PLSC 398B
Data Analytics for Political Science

Onsel Gurel Bayrali
Spring 2023

E-mail: obayral1@binghamton.edu
Office Hours: by appoinment

Course Description:

This course introduces techniques to collect, analyze, and utilize extensive data for social science inferences. The ultimate goal of this class is to improve students' statistical and programming skills. For this purpose, we will cover the recent developments in machine learning techniques and their implementation in political science and public policy.

We will use R, an open-source statistical computing environment. Since this course will introduce the basics of R programming, no prior exposure to the statistical programming language is required. Additionally, this class will provide an overview of how to summarize, visualize, and create reproducible documents in R. Then, students in this course will learn the basics of supervised and unsupervised learning techniques and their empirical implementations in research in politics.

Course Requirements:

The final grade will be evaluated in three major areas: take-home midterm (20%), weekly assignments (50%), and the final project (30%).

1) **Take-home midterm (20%):** The midterm exam will cover the topics discussed in the first stage of the course. Questions will be similar to weekly assignment questions. The tentative take-home exam date is on March 1, 2023. Students are expected to return their answer sheets within 48 hours after the exam questions are released. Students are expected to use R Markdown while answering exams and to return their answer sheets in pdf format. I will share detailed information about the exam format as the exam date approaches.

2) **Weekly assignments (50%):** Students will receive 4-6 homework assignments throughout the course. Questions will be about topics covered in the course and will encourage students to focus on combining R-programming language and machine-learning techniques. Questions will be similar to exercises covered in the class.

3) **Final project (30%):** In the final project, students will have a chance to work on a dataset provided by the course instructor. In this project, students will apply methods covered in the course to address a politically relevant problem. This assignment is designed as a group project. On the final day of the class, each group will present their project to the class. Also, each group is expected to submit a research summary (at most five pages), including the descriptive statistics of the data, graphical illustrations, models, and a brief statement about the results. Details, if necessary, will be shared in the course.

4) **Participation (10%):** Students can earn participation points as a bonus by attending and asking questions in class.

Notes:

- Collaboration in weekly assignments is strongly encouraged. However, it is expected to work out details by yourself, and each student must submit their answers separately. Identical assignments are not allowed and will be penalized.
- Collaboration in the midterm exam is not allowed. You may use any textbooks or online resources, but you must work on the exam yourself.

Computer Software

Throughout this course, we will extensively use R/Posit. You can download it here from the below link.

https://posit.co/download/rstudio-desktop/

Books

There is no required book for this course. I will post readings on Brighspace.

Recommended/Reference Books

Even though there is no required text, the following books might be helpful for references. Each week, I will inform students about which book I use as a reference.

1) James, Witten, Hastie, and Tibshirani "An Introduction to Statistical Learning: with Applications in R."
2) Oswald, Robin, and Viers "Introduction to Econometrics with R."
   https://scpoecon.github.io/ScPoEconometrics/
3) If you need more information about using R, you may consult Venables, Smith, and the R Core Team "An Introduction to R."
   https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
4) A slightly more advanced text but excellent treatment of machine learning methods. Murphy, Kevin P. "Machine Learning: A Probabilistic Perspective."

Other Issues

Late submissions: Assignments handed in within 24 hours of the deadline will have one entire letter grade deducted. Assignments handed in after this time will have one entire letter grade deducted for each additional 24-hour period they are late. Exceptions to these rules will only be made in the case of emergencies and will require documentation. It is up to the instructor to decide what constitutes an emergency.

Disability-related Equal Access Accommodations. Students wishing to request academic accommodations to insure their equitable access and participation in this course should notify the instructor by the second week of class. Authorizations from Services for Students with Disabilities (SSD) are generally required. We encourage you to contact SSD at (607)777-2686 to schedule an appointment with the Director or Learning Disabilities Specialist. Their website (www.binghamton.edu/ssd) includes information regarding their Disability Documentation Guidelines. The office is located at UU-119.

Academic Honesty. Students should review university guidelines regarding academic honesty and ensure that the work they complete for this course is theirs and theirs alone. Students will be asked to submit their written work through Turnitin: a plagiarism detection program. Cases of plagiarism will be taken very seriously and may result in a failing grade for the course and any penalty imposed by the university. If you have any doubt about what constitutes plagiarism or a violation of academic honesty, please consult the university code and contact the instructor.

Information on plagiarism can be found at

http://www2.binghamton.edu/watson/about/academic-honesty.html.

Tutorials on academic honesty can be found at
http://library.binghamton.edu/research/tutorials/WebTutorials/web/index.html.

Resources for Students in Distress. Suppose you are experiencing undue personal or academic stress at any time during the semester or need to talk with someone about a personal problem or situation. In that case, I encourage you to seek support as soon as possible. Significant stress, mood changes, excessive worry, or problems with eating and sleeping can interfere with optimal academic performance. The source of symptoms might be related to your coursework; if so, I invite you to speak with me (or your other professors) directly. However, problems with relationships, family worries, loss, or a personal struggle or crisis can also contribute to decreased academic performance and may require additional professional support. Binghamton University provides various support resources: the Dean of Students Office and University Counseling Center offer to coach on reducing the impact on your grades. Numbers for these and other available resources are provided below.

1. Dean of Students Office: 607-777-2804
2. 2. Decker Student Health Services Center: 607-777-2221
3. 3. University Police: On-campus emergency, 911
4. 4. University Counseling Center: 607-777-2772
5. 5. Interpersonal Violence Prevention: 607-777-3062
6. 6. Harpur Advising: 607-777-6305
7. 7. Office of International Student & Scholar Services: 607-777-2510

Course Outline

Week 1: January 16 Introduction and Preliminaries

Recommended Readings:
Justin Grimmer. We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. Available at https://stanford.edu/~jgrimmer/bd_2.pdf

Week 2: January 25 Basics of Statistical Inference

Recommended Readings:
Chatsiou, K., & Mikhaylov, S. J. (2020). Deep learning for political science. *arXiv preprint arXiv:2005.06540*.

Week 3: February 1 Importing and Visualizing Data

Recommended Readings:
Cranmer, S. J. Introduction to the Virtual Issue: Machine Learning in Political Science.
Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, *24*, 395-419.

Week 4: February 8 Linear Regression I

Week 5: February 15 Linear Regression II

Week 6: February 22 Classification

Recommended Readings:
Rossini, P., J. Stromer-Galley, and F. Zhang (2020). Exploring the relationship between campaign discourse on Facebook and the public comments: A case study of incivility during the 2016 US presidential election. Political Studies, 0032321719890818.

Week 7: March 1   Take-home Midterm (No classes)

Week 8: March 8   Resampling Method

Recommended Readings:
Marcel Neunhoeffer and Sebastian Sternberg. (2019). How Cross-Validation Can Go Wrong and What to Do About It." Political Analysis 27, no. 1 (January): 101-106.

Week 9: March 15   Linear Model Selection

Recommended Readings:
Francisco Cantu and Sebastian M. Saiegh. (2011). Fraudulent Democracy? An Analysis of Argentina's Infamous Decade Using Supervised Machine Learning." Political Analysis. 19: 409-433.

Week 10: March 22 Moving Beyond Linearity

Recommended Readings:
Golder, M., S. N. Golder, and D. A. Siegel (2012). Modeling the institutional foundation of parliamentary government formation. The Journal of Politics 74 (2), 427-445.

Week 11: March 29 Introduction to Unsupervised Learning

Recommended Readings:
Cullen S. Hendrix. (2010). Measuring state capacity: Theoretical and empirical implications for the study of civil conflict." Journal of Peace Research 47 (3): 273-285.

Week 12: April 19   Tree-Based Methods

Recommended Readings:
Mohnen, S. M., A. H. Rotteveel, G. Doornbos, and J. J. Polder (2020). Healthcare expenditure prediction with neighborhood variables-a random forest model. Statistics, Politics and Policy 11 (2), 111-138.
Streeter, S. (2019). Lethal force in black and white: Assessing racial disparities in the circumstances of police killings. The Journal of Politics 81 (3), 1124-1132.

Week 13: April 26   Working with Text as Data

Recommended Readings:
Robinson, D. (2016). Text Analysis of Trump's Tweets Confirms He Writes Only the (Angrier) Android Half.
Robinson, D. (2016). Trump's Android and iPhone Tweets, One Year Later.

Week 14: May 3   Final Project Presentations