# Reflections on Scholarship

My research interests are in the fields of Big Data, Machine Learning, and Data Analytics. As often noted, Data is the new oil; while mining data can be economically valuable, this resource must be handled carefully. Using machine learning and advanced data analytics, we can extract information and intelligence from the vast amount of data being generated. However, care must be taken in how we build and use our models. If these models are going to play an important role in our criminal justice system, healthcare industry, and in maintenance of public safety, it is critical that we have a full understanding of the performance characteristics of these models. In particular, my research focus is on understanding the connection between the source of data and the quality of models that they generate. Large data sets are obtained by combining data from diverse sources, collected over long time periods and in different geographical domains. What is the impact of data sources, time periods and geographical domains on the quality of the models that we build? This is the primary question that I'm trying to address in my research.

I'm working on two specific data-related research projects. The first project is on the development of land-use-regression models for air-quality predictions. The second one is on understanding the impact of air pollution exposure on the severity of COVID-19 infection. Both these projects are in collaboration with a diverse group of researchers from Clarkson, including faculty from Mechanical Engineering, Statistics, and Biology and B.S and PhD students. We meet weekly to discuss different projects that we are working on. My role in the project is to lead all aspects of data, including data access, storage, and processing. This work has been on-going for over a year. The research methodology and results from the first project on air quality predicts has been presented in an international conference (American Association for Aerosol Research) in October 2019 and in a Research and Projects Showcase conference in Clarkson in June 2020. The second project on connection between air quality and the severity of COVID-19 was presented in the same international conference (American Association for Aerosol Research) in October 2020 and in Research and Projects Showcase conference in Clarkson June 2020

In addition to the conference presentations I published a peer-reviewed paper, entitled: "Understanding public response to air quality using tweet analysis" in Big Data and Society Journal in 2019. This is the second paper that I have published since joining SUNY Potsdam. In addition, this work was also presented at the American Association for Aerosol Research Conference in October 2019. Recently (Feb 2021), we submitted a manuscript on our air-quality/COVID-19 study for publication in a peer-reviewed journal.

To support my research activities and bring in our undergraduate students in my work, I have been actively looking for research funding from state and federal agencies. In 2020, I submitted a proposal to NYSERDA for supporting data analysis for a large air quality network. That proposal did not get funded, but I will continue to look for funding opportunities that will sustain my research over a long time period.

I've also been actively involved in efforts to disseminate my research knowledge in other forms such as workshops and seminars. In August 2019, I co-organized a 5-day international workshop on Sensors and Data Analytics in Clarkson University. I was responsible for training the participants on the usage of practical data analytics tools. This involved multi-day instruction in

programming, simple data analytics, and usage of advanced machine learning algorithms. The conference was attended by faculty, researchers, and students from 5 different institutes/universities.  The conference was well received by attendees and a follow-on workshop was planned for August 2020, but had to be cancelled due to the pandemic.

Pursuing research is satisfying from a personal and professional perspective.  My active research collaborative activities are critical for keeping myself updated on the latest happenings in the field.  This allows me to bring in appropriate perspective and depth to the courses I teach.  In particular, my recent research activities have helped me in the development of the new Machine learning course that I'm currently teaching and in constantly updating the department's classical Database course with introduction of the latest tools and techniques being used in the field.  Bringing this expertise and content to the class advantages our students when they look for career options.