

Eric Nicholas Rodriguez

Data Scientist | Data Analyst | Machine Learning Engineer

Boston, MA | (703)-380-7619 | Rodriguez.N.Eric@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

SUMMARY

I'm a data scientist with a background in public health. As a team-player who also excels at working independently, I'm driven to optimize and improve the efficiency of health services and programming through the analysis of healthcare data to increase quality of health outcomes and cost efficiency, particularly within marginalized populations.

SKILLS

Languages: Python (BeautifulSoup, Keras, NLTK, NumPy, Pandas, Requests, Scikit-learn, SciPy, TensorFlow), R (dplyr, GGplot2, tidyr, XGboost), Git, SQL, Stata

Modeling: Linear and Logistic Regressions, Time Series Forecasting, NLP, Neural Networks

Data Visualization: Matplotlib, Seaborn, GGplot, Tableau

PROFESSIONAL EXPERIENCE

GENERAL ASSEMBLY

January 2021 – April 2021

Data Science Immersive Fellow

- 480 hours dedicated to an immersive data science bootcamp
- Applied data collection and cleaning, analysis, modeling, data visualization, and machine/deep learning techniques to solve real-world data problems

DANA-FARBER CANCER INSTITUTE

June 2019 – December 2020

Research Coordinator

- Managed 140-site clinical randomized control trial comparing effectiveness of anticoagulants in cancer patients, increased participant questionnaire follow-up rate from 70% to 82%
- Recruited patients, administered surveys, conducted literature review, monitored adverse events, managed data, analyzed data using excel and Stata, visualized data using Tableau, drafted manuscript
- Acted as a human reviewer to analyze unstructured data from imaging and text reports in electronic health records to train computational deep learning model called PRISSMM

PROJECTS

[COVID-19 Vaccination Time Series Forecasting](#)

- Utilized LSTM recurrent neural networks to forecast the total number of vaccines that will be administered over a one-week period in Massachusetts, USA
- Recursive multi-step one-week forecast achieved a lowest RMSE of 8366, significantly lower than the persistence model baseline of 12900.

[Reddit Web Scraping and Natural Language Processing](#)

- Collected posts from r/MentalHealth and r/CoronavirusUS using Pushshift API
- Used Logistic Regression with Lasso Regularization to properly classify correct posts 92% of the time, deployed a [Streamlit application](#) that predicts if a post indicates need for mental health services

EDUCATION

GENERAL ASSEMBLY

April 2021

Data Science Immersive Program

BLOOMBERG SCHOOL OF PUBLIC HEALTH

January 2020

Master of Science in Public Health in Social and Behavioral Science

Certificate in Health Finance and Management

- GPA: 3.77
- Relevant Course Work: Stata Programming, Introduction to Data Management

JOHNS HOPKINS UNIVERSITY

May 2018

Bachelor of Arts in Public Health Studies