My objective is to pursue a PhD in Biostatistics with a focus on applied machine learning and data science. I am currently studying Biomedical Engineering at Johns Hopkins University (JHU) as a master’s student, doing research under Dr. Brian Caffo. I plan to continue with a career in research after obtaining my PhD.

My motivation to pursue a PhD comes from my experience working on several substantial research projects. I’ve worked in groups at NEU, PolyU, and JHU and have also been a data analysis intern at Johns Hopkins Bloomberg School of Public Health (JHSPH). Through these experiences, I have bolstered my research skills and published several first-author papers [1,2]. I have found that I am most engaged when working on ambitious technical research projects. Although my actual data engineering work has been educational, what really sparked my interest in research was experiencing the first-hand challenges that data engineers face daily. I am interested in projects which, if successful, have the potential to impact a large number of people by improving the state of the art in a particular area.

The interest I’ve developed in Biostatistics first came about while I worked with Professors Brian Caffo and John Muschelli on volumetric analysis of MRIcloud output. By that time, I had completed enough statistics and computer science courses to be able to meaningfully contribute to the MRIcloudT1volumetrics package they were building. With a goal to create tools that can increase researcher productivity and improve package quality, I joined the development team and have improved classification models using principal component analysis and multinomial regression. This package may help find new biomarkers of aging and is available at <https://github.com/bcaffo/MRIcloudT1volumetrics>. Through this experience, I have strengthened my statistics skills and become well-versed in good research culture and practice.

Most of my real-world machine learning experience was gained during my internship at JHSPH. I was working in the Johns Hopkins Data Science Lab (DaSL) and was there introduced to Professor Ciprian M. Crainiceanu. One of our most interesting analyses has been the National Health and Nutrition Examination Survey (NHANES). I led a research project addressing a problem intrinsic to the high-dimensional nature of our data, which would complicate analyses in time series forecasting. My spectrum analysis results and model revealed important associations between daily activity and health outcomes such as BMI and mortality. Ultimately, we collaborated with other PhD students and compiled the package “rnhanesdata” containing the content, the tools, and the context needed to empower users to begin working with NHANES data quickly (accelerometry or otherwise). I submitted a first-author full paper to the American Journal of Epidemiology on this work, which is currently under review [3].

Although I am open to a variety of research, there are several professors at JHU whose projects are especially appealing to me: Brian Caffo, Ciprian M. Crainiceanu, and Jeff Leek. After working with them for a few months, I see a clear fit for my skills and interests at JHU and am confident that it is a great place for me to pursue a PhD.

Thank you for your consideration of my application.

References

[1]

[2]

[3]