

Data Mining and CRM Applications: MKTG 5963, Section 503

(Tentative Syllabus for Spring 2015)

Professor: Dr. Goutam Chakraborty

Office: 420A SSB, Phone: (405) 744-7644.

Class Materials: Most of the class materials will be distributed via the D2L web site for this class (<https://oc.okstate.edu/>). If you are a registered student for this class, you should be able to see this course when you log-in to D2L (the site becomes active one-week before classes begin). If you have problems accessing the D2L class site, please call OSU's IT help desk (405-744-4357) or (toll free) 1-877-951-4836). If you have problems viewing video lectures, please call Administrative/Video lecture support provided by the OSU's Distance Learning office (405, 744-4048, email: spearsdistance@okstate.edu).

Teaching Assistants (TAs): Names and emails of TAs will be announced in the first week of class (check lab videos or email). They will be your primary point of contacts for any issues related to this class. *When writing any email to my TAs, please copy to all TAs.*

E-mail: Please use the class discussion bulletin-board via D2L for any general questions, comments, clarifications about any of the class topic. Use the e-mail to my TAs sparingly. *There is no need to copy me with your email to my TAs – if my TAs are unable to answer your question, they will discuss with me and get back to you.*

Class Discussion via D2L (<https://oc.okstate.edu/>): We will use this format extensively for communication among students as well as between students and the instructor. This will be a bulletin-board type system with specific folders for different aspects of this course. There will be multiple forums (folders) in this bulletin board. Please check these folders regularly. Please post your questions only in the **appropriate forums**. Please use appropriate subject line in your posting and use threaded discussion whenever possible. Do not ask direct questions about how to solve an assignment (asking for clarification or software help is ok).

Required Text:

[Text Mining and Analysis: Practical methods, Examples and case Studies using SAS®](#) by Goutam Chakraborty, Murali Pagolu and Satish Garla, SAS Publishing, 2013, Cary, NC, SAS Institute Inc.

I will also use readings off the web, cases, SAS training materials, chapters from reference books, etc. in this class. I have indicated a number of good books (under reference texts) on this topic that you may find useful. I will announce readings via postings on D2L or via email.

Reference Texts (Optional materials – some of these will be put in reserve at the library)

- [Data Mining Techniques for Marketing, Sales and Customer Relationship Management](#), by Michael J. Berry and Gordon S. Linoff, Wiley Publishing Inc., 2011. (OSU library call number: 658.802 B534d for 2004 version of this book)
- [Data Mining](#), by Witten and Frank, Morgan Kaufman publications, 2011 (OSU library call number: 006.3 W829d 2011).
- [Handbook of statistical analysis and data mining applications](#), by Nisbet, Elder and Miner, Academic Press/Elsevier, 2009, (OSU library call number : 006.312 N724h)

- Principles of Data Mining, by Hand, Mannila, and Smyth. MIT Press 2001. (OSU library call number: 006.3.H236p)
- Data Preparation for Data Mining by Dorian Pyle, Morgan Kaufman publications, 1999. (OSU library call number : 005.74 P996d)

COURSE OBJECTIVES

This course has five major objectives that fit within five of the program learning goals.

Course Objective	Program Learning Goal
Students will be able to engage in analytical reasoning to break problems into their component parts; identify important patterns by analyzing data; and test for assumptions behind models.	• Critical Thinking
Student can apply science and business principles to analyze and interpret data, using analytic and computer-based techniques.	• Critical and Creative Thinking
Students will be able to present written results from their analyses by relating those back to the business issues that demonstrate a mastery of language and mechanics.	• Written Communication
Students will be able to present their results orally using a message that is well organized, concise and quickly understandable by business professionals.	• Oral Communication
Students will be able to use appropriate tools and technologies for data visualization and statistical model building	• Technology Skills

Course Description

Data mining promises to turn business data (both numeric and text) into actionable information. Therefore, this course will focus on learning how to use various data mining tools such as neural networks, decision trees, classification and prediction algorithms etc. in the context of most common applications in business – sales, marketing, and customer relationship management (CRM). Students will be expected to use state-of-the-art industrial strength data mining software (SAS Enterprise Miner) to analyze real-world data and make strategic recommendations for managerial actions. My philosophy in teaching the course is “*you learn by doing*,” that is, you should be prepared to work extensively with SAS Enterprise Miner in analyzing data sets using various techniques such as **neural networks, decision trees, multiple/logistic regression, association rules, sequence detection, ensemble models, text mining, Sentiment mining, content categorization**, etc. The course will use lectures, data analysis using state-of-the-art data mining software, case discussions, and exercises. All class lectures will be handled via video (video links will be posted on the D2L course site) that you can watch at our own convenience (you will need a high-speed Internet connection to watch the lectures). However, non distance learning students must physically attend the data mining lab (at the Gundersen lab) each week to work on exercises/assignments/projects as assigned by the instructor.

Real Office Hours (to talk to me in person)

Monday and Tuesday 8:30-9:30AM, or by appointment (set up via phone or email).

Virtual Office Hours (to get my opinion on any issue related to this class)

Please use the desire to learn (D2L) platform for this purpose. I (and my TAs) will monitor this platform closely and try to answer your questions quickly. I may also set up a SKYPE or Webex based call-in office hours for DL students so you can talk to me (individually or as a group). Details will be communicated via D2L class site.

Course Prerequisites

Students must complete MKTG5983 (Database Marketing) or BADM5513 (Business Analytics) before taking this class. I will assume that all students enrolled in this class have very good ideas of basic probability/statistics, basic statistical models (such as multiple regressions, ANOVAs) and perhaps some exposure to SAS before joining this class.

Course Format:

Please note that this course has a unique format (a combination of all video lectures and lab/discussion sessions).

Lectures: A video for each lecture will be streamed over the Internet. The link for each video lecture will be posted on the D2L. It is your responsibility to **watch the lecture video and do appropriate readings/work** for each week. The class sessions (Mondays, 2:30 PM – 4:20 PM) will be used as labs/discussions (see below).

Labs:

- *Distance learning students: **You do not have to attend labs physically.*** You will however be given access to lab videos as appropriate (where we may discuss solutions to assignments/exercises/cases and/or handle Q&A on lecture topics). I expect you to watch these lab videos as they become available.

Special Note: This course is one of the required courses under the **Graduate Certificate in Business Data Mining** as well as the **SAS and OSU Data Mining (core level)**. If you want to learn more about these certificates, please visit the web sites <http://watson.okstate.edu/datamining/> and (<http://spears.okstate.edu/sasosu/sasosu/program/>) and/or talk to me.

Finally, as an instructor I retain the right to modify this tentative syllabus based on how the class progresses. If I make changes, I will let you know via D2L and/or email.

Class Requirements for DL Section 503 only

Exam: Your score in this exam will count for **40% of the course grade**. The exam will be due on Wednesday of the Finals week.

For those students who may want to work on a comprehensive data mining project on their own using data from their companies or from publicly available sources, I am willing to let you do so as **an alternative to the final exam**. If you pursue this option, an added benefit may be that you could publish a white paper (if it's your company data) or write a joint paper with me for next year's SAS analytics or SAS Global Forum conference. If you choose to work on such a project, then understand the rules as stated below. It is your responsibility to secure project data and arrange for any necessary permission from your company to share the project report with me. You may use numeric, textual or a combination of both types of data. The primary goal for your project should be developing a predictive model to predict your target variable or to develop a segmentation model. If you are interested in this option, then email me a 2-3 page proposal for your project by the end of January 31, 2015. *Your proposal must contain enough details (see end of this document) about the project for me to judge its suitability as an alternative to final exam.* If I accept your proposal, then you will have other interim deliverables (in mid-semester) and your final project report will be due on the Wednesday of the finals week.

Exercises and Assignments (Individual): Because of the emphasis on "hands-on learning" in this course, throughout the semester you will be working on many exercises (using appropriate software) on your own time. These exercises will primarily reinforce the concepts covered in the lectures. These exercises must be done individually and will count for **60% of the course grade**.

Special Note: Although DL students are not required to attend labs on campus, I *strongly suggest that you watch the lab videos* as soon as they become available because in those videos I will do demonstrations, discuss questions related to lecture topics, exercise solutions, etc. These lab videos will enhance your learning and also help you in doing exercises/assignments/cases/exam.

Semester Grades: The final grade for this section will be based as follows: 90% or above will result in A, 80% or more will result in B, 70% or above will result in C, 60% or above will result in D. Those getting less than 60% will get an F. I will look at the distribution of the total scores in this section and use any appropriate normalization as needed.

Late Assignments: Assignments or cases must be turned in by the class time on the due date via D2L drop box (not emails). All late assignments (*even 1-minute late*) must be turned in **via the Late Drop Box** and will be *penalized* as follows:

- One late assignment (within 48-hours of due date and time) – *no penalty*
- All other late assignments will carry following penalty structure:
 - Within 1 hour of due date and time – 15% penalty
 - More than 1 hour but less than 24 hours of due date and time – 30% penalty
 - More than 24 hours but less than 48 hours of due date and time – 50% penalty
 - More than 48 of due date and time – will not be graded (no credit)

I enforce this rule because I believe that part of effective functioning in business is the ability to complete projects on time. **Please do not email/call/contact me or my TAs with excuses (however valid they may be) about making exceptions to my late submission policy.**

Note: More details on the assignments/cases/readings/projects will be posted on the class D2L site. Also, for all other issues such as add/drop policy, academic integrity etc., I will follow OSU guidelines as posted in the site below – look at syllabus attachment for Spring 2015 (<http://academicaffairs.okstate.edu/faculty-a-staff>)

Project (Alternative to Final Exam) Proposal Details for Section 503 only

Any student who wants to do a project as an alternative to the final exam, must submit a 2-3 page proposal via D2L drop box by 11:59 PM US CST on **January 31, 2015**. The purpose of the proposal document is to help me (1) understand the nature and the scope of your project, (2) judge if it is doable within the semester and (3) if it is suitable as an alternative to the final exam, worth 40% of the course grade. I will read your proposal and let you know my acceptance or rejection within a week. If I accept you proposal, I will let you know about interim (mid-semester) and final (Wednesday of the Finals week) deliverables.

When you write the proposal, keep following things in mind:

1. A 2-3 line description of the main idea about your project
2. Describe the business opportunity or problem that your project addresses
3. Describe why anyone (the likely users/clients of your project) should care about this project. Think about what's the best thing that could happen if everything goes according to your plan and how that might benefit the company.
4. What data will you be using for the project?
 - a. I need a metadata for your data that describes variable names, variable types, data types, variable values, variable description, etc.
5. How you will be getting this data?
 - a. If you have to get it from your company, do you currently have access to it?
 - i. Has the company agreed to let you use the data for your project?
 - ii. If you do not have access now, how would you get it?
 - b. If you are using data from publicly available sources, then cite the source and answer following questions.
 - i. Have you downloaded the data? If not, then when?
6. A discussion of following items:
 - a. Your expectation of how much data editing/cleaning will be needed?
 - b. What type of analysis you might be doing – exploratory, predictive, segmentation, others
 - c. What types of recommendations do you expect to make based on this project

In general, the more details you can provide in your proposal, the better will be your chances of acceptance.