Farisayi E Dakwa

fariedakwa.fd@gmail.com linkedin.com/in/farisayi-e-dakwa/ farisayi/project-portfolio

SUMMARY

Data Scientist with +5 years of developing scalable data pipelines and predictive models using Python and R to transform large datasets into actionable insights for research initiatives. Currently at Imvelo Blue, automated workflows and reduced processing time significantly while improving data accuracy by 25%. Prior roles involved end-to-end ML pipeline design achieving >90% predictive accuracy and creating interactive visualizations with Tableau and Shiny. Strong analytical and cross-functional data skills, support collaboration, and innovative problem-solving.

SKILLS

- **Programming Languages**: Python, R, MySQL, SQLite
- **Machine Learning**: Scikit-learn, Supervised & Unsupervised Learning, TensorFlow, PyTorch, Keras, Caret
- **Data Visualization**: Tableau, Shiny, Ggplot2, Matplotlib, Seaborn, Plotly
- **Data Processing**: Pandas, Numpy, Feature Engineering, NLP, tidyverse
- **Data Science & Analytics**: Data Science, Data Analytics

WORK EXPERIENCE

Data Scientist **Mar 2023 - Present**
Imvelo Blue Consultancy *Cape Town, South Africa*

- Develop automated and 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.

Quantitative Researcher (Contract) **Jan 2021 - Aug 2023**
Bayworld Center for Research and Education *Cape Town, South Africa*

- 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 **Aug 2019 - Dec 2021**
Masters by Dissertation, Quantitative Research

IBM **2024**
Professional Data Science Certification

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. Currently 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...