**FRIDAH**

**GECHEMBA MACHANI**

**DATA SCIENTIST**



**SUMMARY**

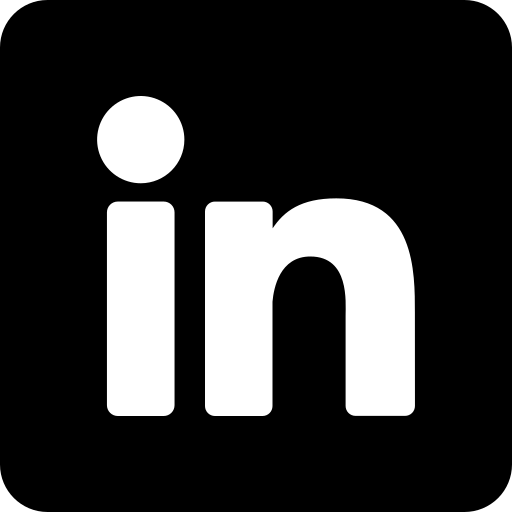
A data-driven professional with experience in using advanced analytical methods to uncover hidden patterns, and transform complex, high-dimensional datasets, into actionable strategic insights. I am skilled in predictive modeling, machine learning, and advanced analytics, and eager to create impactful solutions using Python, R, SQL, Power BI, and modern ML libraries and frameworks.

**CONTACT**











Ludwigsfelde, Germany

[machafrida@email.com](mailto:machafrida@email.com)

+4915216878666

[www.linkedin.com/in/fmachani](http://www.linkedin.com/in/fmachani)

<https://machanig.github.io/>

**WORK EXPERIENCE**

**Additional Training / Full-time Parenting Mar 2024 – April 2025**

* I deepened my data science and machine learning expertise, learned querying databases with SQL, and Power BI for data visualization.

**Research Assistant Aug 2020 – Feb 2024**

**Max Planck Institute MPIMP – Golm, Germany**

* Developed and calibrated regression models (R2 = 0.93) in R to estimate nitrogen concentrations and enzymatic activity from known dilutions; deployed models and made predictions for over 500 samples, demonstrating end-to-end ML model development.
* Implemented multifactor ANOVA models with interactions in R to quantify the effect of mutations and nutrient availability on multiple plant features, showcasing experimental design and statistical modeling.
* Conducted unsupervised learning using K-means, PCA, and PERMANOVA to cluster multivariate metabolomics data (66 features); uncovered tissue-specific and nutrient-specific profiles.
* Utilized threshold-based outlier detection to identify anomalous gene expression profiles in 91K+ high-dimensional Affymetrix data and validated anomalies through wetlab quantification.
* Automated pre-processing pipelines in R to transform raw outputs from qPCR, GCMS, and spectrophotometer systems into structured datasets; by applying custom normalization, log transformation and feature extraction, streamlining data analysis workflows.

**Research Assistant Oct 2019 – May 2020**

**Universität Münster IEB – Münster, Germany**

* Applied time-series analysis to determine optimal exudate collection time for maximizing seed germination rates.

**Lecturer Aug 2018 – May 2019**

**Machakos University – Machakos, Kenya**

* Designed and delivered interdisciplinary courses to students from 3 faculties, adapting content to the audience needs and feedback.

**TECHNICAL SKILLS**

Programming Languages

* Python
* SQL (PostresSQL, MySQL)
* R

Data Analysis and Manipulation

* Pandas, Numpy, Statsmodel , Scipy, Tidyverse

Data Visualization Tools

* Matplotlib, Seaborn, ggplot2, Power BI

Machine Learning

* ML Libraries and Frameworks
* TensorFlow, Scikit-Learn, XGBoost, MLextend, LightGBM
* Supervised Learning
* Regression & Classification
* Anomaly Detection
* Unsupervised Learning
* Clustering (K-means, Hierarchical, K-medoids)
* Dimensionality Reduction (PCA, t-SNE, PCoA)

**Genomics Data Analyst Jan 2017 – May 2018**

**World Agroforestry Center – Nairobi, Kenya**

* Constructed UPGMA and neighbor-joining dendrograms to analyze genetic divergence (max distance = 0.30) and assessed clustering robustness via 1,000 bootstrap replicates.
* Quantified intra- and inter-cluster variances using ANOVA, revealing 86% molecular variation within populations (*p* < 0.001).

**EDUCATION**

**Trainee in Data Analytics Mar 2024 – Jun 2024**

ReDi School of Digital Integration, Berlin, Germany

Coursework:

* Data analytics with Python - fundamentals of the Pandas toolkit, principles of data filtering and the groupby method, data exploration and storytelling.
* Data analytics with SQL - fundamentals of SQL, data analysis and visualization in SQL, creation and presentation of data dashboard.

**Dr.rer.nat. Molecular Genetics**  **Sept 2020 –** **Nov 2023**

Universität Potsdam, Potsdam, Germany

Gained expertise in:

* R programming, data cleaning, data wrangling, data analysis (EDA, PCA, clustering, regression, ANOVA, t-tests, post-hoc tests, anomaly detection), data visualization with ggplot2, and statistical inference using R.
* Creating data reports that effectively meet stakeholder needs.

**Master of Science Biotechnology** **May 2015 –** **Aug 2018**

Kenyatta University, Nairobi, Kenya

Coursework:

* Introduction to Statistics: covered statistical methods such as, analysis of variance, Chi-square, t-test, correlation, regression, probability distributions, post-hoc tests, and nonparametric tests.
* Scientific Data Analysis: covered descriptive, diagnostic, and causal analyses, and making inferences to a larger population.
* Research Methodology for Pure and Applied Science: focused on qualitative / quantitative methods, hypothesis testing and p values.

**ADDITIONAL TRAINING**

* IBM **Data Science** **Specialization** – Coursera. Aug, 2024
* **Machine Learning** **Specialization** by Stanford University & DeepLearning.AI – Coursera. Mar, 2025
* **Machine Learning Fundamentals with Python** Skill Track – DataCamp. Mar, 2025
* **Supervised Machine Learning in Python** Skill Track –DataCamp. Mar, 2025
* University of California Irvine **Predictive Modeling, Cluster Analysis & Association Mining** – Coursera. Apr, 2025
* **Power BI Fundamentals** Skill Track (5 courses) – DataCamp

**PROJECTS**

* **Stacking ML models and feature engineering** – stacked regressors for house price prediction and combined it with advanced feature engineering including polynomial features.
* **Comparing ML models for TB detection –** compared various classifiers in their ability to predict TB from chest X-ray data.
* **Scaling up sentiment prediction with TensorFlow** – built and scaled up a sentiment classifier on 4 million product reviews using TensorFlow and batch processing.

**LANGUAGES**

* English - Native / bilingual
* German - A2
* Ensemble Learning
* Bagging, Voting
* Boosting, Stacking

Statistical Analysis

* Descriptive statistics
* Inferential statistics
* Hypothesis testing
* Diagnostic and Prescriptive analysis
* Predictive analytics
* Causal analytics

Version Control

* Git & GitHub

Professional Strengths

* Analytical thinking & data-driven decision making
* Strong attention to detail
* Effective communicator & collaborative team player
* Creative problem solver with a growth mindset
* Project management