Machine Learning for Public Policy

Spring 2019
Tuesday-Thursday 11:00-12:20

Contact Information:

Rayid Ghani
Office: Crerar 260
Office Hours: Wednesdays 1-2pm, Thursdays 2-3pm (or by appointment)
Email: rayid@uchicago.edu

TAs: TBD

Textbooks:

There are two recommended textbooks:

1) Data Science for Business. Foster Provost and Tom Fawcett
2) Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Peter Flach

You will be expected to read the assigned chapters before each class so we can use class time to have productive discussions and go over any aspects that need further explanations and policy context.

Course Description:

This course will be an *introduction* to machine learning techniques and how to use them to help solve public policy problems. This course is designed for public policy and social science students who are interested in learning modern, scalable, computational data analysis methods (buzzwords include machine learning, data science, big data, AI), and apply them to social and policy problems.

This course will teach students:

1) What role Machine Learning can play in designing, implementing, evaluating, and improving Public Policy.
2) How Machine Learning methods work, how to use them, and how to building machine learning pipelines/systems.
3) How to tackle policy problems using Machine Learning methods and tools

This is a hands-on course where students will be expected to use Python to implement solutions to various policy problems. Prior experience with Python is required. We will cover supervised and unsupervised learning algorithms and will

learn how to use them for public policy problems in areas such as education, public health, sustainability, economic development, and public safety.

Prerequisites:
- Two courses in Computer Programming (Python experience required)
- Two courses in Probability & Statistics
- Courses in Discrete Math and Linear Algebra
- Prior experience with data analysis is highly recommended (using SQL, R, Python)
- Familiarity with Linux and command line tools
- Familiarity with git and github

Topics:

We'll cover the following topics:

1) Overview of the Machine Learning Process
   a) Understanding the Policy Problem
   b) Mapping it to a Machine Learning formulation
   c) Understanding the Data: Data Exploration and Data Stories
   d) Iterative Machine Learning Pipeline Development
      i) Setup the problem
      ii) Feature Development
      iii) Modeling
      iv) Model Selection
      v) Model Interpretation
      vi) Model and System Evaluation
   e) Deployment: Trials, deployment, and maintenance

2) Machine Learning Methods – Spectrum from unsupervised to supervised
   a) Supervised: Regression vs Classification, Binary vs Multiclass vs Multilabel
      i) Logistic Regression
      ii) Nearest Neighbors
      iii) Tree-based models
      iv) Bayesian Methods
      v) Neural Networks
      vi) Ensemble Methods: Bagging, Boosting, Stacking, Random Forests
   b) Unsupervised
      i) Clustering
      ii) Association Rules
      iii) PCA and related methods
3) Applying these methods to Policy Problems
4) Ethics, Bias, Privacy Issues
5) Additional Topics (as time permits)
   a) Text analysis

b) Network Analysis

Lectures:

The lectures are a work in progress. The schedule is subject to change based on class interest and progress. In addition, we may have guest lectures which will cause some of these lectures to be merged. If there are additional topics you'd like to cover or guest lectures you'd like to see, please let me know.

Assignments:

- Regular Programming assignments and quizzes in class covering readings
- Short response/feedback to previous week's lectures due Tuesday before class every week
- There will be a take-home, open-book, open-internet mid-term exam
- There will be no final. Instead, each project team will have a report due and an oral presentation at the end of the quarter.

Project:

Students will form groups (4 students each) and work on a project they'll propose after week 3. I will provide a set of project ideas (and data) to choose from but feel free to bring your own as well and we can discuss if they meet the project criteria for this class.

You will have to submit a project proposal (that will be reviewed by your peers), do a project proposal presentation, submit project update report, a final project report, and do a final project presentation.

Grades:

Assignments and Quizzes: 40%
Mid-Term (extended assignment): 15%
Weekly Class Reviews: 10%
Class Participation: 10%
Project: 25%
   Proposal and Proposal Presentation: 5%
   Progress Report: 5%
   Final Report: 10%
   Final Presentation: 5%