MSA 8200 Predictive Analytics

Class Information:

Location: Room 306, GSU Buckhead Campus.
Website: iCollege http://icollege.gsu.edu
Prerequisites: MSA8190, or consent of the instructor.
You will need to be familiar with statistical programming environments. We will rely on R for the coursework. http://www.r-project.org.

Instructor Information:

Yichen Cheng @ Room 333, GSU Buckhead Campus.
Office hours: (Online after March 20th)
      Monday: 1:30 pm – 3:00 pm
      Wednesday: 4:30 pm – 6:00 pm
Email: ycheng11@gsu.edu

Course Descriptions:

This course introduces students to different predictive models with a focus on real-world applications and datasets. The course covers three primary topics: 1) the analyses of time series data, including estimation and inference for ARIMA models as well as more recent developments; 2) a collection of other important models including Bayesian analyses, survival analyses, models for count data; 3) the set of skills required to analysis real world data, including data pre-processing, data type identification, and different types of models for different data type. In addition, the students will also have hand on experience of working with real world data.

Textbooks:

[S] Shumway, R.H., and Stoffer D.S., Time Series Analysis and Its Applications. Fourth Edition. Springer.
[K] Kleiber, C., and Zeileis, A., Applied Econometrics with R. Springer.
[D] Kleinbaum, D.G., and Klein M., Survival Analysis: A Self‑Learning Text. Third Edition. Springer.
[P] Class notes will be posted on iCollege.

Other Useful References:

[R] https://www.econometrics-with-r.org/
[E] The elements of statistical learning: https://web.stanford.edu/~hastie/Papers/ESLII.pdf

Course Learning Outcomes:

At the end of this course students will be able to:
1. Understand different data types and identify the correct model for data analysis.
2. Model and forecast time series data.

3. Deal with messy data that has missingness or censored observations.
4. Feel comfortable with analyzing real world data set and providing business insights.

Methods of Instruction:

The material will be presented in lecture form. As a general approach, I first discuss models and methods conceptually, and I then provide and discuss a variety of example problems/programs that illustrate the concepts.

Attendance Policy:

Attendance is not formally taken. However, it is strongly suggested that students do not miss class as most students will have difficulties completing the assignments without attending the lectures.

Homework

There will be homework assignments to clarify and deepen concepts. Homework assignments will be posted on iCollege. Please pay attention to the due dates. No late homework will be graded.

You may discuss the assignments among each other, but every student has to write up the assignment on her/his own. Students copying from their classmates and the student who let others copy from her/his assignment will receive a zero score for the corresponding homework or possibly the ENTIRE course.

Exam:

The mid-term exam will be closed book. You may bring a cheat sheet (one-sided A4-sized paper).

Course Project:

There are two projects for the course: one mini project and one final project. Students will form into groups of 3-4 people and work on the projects. The same group should work on both the mini project and the final project.

Mini project: A list of the mini project topic will be posted on icollege. Each group will sign up for a different topic. The basic requirement is to follow the analyses in the materials provided. Each team will have 15 mins to present their work (10 mins presentation + 5 mins Q&A). The presentation should address the following items: 1. What was done? 2. Why you do it this way? 3. Are there other better alternatives? The presentation will be evaluated by: 1. Quality of the analyses. 2. The presentation skills. 3. The Q&A session.

Final project: Each team will formulate a problem about predictive analytics and discuss the idea with the instructor by Week 10. **Each team will have 20 mins to present their work (a 20 mins presentation)**. The presentation should address the following items: 1.

Problem Statement. 2. Different approaches you have tried. 3. Conclusions and business insights. The presentation will be evaluated using similar criteria as the mini project.

Grading:

Quizzes---15%
Homework--15%
Mini project ---10%
Midterm--30%
Final project---30%

The anticipated grading scales for this class is as follows:

A+	A	A-	B+	B	B-	C+	C	C-	D+	D	F
98	95	90	85	80	75	70	65	60	55	50	<50

Please advise the instructor if you have a documented disability that needs to be accommodated.

As members of the academic community, students are expected to recognize and uphold standards of intellectual and academic integrity. See the University's policy on Academic Honesty (Section 409, http://www2.gsu.edu/˜wwwfhb/sec409.html) for details. Accommodations for students with disabilities: Georgia State University complies with Section 504 of the Rehabilitation Act and the Americans with Disabilities Act. Students with disabilities who seek academic accommodations must first take appropriate documentation to the Office of Disability Services locate in Suite 230 of the New Student Center.

Detailed Outline of the class:

Date	Topics	References	Readings	Due
Week 1 1/15/2020	Introduction of time series models ACF Stationary process	[S] Ch 1-3	https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/	
Week 2 1/22/2020	AR models	[S] Ch 3	https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/	
Week 3 1/29/2020	**Quiz 1** MA model PACF	[S] Ch 3	http://r-statistics.co/Time-Series-Analysis-With-R.html	HW1
Week 4 2/5/2020	ARIMA model Alternative ways?	[S] Ch 3	https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-	

			learningnd-deep-learning-techniques-python/	
Week 5 2/12/2020	Regression with time series data. Case Study: Sales prediction	[K] Ch 3.5	https://rpubs.com/spillai/walmart_store_sales_forecast	HW2
Week 6 2/19/2020	**Quiz 2** Bayesian analysis Mini projects overview Review session	[E] Ch 8.3 [P]		
Week 7 2/26/2020	Midterm			
Week 8 3/4/2020	Omitted variable bias Panel data Fixed effect model Random effect model Diff in diff model	[R] Ch 6,10 [P]		
Week 9 3/11/2020	Mini project Presentation Guest speaker			Report Due
Week 10 3/18/2020	Spring Break			
Week 11 3/25/2020	No Class			
Week 12 4/1/2020	Regression models for count data	[K] Ch 5.1 & Ch 5.3	https://corporatefinanceinstitute.com/resources/knowledge/other/poisson-distribution/ https://bookdown.org/roback/bookdown-bysh/ch-poissonreg.html	
Week 13 4/8/2020	Predicting arrivals: Survival analysis, Censored dependent variables	[D] Ch 2-3 [P]	https://blog.exploratory.io/introduction-to-survival-analysis-part-1-survival-curve-9364ea642114 https://www.emilyzabor.com/tutorials/survival_analysis_in_r_tutorial.html	
Week 14 4/15/2020	Dimension reduction;	[E] Ch 14 [P]		HW4

	Graph analytics: PageRank, etc.			
Week 15 4/22/2020	Final presentation			HW5 Report Due

Note that this course syllabus provides a general plan for the course; deviations may be necessary.