Machine Learning for Public Policy

Spring 2017
Tuesday-Thursday 10:30-11:50 (Ryerson 276)

Contact Information:

Rayid Ghani
Office: Searle 219
Office Hours: Tuesday and Thursdays 2-3pm (or by appointment)
Email: rayid@uchicago.edu

TA: TBD
Email: 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

Course Description:

This course will be an introduction to machine learning techniques and tools as well how they can be used in public policy problems. This course is designed for public policy and social science students who are interested in learning how to use modern, scalable, computational data analysis methods and tools, 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) Machine Learning methods and tools, with a focus on supervised and unsupervised methods, practical aspects of how to use them, and building machine learning pipelines using Python
3) How to solve 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. 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.
- Prior experience with data analysis is highly recommended (using SQL, R, Python)
- Familiarity with linux and command line tools
- Familiarity with git and github
- Discrete Math and Linear Algebra knowledge is highly recommended

Topics:

We'll cover the following topics:

1) Overview of the Machine Learning Process
    a) Understand Problem
    b) Map to Machine Learning formulation
    c) Understand the Data, Data Exploration, Data Stories
    d) Machine Learning Pipeline Development
        i) Setup the problem
        ii) Feature Development
        iii) Modeling
        iv) Evaluation
    e) Deployment

2) Machine Learning Methods
    a) Supervised
        i) Regression
        ii) KNN
        iii) Trees
        iv) NN
        v) SVM
        vi) Random Forests
        vii) Ensemble Methods
    b) Unsupervised
        i) Clustering
        ii) Association Rules
        iii) PCA ad related methods
    c) Semi-Supervised (Not covered in this class)
3) Feature creation/engineering
4) Model Selection
5) Evaluation
6) Applying these methods to Policy Problems

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
- Short response 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 will have a report due and oral presentation

Project:

Students will form groups (3-4 students each) and work on a project they'll propose after week 3.

Grades:

Assignments: 25%
Mid-Term (extended assignment): 15%
Weekly Class Reviews: 15%
Class Participation: 15%
Project: 30%
    Proposal: 5%
    Proposal Presentation: 5%
    Progress Report: 5%
    Final Report: 10%
    Final Presentation: 5%