MICHAEL NSIAH-NIMO, M.Sc.

DATA SCIENCE | ANALYTICS

CONTACT

GitHub

michael20129648@gmail.com

Linkedin

Project Portfolio

QUALIFICATION HIGHLIGHTS

- Possessing a strong foundation in theoretical probability theory and applied statistics, I bring extensive expertise in cutting-edge machine learning techniques, including AI, deep learning, and regression methods
- Proficiency in study design and Statistical Analysis Plan creation, protocol reviewing, grant writing, and report writing
- 6 years of hands-on experience in Python and R, and 4 years of experience in SAS, and robust statistical analysis processes
- Proficient in critical thinking and effective team collaborations
- Avid skills in creativity, problem solving, active listening and communication

EDUCATION

MSc Statistics

University of Texas at El Paso

2015 - 2017

BSc Actuarial Science

Kwame Nkrumah University of Science and Technology

2010 - 2014

PROFILE

Drawing upon more than six years of extensive experience in statistical analysis and machine learning, I am deeply committed to addressing complex health challenges through the strategic application of data-driven insights from AI and Machine Learning.

.

SKILLS & TOOLS

- **Programming**: Python (Base, Pandas, Numpy, Matplotlib, Statsmodels Scikit-Learn, Keras, TensoFlow), SQL, R, SAS
- **Probability and Statistics**: Experimental Design, Multivariate Analysis, Study Design and Sample Size Estimation, Probabilty Theory, Statistical Modeling, Parametric and Non - Parametric Statistics , Multilevel Modeling, Survival Analysis, Longitudinal Data Analysis, Hypothesis Generation and Testing, Randomization and Clinical Trials, Statistical Analysis Plan (SAP) Creation , Data Mining and Visualization, Report Generation and Interpretation
- **Artificial Intelligence and Machine Learning**: Linear Regression, Logistic Regression, Decision Trees, Random Forest, KNN, k-means, PCA, Association Rule Learning, Causal Impact Analysis, Artificial and Convolutional Neural Networks, Deep Learning, Natural Language Processing, Generative AI and Large Language Models
- **Professional Skills**: Statistics, Github, Data Visualization, MS Office, Tableau, Jupyter Notebook, AWS, Google Cloud Platform

WORK EXPERIENCE

Statistical Analyst and Data Scientist

University of Texas at El Paso,
College of Science , Research Enterprise

JULY 2020 - PRESENT

- Lead a team of statisticians and data scientists in developing a Python-based Deep Learning model to predict immunological age and inflammation levels in Hispanic Origin Individuals (HOI) for a national grant
- Foster collaborations with clinicians , bioinformaticians and professional faculty members to design and study robust disease risk assessment models contributing to enhanced healthcare outcomes
- Performed comprehensive data cleaning, variable selection, and feature engineering initiatives across projects, creating a remarkable 30% reduction in costs associated with measuring social factors and molecular features in randomized studies for federal grant applications
- Implement Big Data methodologies for the UTEP Center for Integrative and Translational Research (CITR) driving transformative initiatives for healthcare equity in the Ciudad Juárez and United States Borderplex

Data Research Analyst

College of Science, University of Texas at El Paso

JULY 2018 - JULY 2020

- Worked with a dynamic team of scholars and analysts to explore and diverse datasets, leveraging both parametric and non-parametric statistical methodologies. for NIH and philanthropic foundations' grants.
- Developed comprehensive social determinants of health trends, in prevalent issues like obesity, cancer, and metabolic syndrome in the Borderplex region.
- Used SQL, Tableau, and MapBox, to extract, model and visualize demographic data from El Paso County, boosting targeted grant applications for health disparities research in the College of Science by 50%.
- Led data science initiatives to assess and implement appropriate statistical models, including linear mixed models and logistic regression to advance clinical and socio-behavioral research.
- Presented and participated in national data science and statistical conferences contributing to bioinformatics, health equity and statistical knowledge and health outcomes.

MICHAEL NSIAH-NIMO, M.Sc.

DATA SCIENCE | ANALYTICS

COURSES & CERTIFICATIONS

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

Biostatistics in Public Health Specialization
(Coursera - Johns Hopkins University)
Actionable Learnings:

- Adapted and introduced randomly selected samples in reducing bias and acknowledging the potential for non-random samples to introduce systematic biases
- Employed diverse study designs, like randomized and , observational cohort, and case-control studies, to reveal their respective strengths and limitations in drawing conclusions from data
- Tackled the analytical challenge of comparing outcomes across non-randomized groups to highlight confounding variables that may influence clinical and statistical relevance and results
- Completed projects with diverse research data, including continuous, binary/categorical, and time-to-event data adapting statistical methods for appropriate analysis and interpretation

Design and Conduct of Clinical Trials
(Coursera - Johns Hopkins University)
Actionable Learnings:

- Assessed and evaluated clinical trial designs, bias control measures, and randomized participants into groups while defining trial outcomes
- Applied principles randomization, and masking, and to evaluate various bias control approaches and errors in randomization procedures

Clinical Data Analysis with SAS (O'Reilly)
Actionable Learnings:

- Explored key components of the pharmaceutical industry, phases of clinical trials
- Utilized of SAS for data analysis to comprehend data encountered in clinical trials

WORK EXPERIENCE

Statistics and Data Science Intern

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

- Worked as a statistician in 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

Projects Utilizing Machine Learning and Statistics

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

- Reimplemented a Stanford Research Group project to develop an Inflammatory Aging Clock (iAge) utilizing artificial neural networks and auto-encoded features modeling and tracking aging-related processes and health disparities

Sample Size Estimation for Linear Mixed models with Dependent End Points

- Developed a tree spanning algorithm for estimating sample size in linear mixed models (LMMs) with arbitrary dependency structures that prioritized control over family-wise error, offering more powerful bounds and smaller sample sizes compared to Bonferroni sample size estimations

Efficacy of Treatment Sessions in a Medical Emergency Department

- Utilized Interaction Plots, Hasse Diagrams and Linear Mixed Models to analyze and conduct hypothesis tests to reveal patterns of illicit drug use, such as cocaine and cannabis, among individuals, gender and age

Compressing Feature Space For Classification Using PCA

- Applied Principal Component Analysis (PCA) to reduce the dimensionality of 50 chemokines and cytokines to capture variation in estimated biological ages among Hispanic Origin Individuals

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

Federal and Foundational Grants Projects

- 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
- Robert Wood Johnson Foundation Hispanic Health Disparities Grant
- Bill Gates Foundation Health Equity Grant
- Arnold Ventures Health Services Inequity and Disparities Grant
- Paso Del Norte Health Research Grants

Conferences Attended

- JSM Conference 2017, Diversity Scholar
- Cardwell Foundation Seminar, 2016
- College of Science Stakeholders' , 2016 ,2017, 2018, 2019, 2020, 2021, 2022
- 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
- Microsoft AI Research Forum, 2024