Farisayi E Dakwa

✉ fariedakwa.fd@gmail.com in linkedin.com/in/farisayi-e-dakwa/ ⬡ farisayi/project-portfolios

SUMMARY

Data Scientist with 5+ years of experience applying machine learning and statistical modeling to real-world problems. Proven ability to build scalable data pipelines, machine learning models, and deliver actionable insights from large datasets.

See more project portfolio…

I'm proficient in:

- **Languages**: Python, R
- **Databases**: MySQL, PostgreSQL
- **Data Visualization**: Tableau, Matplotlib, Seaborn, ggplot2, Plotly
- **Machine Learning & Statistics**: Supervised/unsupervised learning, NLP, GLMs, Random Forests, classification, clustering
- **Tools & Frameworks**: Pandas, Numpy, Scikit-Learn, TensorFlow, Caret, Git, Rshiny
- **Specialities**: Data Science, Visualisation, Machine Learning, Web Scraping, Experimental Design

WORK EXPERIENCE

InscopeData   **Mar 2025 - Present**
Data Scientist   *Remote*
- Designed and automated data mining pipelines in Python using web scraping tools to consolidate sales data across multiple e-commerce platforms.
- Performed ETL, data cleaning, manipulation, and transformation of raw data into query-ready datasets in SQL
- Developed product matching algorithms using LLM and NLP techniques, integrating agentic AI and cloud APIs

Imvelo Blue Consultancy   **Mar 2023 – Present**
Data Consultant   *Part Time*
- Develop and maintain scalable data pipelines in R (tidyverse) to clean, preprocess, and analyze large datasets, cutting processing time from 15 hours/week to 5 hours/week and enhancing error detection.
- Developed interactive visualizations in R (ggplot2, shiny), improving report techniques that became the standard for company reporting practices.
- Leveraged statistical modeling and geospatial analysis in R (terra, sf, geosphere) to identify trends, boosting data accuracy by 25% and improving data-driven decisions.

Bayworld Center for Research and Education   **Jan 2021 - Aug 2023**
Quantitative Researcher
- Designed end-to-end ML pipelines (GLMs, Random Forests, GAMs) using caret and H2O, achieving over 90 % predictive accuracy.
- Processed and analyzed over 50k geospatial data points using sf, terra, geosphere, and ggplot2, building publication-grade visualizations (mapview, leaflet) to communicate complex trends.
- Automated data wrangling (tidyverse), feature engineering, and model evaluation (RMSE, R2, AUC, Imodel2), implementing clustering, classification, and multivariate analysis for predictive modeling.

EDUCATION

University of Cape Town, South Africa   **2019 - 2021**
Master's by research degree in Quantitative Research
- Research thesis in Applied Science, designed experiments analysing 50 years of data
- Trend analysis, probabilistic modelling, clustering, and regression
- Applied Bayesian techniques and multivariate statistics on demographic data

IBM - *Professional Data Science Professional Certificate*   **2023/24**
- Data wrangling, ETL and SQL proficiency

                                                                         2025

Stanford University Online
Machine Learning Specialisation
- Statistical foundation of machine learning frameworks
- Advanced learning algorithms implementation in Python, deep learning, and reinforcement learning

PROJECTS

Machine Learning-based Classification & Sales Analysis Model

- Developed a machine learning pipeline in Python using NLP techniques, TF-IDF vectorization, and Naïve Bayes to classify over 700,000 product descriptions into Google taxonomy categories, achieving 88% accuracy. Currently exploring deployment strategies (APIs) to support client-facing product features Learn more…
- Summarized and visualised sales data using Python (pandas, seaborn, ggplot2) and created an interactive Shiny dashboard in R to visualize product associations and trends based on Black Friday sales week data.