MICHAEL NSIAH-NIMO, M.Sc.

SENIOR RESEARCH STATISTICIAN | DATA SCIENTIST
DEVOPS | MLOPS | CLOUD & AI SOLUTIONS

My Website

CONTACT

Lincoln, Nebraska

michael20129648@gmail.com

Linkedin

EDUCATION

University of Nebraska, Lincoln
PhD Statistics
2025 - 2030

University of Texas at El Paso
MSc Statistics
2015 - 2017

Kwame Nkrumah University of Science and Technology
BSc Actuarial Science
2010 - 2014

TOOLS AND TECHNOLOGIES

- **Languages:** Python, R, SAS Data and Proc Steps, SQL, Bash, Java, Go
- **ML/Stats:** Scikit-learn, TensorFlow, PyTorch, glmnet, XGBoost, SHAP, Bayesian models
- **Cloud/MLOps:** GCP (BigQuery, Cloud Functions), AWS SageMaker, MLflow, Docker, Kubernetes, GitHub Actions, Helm
- **NLP/CV:** Hugging Face Transformers, LangChain, OpenCV
- **Dashboards & Visualization:** R Shiny, Dash, Power BI,Tableau, ggplot2, Matplotlib, Seaborn
- **Databases:** REDCap, All of Us, PHIX, EHR systems
- **Compliance:** HIPAA, IRBManager, CRF design, Data governance, CDISC Data Standards

CERTIFICATIONS

- **Decision Intelligence : LinkedIn**
- **Probability Theory, Generalized Linear Models and Inferential Statistics (Duke University)** | May 2018 – September 2019
- **Supervised Machine Learning**: Regression and Classification | October 2022
- **Data Analysis with SAS** (Oreilly)

Behavioral Skills

- Adept at using simple language without watering down the key ideas of the research and focus on the purpose and results of the study to capture audience comprehension and relatability
- Proven track record of seamless collaboration across interdisciplinary teams, contributing to project success.
- Adept in identifying challenges and implementing innovative solutions ensuring the integrity of the analyses.
- Experience in providing guidance and mentorship to junior team members fostering a collaborative work environment.
- Flexibility with changing priorities, strong attention to detail, ability to work well under pressure and take on unfamiliar tasks

SUMMARY

Research Statistician, Full-Stack Data Science and Analytics Professional with **7+ years of experience** in **research, advanced analytics, and machine learning.** Skilled in **building predictive models,** uncovering **patterns and trends in high-dimensional datasets**, and **developing scalable AI/ML pipelines** on **cloud platforms (AWS, Azure, GCP)**. Proficient in **statistical modeling, supervised** and **unsupervised ML,** and **data engineering** for **large, complex datasets**. Experienced in **data integration and standardization from diverse sources** to ensure **high data quality, interoperability, and reproducible analytics.** Contributed to **12+ large-scale federal data projects**, delivering **actionable insights**, **workflow automation, and decision-support tools** that drive **business efficiency** and empower **cross-functional teams.**

WORK EXPERIENCE

Senior Research Statistician | Data Science

Integrative Analytic Unit, Border Biomedical Research Center(BBRC) JULY 2020 - PRESENT
El Paso, TX

- **Collaborated with cross-functional teams** of **data scientists and analysts, engineers**, and **statisticians on enterprise-wide analytics projects,** contributing to **over $1M in project funding** through **data-driven strategies.**
- **Mined and harmonized data from disparate sources, mapping variables to standardized taxonomies to ensure interoperability, high data quality, and readiness for predictive modeling.**
- **Created interactive dashboards** and **real-time visualizations using Power BI, Tableau, and R Shiny (Posit), enabling data-driven decision-making for executive and technical stakeholders.**
- **Developed non-parametric predictive models** leveraging **high-dimensional datasets**, utilizing **robust regression, penalized models (LASSO, MCP, SCAD)**, and **deep neural networks (DNNs)** to uncover key **drivers of business outcomes**.
- **Applied unsupervised ML techniques** (e.g., **clustering, PCA**) to reveal hidden patterns and create **actionable segmentation strategies**.
- **Built and optimized supervised ML pipelines** to identify **variable interactions** and predict critical outcomes such as **risk, performance, and segmentation.**
- **Engineered cloud-native ML workflows** on **AWS**, automating **model training, deployment, and monitoring** using **SageMaker, Lambda, Step Functions, and S3** for **scalable** and **secure data processing.**
- **Deployed reproducible ML environments** with **Docker and Kubernetes (GKE),** integrated **MLflow for experiment tracking**, and implemented **CI/CD pipelines** with **GitHub Actions** to streamline **MLOps.**
- **Developed NLP pipelines** with **transformer-based LLMs** (e.g., **GPT-3.5, LangChain**) to **extract, summarize, and analyze unstructured text data**, producing **business insights** from documents and communications.
- **Designed and executed data and statistical analysis plans** for **experimental and longitudinal studies**, incorporating **confounding adjustments, stratification**, and **power calculations** using **R and SAS.**
- **Conducted quality control and validation** of **analytical outputs** (tables, visualizations, reports), created **data validation specifications (DVS),** and ensured **compliance with data standards** for **regulatory and business reporting.**

Data Analyst

College of Science Research Enterprise, University of Texas at El Paso, El Paso, TX JULY 2018 - JULY 2020

- Performed **data extraction, cleaning, and transformation of complex**, multi-source datasets including structured, survey-based, and transactional records to support exploratory and advanced analytics.
- Conducted **descriptive, bivariate, and multivariate analyses** using SAS and R, applying **GLMs, logistic regression, and survival analysis** to identify patterns and key drivers of outcomes across diverse demographic and socioeconomic groups.
- Led **stratified analyses** on large-scale datasets to uncover disparities and trends, generating actionable insights for strategic planning and performance improvement initiatives.
- Contributed to the **design and evaluation of experimental studies, hypothesis generation and testing**, assessing pre/post intervention effects using **mixed-effects modeling and time-series analysis.**
- Conducted **sample size estimation, randomization, and statistical validations** to ensure methodological rigor and reliable results for high-stakes projects.
- Delivered **publication-ready visualizations, summary tables, and dashboards,** collaborating on technical reports, white papers, and executive presentations.
- Mentored junior analysts and team members, providing training on **SAS macros, Python and R functions for data mining, and data visualization and BI** best practices to enhance team analytics capabilities.

MICHAEL NSIAH-NIMO, M.Sc.

SENIOR RESEARCH STATISTICIAN | DATA SCIENTIST
DEVOPS | MLOPS | CLOUD & AI SOLUTIONS

PUBLICATIONS

- *[Publication] Exploring Socio-Behavioral Correlates of Metabolic and Inflammatory Risk in Hispanics Living Along the U.S./Mexico Border: A Pilot Study Concomitantly Collecting Survey Data, Blood and Hair Samples, and Physical Measures, by Gabriel A. Frietze, Cai Xu, Bibiana Mancera, Elisa Robles, Escajeda, Alyssa A. Martinez,Michelle Gil, Diana P. Flores, Khodeza Begum, Panfeng Liang, Abhijit Mandal,* ***Michael Nsiah-Nimo****, Nilotpal Sanyal, Ming-Ying Leung, Michael J. Kenney and Robert A. Kirken*

In Review

- *Biological Age Index of Chronic Inflammation as a Predictor of Cardiometabolic Phenotypes: Differential Associations with General, Central Obesity and Diabetes in Mexican Origin Hispanics, Robles, Elisa, Leung, Ming-Ying,* ***Nsiah-Nimo, Michael,*** *Sanyal, Nilotpal, Xu, Cai, Frietze, Gabriel, Begum, Khodeza , Flores, Diana P , Mandal, Abhijit, Liang, Panfeng*

- *Associations Between Social Support, Perceived Stress, and Metabolic Health Outcomes: A Cross-Sectional Analysis of General Obesity, Central Obesity, and Diabetes in the Hispanic Health Hub. Robles, Elisa, Leung, Ming-Ying,* ***Nsiah-Nimo, Michael,*** *Sanyal, Nilotpal, Xu, Cai, Frietze, Gabriel, Begum, Khodeza , Flores, Diana P , Mandal, Abhijit Liang, Panfeng*

PROFESSIONAL CONFERENCES

- Microsoft AI Research Forum, 2024
- University of Hawaii Bioinformatics Data Science Workshops, 2023
- NIH NIGMS Cloud Based Biomedical Research, 2023
- Better Data for More Equitable Research : Research America Alliance, 2023
- RCMI Seminars at Meharry Medical College, 2023
- College of Science Stakeholders' on Health Disparities and Cancer Research, 2016 ,2017, 2018, 2019, 2020, 2021, 2022, 2023
- JSM Conference 2017, Diversity Scholar
- Cardwell Foundation Seminar, 2016

Lecturer / Research Associate

AUGUST 2015 - JULY 2017

Alcohol Research Group / Mathematical Sciences Department, UTEP

- **Analyzed treatment efficacy in emergency department settings** by **employing linear mixed models and interaction plots, uncovering significant patterns of illicit drug use, such as cocaine and cannabis, across** different age groups and genders, to inform tailored intervention strategies.
- **Applied Hasse Diagrams and advanced statistical techniques** to conduct hypothesis testing, providing actionable insights on substance use behaviors, supporting clinical teams in optimizing treatment sessions for Hispanic health initiatives.
- Collaborated on statistical research projects, enhancing methodologies for data interpretation and **contributing to manuscripts reviewed for publication in peer-reviewed journals**.
- **Applied Design of Experiments (DOE) techniques to clinical data analysis,** increasing experimental accuracy and reliability by 20%, resulting in more consistent and valid research outcomes.
- **Instructed over 150 students in advanced statistical methods**, achieving a 90% pass rate **while mentoring students on real-world statistical applications.**

Independent Projects & Grant-Funded Collaborations

Data Science and Machine Learning Projects

Compressing Feature Space For Classification Using PCA

- Applied Principal Component Analysis (PCA) ***to reduce the dimensionality of 50 cytokines to capture variation in estimated biological ages among Hispanic Origin Individuals leading to an increase in model performance***

Enhancing Targeting Accuracy Using ML

- Built a model that would accurately predict the customers that would sign up for a delivery club. ***This allowed for a much more targeted approach when running the next iteration of the campaign***

Predicting Customer Loyalty Using ML

- Built a predictive model using Random Forest in Python to estimate customer loyalty scores for a company's data agency ***leading to a 30% increase in the ability to contact customers with promotional material***

"You Are What You Eat" Customer Segmentation

- Used k-means clustering on grocery transaction data to split out customers into distinct "shopper types" ***to accurately target customers with relevant content & promotions***

Fruit Classification Using A Convolutional Neural Network

- Built & optimized a Convolutional Neural Network to classify images of fruits, with ***the goal of helping a grocery retailer enhance & scale their sorting & delivery processes.***

Quantifying Sales Uplift With Causal Impact Analysis

- Analyzed customer retail practices to quantify the increase in sales attributed to customers joining the company's delivery club compared to what they would have spent without the club***. This analysis revealed a significant 41.1% uplift in sales among club members, indicating the club's positive impact on customer spending***

NIH & Health Research Contributions

NIH U54 & RCMI Research Projects

- Contributed advanced statistical modeling and multi-source data harmonization to NIH funded studies on inflammatory aging, metabolic disease, and health disparities. Projects included:
  - NIH Project Nos. 5U54MD007592-27 / -29, 1C06OD032074-01, 3U54CA272167-02S1
  - NIH U54 Clinical, Socio-behavioral, and COVID-19 Supplements
  - RCMI and NIMHD-funded research on Hispanic chronic disease risk and inflammation.

NIH All of Us Research Program

- Conducted comparative analyses across racial/ethnic groups using SDOH, biomarker, and survey data to inform hypotheses on healthspan, resilience, and inflammatory risk factors in underserved populations.

Paso del Norte Health Information Exchange (PHIX)

- Mined and standardized EHR data using ICD-10, LOINC, SNOMED CT, and OMOP CDM vocabularies to support downstream modeling of immunological aging, hepatocellular carcinoma and cardiometabolic outcomes.

Health Disparities Research Consulting

- Collaborated remotely with faculty and clinical teams to design studies, draft Statistical Analysis Plans (SAPs), and model immune and physiological signatures for NIH submissions and publications.