-13: 978-0124166028

ELECTRONIC RESOURCE:

Canvas: Lectures/homework assignments/handouts etc. will be posted on Canvas. It is extremely important to check Canvas and email regularly to stay informed about the class.

SAS Enterprise Miner:



SAS Enterprise Miner will be primarily used throughout the course and final project. At the end of the course, students have a firm understanding and mastery of the functionalities for predictive modeling available in SAS Enterprise Miner 7. The software can be accessed from classroom and labs. Necessary materials for using the software will be provided by the instructor. Students will have an opportunity to take the SAS Certified Predictive Modeler using SAS Enterprise Miner 7 Credential. This credential is designed for SAS Enterprise Miner users who perform predictive analytics. The exam will be provided at a discounted price for students (<http://support.sas.com/certify/creds/pm.html>). Class lectures, tests, homework assignments, and the final project will prepare students for the certification exam.

The instructor will also have an in-class review session (see course schedule) before the exam. The certification exam is expected to be scheduled at the end of the semester at one of the computer labs in the college of business. Students who pass the exam and wish to share their results will be awarded bonus points.

GRADING:

The final grade will comprise

- Test 1 100
- Test 2 100
- Final Project 150
- Homework 150
- Bonus (SAS Predictive Modeling Certification) 50

Numerical scores will be converted to letter grades as follows:

- A = 450– 550
- B = 400– 449
- C = 350– 399
- D = 300– 349
- F = 0–299

TESTS:

There will be two tests in the course. Each test is marked on a scale of 100. The tests will be held in-class and within the class time. Students will use the SAS Enterprise Miner software for answering questions on the test. The tests are closed book and necessary datasets and other supplemental material will be provided. These tests are comprehensive and will prepare students towards the SAS Certified Predictive Modeler using SAS Enterprise Miner 7 Credential. The instructor will have a review session before the test (see schedule).

Make up exams: In the event that a student misses an exam and provides a **University approved excuse** in accordance with the following guidelines, the student will be given the opportunity to make up the missed exam. Students must notify the instructor prior to the test and turn in their excuse to the instructor on the first day upon returning to class. Lack of preparation for an exam is not a valid excuse and students should not be tempted to “fake” a doctor’s excuse due to lack of preparation. The penalty for this is severe and could affect you for the rest of your career.

All **AU approved excuses** must be turned in by the **first class day** after the missed day.

HOMEWORK:

There are 6 computer assignments, each marked on a scale of 25. Assignments will involve answering a real application problem. Instructor will provide the necessary datasets and explain the business problem and students are expected to analyze and report findings using SAS Enterprise-Miner.

HW	Topic Covered	Due Date
1	Decision Tree	M, Sep 15
2	Cluster Analysis and Association	M, Sep 22
3	Neural Networks	M, Oct 13
4	Model Assessment and Implementation	M, Oct 27
5	Time Series Analysis	M, Nov 3
6	Social Media Analysis	M, Dec 1

Every homework assignment **MUST** have the following sections:

- a. Title Page – Name of the student, assignment Number, and title
- b. Executive Summary – A brief description of the business problem and conclusions from analysis. 1-2 paragraphs
- c. Table of Contents -
- d. Introduction/Business Problem – Complete description of the business problem. 1-2 paragraphs
- e. Methodology – Type of analysis conducted and rationale for using the technique. 1-2 paragraphs
- f. Analysis Results – Tables and Figures from analysis. Note: Do not copy paste from tables from software output. Create tables in excel/word. You are use figure from the software output.
- g. Conclusions – Conclusions from the analysis. 1- 2 paragraphs
- h. Appendix – Output from software and screen shots of all the steps involved.

The document must contain page numbers starting from the business problem section. Title page and executive summary do not need page numbers. Writing must be clear and concise. Anyone needing assistance with writing must see the instructor or visit on campus writing center. The document must be professional and mimic a final report submitted to the client. **Instructor will accept electronic or paper copy of the homework. All homework's are due at the beginning of class lecture.** Electronic copies will only be accepted before class and not accepted during/after class.

PROJECT:

The final project is the culmination of all the data mining methods and techniques discussed during the entire course of the semester. The project work will involve working with real industry data. Students will be acting as Business Analysts trying to solve a real business application. In week 10, the instructor will introduce the final project, including the data, business problem and expected outcomes. Students will be working in groups under the guidance of instructor. The final project report should provide a clear and through explanation and documentation discussing all decisions about data mining analyses. Reports must be prepared in a professional manner and appearance. Students will present (PowerPoint) the findings of the project in-class. Students are encouraged to use the project findings at data mining/analytics poster presentations. The project work will provide students with real industry experience. The report **MUST** contain the following sections:

- a. Title Page – Name of the student, assignment Number, and title
- b. Executive Summary – A brief description of the business problem and conclusions from analysis. 1-2 paragraphs
- c. Table of Contents
- d. Literature Review/Background – Literature and background of the study
- e. Introduction/Business Problem – Complete description of the business problem.
- f. Methodology – Type of analysis conducted and rationale for using the technique.
- g. Results – Tables and Figures from analysis. Note: Do not copy paste from tables from software output. Create tables in excel/word. You are use figure from the software output.
- h. Conclusions/Discussion – Conclusions from the analysis.
- i. Appendix – Output from software and relevant screen shots.

COMMUNICATION:

I will use your Auburn University e-mail address ([userid@auburn.edu](mailto:user@auburn.edu)) for course communication. It is your responsibility to contact the Information Technology Help Desk to have this address forward mail to the e-mail address that you regularly check.

ACADEMIC HONESTY:

All portions of the Auburn University student academic honesty code (Title XII) found in the AU Office of Provost website will apply to university courses. All academic honesty violations or alleged violations of the SGA Code of Laws will be reported to the Office of the Provost, which will then refer the case to the Academic Honesty Committee.

CLASS PROCEDURES:

- Turn off all wireless communications devices in class. Do not make or receive calls during class. Further, making or receiving calls during an exam will be viewed as an Academic Honesty violation.
- The instructor has the discretion to modify course content and assignment/test dates.

SPECIAL ACCOMMODATIONS DUE TO DISABILITIES:

Students who need special accommodations, as provided for by the American with Disabilities Act, should make an appointment as soon as possible with the faculty member to discuss their Accommodation Memo. It is essential that the faculty member be aware of necessary accommodations at the beginning of the course. The student must bring a copy of his/her Accommodation Letter and an Instructor Verification Form to the meeting. If the student does not have these forms but needs special accommodations, he/she should contact the Program for Students with Disabilities, 1288 Haley Center, 334-844-2096 (V/TT).

SCHEDULE:

Please read each chapter prior to the date on which it is covered. Test dates are fixed, but the chapter coverage dates and chapters covered on each test are subject to change.

Week	Date	Topics
1	Aug 18, 20	Syllabus; Chapter 1: Introduction to Predictive Modelling & SAS Enterprise Miner Chapter 2: Accessing and Assaying Prepared Data
2	Aug 25, 27	Chapter 3: Decision Trees; Homework 1
3	Sep 1, 3	Chapter 3: Decision Trees (2)
4	Sep 8, 10	Chapter 4: Cluster Analysis; Self-organizing Maps.
5	Sep 15, 17	Chapter 4: Association Rule Analysis; Market Basket Analysis. Homework 2
6	Sep 22	Catch up and Review
	Sep 24, W	Test 1 (Chapter 1-4)
7	Sep 29, Oct 1	Chapter 5: Neural Networks (1)
8	Oct 6, 8	Chapter 5: Neural Networks (2); Homework 3
9	Oct 13, 15	Chapter 6: Model Assessment
10	Oct 20, 22	Chapter 7: Model Implementation; Homework 4; Assign Final Project
11	Oct 27, 29	Chapter 8: Time Series Analysis; Homework 5
12	Nov 3	Catch up and Review
	Nov 5, W	Test 2 (Chapter 5-8)
13	Nov 10, 12	Chapter 9: Special Topic: Text and Web Mining;
14	Nov 17, 19	Chapter 9: Special Topic: Social Media Analysis; Homework 6
15	Thanksgiving	
16	Dec 1, 3	Final Project Work and Presentation
	Dec 3-8	Optional SAS Certificate Exam

Note: No class on Sep 1 (*Labor Day*)