MICHAEL NSIAH-NIMO, M.Sc.

DATA SCIENCE | ANALYTICS

CONTACT

GitHub

michael20129648@gmail.com

Linkedin

Project Portfolio

SKILLS & TOOLS

- Programming: Python (Base, Pandas, Numpy, Matplotlib, Scikit-Learn, Keras, TensoFlow), SQL, R, SAS
- Machine Learning: Linear Regression, Logistic Regression, Decision Trees, Random Forest, KNN, k-means, PCA, Association Rule Learning, Causal Impact Analysis, Deep Learning, NLP, Generative AI and LLMs
- Statistics: Experimental Design, Multivariate Analysis, Study Design and Sample Size Estimation, 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
- Professional Skills: Statistics, Github, Data Visualization, MS Office, Tableau, Jupyter Notebook, AWS, Google Cloud Platform

EDUCATION

MSc Statistics

University of Texas at El Paso

2015 - 2017

BSc Actuarial Science

Kwame Nkrumah University of Science and Technology

2010 - 2014

PROFILE

I am driven by a deep passion for addressing complex health challenges and leveraging data-driven insights to deliver tangible value and drive impactful business decisions through Data Science, Statistics, and Machine Learning. With over 6 years of experience, in AI, machine learning, and statistical data processing, I develop predictive models to address complex healthcare challenges. I maintain a continuous learning mindset, constantly seeking ways to enhance my expertise and make meaningful contributions.

WORK EXPERIENCE

Data Scientist and Research Analyst

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

- Led a team of 5 data scientists to develop a Python-based Deep Learning model for a 5 year national grant award to predict the biological age of Hispanic Origin Individuals (HOI) using proteomics data on inflammatory cytokines, chemokines, and growth factors. The model's output provided participants with an estimated biological age - a proxy of the inflammatory burden of the individual, enabling them to compare it with their chronological age and make informed decisions about their health and nutrition habits.
- Employed seaborn visualization and regression models from the Python statsmodels package to explore correlations and examine the connections between estimated biological ages and various socio-behavioral, biomarker, proteomic, metabolomic, physiological, and clinical factors among individuals of Hispanic origin. This comprehensive analysis helped identify key features essential for our predictive model, thereby bolstering our federal grant application and subsequent analytical endeavors. As a result of our approach, we achieved a significant 30% reduction in the costs associated with measuring social factors, enabling more efficient allocation of resources for thorough investigations.
- Pioneered and Spearheaded the implementation of advanced Big Data methodologies at the UTEP Center for Integrative and Translational Research (CITR) with the Dean of Research to foster collaboration and innovation among interdisciplinary researchers. Through the Analytics Core, we established a comprehensive data repository to address health disparities for Mexican-origin Hispanics by integrating diverse determinants of health. This led to and advanced continuous improvement in healthcare equity which underscored tremendous transformative initiatives aimed at positively impacting underserved communities.

Data Analyst

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

- Used SQL, Tableau, and MapBox to extract and model demographic data from El Paso County and created dynamic visualization reports which uplifted more targeted grant applications for health disparities research in the College of Science by 50%
- Originated and analyzed comprehensive social determinants of health trends to identify prevalent issues related to obesity, cancer and metabolic syndrome in the Borderplex region with the Assistant Dean of Research and Leadership.
- Lead a team of scholars and analysts to mined, explore and analyze several data sources including electronic health records, biosamples, surveys, physical measurements and wearable devices, utilizing the NIH All of Us Data Repository to secure federal grants from the National Institute of Health (NIH) and foundation grants from organizations like the Robert Wood Johnson Foundation and the Bill Gates Foundation. I supervised and contributed to critical research initiatives focused on health disparities, clinical research, and socio-behavioral research.
- Presented and participated in cutting - edge data science and statistical conferences and seminars to improve knowledge in bioinformatics, data science, and health equity, in the College of Sciences to drive impactful research outcomes.

MICHAEL NSIAH-NIMO, M.Sc.

DATA SCIENCE | ANALYTICS

COURSES & CERTIFICATIONS

DSI Data Science Professional Certification
Actionable Learnings:

- Extracting & manipulating data using SQL. Application of statistical concepts such as hypothesis tests for measuring the effect of AB Tests. Utilizing Github for version control, and collaboration. Using Python for data analysis, manipulation & visualization
- Applying data preparation steps for ML including missing values, categorical variable encoding, outliers, feature scaling, feature selection & model validation
- Applying Machine Learning algorithms for regression, classification, clustering, association rule learning, and causal impact analysis for measuring the impact of an event over time.
- Machine Learning pipelines to streamline the ML pre-processing & modelling phase. Deployment of a ML pipeline onto a live website using Streamlit.
- Using Tableau to create powerful Data Visualizations. Turning business problems into Data Science solutions.

Machine Learning Specialization
(Deep Learning.ai - Coursera)
Actionable Learnings:

- Built machine learning models in Python using popular machine learning libraries NumPy and scikit-learn. Buillt and trained supervised machine learning models for prediction and binary classification tasks, including linear regression and logistic regression.
- Built and trained a neural network with TensorFlow to perform multi-class classification. Applied the best practices for machine learning development so that the models generalized to data and tasks in the real world

Conferences Attended

- 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
- JSM Conference 2017, Diversity Scholar

WORK EXPERIENCE

Research Analyst

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

- Served as a lead statistician in project planning. design, and analysis, identifying research designs and pertinent data : to address institutional and educational policy issues at UTEP.
- Retrieval and mined higher educational research databases, which generated reports focused on institutional research and continuous improvement.
- Contributed to the development of academic publications and presentations showcasing the utilization of Big Data for institutional research, which facilitated informed decision-making among university stakeholders.

Projects

Enhancing Targeting Accuracy Using ML

- Built a ML model that accurately predicted whether or not customers would sign up for a delivery club. This allowed a critical targeted approach essential to comprehending and running the next iteration of the campaign.

Predicting Customer Loyalty Using ML

- Built a predictive model using a Random Forest in Python that estimated customer loyalty scores for customers that a company's data agency couldn't tag (r-squared 93%). This led to a 30% increase in customers we could analyze, and contact with promotional material.

Quantifying Sales Uplift With Causal Impact Analysis

- Employed causal impact analysis 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. The analysis revealed a significant 41.1% uplift in sales among club members, indicating the club's positive impact on customer spending.

"You Are What You Eat" Customer Segmentation

- Used k-means clustering on grocery transaction data to split out customers into distinct "shopper types" that could be used to better understand customers over time, and to target customers more accurately with relevant content & promotions.

Efficacy of Treatment Sessions in a Medical Emergency Department

- Interaction Plots and Linear Mixed Models were developed and implemented to provide appropriate services to individuals having problematic patterns and use of illicit drugs (cocaine, cannabis). Hypotheses tests were conducted to investigate if there were differences in means amongst different treatment sessions as well as interactions between levels of age, sessions and gender.

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 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