[PADP 9200] Ethics and Algorithms in Public Policy

Professor Jason Anastasopoulos
Email: ljanastas@uga.edu

Office Hours: Tuesday, Thursday, 3:30-4:30 PM

Course Objectives

The big data revolution is transforming public policy and governance. The goal of this course is to provide a non-technical overview of some of the methods driving big data methodologies and to explore how these technologies are shaping and will shape the future of public policy and government with an eye towards the ethical dilemmas that these technologies raise. We begin the course with a discussion of some of the fundamental theories and applications of machine learning methods, which form the basis of big data and artificial intelligence technologies. We then move on to an in-depth discussion of the ethical and societal promises and perils that these technologies pose for decision making in government more broadly, focusing on the potential of these techniques to shape government and policy.

Attendance and Participation

The most important content from this class will come from the lectures and group assignments during lecture time. Because of this and the rather technical nature of this class, attendance and participation in class is extremely important. If you cannot attend a lecture you must present me with a valid excuse at least 24 hours prior to the start of class unless the situation you encountered was an emergency. Either way, absence requires explanation and documentation if you do not want want points taken off your final participation grade.

Grading and Requirements

- *Participation/discussion leader: 40%.*
- *Three (3) problem sets: 20% (lowest of 3 is dropped)*
- *Final group project proposal: 5%*
- *Final group report and presentation: 35%.*

Key Assignment Dates (to be submitted via the ELC)

- **TBD**

Key Discussion Leader Dates (TBD)

- *Algorithms and government: overview.*
- *Algorithms and government - decision making in theory.*
- *Algorithms and government - decision making in practice - police and judges*
- *Strengths and Weaknesses of Machine Learning Systems for Public Policy*
- *Machine Learning and Bias I: Overview*
- *Machine Learning and Bias II: Sources and Pathways in Practice*
- *Machine learning, big data and ethics: the modern panopticon?*
- *Machine learning and political institutions I: accountability and transparency*

Discussion Leaders

Groups of students will be assigned the role of "discussion leaders." Discussion leaders will lead the class discussion for that week by reading the assigned content, preparing a 15-30 minute presentation summarizing the readings and will propose a series of 3-5 and discussion points to start off our discussion about the content. Every student MUST participate in a group as a discussion leader at least once. Discussion leader groups can have a maximum of 2 people and if you do not sign up for one week as a discussion leader you will be assigned to a week.

Problem Sets

During the first three weeks of the course, there will be three problem sets which you can work on in groups and which are designed to give you some hands on experience with machine learning algorithms in a policy context. These problem sets will involve some rudimentary programming in the statistical language **R** and will teach you about some of the basic machine learning algorithms used in practice today.

Group Project

A major portion of your grade will involve writing a policy memorandum in response to a question that I will assign two weeks prior to the memorandum due date. For guidelines on how to write a policy memorandum, please see this excellent guide by Iris Malone: http://web.stanford.edu/~imalone/Teaching/ps1winter17/PS1-PolicyMemo.pdf. Policy memoranda will be graded on the basis of the quality and clarity of your writing and the quality and clarity of the ideas that you present.

Required and Recommended Texts*

Foucault, Michel. 1995 edition. “Discipline & Punish: The Birth of the Prison”.

Machine Learning: The New AI Ethem Alpaydin (2016) MIT Press. Referred to in the schedule as **EA**.

James, Witten, Hastie and Tibshirani. 2015. *An Introduction to Statistical Learning with Applications in R*. Springer Science. Available for free here: http://www-bcf.usc.edu/~gareth/ISL/. Referred to in the schedule as **JWHT**.

Monogan III, James E. 2015. Political Analysis Using R , Springer. http://link.springer.com/book/10.1007%2F978-3-319-23446-5 .
Referred to in the schedule as **M3** .

* Most texts will be available online on the course site.

Course Outline

1. Introduction to machine learning and artificial intelligence.
2. Overview of algorithms and government.
3. Algorithms, ethics and public policy.
4. Algorithms, behavior and decision-making.
    - Decision making by humans and machines.
    - Strength and weaknesses of machine learning and AI.
    - Machine learning and fairness.
5. Algorithms and political institutions.
    - War and international affairs
    - Policymaking.
    - Accountability.

COURSE SCHEDULE

Class 1 : *Introduction to Machine Learning and Artificial Intelligence: Fundamentals*

Programming fundamentals.

- Introduction to programming in **R**.

Overview of machine learning

- Machine learning in public organizations.
- What is machine learning?
- Supervised & unsupervised learning.
- Inference versus prediction.

Readings:

- Larson et al "How We Analyzed the COMPAS Recidivism Algorithm" https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
- **EA** Chapter 1.
- **M3** Chapters 1, 2, 10, 11.1-11,4.
- Kleinberg, J., Ludwig, J., Mullainathan, S. and Obermeyer, Z., 2015. Prediction policy problems. American Economic Review, 105(5), pp.491-95. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4869349/pdf/nihms776714.pdf .
- **JWHT** – Introduction, pp 1-15.

Class 2 : *Introduction to statistical learning theory*

- Training, testing and cross–validation.
- Assessing model accuracy.
- Overfitting.
- Regression vs. classification problems.
- The Bias–Variance tradeoff.
- **Application**: H1–B Visa Certification. H1–B Application Data.

Readings

- ❖ **JWHT –** Statistical Learning, pp 15-37, 176–184.
- ❖ **EA** - Chapter 2.

Class 3 : *Understanding supervised machine learning through examples: decision trees and regression.*

- "Pure ML": Decision tree algorithms and CART.
- "Statistical ML": Linear regression as a machine learning algorithm.
- Application 1: Preventative policing: pre–crime targeting and detection. NYC Stop and Frisk Data: 2003–2016

Readings

- ❖ **JWHT –** Chapter 3.
- ❖ **EA -** Chapter 3.

Class 5 : *Algorithms and government - overview.*

- Overview of administrative decision making and machine learning.
- Bureaucracy and technology.

Readings

- ❖ **JW** - Chapters 1 & 2.
- ❖ Lee, Ronald M. "Bureaucracies, bureaucrats and information technology." *European Journal of Operational Research* 18, no. 3 (1984): 293-303.
- ❖ Coglianese and Lehr. 2017. "Regulating by Robot Administrative Decision Making in the Machine-Learning Era". Georgetown Law Journal.

Class 6: *Algorithms, ethics and public policy I: Approaches*

- Approaches to algorithmic fairness.

Readings

- Binns, Reuben. "What Can Political Philosophy Teach Us about Algorithmic Fairness?." *IEEE Security & Privacy* 16, no. 3 (2018): 73-80.

- Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. "From What to How. An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices." *arXiv preprint arXiv:1905.06876* (2019).

- Heidari, Hoda, Michele Loi, Krishna P. Gummadi, and Andreas Krause. "A moral framework for understanding of fair ml through economic models of equality of opportunity." *arXiv preprint arXiv:1809.03400* (2018).

- Glymour, Bruce, and Jonathan Herington. "Measuring the Biases that Matter: The Ethical and Casual Foundations for Measures of Fairness in Algorithms." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 269-278. ACM, 2019.

Class 7: *Algorithms and government - decision making in theory.*

- Decision making by humans v. machines.

Readings

- *Lipsky, Toward a theory of street-level bureaucracy.*
- *Keiser, Lael. 2010. "Understanding Street Level Bureaucrats Decision Making," Public Administration Review. 70 (02) pp.247-57.* ***JSTOR***
- *Dawes et al. Clinical versus actuarial judgement*
- *Dietvorst. Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*

Class 8: *Algorithms and government - decision making in practice - police and judges*

- Technology and street level bureaucracy.
- Applied examples: judges and police.

Readings

- Buffat, Aurélien. "Street-level bureaucracy and e-government." *Public Management Review* 17.1 (2015): 149-161.
- Kleinberg, Jon, et al. "Human decisions and machine predictions." *The quarterly journal of economics* 133.1 (2017): 237-293.
- Harcourt, Bernard E. "Against prediction: Sentencing, policing, and punishing in an actuarial age." (2005).

Class 9 : *Strengths and Weaknesses of Machine Learning Systems for Public Policy*

- Strengths and weaknesses of the machine learning approach and how it might apply to public policy.

- *Breiman. Statistical Modeling: The Two Cultures*
- *Norvig. On Chomsky and the Two Cultures of Statistical Learning*
- *Lazer et al. The parable of Google Flu.*
- *Olteanu et al. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries*

Class 10 : *Machine Learning and Bias I: Overview*

- Defining bias in the machine learning context.

- *Angwin et al. Machine Bias*
- *Angwin & Larson. Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say*
- *Chouldechova. Fair Prediction with Disparate Impact: A study of bias in recidivism prediction instruments.*
- Kleinberg et al. Inherent Trade-Offs in the Fair Determination of Risk Scores
- *Corbett-Davies et al. Algorithmic Decision Making and the Cost of Fairness.*

Class 11 : *Machine Learning and Bias II: Sources and Pathways in Practice*

- Machine learning bias in practice.

- *Pierson et al. A large-scale analysis of racial disparities in police stops across the United States*
- *Caliskan et al. Semantics Derived Automatically from Language Corpora Contain Human-like Biases*
- *Torralba & Efros. Unbiased Look at Dataset Bias*

Class 12 : *Machine learning, big data and ethics: the modern panopticon?*

- Ethical considerations of big data and machine learning.

- *Foucault. “Panopticism” in Discipline and Punishment. Pp 195-228*
- *Ohm & Peppet. What if Everything Reveals Everything?*
- *Kosinski, Stillwell, and Graepel. Private Traits and Attributes Are Predictable From Digital Records of Human Behavior*
- *Wu & Zhang. Automated Inference on Criminality using Face Images*
- *Wang & Kosinsky. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images*

Class 13 : *Machine learning and political institutions I: accountability and transparency*

- Coglianese and Lehr, Transparency and Algorithmic Governance
- *Sandvig, Hamilton, Karahalios, and Langbort, Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*
- *Ananny & Crawford, Seeing without knowing Limitations of the transparency ideal and its application to algorithmic accountability*