

Vusumuzi Nkosi

Data Scientist | Newcastle, KwaZulu-Natal, South Africa
vusizwe@rocketmail.com | +27 72 641 2519
[linkedin.com/in/ngwanelegacie](https://www.linkedin.com/in/ngwanelegacie) | github.com/ngwanelegacie | ngwanelegacie.github.io

S U M M A R Y

Data science postgraduate student (STADIO, NQF 8) and **ALX Data Science graduate** who builds machine learning models, cleans and analyses datasets, and creates dashboards that drive decisions. Built a credit risk prediction model achieving **87% accuracy and 0.92 AUC-ROC** using Python and scikit-learn. Nine years of mathematics teaching developed strong problem-solving, statistical thinking, and communication skills — the ability to translate complex data into clear insights for non-technical stakeholders. Seeking data scientist, data analyst, or machine learning roles in South Africa.

T E C H N I C A L S K I L L S

Programming: Python, SQL, R (basic)

ML & Modelling: scikit-learn, XGBoost, Classification, Regression, Logistic Regression, Random Forest, Gradient Boosting, SMOTE, Feature Engineering, Hyperparameter Tuning, Cross-Validation, GridSearchCV

Data Analysis: Pandas, NumPy, Data Wrangling, Data Cleaning, Exploratory Data Analysis, Descriptive Statistics, Hypothesis Testing, A/B Testing, Correlation Analysis, Regression Analysis

Visualisation & BI: Matplotlib, Seaborn, Power BI, Tableau, Excel, Data Visualisation, Dashboard Design, Storytelling with Data

Tools & Platforms: Git, Jupyter Notebook, Google Colab, Google Sheets, AWS SageMaker, Prompt Engineering, Generative AI

P R O J E C T S

Credit Risk Prediction Model

Python, scikit-learn, XGBoost

github.com/ngwanelegacie/credit-risk-prediction

- Built an end-to-end ML pipeline predicting loan defaults on a 10,000-record dataset with 18% class imbalance — best model achieved 87% accuracy and 0.92 AUC-ROC
- Engineered 8 predictive features (income-to-loan ratio, high-risk credit flags, DTI thresholds) and applied SMOTE oversampling to balance training data
- Compared 3 classifiers with 5-fold cross-validation; optimised XGBoost across 36 hyperparameter combinations using GridSearchCV

Excel Sales Performance Analysis

Excel

github.com/ngwanelegacie/excel-sales-analysis

- Analysed sales data using pivot tables, VLOOKUP, and conditional formatting to identify top products, seasonal trends, and revenue drivers
- Built an interactive dashboard with charts and slicers to support data-driven business decisions

US Household Income Analysis

SQL

github.com/ngwanelegacie/us-household-income-sql

- Cleaned and analysed household income data using advanced SQL (joins, window functions, CTEs, aggregations), uncovering regional income trends and demographic patterns

World Life Expectancy Analysis

SQL

github.com/ngwanelegacie/world-life-expectancy-sql

- Investigated correlations between life expectancy and health variables across 190+ countries using SQL data cleaning and exploratory analysis

WORK EXPERIENCE

Senior Lecturer — Mathematics & IT

Feb 2017 – Present

Amajuba TVET College — Newcastle, KwaZulu-Natal

- Deliver mathematics and IT instruction to 100+ students annually across multiple cohorts, designing curriculum and assessments aligned with national standards
- Analyse student performance data to identify at-risk learners and adjust teaching strategies — improved pass rates through targeted, data-driven interventions
- Developed standardised assessment rubrics adopted across the mathematics department for consistent grading
- Mentored students in IT competitions at regional and national level, with multiple teams advancing to national rounds

Special Educator — Mathematics

Jan 2012 – Mar 2016

Mdumiseni High School — KwaZulu-Natal

- Delivered mathematics lessons for diverse learner groups, tracking individual progress and adapting materials to different learning needs

EDUCATION

Postgraduate Diploma in Data Science (NQF 8)

2026 (in progress)

STADIO Higher Education — Remote

- 120-credit qualification (SAQA ID: 122721) — modules: Introduction to Data Science, Working with Data, Statistical Modelling, Applied Machine Learning

Data Science Programme

Mar 2025 – Apr 2026

ALX Africa — Remote

- Comprehensive programme covering data analytics, Python, machine learning, and professional foundations

BSc Mathematics & Computer Science

2022 – 2025

University of South Africa

BEd Mathematics & Computer Science

2012 – 2017

University of South Africa

CERTIFICATIONS

- Data Science Certificate — ALX Africa (2025–2026)
- Data Analyst Certificate — ALX Africa (2025) — SQL, Power BI, A/B Testing, Regression Analysis
- AI Engineering Bootcamp — Zero To Mastery Academy (2025)
- Prompt Engineering Bootcamp — Zero To Mastery Academy (2025)