

# Farisayi E Dakwa

fariedakwa.fd@gmail.com | [linkedin.com/in/farisayi-e-dakwa/](https://www.linkedin.com/in/farisayi-e-dakwa/) | [farisayi/project-portfolio](#)

## WORK EXPERIENCE

### Data Scientist

**Imvelo Blue**, Cape Town South Africa | 03/2023 – Present

- Develop automated and scalable data pipelines in R (*tidyverse*) to clean, preprocess, analyze large datasets, cut processing time from 15 hours/week to 5 hours/week, and enhance error detection.
- Developed interactive visualizations in R (*ggplot2*, *shiny*), drove operational efficiency, improving report techniques which became the gold 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.

### Quantitative Researcher (Contract)

**Bayworld Center for Research and Education**, Cape Town, South Africa | 01/2021 – 08/2023

- Designed end-to-end ML pipelines (GLMs, Random Forests, GAMs) using *caret* and *H2O*, achieving >90% predictive accuracy.
- Processed and analyzed 50k+ geospatial data points using *sf*, *terra*, *geosphere*, *gganimate* and *ggplot2*, building publication-grade visualizations (*mapview*, *leaflet*) to communicate complex trends.
- Automated data wrangling (*tidyverse*), feature engineering, and model evaluation (RMSE,  $R^2$ , AUC, *Imodel2*), implementing clustering, classification, and multivariate analysis for predictive modeling.

## PROJECTS

### Machine Learning based Product Classification & Sales Analysis Model

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

## SKILLS

Python (Pandas, Numpy, Scikit-Learn), R, MySQL, SQLite, Microsoft, Machine learning (TensorFlow, PyTorch, Keras, caret), Feature Engineering, NLP, Tableau, Shiny, ggplot2, Matplotlib, Seaborn, Plotly

## EDUCATION

**University of Cape Town**, South Africa

**Masters by Dissertation in Quantitative Research** | 08/2019 – 12/2021

- Thesis on applying regression, time series modeling, and multivariate analysis to big datasets (50 years), identifying key predictors and trends for ecological insights and forecasting.

**IBM | Professional Data Science Certification** | 2024

- Improved expertise in Python programming, SQL, and API integration, applying machine learning workflows, ETL pipelines, and MLOps best practices for end-to-end data-driven decision-making.