Course Name and Number: DATA 624 Predictive Analytics - SPRING 2019
Credits: *3 cr.*
Prerequisite(s): DATA 621, Sound Knowledge of R

Course Summary:
This course teaches students to use advanced machine learning techniques that are focused on predictive outcomes. Topics will include time series analysis and forecasting, recommender systems, and advanced regression techniques. In addition, students will learn how to evaluate the predictions that result from these techniques, how to assess model quality, and how to improve models over time.

Course Learning Outcomes:
At the end of this course, students will be able to:

- Apply advanced regression techniques such as constrained linear (PLS, NIPALS, Ridge, LARS), nonlinear (MARS, SVM, KNN), Trees (RF, Boosted).
- Utilize various forecasting techniques to produce reliable and robust forecast models.
- Develop recommendation systems using knowledge-based and content-based approaches.
- Evaluate the quality of models produced and make recommendations for improvement to models.

Program Learning Outcomes/Competencies addressed by the course:

- Business Understanding. Students will learn how predictive modeling and forecasting techniques can add value to existing business analytics.
- Data Understanding. Students will learn how to explore data to find patterns that allow for forward-looking forecasts and recommendations.
- Model Implementation. Students will learn to implement models for the various predictive modeling techniques covered in the course, with a focus on recommendations, estimation, and forecasting techniques.

How is this course relevant for analytics professionals?
Predictive modeling and forecasting, are mainstays of the analytics profession. Predictive modeling spans numerous fields and approaches. Indeed, within this course the student will be introduced to a multitude of techniques, some of which fall under the moniker "statistical modeling" while others are referred to "machine learning." For this course it's less important the lineage of a particular technique, but rather the classes of problems to be solved.

Each class of problems introduce multiple techniques. It's likely that the student has encountered many of these approaches in the past. This is both unavoidable and also fortuitous as the bulk of the course can thus focus on applying these techniques to the problem classes as opposed to learning the theory of the techniques

Assignments and Grading:
Each section comprises an introduction, a reading assignment, and book exercises. In addition, the bulk of the grading is focused on two course-specific projects along with paper submissions detailing methodology and results, including data visualizations.

NOTE: Most assignments will be a team effort. I will assign the team membership. Your team will elect a 'point person' or representative. This representative will be the person responsible for team submissions (projects and homework assignments). IMPORTANT: At the end of the class your teammates will rate every person on their team. This rating will determine the final grade for individuals on assignments. Thus, it is possible for a team to be awarded a B+ on a project and an individual to receive a C if all her/his teammates rate allot them low points.

Book Exercises 30% - Team Submissions, 2X

Completion of exercises must include working R code along with a brief discussion of the approach and results. Explicit instructions will be given. Assignments will be submitted in two batches. NOTE: You will work on these and submit them as a team. These will be collected in 2 batches (midterm and end of term). However you need to keep up with the assignments weekly. NOTE: you may turn in your work early, but not late without penalty. There will be a 5 point a day late penalty for the midterm assignment and no exceptions at the end of semester (zero). So, if you get sick often or other things impair you meeting deadlines you should consider turning them in early.

Discussion Board Authoring and Participation 10% - Solo Author, Class Response

You will be responsible to research and submit an engaging discussion topic. These will require citations and attribution where necessary. You will also respond to other discussions weekly.

Project A 15% - Team Submission

The first student project will be a time series and forecasting problem that you will tackle as a team. A professionally written report will be required. Details in Announcements. NOTE: you may turn in your work early, but not late without penalty. There will be a 10 point a day late penalty. So, if you get sick often or other things impair you meeting deadlines you should consider turning them in early.

Project B 30% - Team Submissions, 2X

The second student project will be a team predictive modeling problem. A professionally written report will be required. Part of your grade will be determined by your peers on your contribution to the submission. Details in Announcements. This is an end of term project and must be turned in on time, no exceptions.

Presentation 15% - Solo or Couple

You will be required to prepare and give a presentation for the class. Due to the size of the class I will allow 2 people to present together. We will be meeting in GoToMeetings and you will present from your workstation. Initial assignments will be voluntary and early volunteers receive extra credit. You can opt to co-present with a classmate. Details in Announcements and the 1st meetup.

Required Texts and Materials:
Reading assignments span two primary texts. These are

- Hyndman & Athanasopoulos. "Forecasting: Principles and Practice." https://www.otexts.org/fpp/
- Kuhn & Johnson. "Applied Predictive Modeling." http://appliedpredictivemodeling.com/ A

third book can be used for supplemental reading, which is

- Hastie, Tibshirani, & Friedman. "Elements of Statistical Learning." https://web.stanford.edu/~hastie/Papers/ESLII.pdf

Some of the reading will overlap across the two books. Where there is overlap, HA is generally more accessible, acting as an introduction, while KJ is a bit more theoretical. The student is encouraged to exercise judgment as to whether to skip the overlapping content.

NOTE: Books are referenced by abbreviation for convenience. Hyndman & Athanasopoulos is abbreviated HA, and Kuhn & Johnson is abbreviated KJ.

Relevant Software, Hardware, or Other Tools:
This course requires using the R language. Students must be familiar with the language and know how to install packages. All homeworks must be written in R and submitted as code that can easily be cut and copied into R Studio to run. Students must describe in written form their approach and analysis for all problems. The exposition is used to not only determine whether thought processes are sound but also to provide partial credit on problems.

My Contact Information:
Please address me as Dr. Burk or Dr. B during the course/semester. You are encouraged to ask me questions on the "Ask Your Instructor" forum on the course discussion board where other students will be able to benefit from your inquiries. I generally check the forums early in the mornings and in the evenings.

I am available by email (scott.burk@sps.cuny.edu). We can also set up an interactive session for screen share. For the most part, you can expect me to respond to questions by email within 24 to 48 hours. If you do not hear back from me within 48 hours of sending an email or have an emergency, you may send a text message or call my mobile phone at 254-563-6909 (text is better).

Course Outline:

This is the planned course schedule, there may be minor modifications required, therefore, be sure to keep up with announcements and attend meetups. Note that the first week is an introductory week and I expect you to use it to get a jump start on the class as well as make relationships with fellow students. You will greatly benefit by discussing and working together with your classmates. The projects are a major part of the class and will require early planning and organization.

CUNY Spring 2019	Presentation/Units	Assignments	
Friday, January 25, 2019	Class Starts (Saturday)	See Get Started, Read KJ#1, KJ#2	Due
Tuesday, January 29, 2019	Meetup, Introduction, Course Format and Objectives, KJ#1, KJ#2	Read HA #1, HA#2, HA#3 HW - HA(2.3, 2.7, 2.10, 3.1, 3.8) Prepare Your Presentation	Monday, February 4, 2019
Tuesday, February 5, 2019	Meetup, Time series decomposition (Presentation #1, Discussion #1 Opens)	Read HA #6 (HW 6.2 and 6.6), Prepare Your Presentation	Monday, February 11, 2019
Tuesday, February 12, 2019	Meetup, Data Pre-processing (Presentation #2, Discussion #2 and #3 Open)	Read KJ #3 (HW 3.1, 3.2), Prepare Your Presentation	Monday, February 18, 2019
Tuesday, February 19, 2019	Meetup, Exponential Smoothing Presentation, Discussion #4)	Read HA #7 (HW 7.5, 7.6, 7.10), Prepare Your Presentation	Monday, February 25, 2019
Tuesday, February 26, 2019	Meetup, Project #1 Assigned, Presentation, Discussion #5		Monday, March 4, 2019
Tuesday, March 5, 2019	Meetup, ARIMA Models Presentation, Discussion #6 Opens) Project #1 Assigned	Read HA #8 (HW 8.1, 8.2,8.7, 8.12)	Monday, March 11, 2019
Tuesday, March 12, 2019	No Presentation Discussion #7 and Catchup	COMPLETE HW - DUE NEXT WEEK	Monday, March 18, 2019
ALL HA Assignments Due by 3/18 @ 8 ET			
Tuesday, March 19, 2019	Linear Regression and Its Cousins, (Presentation/Discussion)	Read KJ #6 (HW 6.3), HW	Monday, March 25, 2019
Tuesday, March 26, 2019	Nonlinear Regression Models (Presentation/Disc)	Read KJ #7 (HW 7.2 and 7.5)	Monday, April 1, 2019
Tuesday, April 2, 2019	Regression Trees and Rules-Based Models (Presentation/Disc)	Read KJ #8 (HW# 8.1, 8.2, 8.3, 8.7)	Monday, April 8, 2019
Tuesday, April 9, 2019	Case Study - Strength of Concrete (Presentation/Disc)	Read KJ #10, finish chapter 8 problems	Monday, April 15, 2019
Friday, April 19, 2019	**Spring Recess**		Sunday, April 28, 2019
(Project #1 DUE 4/30/19 by 8 ET)			
(KJ Problems DUE 4/30/19 by 8 ET)			
Tuesday, April 30, 2019	Presentation/Disc Project #2 Assigned		Monday, May 6, 2019
Tuesday, May 7, 2019	Intro to Recommendation Systems (Presentation/Disc)	Assigned via Announcement	Monday, May 13, 2019
Tuesday, May 14, 2019	Clustering Applied to Recommender Systems (Presentation/Disc)	TBD	Monday, May 20, 2019
Tuesday, May 21, 2019	Class Wrap		Wednesday, May 22, 2019
(Recommender Problems and Team Project DUE 5/22/19 by 8 ET)			
Wednesday, May 22, 2019	Spring Term Ends		

ACCESSIBILITY AND ACCOMMODATIONS

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the
resources and accommodations provided. Please see:
http://sps.cuny.edu/student_services/disabilityservices.html

ONLINE ETIQUETTE AND ANTI-HARASSMENT POLICY

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see:
http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf

ACADEMIC INTEGRITY

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see:
http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf

STUDENT SUPPORT SERVICES

If you need any additional help, please visit Student Support Services:
http://sps.cuny.edu/student_resources/