Reflections on Scholarship

My research area of focus is Big Data, Machine Learning, and Data Analytics. By utilizing machine learning and advanced data analytics, we hope to extract meaningful information from the vast amount of data generated. Nevertheless, caution must be taken when constructing and implementing these models, especially if they are intended to play a significant role in our daily life.

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 several data-related research projects focused on statistical analysis, natural language processing, and machine learning. In one of these projects, we use big data analysis to understand the accuracy of air-quality data coming from low cost sensor networks. The current state of the art approach in correcting data from these networks is using linear regression. With our data-based analysis we were able to show that the current calibration and correction approaches favor certain sources of air pollution over others, thus biasing any conclusions that might be derived from them. In a second area of research, I worked with an undergraduate student in SUNY Potsdam to analyze 20 years of presentations made in the American Association for Aerosol Research Conferences. In this research, we used natural language processing (NLP) tools with machine learning to automatically classify the publication by topic and then analyzed how topics evolved over the years.

As an advisor for several senior projects over six years, I was struck by the diverse range of interests and abilities among our students. Some students excelled in technical skills, while others demonstrated strong creative abilities. This experience has convinced me that there are a section of our students who are capable of independent research with proper guidance and I need to give them more freedom to explore and make their own decisions, which will help them in developing critical thinking and problem-solving skills.

In the third area of research, we are combining our statistical tools with machine learning to understand the impact of air pollution exposure on the severity of COVID-19 infection.

All of these projects are in collaboration with a diverse group of researchers from Clarkson, including faculty from Mechanical Engineering, Statistics, and Biology and includes B.S and PhD students. 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 several years. Our research has been presented in several international conferences, including the American Association for Aerosol Research, Air Sensors International Conference and locally in the Research Projects Showcase conferences in Clarkson.

In addition to conference presentations, my research activities at SUNY Potsdam has resulted in three peer-reviewed publications, one pre-print, and one under review. To support my research

activities and bring in our undergraduate students in my work, I have been actively submitting proposals for research funding from state and federal agencies.

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.