Machine Learning for Population Health Management

Course Description

Population health management aims to improve individual health outcomes, improve health equity, and reduce overall cost of care for groups of people. The most successful organization will be those who are able to leverage data in order to improve decision making and guide successful health interventions. The purpose of this course is to equip students with the ability to build machine learning models whose purpose is to guide population health interventions.

The course is designed to be truly interdisciplinary, and will combine information from healthcare policy, programming, and statistical methods of learning. The course emphasizes a human-centered approach to data science, with specific emphasis placed on the interpretation of models and the measurement and mitigation of algorithmic bias.

Learning Objectives

This course covers a wide range of topics, all of which are salient for analytic professionals interested in the field of healthcare. By the end of the course, successful student will understand:

- How programs of population health management can be applied to patient populations to avoid adverse outcomes, and the role of data in driving program efficiency and effectiveness.
- How to leverage SQL to process medical claims records into data appropriate for visualization and predictive analytics.
- The role of interpretable machine learning models, and how these models should be assessed for algorithmic bias.

Prerequisites

Required

- I 304 (Programming for Informatics). Note, while the I 304 requirement can be satisfied with other courses, this course will assume a working understanding of Python.
- I 306 (Statistics for Informatics). As Statistics for Informatics covers Linear and Logistic Regression, this course will assume the students have a strong command of the theoretical underpinnings to these models.

Course Requirements and Grading

Reading Materials

As this is a multidisciplinary course, reading materials will be pulled from publicly available documents, including peer reviewed journal articles, data dictionaries, code documentation, and policy announcements. For concepts about machine learning, we will be using “An Introduction to Statistical

Learning". The book is available for free download here; the book is excellent, so I do recommend purchasing a copy if you are able.

The course will alternate between discussions of policy in healthcare, and the practical application of data processing to pursue this work.

Required Devices

You will need a laptop computer capable of running Python & Jupyter notebooks, access to the internet, and able to open PDFs. Students should bring their laptops to class.

Grading

Grading will consist of three components:

- 30% Weekly Quiz
- 35% Midterm Exam
- 35% Final Exam

Schedule

The schedule will consist of two weekly sessions. The first session of the week will focus on code, manipulation of data, and putting lecture topics into action. The second session of the week will introduce healthcare and machine learning concepts, and be a more traditional lecture format. The following outlines the intended schedule for the semester.

Week	Session 1	Session 2
1	Introduction & Overview	Introduction to Population Health Management
2	*Accessing medical claims data*	Emergency Room High Utilization
3	*Identifying emergency room use in claims*	ML Basics
4	*Introduction to SKLearn*	NYU Avoidable ED Model
5	*ML pipeline on claims data*	Tree based Algorithms
6	*Building an EDHU Random Forest*	How population choices change models
7	*Creating populations on data*	HIPAA, data privacy, claims limitations
8	Midterm	Model Interpretability
9	*Implementing SHAP*	Algorithmic bias
10	*Measuring & mitagating ML bias*	Social Determinants of health
11	*Merging SDOH data into your pipeline*	Avoidable Readmissions
12	*Building admit periods on claims data*	Avoidable Hospitalizations & CCM
13	*Building readmission models*	Advanced feature creation
14	*Feature engineering*	Survey of additional use cases
15	Review	Final

All lecture materials will be made available for review after the session for that week. In addition to class and programming assignments, most weeks you'll be asked to read 1-3 documents before the lecture

session (session 2) of the week. For reading assignment from "An introduction to statistical Learning", I will use the shorthand ISL followed by the section numbers.

Week	Reading
1	• A strategy for health care reform—toward a value-based system (pdf) • What is Population Health? (pdf) • Accountable Care Organization Realizing Equity, Access, and Community Health Model (link)
2	• Frequent users of US emergency departments: characteristics and opportunities for intervention (pdf) • ISL: 2.1, 2.2
3	• ISL: 3.1, 3.2 • The NYU algorithm "patch" (pdf)
4	• Revisiting the NYU algorithm (pdf) • ISL: 4.1-4.3
5	• ISL: 8.1, 8.2
6	• Patient Segmentation (pdf) • Window Function Concepts (link)
7	• HIPAA Basics for Providers (pdf) • Claim and Claim Line Feed Information Packet, (page 1 – 23) (pdf)
8	• Welcome to the SHAP documentation (link)
9	• Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities (pdf)
10	• Social Conditions as Fundamental Causes of Health Inequalities (pdf)
11	• Unplanned Hospital Visits (link)
12	• Primary Care practices' ability to predict future risk of expenditures and hospitalization using risk stratification and segmentation (pdf)
13	• Charlson Comorbidity Index (link)
14	• ISL 11.3, 11.4
15	