FRIDAH GECHEMBA MACHANI

DATA SCIENTIST

CONTACT

+4915216878666

machafrida@gmail.com

Ludwigsfelde, Germany

www.linkedin.com/in/fmachani

https://github.com/machaniG

https://machanig.github.io/

SKILLS

Professional Competencies

- Data Analysis & Visualization
- Machine Learning
- Statistical Modeling
- Data Preprocessing & Cleaning
- Analytical Problem-Solving
- Project Management
- Communication & Data Storytelling
- Cross-Functional Collaboration

Technical Skills

- Python (Pandas, Scikit-Learn)
- SQL
- TensorFlow
- Power BI
- R

SUMMARY

Data-driven professional with hands-on experience in predictive modeling, machine learning, and advanced analytics. Skilled in transforming complex datasets into actionable insights using Python, R, SQL, and Power BI. Proven ability to deliver end-to-end data solutions and communicate findings to diverse stakeholders.

WORK EXPERIENCE

Full-Time Parenting Mar 2024 – Present

- Completed advanced training in data science, machine learning, and analytics to strengthen my technical skills.
- Developed strong time management, multitasking, and prioritization skills while balancing learning with full-time childcare responsibilities
- Enhanced adaptability, resilience, and problem-solving through managing competing demands and continuous upskilling.

Doctoral Researcher Aug 2020 – Feb 2024

Max Planck Institute MPIMP – Golm, Germany

- Developed and deployed regression models ($R^2 = 0.93$) in R to predict key outcomes from complex datasets, high-dimensional datasets; delivered actionable insights for over 500 real-world samples.
- Designed and implemented multifactor statistical analyses to evaluate the impact of multiple variables on measurable outcomes, supporting data-driven decision-making.
- Applied unsupervised machine learning techniques (K-means, PCA) to identify patterns and segment multivariate datasets, uncovering actionable insights for further analysis.
- Automated data preprocessing pipelines in R, transforming raw experimental outputs into structured, analysis-ready datasets using normalization, log transformation, and feature extraction.
- Developed anomaly detection workflows to identify and validate outliers in large datasets (91,000+ features), enhancing data quality and reliability.
- Conducted exploratory data analysis and statistical summarization to uncover correlations and generate hypotheses for further investigation.
- Effectively communicated complex analytical findings to cross-functional teams, translating technical results into clear, actionable recommendations.

EDUCATION

Data Analytics
ReDi School of Digital Integration, Berlin, Germany
Mar 2024 – Jun 2024

- Data analytics with Python: Data manipulation (Pandas), filtering, grouping, exploration, and storytelling.
- Data analytics with SQL: Querying, data analysis, visualization, and dashboard creation.

Dr.rer.nat. Molecular Genetics
Universität Potsdam, Potsdam, Germany
Sept 2020 – Nov 2023

- Gained expertise in R programming, data cleaning and wrangling, exploratory data analysis, regression, clustering, statistical modeling (ANOVA, t-tests, post-hoc tests), anomaly detection, data visualization (ggplot2).
- Led independent research projects requiring advanced data analysis, statistical inference, and clear communication of results to technical and non-technical audiences.

Master of Science Biotechnology
Kenyatta University, Nairobi, Kenya
May 2015 – Aug 2018

- Relevant Coursework: Statistics (regression, probability distributions, ANOVA), scientific data analysis, research methodology, inferential statistics.

LANGUAGES

- English – Native / bilingual
- German – A1

PROJECTS

Optimizing Marketing KPIs and Predicting Customer Conversion

- Built customer conversion funnels for new and existing users using SQL and pandas.
- Analyzed CTR, CR, CPC, and CAC across campaign types and channels; visualized trends with Plotly.
- Predicted conversion using XGBoost and Random Forest (Precision: 0.92 and 0.89 respectively). Compared model performance and feature importances to inform campaign strategy.

Understanding E-Commerce Consumer Buying Habits

- Segmented customers using rule-based thresholds (e.g., VIP: > $1,000 per transaction).
- Analyzed VIP purchasing patterns by gender, location, product category, and time. Identified key temporal insights (e.g., no VIP transactions on Tuesdays, sales spike in April).
- Created comparative dashboards in Power BI across segments: customer types, products, and KPIs.

Sentiment-Enhanced Hybrid Recommendation System

- Built and trained a custom sentiment classifier using LSTM with Bidirectional layers in TensorFlow on raw product reviews, achieving 99% overall accuracy and 88% recall for negative reviews.
- Integrated predicted sentiment scores into recommendation logic to enhance recommendation quality and diversity.
- Developed a hybrid recommendation engine that combines collaborative filtering (Surprise SVD), content-based filtering, and LSTM-based sentiment analysis of product reviews, tailoring recommendations to individual preferences.

CERTIFICATIONS & COURSES

- IBM **Data Science Specialization** – Coursera. Aug, 2024
- **Machine Learning Specialization** (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
- **Predictive Modeling, Cluster Analysis & Association Mining (**UC Irvine) – Coursera. Apr, 2025
- **Power BI Fundamentals** Skill Track – DataCamp