MICHAEL NSIAH-NIMO, M.Sc.

DATA SCIENCE | RESEARCH AND ANALYTICS

CONTACT

My Website

michael20129648@gmail.com

Linkedin

GitHub

EDUCATION

MSc Statistics

University of Texas at El Paso

2015 - 2017

BSc Actuarial Science

Kwame Nkrumah University of Science and Technology

2010 - 2014

SKILLS & TOOLS

- **Experimentation Skills**: Proficient in causal analysis, A/B testing, hypothesis testing, statistics, and quantitative analysis, with experience in experimental design.
- **Programming Languages**: Skilled in Python, R, SQL, NoSQL, Linux, Tableau, PowerBI, Excel, and PowerPoint.
- **Database Technologies**: Experienced with Google BigQuery, AWS, Spark, MongoDB, Hadoop, Snowflake, Azure, Databricks, and other cloud-based database platforms.
- **Libraries and Frameworks**: Proficient in Git, GitHub, Docker, PySpark, PyTorch, TensorFlow, NumPy, Pandas, Plotly, Scikit-Learn, Matplotlib, Seaborn, Spacy, Scipy, Selenium, Streamlit, AI, machine learning, and various algorithms
- **Probability and Statistics**: Proficient in 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, Data Mining and Visualization, Report Generation and Interpretation
- **Artificial Intelligence and Machine Learning Algorithms**: Skilled in Linear Regression, Logistic Regression, Decision Trees, Random Forest, KNN, k-means, PCA, Association Rule Learning, Causal Impact Analysis, Artificial and Convolutional Neural Networks, Transfer Learning, Deep Learning, Natural Language Processing, Generative AI and Large Language Models

PROFILE

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

.

WORK EXPERIENCE

Data Scientist

University of Texas at El Paso,
Integrative Analytic Unit, Border Biomedical Research Center

JULY 2020 - PRESENT

- Lead a team of statisticians and data scientists in developing a **deep learning model to predict immunological age and inflammation levels, tracking aging-related processes and health disparities in Hispanic Origin Individuals (HOI)** to secure a national grant
- Foster collaborations with clinicians , bioinformaticians and professional faculty members to **design and build robust disease risk assessment models utilizing neural networks and tree models** contributing to enhanced healthcare outcomes
- **Wrote python codes to build ML pipelines for comprehensive data cleaning, variable selection, and feature engineering** across projects, creating a remarkable 30% reduction in costs associated with measuring social factors and molecular features in randomized studies for federal grant applications
- Conducts **A/B Tests on identifying relevant metrics** that will boost and extend current strategic initiatives focused on increasing extramural grant and philanthropic support.
- Analyzed and mined **diverse datasets** including surveys, health records, biomedical, genomic and physiological biomarkers, **from a relational data base adapting SQL queries**, to generate actionable insights, advancing research objectives effectively.
- Spearheaded the implementation of **Big Data technologies with the integration of high performance computing** for the UTEP Center for the Integrative and Translational Research (CITR) driving transformative initiatives for healthcare equity in the Ciudad Juárez and United States Borderplex
- Designed **protocols and calculated sample sizes for impactful grant projects, driving research in health disparities**, clinical studies, and COVID-19 mitigation.
- Provided **statistical and data science expertise** for major initiatives like NIH U54 projects, using advanced methodologies including longitudinal analysis, mixed-effect modeling, logistic regression analysis, survival analysis, to uncover insights and meet research goals.

Statistical and Data Analyst

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.
- Used **data analytics to extend strategic initiatives** focused on increasing extramural funding
- 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 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.

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

MICHAEL NSIAH-NIMO, M.Sc.

DATA SCIENCE | RESEARCH AND ANALYTICS

SKILLS & TOOLS

Behavioral Skills

- Demonstrated 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 at identifying challenges and implementing innovative solutions, ensuring the integrity of the analyses.
- Experienced in providing guidance and mentorship to junior team members, fostering a collaborative work environment.
- Flexibility in adapting to evolving project requirements, delivering high-quality analyses in dynamic research environments

COURSES & CERTIFICATIONS

Standford Online - Coursera

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

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

PROJECTS My Website

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***

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***

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

- 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***.