**Reflections on Scholarship**

My research interests are in the fields of Big Data, Machine Learning, and Data Analytics. As is 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. To obtain large data sets, we often integrate measurements made over long time periods and in different geographical domains. As we broaden the spatio-temporal origin of data, how does that impact the quality of the models that we build? This is the primary question that I’m trying to address in my research.

One of my current related research projects uses a high-resolution, rich air quality data set from a dense sensor network in Chicago. Using this data set, we are building models to relate air quality to land use parameters such as traffic, building density, etc. We are then testing to see if the accuracy of models for predicting air quality over a selected time period varies with the size of the data used for model generation. We will use this model to then predict air quality in different cities and conduct tests to determine how to improve its robustness for wide spatial-scale application. This research is important as it will allow us to clearly establish the global applicability of big data models built with local data sets.

These research activities require collaboration with scientists across fields and universities. I have fortunately been invited to be part of a large team at Clarkson University to work together on this Big data problem. This team consisting of faculty from Mechanical engineering, Statistics, and Biology along with me and several PhD students and undergraduates meets weekly to discuss all aspects of the project. In this team, my role is to lead all aspects of data, including data access, storage, and processing. This work has been on-going for 1 year and has resulted in two presentations in national and local conferences, with another presentation just accepted for an international conference (Air Sensors International Conference) in May.

Last year, I published a paper, entitled: “Understanding public response to air quality using tweet analysis” was published in Big Data and Society Journal. This is the third paper that I have published since joining SUNY Potsdam. In addition, related work was also presented at the American Association for Aerosol Research Conference in October 2019 as a prestigious platform presentation.

I’m also keen on disseminating 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 is planned for August 2020.

I presented a seminar on Data Visualization in a seminar in our department. This seminar focused on using visualization to effectively present research results to a diverse audience. Visualization is a critical element of data analytics and is an emerging research topic. The seminar was well attended with almost all our faculty members and students being present and actively participating.

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.