

# Anele Asanda Maphalala

Junior Data Scientist

+27 78 105 8334 • [maphalalaanele@gmail.com](mailto:maphalalaanele@gmail.com) • <https://www.linkedin.com/in/anele-maphalala-04b02921a/>  
• <https://github.com/AneleMaphalala> • <https://anelemaphalalaportfolio.netlify.app/>

## Summary

An enthusiast driven by a passion for problem-solving, continuous learning, and extracting meaningful insights from data to drive well-informed decision-making and innovation across various industries. Proficient in data wrangling, database management, data visualization using Power BI and Tableau, statistical analysis, machine learning, and deep learning, leveraging languages such as Python, R and SQL. Additionally, adept in test-driven development, software development, and web development using languages such as C#, Java, JavaScript, HTML and CSS.

## Skills

Languages: Python • R • SQL • Java • HTML • CSS • JavaScript • C#

Frameworks: Pandas • Numpy • Scikit-Learn • Matplotlib • PyTorch • TensorFlow • Keras

Tools: Power BI • Microsoft Excel • Tableau • Tableau Prep Builder • SQLite • MySQL • Power Query • SQL Server Integration Services

Platforms: PyCharm • Jupyter Notebook • Visual Studio Code • IntelliJ IDEA • Google Colab • SQL Server Management Studio • RStudio • Streamlit

Soft Skills: Problem Solving • Team Collaboration • Communication • Storytelling • Adaptability • Interpersonal

## Experience

Sand Technologies

Data Scientist Intern

03/2024 - 06/2024

Tools utilized: Python, Jupyter, Visual Studio Code, Keras, TensorFlow, PyTorch, Google Colab, Streamlit, Git, GitHub

- Contributed to a data science initiative addressing water loss in distribution networks using acoustic loggers.
- Collaborated with domain experts to align technical decisions with business objectives.
- Conducted comprehensive Exploratory Data Analysis to ensure model reliability, identifying false negatives and false positives.
- Implemented rigorous data preprocessing techniques for optimal model training.
- Designed deep learning models to classify sound files from acoustic loggers, distinguishing between leak and no-leak conditions with high accuracy.
- Selected, trained, and evaluated PyTorch Convolutional Neural Network model achieving 84% accuracy and TensorFlow/Keras Recurrent Neural Network model achieving 85% accuracy.
- Integrated trained models into a Streamlit application, preprocessing audio files, extracting features, and determining leak presence.

InternCareer

Data Analyst Intern

01/2024 - 02/2024

Tools utilized: Python, Visual Studio Code, Jupyter, Power Bi, Git, GitHub

- Utilized Python for comprehensive data analysis and manipulation, including data cleaning, transformation, and exploration to extract actionable insights.
- Developed and implemented a content-based filtering algorithm of YouTube content creators.
- Significantly enhanced user engagement by delivering personalized recommendations improving viewer satisfaction.
- Developed insightful visualizations using Power BI to facilitate data-driven decision-making and reporting.
- Improved data transparency and accessibility, aiding stakeholders in understanding key trends and metrics.

## Projects

Credit Card Approval

<https://github.com/AneleMaphalala/Credit-Card-Approvals>

A machine learning system that evaluates applicants' creditworthiness based on demographic and financial factors using Python.

Tools utilized: Python, Jupyter, Visual Studio Code, Git, GitHub

- Implemented Logistic Regression and Random Forest Classifier both achieving 90% accuracy and ADA Boost Classifier (91% accuracy) to predict credit card approval outcomes.
- Conducted data preprocessing, feature selection, and model evaluation to ensure reliable and interpretable predictions.
- Optimized models for performance, achieving high accuracy and improving decision-making in the credit card approval process.

Healthcare Analytics Dashboards

<https://github.com/AneleMaphalala/Healthcare-Analytics-Dashboards>

Developed interactive dashboards to analyze patient data and derive actionable insights for healthcare management.

Tools utilized: Power BI, Power Query, Git, GitHub

- Observed a high prevalence of chronic conditions like Arthritis (17%) and Diabetes (17%) indicating key areas for targeted health interventions.
- 34% of admissions were classified as "Elective," indicating proactive healthcare planning and opportunities for more efficient resource allocation in hospitals.
- Patient demographics dashboard revealed an age distribution skewed towards patients aged 36-64, indicating a higher demand for age-specific medical services.
- A- (34%) and A+(34%) blood types were the most common among patients, highlighting the need for hospitals to prioritize blood supplies and tailor resources for blood-related treatments and interventions.

## Projects

### E-Commerce Sales Analysis

<https://github.com/AneleMaphalala/E-Commerce-Sales-Analysis>

An in-depth sales analysis on an e-commerce dataset using Standard Query Language (SQL).

Tools utilized: SQL, Git, GitHub, SQL Server Management Studio, SQL Server Integration Services

- Implemented the RFM (Recency, Frequency, Monetary) model to categorize customers based on purchasing behavior, helping to identify loyal customers and potential market segments.
- Generated insights into top-performing products, profitable cities, and states to guide business expansion strategies.
- Analyzed monthly profitability trends and identified areas of improvement for sales targets, customer engagement, and marketing efforts.

### IMDB Sentiment Analysis

<https://github.com/AneleMaphalala/IMDB-Sentiment-Analysis>

A sentiment analysis in natural language processing(NLP) and machine learning project using the IMDB Dataset leveraging Python.

Tools utilized: Python, Jupyter, Visual Studio Code, Git, GitHub

- IMDb's sentiment analysis involves classifying reviews as either positive or negative, a binary outcome.
- Demonstrated that the Logistic Regression model (achieved an 88% accuracy) is a reliable and effective classifier for IMDb movie review sentiment analysis, with balanced performance across both positive and negative sentiments and high overall accuracy.
- Enhances user experience by 20% and provides valuable insights but also sets the stage for future advancements in natural language processing and machine learning within the IMDb platform.

## Education

### ExploreAI Academy

National Certificate: Information Technology (Systems Development)

06/2023 - 07/2024

- Data Science

### WeThinkCode\_

National Certificate: Information Technology (Systems Development)

09/2022 - 12/2023

- Software Engineering

### University of KwaZulu-Natal

Bachelor of Medical Science Honours

02/2021 - 12/2021

- Microbiology

### University of KwaZulu-Natal

Bachelor of Medical Science

02/2018 - 12/2020

- Physiology

## Certifications

SQL: A Practical Introduction for Querying Databases — IBM

Excel for Data Science — Great Learning

Data Analytics using Excel — Great Learning

Multivariate Time Series Forecasting in R — Great Learning

R for Data Science — Great Learning

Data Fundamentals — IBM

Machine Learning Regression in Python — Udemy

Java Intermediate — Sololearn

## References

Available upon request.