Research Statement

Joseph C. Robertson, PhD
Mato Ohitika Analytics LLC
August 2020

I am currently the Chief Data Scientist of Mato Ohitika Analytics LLC, a Native owned startup analytics company I founded during the final stages of my doctoral work: *The Impact of Data Sovereignty on American Indian Self-Determination: A Framework Proof of Concept Using Data Science*. I designed my doctoral dissertation as the cornerstone and blueprint of my company goals to create SMART solutions for communities by gathering appropriate intelligence and leveraging that information for strategic planning.

This company is one of the final components in my comprehensive plan to merge my professional and educational portfolio to serve a higher purpose moving forward. Mato Ohitika Analytics LLC was designed as a long-term research and development platform for a lifetime of designing open source artificial intelligence (AI) solutions that will help me realize a personal goal of mine: to develop an indigenous artificial intelligence for language and cultural preservation and for practical smart solutions for community economic development.

I have outlined this brief to highlight how my research interests and work contribute to a greater body of work in AI and machine learning (ML), not only in creating practical solutions that serve economically challenged communities, but serves to educate students, colleagues, and organizations that without collaboration and cooperation, AI will only serve the privileged.

I have finished the first stage of the intelligence I have gathered by visiting communities and promoting data sovereignty and data governance to develop indigenous protocols for AI, smart solutions, and education outreach. I am deeply committed to contributing my expertise in forwarding AI and data science in the academic sector to collaboratively create partnerships that are meaningful and equitable.

I have a great interest in contributing to the strategic goals of creating a world class company in designing data science, AI, and research and development in the community context. I look forward to serving our communities through this approach for many years to come.

Summary of Previous and Ongoing AI / ML Research and Development

Data science has become the foundation of the continuing evolution between how data is actualized and then applied to actual phenomena. The American Statistical Association (ASA) maintains there are three professional communities that are emerging and foundational to data science:

- Database Management enables transformation, conglomeration, and organization of data resources;
- Statistics and Machine Learning convert data into knowledge; and
- Distributed and Parallel Systems provide the computational infrastructure to carry out data analysis.

As data science has become increasingly diversified into many technology fields, I felt it was important to contribute to this body of work through the lens of computational statistics. I identified early on in my

master's thesis work in statistics that artificial intelligence and machine learning would be one of the most critical aspects governing data science.

My current research interests have been a culmination of carefully designed pedagogical foundations in higher education, citizen science, company entrepreneurship, business acumen, and community development. The primers of my doctoral dissertation were designed to function as a living document to continually assess and reassess the state of how data science, machine learning, and AI affect communities once smart solutions are created.

The premise of my doctoral work first focused on designing the *Data Sovereignty Initiative* (DSI), which is a collection of ideas that were designed to create SMART solutions for tribal communities. This concept was to develop a horizontal governance framework to create a strategic act of sovereignty using data science. The core concept of this idea was to present data sovereignty as a way for tribal communities to take ownership of data in order to affect policy and strategic decisions that are data driven in nature.

My doctoral case studies were developed around statistical theories of spatial statistics, exploratory data analysis, and machine learning. And although these case studies are first, scientific in nature, the data sovereignty framework was designed around these concepts to leverage nation building, cultural capital, and citizen science for economic development and planning. Although this framework was written to better understand American Indian community's needs, the key indicators of the design can be easily modified to work in any community or organization. It is important to understand that this framework is an inclusive exercise to promote nation building and collaboration.

As my portfolio has grown during my graduate studies and beyond, I have gathered valuable intelligence that have formed a diversity of research interests such as serving as a statistical consultant, data scientist, educational community partner, and grass roots advocate.

I want briefly outline the three topics of my current research interests that are not just exclusive to American Indian tribal communities, but to a growing body of data sovereignty issues, data ethics principles, and applications of these concepts:

Geo-Artificial Intelligence (GEOAI): The second case study of my doctoral dissertation was to present a comprehensive analysis of the construction of a support vector machine (SVM) for image classification of tribal infrastructure using satellite and drone imagery. During my studies, I conducted a statistical analysis of my home tribal community the Sisseton Wahpeton Oyate of the Lake Traverse Reservation's first tribal census in over a decade.

After assessing the methods and data collection processes, it became apparent that sending five census takers to cover the entire reservation boundaries of nearly one million acres was not only unfeasible, but impractical. This was the basis for designing a SVM for future tribal censuses. The results proved that a machine learning algorithm could not only map a master address file (MAF) for future census, but could be designed for any community to take ownership of their infrastructure using image classification.

My current research and development have stemmed from the rapid advancement of provisioning a data science virtual machines for GEOAI using python's deep learning libraries where the technological shift

from SVM applications (which are pixel based) has shifted to object-oriented classification using GEOAI. This work is ongoing.

Natural Language Processing for Cultural and Language Preservation: The loss of Native speakers in our Dakota language has always been a great concern to me. The acculturation process for many Native people has resulted in the denial of access to our language from an early age. With the advent of the NLTK libraries in python, it has been possible to create ontologies and knowledge graphs that teach a recursive neural network (RNN) about the Dakota language.

It is one of my goals to unify a language framework that provides open source code in python to collect, analyze and develop an AI for preserving Native languages using deep learning and eventually creating an education tool for students who can contribute to a project similar to the Amazon Alexa, but instead this assistant speaks in our Native languages.

Internet of Things (IoT) for Cultural Preservation: In September of 2018, I continued my research by invitation to the Supporting Tribal Data Governance for Community Resilience: A Southwest Indigenous Climate Summit which was sponsored by the US Indigenous Data Sovereignty Network (USIDSN) at the University of Arizona to discuss how data could be used in use cases where environmental issues were affecting the health of tribal citizens.

In addition, in November of 2018 I was invited by the water protectors in the Brave Heart Society on the Yankton Sioux Reservation to discuss issues of the Dakota Access Pipeline protests and how the priorities have shifted to mapping the Missouri River for water collection and analysis, but also for maintaining these maps for preserving sites for collection of herbs and plants for medicinal and ceremonial purposes.

This intelligence was not only informative, but powerful as it presented a need for designing IoT devices for this data collection. If the design of IoT in real time can measure pollution or water quality in any of these use cases, then we are obligated to use our technology to assist these communities with solutions that can make a real impact. This research is also ongoing.

In addition, I was given the opportunity to apply these real-world outcomes when I testified as to risks of the construction of the Keystone XL project would bring hundreds of workers from the man camps into tribal rural areas of South Dakota that have only handfuls of law enforcement officers to handle big areas.

I completed a preliminary risk assessment to study harm reduction related to Murdered and Missing Indigenous Women (MMIW) on behalf of the Brave Heart Society and the Yankton Sioux Tribe to study police capacity and the impact of the Keystone XL pipeline man camps. This important work continues even to this day.

Future Research

The future of AI revolves around the development of open source methods and techniques that empower organizations to use AI responsibly.

The premise of my doctoral work in the previous section is a data governance framework designed specifically to address how AI / ML is to work specifically in real world situations. Tribal organizations have fears that data collected will not be under their direct control or used against them without their

consent. In the past two years, I have been gathering intelligence in American Indian communities to assess their operational data capacity in data science. The results have been quite extraordinary in that although many tribal communities do not have an advanced digital footprint; many communities embrace these technological advances provided they have a stake in how this data is collected, shared, and used for the communities' benefit.

There were many ideas presented to me that would make a direct impact if data science could be understood in the context of designing a platform that is *not* in search of a problem; rather the intelligence gathered informs solutions of a real world problem expressed by the community it is designed to serve.

With this in mind, the future of my research interests is divided into three areas of research and development, not including teaching goals that are outlined in the accompanying teaching statement:

1. The research into protocols of artificial intelligence that include well-designed principles of the ethics of AI, privacy implications, data use policies, and data governance strategies.
2. A strategic assessment and classification of machine learning techniques that are viable in solving specific types of problems such as natural language processing, image classification, and data mining. In addition, this can also include how these techniques apply to computational statistical inference modeling techniques that are the underlying principles governing quadratic programming, vector gradient descent, and optimization theory.
3. Developing fundamental techniques in the creation, testing and deployment of AI and ML that have real world applications.

The DSI has tremendous potential for growth and could benefit from a more robust examination of these ideas to provide students and communities with real world context when constructing data science projects to gain a fundamental understanding of AI and machine learning techniques in such a way that interdisciplinary concepts meets the technology in applied situations.

In addition, it is extremely important that students and community stakeholders gain a fundamental understanding of the underlying mathematical and statistical concepts that are interconnected with how data science intersects our everyday lives so communities are able to take ownership of these solutions they design from the ground up. These are the challenges we face in honoring our culture, rebuilding our communities, and securing an intelligent workforce to prepare for the real-world challenges we will face together moving forward.

Applications of Future Research and Design

Data science is an interdisciplinary construct of many systems working to solve real world problems. The techniques I outlined in my doctoral dissertation focused on creating SMART solutions using the *Data Sovereignty Initiative* framework as a flexible way to leverage data in the context of data governance, data ownership, and privacy. A SMART solution contains three interdisciplinary constructs: higher education, citizen science and data science.

Combined with the intelligence gained from indicators that honor an organization's governance and community and culture is how this entire construct informs a new praxis in education and community development. The strategic interlocking of these ideas is how we move from data science in theory to real practical solutions.

The DSI aligns itself with the directive to ***develop and inspire healthy global citizens and create a smart, sustainable world.***

Applying educational pedagogy to ongoing research and development can be related directly to the information I have gathered and increase breadth of knowledge such as:

1. **Modifying Internet of Things to Internet of Robotic Things**: Internet of things (IoT) devices that can be designed and deployed to solve a specific problem and using python to create code to connect, gather and upload data collected to cloud services such as AWS Sagemaker or Azure IoT for real time analytics is essential. This concept once perfected could be used to integrate IoT into the Internet of robotic things (IORT), a natural progression from the base concepts to the more advanced.
2. **Smart and Sustainable Cities**: My previous work in support vector machines for image classification has been an extremely important project for me personally. The second case study in my dissertation was how to use a SVM to classify housing infrastructure and create polygons and point patterns in order to create a master address file for tribal governments to conduct their own census. Intelligent design comes from an experiential approach. This can be further enhanced by provisioning a data science virtual machine to create a GEOAI using python's deep learning libraries in the ArcGIS Enterprise system.
3. **STEM Education**: I have spent over a decade advocating for underrepresented minorities to promote STEM as a teacher of record, school leadership team member, and mathematic coach. In addition, the work I designed in graduate statistics to non-major graduate students was to understand the multiple intelligences students have in studying and mastering statistical principles that serve their field of study. I would like to further this concept by teaching statistics and probability within the context of a data science program aimed at producing global citizens.
4. **AI and Data Science**: I have outlined how the expertise I have in statistical principles are fundamental in exploring the foundations and history of data science. I envision teaching, mentoring, and advocating for students to become global citizens in the role data science plays theory, coding, interpretation, and personal growth.

In summary, this brief outlines my commitment to many interdisciplinary ideas that are fundamental to critical thought, design, and pedagogy in higher education, and particularly data science. I have been on a number of grant writing committees and have consulted on revisions of NSF grants that require sound statistical methodologies given the competitive nature of data driven funded research opportunities. In addition, part of my comprehensive planning in AI research was to design open source systems that minimize the cost of for underrepresented organizations. This creates a balance with funding opportunities and community service.

I have great interest in collaborating with interdisciplinary colleagues who can lend their expertise to my growing portfolio in database management, machine learning, and AI solutions. My expertise in the R programming language, as well as the research I am currently conducting in developing *Distributed GIS Systems* using deep learning is the primary direction I am current moving in space. Research should inform praxis, and in turn praxis informs educational decisions that provide communities with the best tools for them to succeed in the growing world of technology and beyond. I look forward to working in this exciting interdisciplinary field of experts to make a more sustainable world.