HABEEB ABOLAJI BASHIR

843 S Broadway Lexington, Kentucky
Phone:(+1) 9157310967 Email:habeebabolaji@gmail.com LinkedIn:
www.linkedin.com/in/bashir-abolaji-71885411b/

Summary

As a dedicated PhD student in Statistics and Data Science, I excel in applying SAS, SQL, Pyspark/Spark, Hadoop, Python, R, Databricks, and C++ to tackle complex data analysis, machine learning, and statistical modeling challenges. My expertise spans data structures, algorithms, and NLP, aiming to drive innovation and enhance decision-making in data-driven projects. Committed to contributing to advancements and fostering knowledge exchange in dynamic environments, I am poised to leverage my comprehensive technical skills for continuous improvement in the field of data science.

Professional Experience

Research Assistant | University of Texas - El Paso, TX | 06/01/2022 - 05/13/2024.

- Implemented supervised and unsupervised machine learning algorithms, text analytics, and time series forecasting to predict traffic patterns and delays, and optimize routing strategies, enhancing prediction accuracy by 25%.
- Collaborated with a research supervisor to devise strategies for improving transportation efficiency data, leading to a 46% increase in operational speed and reliability.
- Designed and implemented Python modules and R scripts, alongside machine learning models, for forecasting traffic patterns and improving logistical efficiency in daily transport operations, achieving a 35% boost in predictive accuracy.
- Employed dimensionality reduction techniques like Principal Component Analysis and k-fold cross-validation to optimize transportation models, reducing computational load by 40%.
- Conducted research and generated rapid plots using advanced data mining and statistical modeling methods, resulting in a 50% improvement in data visualization clarity.
- Established the necessary infrastructure for efficient extraction, transformation, and loading of transportation data from various sources using SQL and AWS big data technologies, cutting data processing time by 45%.
- Utilized techniques such as confusion matrix, cross-validation, and AUC-ROC for parameter tuning and model evaluation, leading to a 30% improvement in model precision and recall.

Data Science Intern | Nigeria Bureau of Statistics (NBS) | 01/05/2019 – 05/12/2021

- Played a key role in cross-functional team projects that enhanced data-driven decision-making processes by 40%.
- Led data quality assessments and preprocessing initiatives, resulting in a 25% improvement in data accuracy and usability.
- Contributed to the deployment of predictive models, increasing operational efficiency by 30% in targeted use cases.
- Developed and implemented a robust supervised machine learning algorithm for optimizing treatment decisions for a dataset of 200,000 diabetes patients, achieving a 95% improvement.
- Applied predictive modeling techniques, including Rpart, Lasso, Boosting, and Random Forest, to inform clinical trial treatment decisions through an average of 100,000 simulations, yielding interpretable results.
- Use Infrastructure as Code tool such as Terraform to provision infrastructure necessary to host and run applications on AWS
- Compile and build software source code using tools such as Jenkins, GitHub, Git, and Maven

Educational Background

PhD, Statistics and Data Science | University of Kentucky, Lexington | July 2024 – Till Present

- Relevant Coursework: Data Collection and Cleaning, Exploratory Data Analysis, Statistical Modeling and Machine Learning, Advanced Data Visualization, Big Data Analytics, and Computational Statistics.

Master of Science, Statistics and Data Science | University of Texas at El Paso, Texas USA | August 2021- May 2023

- **Relevant Coursework:** Machine Learning and Predictive Modeling, Exploratory Data Analysis, Advanced Statistical Inference, Operation Research, Simulation Modelling, Statistical Quality Control.

Bachelor of Science, Statistics | Usmanu Danfodiyo University Sokoto, Nigeria | December 2012- October 2017

Skills

- Demonstrated proficiency in R, SAS, Pyspark/Spark, Hadoop, and Python, improving data analysis efficiency by over 60%.
- Applied expertise in analytical methods to enhance model accuracy and reliability by up to 25%.
- Advanced knowledge in relational databases and SQL, streamlining data retrieval processes by 20%.
- CICD tool and Build Tools: Maven, Jenkins.
- AWS tools: EC2, Kafka brokers, ALB, Lambda, S3, Route 53, Elastic Cache.
- Improved data visualization impact by 25% using innovative approaches with Tableau and ggplot.
- Proficient in C++ with RCPP, enhancing computational performance and efficiency in statistical computing tasks.
- Successfully integrated C++ code into R environments for high-performance computing applications, yielding a 40% reduction in computation time for complex data analysis and simulation tasks.

Additional Information

- Master's Thesis Project: Developed flexible models for treatment effect estimation, increasing methodological adaptability and applicability by over 75%