Benson Cyril Nana Boakye

Portfolio Website | +1 541-286-0930 | nanaboab@gmail.com | LinkedIn | Github

Summary

Data scientist and statistician with 4+ years of experience in machine learning, data analysis, and predictive modeling across biomedical research, experimental design, and public health. Proven expertise in statistical programming and clinical trials, with 2+ years of hands-on experience specializing in Python, R, SAS, and SQL, ensuring compliance with industry standards. Proactive and innovative, eager to apply theoretical knowledge and practical skills to deliver data-driven solutions that drive business growth, boost profit, and optimize operational efficiency.

Skills

Programming Languages: Python, R, SQL, SAS.
Machine Learning: Neural Networks, Tensorflow, SVM, Logistic Regression, Random Forest, PCA, Cross-validation
Statistical Methods: GLM, Bayesian Inference, Survival & Multivariate Analysis, Hypothesis Testing, linear Models
Tools & Technologies: Tableau, Excel, GCP, DESeq2, scRNA-seq, Data Structures, CDISC (SDTM, ADaM), MongoDB.

Experience

Statistical Consultant — **September 2024 – June 2025**
Oregon State University — *Corvallis, Oregon*

- Optimized statistical designs for 10+ biomedical studies, including RCTs and longitudinal studies, reducing required sample sizes by 20% while preserving statistical power and study integrity.
- Analyzed genomic data (RNA-seq) using R and Python, identifying 5 biomarkers and contributing to publications.
- Advised clients on analyzing immune responses to TB antigens and their links to malnutrition, using logistic and GLM's.
- Evaluated the correlation between breath condensate pH, reflux events, and tracheal aspirate pepsin levels using SAS.
- Applied logistic regression (GLM) to predict pollen fitness from genomic data, using LASSO for feature selection, improving model AUROC by 70% and interpretability.

Graduate Teaching Assistant — **September 2023 – June 2025**
Oregon State University — *Corvallis, Oregon*

- Introduced experimental design principles (randomization, blocking, factorial designs) to optimize data collection and analysis (ETL processes), improving students' ability to evaluate and apply research methodologies by 20%.
- Integrated SVMs, Decision Trees, and Python libraries to streamline practical exercises, cutting preparation time by 40%
- Led recitation sessions for 500+ students on statistical methods using Python & Excel, boosting performance by 15%.
- Applied PCA and t-SNE for dimensionality reduction, enhancing model interpretability and revealing data clusters.
- Ensured confidentiality, evaluated and gave constructive feedback, fostered critical thinking, and promoted collaboration.

Actuarial Intern — **September 2022 – December 2022**
National Insurance Trust — *Accra, Ghana*

- Applied unsupervised machine learning in python for anomaly detection on 2TB of data, reducing process time by 20%.
- Built predictive models in Python(Tensorflow) to improve policyholder behavior, improving risk profiling accuracy 24%.
- Created visual reports using Power BI and Tableau, applied run-off triangles and pricing models to claims, generated insights to ascertain competitiveness, and communicated with executives to finalize statistical analyses and TLFs.
- Assisted in utilizing Monte Carlo simulation methods to optimize reinsurance strategies and deployed asset-liability management strategies to enhance investment portfolios resulting in a 5% higher ROI for clients.

Projects

Survival analysis of treatment efficacy in Primary Biliary Cirrhosis (PBC) | *SAS*

- Analyzed Cyclosporin A's effect on PBC survival using Kaplan-Meier, improving prediction accuracy by 15%.
- Implemented AFT and Cox models to assess survival, identifying high bilirubin and low albumin as key risk factors.

Predictive modeling and data analysis of breast cancer | *Python*

- Designed and assessed 5 classification models (Logistic Regression, KNN, Random Forest, SVM with RBF kernel, and Gradient Boosting), achieving up to 94% precision in breast cancer through diagnosis using FNA image features.
- Achieved 98.25% accuracy in classifying breast cancer as malignant or benign using an SVM with an RBF kernel.

Modeling Female Sex Worker distribution in Sub-Saharan Africa for HIV prevention | *R-studio*

- Deployed mixed-effects models (negative binomial, zero-inflated, poison, etc.) to analyze data, with an AIC of 2863.208.
- Applied REML estimation to confirm the inclusion of country, region, and dataYear as random effects ($p < 0.0001$).

Air quality analysis with R-shiny app | *R-studio*

- Developed an interactive app to explore PM2.5 levels in cities across the USA and India, allowing users to analyze trends.
- Added event annotations, like COVID-19, to highlight air quality changes and guide public health decisions..

Education

Oregon State University — **June 2025**
Master of Science degree in Statistics — *Corvallis, Oregon*

Kwame Nkrumah University of Science and Technology (KNUST) — **August 2018 – September 2022**
Bachelor of Science in Actuarial Science — *Kumasi, Ghana*