MICHAEL NSIAH-NIMO, M.Sc.

DATA PROFESSIONAL | RESEARCH AND BUSINESS ANALYTICS

My Website

CONTACT

Lincoln, Nebraska

michael20129648@gmail.com

Linkedin

EDUCATION

MSc Statistics

Master's Thesis : Sample Size Estimation for Linear Mixed Models with Dependent End Points

University of Texas at El Paso

2015 - 2017

BSc Actuarial Science

Kwame Nkrumah University of Science and Technology

2010 - 2014

EXPERT LEVEL SKILLS

- **Programming & Development:** Python (7+ years), R (6+ years), SAS (5+ years), Jupyter Notebook, Visual Studio Code, Spyder, Git, Microsoft Office Suite (Word, Excel, PowerPoint, Access, Outlook)
- **Data Cleaning & Manipulation:** Python (Pandas, NumPy), R (Tidyverse, dplyr), SQL, Data Wrangling Techniques, Feature Engineering & Dimensionality Reduction (PCA, t-SNE)
- **Statistical & Machine Learning Tools:** Python (statsmodels, Scikit-Learn, Pytorch), R (glm, lmer), etc.
- **Databases & Data Storage:** SQL Server, Google BigQuery, Amazon S3, Snowflake, REDCap, QuestionPro, Hadoop Ecosystem
- **Data Visualization:** Tableau, Power BI, Python (Seaborn, Matplotlib), R (ggplot2, Shiny)
- **Operating Systems:** UNIX, WINDOWS
- **Study Design & Optimization:** Observational & Experimental Designs, Clinical Trials, Regression Techniques, Classification Methods, A/B Testing, Randomized Controlled Trials, Hypothesis Testing
- **Robustness & Validation:** Advanced Sampling Schemes, Sample Size Estimation, Power Analysis, Cross-Validation, Residual Analysis, Accuracy Metrics, Bootstrap Methods
- **Multivariate & Causal Analysis:** General Linear Models (GLM), Bayesian LASSO Models, Multi-Level & Survival Analysis, Longitudinal Data, Time Series, Propensity Score Matching, Instrumental Variable Analysis
- **Algorithmic Modeling & Machine Learning:** Gradient Boosting, Neural Networks, Random Forest, SVM, Ensemble Methods, Decision Trees, K-Nearest Neighbors, Clustering (K-Means, Hierarchical)

DATA MANAGEMENT, COMPLAINCE AND REPORTING

- **Data Management:** Missing Data Analysis, Outlier Detection, Feature Scaling, Normalization, Encoding, Clean Dataset Creation, Summary Tables, Data Listings
- **Compliance & Standards:** Regulatory Adherence (EMA, ICH, CDISC, IRB, HIPAA, PHI), NIH/NIMHD Guidelines, Statistical Analysis Plans (SAPs), Regulatory Submissions
- **Reporting & Interpretation:** Model Interpretation, Hypothesis Testing, Report Writing, Dynamic Dashboards, Publication-Ready Manuscripts, Conference Presentations

PROFILE

Over the past 7 years, I have developed expertise in designing and adapting statistical, machine learning/AI, deep learning, prevalence informatics, and data visualization models focused on business, healthcare, and marketing outcomes. **I've consulted on over 100 projects across healthcare, industry, and consumer services, simplifying complex challenges to drive organizational success and profitability.**

WORK EXPERIENCE

Staff Research Scientist | Research and Business Analytics

University of Texas at El Paso, Integrative Analytic Unit — JULY 2020 - PRESENT

- Design and execute **advanced statistical analyses for observational and experimental studies, including sampling strategies, recruitment methodologies, and sample size calculations, ensuring precise and reliable results that enhanced study validity and credibility**
- Lead a team of 7 statisticians and data scientists to develop a **deep learning model to create risk indices using storyboarding and data visualization tools to predict and track business risk levels, market and consumer trends**, and financial disparities in relationship to the Hispanic population affecting industrial and retail production **(MAD: 8 years, R =0.92, p value <0.0001)**
- Direct and manage **comprehensive exploratory data analyses, hypothesis testing and statistical modeling for federal grant projects, informing model building** and significantly advancing research in socioeconomic disparities and finical metrics of insurance and billing costs related to financial inequity across the US-Mexico Borderplex , **leading to the successful acquisition of over $2 million in research funding**
- **Manage and lead** over 10 financial/business analysts, data engineers, and professional faculty to design and build business risk assessment and stratification models utilizing neural networks and tree models , contributing to a **25% reduction in operational risks and a 20% in crease in project funding**, enhancing financial outcomes for the research enterprise
- **Develop reliable and reusable Python scripts** across **diverse business and large upscale industrial projects** for numeric and text preprocessing, assessing data drift distribution across time points, driving transformative changes **including a 40% increasing data processing efficiency and a 30% reduction in data storage costs**
- **Develop and integrate Bayesian analysis with frequentist hypothesis testing, model evaluation** in measuring economic factors and relevant population features in randomized studies for federal grant applications **boosting a remarkable 30% reduction in the related costs**
- Conducted A/B tests to identify relevant metrics and timelines to boost and extend strategic financial, and business initiatives, resulting in a **35% increase in extramural grant funding, a 25% boost in stakeholder advertisement reach, and a 20% rise in philanthropic support.**
- **Analyze and mine diverse datasets and customer surveys** to advance records, marketing data, and outreach economic indicators from a relational database **using advanced SQL and python pandas queries to generate actionable insights; advancing research and stakeholders objectives effectively.**

Data Scientist | Statistician | Analytics

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

- **Conducted time series analysis and forecasting on expenses, revenue, and grant funding data**, resulting in improved financial planning and resource allocation strategies
- **Formulated and applied mathematical models and other optimizing methods to develop and interpret information** that assists management with decision making, policy formulation, and other functions associated with the College of Science research enterprise.
- Worked on diverse datasets, leveraging both **parametric and non-parametric statistical methodologies** for federal and philanthropic foundation grants and increasing external funding
- Developed **comprehensive metrics to serve as financial determinants, analyzing key issues such as revenue, expenditure, and extramural funding performance, market trends, and economic disparities in the Borderplex region**
- Used **SQL, Tableau, and MapBox** to **extract, model, and visualize demographic data f**rom El Paso County, assessing socioeconomic divides within zip codes and their related financial implications **boosting targeted grant applications for economic disparity research by 50%.**

Statistics and Data Science Intern

University of Texas at El Paso, Center for Institutional, Research and Planning — SUMMER 2018

- Collaborated with a team of researchers on multiple **project planning, design, and analysis** to address institutional and educational policy issues at UTEP
- Contributed expertise in **mining higher educational research databases** which generated reports focused on **institutional research end products**
- Facilitated academic publications presentations utilizing Big Data for institutional research and university stakeholders

MICHAEL NSIAH-NIMO, M.Sc.

DATA PROFESSIONAL | RESEARCH AND BUSINESS ANALYTICS

Behavioral Skills

- Ability to convey complex machine learning and statistical concepts to non-technical stakeholders, fostering understanding and collaboration.
- 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

COURSES & CERTIFICATIONS

Standford Online

- Advanced Learning Algorithms
- Supervised Machine Learning - Regression and Classification
- Unsupervised Machine Learning
- Reinforcement Learning

Duke University Online

- Inferential Statistics
- Probability and Data with Python and R
- Generalized Linear Models (Linear regression and Modeling)

IBM Devops

- Continuous Integration and Continuous Delivery (CI/CD) DevOps · Agile Methodologies · Cloud-Native Architecture · Unit Testing · Infrastructure as code (IaC) · Automated Continuous Deployment · Microservices

DSI Data Science Professional Certification

Actionable Learnings:

- Learned how to extract & manipulate data using SQL. Application, Python for statistical concepts such as hypothesis tests for measuring the effect of AB Tests
- Applied data preparation steps for ML including missing values, categorical variable encoding, outliers, feature scaling, feature selection & model validation
- Utilized Machine Learning and Deep Learning algorithms for regression, classification, clustering, association rule learning, and causal impact analysis for measuring the impact of an event over time
- Deployed ML pipelines onto a live website using Streamlit and adapted Github for version control and collaboration
- Utilized Tableau to create powerful Data Visualizations and turn business problems into Data Science solutions

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 Statistician

University of Texas at El Paso — AUGUST 2015 - JULY 2018

- Collaborated with Dr. Amy Wagler on multiple statistical research projects, **leading to advancements in statistical methodologies that improved data interpretation and contributed to manuscripts reviewed for publications.**
- **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 mathematics, **achieving a 90% pass rate while mentoring students on end-of-semester projects,** enhancing their understanding and application of statistical techniques in real-world scenarios.
- **Assisted in developing and publishing a novel statistical method** for calculating sample sizes for multiple hypotheses in linear mixed model settings. This method produced non-conservative bounds compared to the Bonferroni method, increasing research efficiency
- **Created and managed advanced statistical codes using R , SAS ad Python software** for data cleaning, validation, analysis, and report generation, streamlining data processes and **reducing analysis time by 40%**

BUSINESS PROJECTS

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 (Accuracy : KNN = 0.936, RF = 0.935, Logistic Regression = 0.866)***

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 (4-Fold CV R-squared: Random Forest = 0.925, Decision Tree = 0.871)***

"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***

Understanding Alcohol Product Relationships Using Association Rule Learning

- Utilized Association Rule Learning to analyze transactional relationships and dependencies between products in the alcohol section of a grocery store, ***resulting in a 20% increase in cross-selling effectiveness and a 15% boost in overall revenue***.

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.***

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***
- Used Principal Component Analysis (PCA) to compress 100 unlabelled, sparse features ***into a more manageable number for classifying buyers of Ed Sheeran's latest album.***

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***

Development of an Inflammatory Aging Clock (iAge) for Individuals of Hispanic Origin

- Developed a biological age metric engineered from sociobehavioral surveys and molecular data utilizing **artificial neural networks and auto-encoded features modeling and tracking health and insurance billing and financial outcomes** in health disparities

Professional Contributions to Financial and Healthcare

Federal and Non - Federal Grants Projects

NIH Project No : 5U54MD007592 -27, SUB IDS : 8320, 8321, 8317
NIH Project No : 5U54MD007592 -29, SUB IDS : 8320, 8321, 8317
NIH Project No: 1C06OD032074-01 & 3U54CA272167-02S1

- National Institute of Health (NIH) U54's for the College's Research Core :
- NIH All of Us Data for Preliminary Data for Health Disparities Research
- NIH U54 Clinical Research
- NIH U54 Socio behavioral Research
- NIH Covid 19 U54 Supplements
- NIH C06 Grants for the Borderplex Biomedical Research Core
- NIMHD Grants for Health Disparities, RCMI Grants for Hispanic Health Disparities
- Paso Del Norte Health Research Grants