Data Mining and CRM Applications: MKTG 5963, Section 513(Online)

(Tentative Syllabus for Spring 2016)

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 (Via D2L or email). They will be your primary points of contact 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 TAs/instructor. Please use appropriate subject line in your posting and use threaded discussion whenever possible. Do not ask direct questions about how to solve an exercise (asking for clarification or help with software is ok).

Required Text:

No required text. I will primarily 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)

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 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, survival models,** etc. The course will use lectures, data analysis using state-of-the-art data mining software 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). It is your responsibility to keep up with lectures and complete the course related home work as mentioned in the syllabus and schedule.

Real Office Hours (to talk to me in person)

Monday and Tuesday 8:00-9:00AM, 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.

Course Prerequisites

Students must complete MKTG5983 (Database Marketing) 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 software before joining this class.

Course Format:

This section is entirely online – there is no physical class attendance required in this course. However, you must take the midterm and the final exam in Stillwater campus at the designated date and time (see course schedule)

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

Class Requirements for Non-DL section 001 only

Exams: One midterm exam (**20% of the course grad**e) and one comprehensive final exam (**40% of the course grade)**. *Please note on your calendar that this exam will be held on Stillwater campus on Monday evening.* Exact time and location will be communicated to you. You must take the exam on that day and time.

Individual Exercises: Because of the emphasis on "hands-on learning" in this course, there will be eleven individual exercises in this class. I will drop your worst exercise. **You must do these individual exercises alone and not seek help from others**. These exercises will primarily reinforce the concepts covered in the lectures. These exercises will count for **20% of the course grade**.

Group Assignment: There is one major group assignment (**worth 20% of the course grade**). You will work in groups of **about 4 students** to do this assignment. The actual number of students in each group will depend on the number of students enrolled in the class. From time to time, peer evaluation forms may be used to measure each group member's contribution to group work. I may also use the peer evaluations to adjust assignment grades for a group member, if necessary.

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 within this section and use any appropriate normalization as needed.

Late Assignments: Exercises or assignments must be turned in by the designated 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 1-hour 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: For all other issues such as add/drop policy, academic integrity etc., I will follow OSU guidelines as posted in the website (http://academicaffairs.okstate.edu/content/resources-faculty-staff)

Tentative Schedule of Topics (Spring 2016)

General: The schedule *below is tentative and subject to change* based on the pace of progress in class. The exercises/assignments will be announced via D2L class site. Additional exercises may be announced during lab sessions. All changes in the schedule will be communicated to you via email or D2L class site. Readings, exercises, assignments, projects, etc. will be assigned and posted on the D2L class site. It is your responsibility to check D2L site and your email every week for changes/announcements with respect to schedule/exercises/assignments etc.

Note about Exercises: Expect something to do (exercise, assignments, cases, etc.) in each week. In general, exercise/assignment corresponding to lecture topics covered in any week is due on **Monday of the next week by 1159 PM US CST** via appropriate drop box unless mentioned otherwise in the schedule or announced on D2L or via email. So, for example, Individual Exercise 1 mentioned in week 2 is actually due on Monday of week 3. Therefore, I will assume that before doing Individual Exercise 1, you must have reviewed lectures up to week 2. Exception of this rule is the exam weeks – in those weeks, exercises may be due earlier (see schedule below).

Week 1 (Week begins Jan. 11):

Video lecture: Course and faculty introduction, overview of data mining, a recap of basic statistical concepts, and overview of software access and interface.

Readings: What is data mining? (http://www.twocrows.com/intro-dm.pdf)

Week 2 (Week begins Jan. 18):

Video lecture: Overview of analytical methodology, Overview of data management and integration, Data preparation for data Mining, Overview of initial challenges in data mining, Overview of honest assessment in predictive modeling, Overview and demonstration of RFM Analysis.

Readings: Quick profits with RFM Analysis by Arthur Hughes (http://www.dbmarketing.com/articles/Art149.htm)

Exercise

Complete Individual Exercise 1 and upload your solution by Jan. 25, 1159 PM CST.

Week 3 (Week begins Jan. 25):

Video lecture: Exploring and preparing data with summary statistics, plots. Handling missing values, handling transformations and handling extreme values (outliers) in the data.

Readings:

- http://www.togaware.com/datamining/survivor/Exploring_Data.html

- http://en.wikipedia.org/wiki/Exploratory_data_analysis
- http://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm

Exercise:
Complete Individual Exercise 2 and upload your solution by Feb. 1, 1159 PM CST.

Week 4 (Week begins Feb. 1):

Video lecture: Introduction to predictive modeling via decision trees. Predictive modeling essentials, Understanding how split search works in decision trees, building and pruning decision trees, understanding tree variations with different splitting rules.

Readings:

- http://www.wuss.org/proceedings10/analy/3055_2_ANL-Hobbs.pdf
- http://www.statsoft.com/textbook/classification-trees/?button=1

Exercise:
Complete Individual Exercise 3 and upload your solution by Feb. 8, 1159 PM CST.

Week 5 (Week begins Feb. 8):

Video lecture: Introduction to predictive modeling via regression models. Overview of Logistic regression, handling sequential variable selection in regression, optimizing complexity in regression, using nonmetric inputs in regression, accounting for nonlinearities in regression.

Readings:

- Basics of logistic regression: http://faculty.cas.usf.edu/mbrannick/regression/Logistic.html
- Odds ratio : http://en.wikipedia.org/wiki/Odds_ratio
- Regression vs. Decision Trees: http://www.forecastingprinciples.com/paperpdf/exploratory.pdf

Exercise:

Complete Individual Exercise 4 and upload your solution by Feb. 15, 1159 PM CST.

Week 6 (Week begins Feb. 15):

Video lecture: Introduction to predictive modeling via Neural Networks. Background of ANN, Understanding how ANN works, Optimizing ANN models, other modeling tools in SAS EM, different types of ANN (MLP vs. RBF) models.

Readings:

- http://www.statsoft.com/textbook/neural-networks/?button=2
- http://en.wikipedia.org/wiki/Artificial_neural_network
- http://ulcar.uml.edu/~iag/CS/Intro-to-ANN.html

Exercise:
Complete Individual Exercise 5 and upload your solution by Feb. 22, 1159 PM CST.

Week 7 (Week begins Feb. 22):

Video lecture: Model assessment, model implementation and special topics. Assessment and comparison of models via summary metrics, graphs and charts. Adjustments for oversampling, creating and using profit matrices for model selection and scoring of new data. Use of decision tree to consolidate large number of categories into a smaller number and to understand predictions from a Neural net model.

Readings:

- http://homepage.cs.uri.edu/faculty/hamel/pubs/hamel-roc.pdf
- http://gim.unmc.edu/dxtests/ROC1.htm
- http://www.scholarpedia.org/article/Ensemble_learning

Exercise (Special Deadline):
Complete Individual Exercise 6 and upload your solution by Friday, Feb. 26, 1159 PM CST.

Week 8 (Week begins Feb. 29:

No lecture – Midterm Exam on Monday, Feb. 29 evening (block your calendar from 530 to 930 PM) at Stillwater Campus. Exact time, place for exam will be communicated to you. All students must take the midterm exam in Stillwater at the designated time.

Week 9 (Week begins Mar. 7):

Video lecture: Pattern discovery techniques (market segmentation via clustering)

Readings:

- http://en.wikipedia.org/wiki/Market_segmentation
- http://en.wikipedia.org/wiki/Cluster_analysis

Exercise (Special Deadline)
Complete Individual Exercise 7 and upload your solution by Mar. 21, 1159 PM CST.

Week 10 (Mar. 14 – Mar. 18): SPRING BREAK WEEK (NO LECTURE)

Week 11 (Week begins Mar. 21):

Video lecture: More on clustering, segmentation and profiling.

Readings:

- http://support.sas.com/resources/papers/proceedings13/068-2013.pdf
- http://support.sas.com/resources/papers/proceedings12/200-2012.pdf

Exercise:
Complete Individual Exercise 8 and upload your solution by Mar. 28, 1159 PM CST.

Week 12 (Week begins Mar. 28):

Video lecture: Association Rule discovery techniques.

Readings:

- http://en.wikipedia.org/wiki/Association_rule_learning
- http://www.statsoft.com/textbook/association-rules/?button=1
- Teknomo, Kardi (2008) **Market Basket Analysis.** http://people.revoledu.com/kardi/tutorial/MarketBasket/

Exercise:
Complete Individual Exercise 9 and upload your solution by Apr. 4, 1159 PM CST.

Week 13 (Week begins Apr. 4):

Video lecture: Introduction to survival analysis, exploratory analysis using survival curves, Using LIFETEST procedure and Life Tables method for survival analysis.

Readings: To be announced
Exercise:
Complete Individual Exercise 10 and upload your solution by Apr. 11, 1159 PM CST.

Week 14 (Week begins Apr. 11):

Video lecture: Parametric survival analysis using LIFEREG, Non-parametric survival analysis using PHREG.

Readings: To be announced
Exercise:
Complete Individual Exercise 11 and upload your solution by Apr. 18, 1159 PM CST.

Week 15 (Week begins Apr. 18):

Video lecture: None – work on group assignment
Exercise: None – work on group assignment

Week 16 (Week begins Apr. 25):

Video lecture: None – finish group assignment
Group Assignment:
Complete group assignment and upload your solution by Apr. 25, 1159 PM CST.

Week 17 (Week begins May 2):

Final Exam Week:

All students must take final exam on Stillwater campus between 530-930 PM on Monday, May 2. Exact time, place for exam will be communicated to you.