KOGOD SCHOOL of BUSINESS
AMERICAN UNIVERSITY • WASHINGTON, DC

MS Analytics Course

ITEC 621 Predictive Analytics

Last updated 6/12/2017

[Go to Class Schedule]

Professor: J. Alberto Espinosa, Ph.D.
alberto@american.edu
http://auapps.american.edu/alberto/www/

Office: KSB 33

Office Hours: T-Th 4:00 - 5:30 PM

Term: Summer 2017 (7-week module)

Class Schedule: Tu & Th 5:30 - 8:40 PM

Room: KSB T-61

Textbook

Required:

- "An Introduction to Statistical Learning: with Applications in R" by James, Witten, Hastie and Tibshirani, Springer, 1st. Edition, 2013. Please note that the authors of this book have a free PDF version on their website: http://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf
- ISLR Textbook authors' lectures and videos: http://fs2.american.edu/alberto/www/analytics/ISLRLectures.html

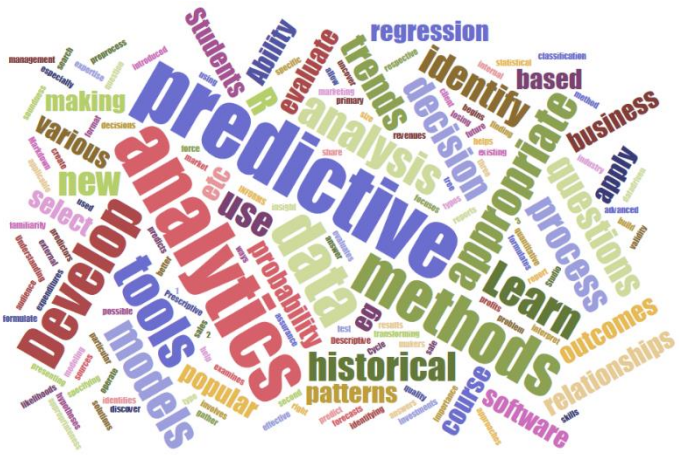
Optional (Recommended for R):

- "R for Everyone: Advanced Analytics and Graphics" by Lander, J., Addison-Wesley Data & Analytics Series, 1 edition, 2013

Analytics Resources: http://fs2.american.edu/alberto/www/analytics/AnalyticsResources.html

Course Overview

Analytics is the process of transforming data into insight for making better decisions (INFORMS). There are three primary types of analytics: "Descriptive," which examines historical data and identifies and reports historical patterns and trends; "Predictive," which predicts outcomes and future trends from existing data to help discover new relationships; "Prescriptive," which formulates and evaluates new ways for a business to operate. This course focuses on the second type, Predictive Analytics, which is of particular importance for business because it helps decision makers evaluate possible



outcomes (e.g., revenues, profits, market share, probability of making a sale, probability of losing a client, etc.) based on other historical data predictors (e.g., marketing expenditures, quality assurance investments, sales force size, etc.). The process of analytics involves specifying a question, problem, or decision, and finding the right answers using data. The process begins with identifying the appropriate data sources (internal or external, data format), and the appropriate models, tools, and methods for analysis. In this course, students are introduced to predictive modeling methods, approaches and tools. Students develop skills in predictive analytics that will allow them to: (1) develop and use advanced predictive analytics methods; (2) develop expertise in the use of popular tools and software for predictive analytics; (3) learn how to develop predictive analytics questions, identify and select the most appropriate predictive analytics methods and tools, apply these methods to answer the respective questions and presenting data-driven solutions.

Course Learning Objectives

After completing this class, the student will develop the following competencies.

- Competency-1: Predictive Analytics Methods
  - ✓ Ability to apply specific statistical and regression analysis methods applicable to predictive analytics to identify new trends and patterns, uncover relationships, create forecasts, predict likelihoods, and test predictive hypotheses.
  - ✓ Ability to develop and use various quantitative and classification predictive models based on various regression and decision tree methods.
- Competency-2: Predictive Analytics Tools
  - ✓ Develop familiarity with popular tools and software used in industry for predictive analytics, especially R, R Studio and R Markdown.
- Competency-3: The Predictive Analytics Cycle
  - ✓ Understanding of how to formulate predictive analytics questions.
  - ✓ Learn how to select the appropriate method for predictive analysis, and how to build effective predictive models.
  - ✓ Learn how to search, identify, gather and pre-process data for the analysis.
  - ✓ Learn how to evaluate the soundness, appropriateness and validity of their models and how to interpret and report on results for a management audience.

Student Requirements and Responsibilities

- Students need to be familiar with this syllabus and the weekly class schedule below. All assignments and class events will be posted either in the class schedule or on Blackboard.
- Similarly, students need to check all announcements posted on Blackboard before each class.
- Students are required to check their American University e-mail regularly for class announcements. Students who do not use their AU e-mail regularly need to either forward their AU e-mail to their personal e-mail accounts or change their e-mail address in Blackboard.
- Students are required, per University policy, to be familiar with AU's Academic Integrity Policy. Please read carefully the policies and read the Academic Integrity Policy section below. These policies will be strictly enforced in this course.
- Students are required to read all assigned material prior to class, prepare for class as instructed, participate actively in class discussion, and take a proactive role to maximize their learning from this class and in helping others benefit from the course. Students must read the assigned material before class and review the R code and related instructions before the corresponding R sessions.
- A good portion of the class lectures will come from sources other than the textbooks. Therefore, this class requires regular attendance and consistent week-to-week commitment on the part of the student. The material in this course is sequential in nature, so missing a lecture will not only affect the student's learning on the missed lecture, but also on subsequent material covered.

Grading Structure

Course Component	Weight	Composition
4 Homework	20% (4 @ 5% each)	Individual
Exam	30%	Individual
Term project	30%	Individual or Team
Quizzes and class exercises	10%	Individual
Attendance and participation	10%	Individual
Total	**100%**	

Grading Legend:

A: 93 or above; A-: 90 to less than 93;
B+: 88 to less than 90; B: 83 to less than 88; B-: 80 to less than 83;
C+: 78 to less than 80; C: 73 to less than 78; C-: 70 to less than 73;
D: C-: 60 to less than 70;
F: less than 60.

Course Components (all work in this course is individual)

1. **Homework:** 4 homework assignments on predictive analytics modeling. The homework will focus on hands-on use of R software to develop predictive models. The homework will be prepared in R Markdown and submitted as an HTML file (Produced by knitr HTML from R Markdown).

2. **Exam:** There one in-class exam towards the end of the semester. The exam will be conceptual. The exams will aim at testing your ability to process various business scenarios/problems/questions and

select and justify specific predictive modeling method. The exam covers all lectures and ISLR textbook readings up to and including the last class. R coding will not be covered in the exam but you need to be able to interpret plots and other outputs I prepared in R. In each question you will be presented with an analytics scenario. This scenario will contain one or more of the following:

2.1. A particular problem to resolve or business question to answer with predictive analytics (important note: your goal will NOT be to solve the problem or answer discussion, but to discuss your approach to do that);

2.2. An analysis goal (i.e., interpretation, inference or prediction); and

2.3. Relevant exhibits, which may include things like: model summary outputs; plots; distributions; data descriptions or displays. I will not ask any questions on R coding, but I will include a few R plots and outputs for interpretation. Each question will require a short, concise and precise answer demonstrating your knowledge of the material understanding of the specifics of the scenario, rather than long essays with vague generalities.

3. **Term Project:** The project will be done in teams of maximum 3 students. Students are also welcomed to work in pairs or even individually. For students working in teams, it is expected that all team members will contribute equally and that everyone will take the opportunity to learn from each other. Students will identify a business problem to address through predictive analytics. The goal is to select appropriate models and model specifications, and apply the respective methods to enhance data-driven decision making related to the business problem. Students will identify potential use of predictive analytics, formulate the problem, identify the right sources of data, analyze data, and prescribe actions to improve not only the process of decision making but also the outcome of decisions. See further instructions on Blackboard.

4. **Quizzes and Class Exercises:** You will complete several quizzes and class exercises during the semester. The quizzes will be short (10 minutes or so) based either on assigned material or on material already covered in class. The class exercises will involve short R assignments in class. There will be 8 to 10 quizzes and/or exercises during the semester. The lowest grade will be dropped from your average. So try not to miss more than one of these.

5. **Class attendance, participation and exercises:** Attendance is a straight percentage of the classes you attended, adjusted for lateness and early departures; In-class participation is measured by the ability of students to bring quality discussion into the class. This course is based on a model of active learning, with class discussions and exercises playing a central role. Students are expected to read the assigned material and to carefully prepare for all cases and exercises before coming to class and completing the required class exercises, when assigned. Students will be called upon to respond

to faculty questions. This course is very hands-on and the only way to learn the material well is through intensive exercises. About 50% of the class will be focused on hands on demonstrations and graded exercises.

Class Schedule

Note: The textbook authors have a nice series of video lectures where they narrate the book themselves. While these videos are not required, they really help understand the readings. You can find all the video lectures associated with each chapter at:
http://auapps.american.edu/~alberto/analytics/ISLRLectures.html

Week	Student Learning Objectives	Topic	Activity (R) Readings; (HW) Homework; (E) Exam; (P) Project; (V) Watch Video Lecture; (o) Other
1A Tu 5/16	Develop a deep understanding of the predictive analytics life cycle and several foundational concept that will be used throughout the course.	**1. Introduction** • Syllabus Overview • Course Introduction • The Analytics Life Cycle • Introduction to Predictive Analytics • Matrix Notation • Basic Foundations • (B) Model, Method and Feature Selection	All authors' slides and video lectures available at: http://auapps.american.edu/alberto/www/analytics/ISLRLectures.html (R) ISLR Ch.1 Introduction (V) ISLR Ch.1 (O) Download and install R and R Studio
1B Th 5/18	Overview of R for Predictive Modeling Overview of basic statistics and the Ordinary Least Squares (OLS) regression model	**2. R Refresher** **3. Regression Refresher** • Covariance, Correlation and ANOVA review. • Simple Linear Regression • OLS Model Diagnostics	(R) ISLR 2.3 Lab: Introduction to R (O) Recommended R book: R for Everyone
2A Tu 5/23	Further insights into the OLS regression model and its assumptions and limitations. Exploring the first departure from OLS due to heteroscedasticity WLS. Taking a first look at GLM	**3. Regression Refresher (cont'd.)** • Dummy Variables • Multivariate Regression • OLS Assumptions • Weighted Least Squares (WLS) • Generalized Linear Models (GLM)	(R) ISLR Ch.3 Linear Regression HW1 Due R Practice

2B Th 5/25	Learning to work with various data types and how to pre-process the data for analysis, including popular transformations like Box-Cox, standardized data, log transformations and lagging time series data.	**4. Data Pre-Processing** • Overview • Variable Types • Introduction to Data Transformations • Data Transformations: 1. Categorical to Dummy Variables 2. Polynomials 3. Box-Cox Transformation	
3A Tu 5/30		**4. Data Pre-Processing (cont.d)** 4. A) Log & Elasticity Models B) Logit Transformation 5. Count Data Models 6. Centering 7. Standardization	(R) ISLR Ch.2 Statistical Learning (R) ISLR 5.1 Cross-Validation
3B Th 6/1		**4. Data Pre-Processing (cont.d)** 8. Rank Transformations 9. Lagging Data (Causal Models) 10. Data Reduction	HW2 Due
4A Tu 6/6	Learning the basic concepts behind "machine learning" and the various ways of evaluating the predictive accuracy of models.	**5. Machine Learning** • Machine Learning Overview • Bias vs. Variance Tradeoff • Error Measures • Cross-Validation	
4B Th 6/8	Learning how to select the number of predictors in a model and address issues of dimensionality, like multi-collinearity.	**6. Variable Selection** • Dimensionality Issues • Multi-Collinearity • Variable Selection Methods • Step Methods	(R) ISLR Ch.6.1 Linear Model Selection HW3 Due
5A Tu 6/13	Learn how to build predictive models when the relationship between the predictors and the outcome variable don't appear to follow a linear trend.	**7. Non-Linear Models** • Non-Linearity Overview • Interaction Models • Polynomial Models • Step Models • Piecewise Models • Piecewise Linear Models • Piecewise Polynomial Models • Spline (MARS) Models	(R) ISLR Ch.7 Beyond Linearity **(P) Term Project Proposal Due**

5B Th 6/15	Learn how to build predictive models when the outcome is binary (e.g., yes/no, success/failure, approve/decline), using popular methods like logistic regression and discriminant analysis.	**8. Classification Models** • Introduction • Binomial Logistic Regression • Multinomial Logistic Regression • Linear Discriminant Analysis • Quadratic Discriminant Analysis	(R) ISLR Ch.4 Classification HW4 Due
6A Tu 6/20	Learn the various methods to build predictive classification models using decision trees, rather than regression models.	**10. Decision Trees** • Decision Trees • Regression Trees • Growing Trees • Regression Tree Issues • Classification Trees • Pruning Trees • Bootstrap Aggregation (Bagging) • Random Forest Models **Review for Exam**	(R) ISLR Ch.8 Tree-Based Methods
6B Th 6/22	**Exam** (up to and including 8 above)		
7A Tu 6/27	Learning how to handle predictive with large number of predictors, and how to reduce the set of predictors using regularization, penalized models and other dimension reduction methods like principal components and partial least squares. Course Wrap-Up	**9. Dimensionality** • (D) Regularization (Penalized or Shrinkage Models) ✓ Ridge Regression ✓ LASSO • (D) Dimension Reduction Models ✓ Principal Components Regression (PCR) ✓ Partial Least Squares (PLS) **Model/Method Selection Review** **Fly Solo** **Reviews for Term Project**	(R) ISLR Ch.6.2 Shrinkage Methods
7B Th 6/29		**Project Reports in Class**	**(P) Term Project Due** **IN CLASS**

Academic Integrity Code

Academic integrity is paramount in higher education and essential to effective teaching and learning. As a professional school, the Kogod School of Business is committed to preparing our students and graduates to value the notion of integrity. In fact, no issue at American University is more serious or addressed with greater severity than a breach of academic integrity.

Standards of academic conduct are governed by the University's Academic Integrity Code. By enrolling in the School and registering for this course, you acknowledge your familiarity with the Code and pledge to abide by it. All suspected violations of the Code will be immediately referred to the Office of the Dean. Disciplinary action, including failure for the course, suspension, or dismissal, may result.

Additional information about the Code (i.e. acceptable forms of collaboration, definitions of plagiarism, use of sources including the Internet, and the adjudication process) can be found in a number of places including the University's *Academic Regulations*, *Student Handbook*, and website at <http://www.american.edu/academics/integrity>. If you have any questions about academic integrity issues or about standards of conduct in this course, please discuss them with your instructor.

Academic Support Services

If you experience difficulty in this course for any reason, please don't hesitate to consult with me. In addition to the resources of the department, a wide range of services is available to support you in your efforts to meet the course requirements.

Academic Support Center (x3360, MGC 243) offers study skills workshops, individual instruction, tutor referrals, and services for students with learning disabilities. Writing support is available in the ASC Writing Lab or in the Writing Center, Battelle 228.

Counseling Center (x3500, MGC 214) offers counseling and consultations regarding personal concerns, self-help information, and connections to off-campus mental health resources.

Disability Support Services (x3315, MGC 206) offers technical and practical support and assistance with accommodations for students with physical, medical, or psychological disabilities. If you qualify for accommodations because of a disability, please notify me in a timely manner with a letter from the Academic Support Center or Disability Support Services so that we can make arrangements to address your needs.

Kogod Center for Business Communications (x1920, KSB 101) To improve your writing, public speaking, and team assignments for this class, contact the Kogod Center for Business Communications. You can get advice for any written or oral assignment or for any type of business communication, including memos, reports, individual and team presentations, and PowerPoint slides. Hours are flexible and include evenings. Go to http://www.kogod.american.edu/cbc and click on "make an appointment," visit KSB 101, or email cbc@american.edu. You may also call x1920.

Financial Services and Information Technology Lab (FSIT) (x1904, KSB T51) to excel in your course work and to maximize your business information literacy in preparation for your chosen career paths, we strongly recommend to take advantage of all software applications, databases and workshops in the FSIT Lab. The FSIT Lab promotes action-based learning through the use of real time market data and analytical tools used by business professionals in the market place. These include Bloomberg, Thomson Reuters, Argus Commercial Real Estate, Compustat, CRSP, @Risk etc. For more information, please check out the website at Kogod.american.edu/fsit/ or send us an email to fsitlab@american.edu.

EMERGENCY PREPAREDNESS FOR DISRUPTION OF CLASSES

In the event of an emergency, American University will implement a plan for meeting the needs of all members of the university community. Should the university be required to close for a period of time, we are committed to ensuring that all aspects of our educational programs will be delivered to our students. These may include altering and extending the duration of the traditional term schedule to complete essential instruction in the traditional format and/or use of distance instructional methods. Specific strategies will vary from class to class, depending on the format of the course and the timing of the emergency. Faculty will communicate class-specific information to students via AU e-mail and Blackboard, while students must inform their faculty immediately of any absence. Students are responsible for checking their AU e-mail regularly and keeping themselves informed of emergencies. In the event of an emergency, students should refer to the AU Student Portal, the AU Web site (http://www.american.edu/emergency/) and the AU information line at (202) 885-1100 for general university-wide information, as well as contact their faculty and/or respective dean's office for course and school/ college-specific information.