Emmanuel Sifiso Nkosi

Gauteng, South Africa • (+27) 69 182 3052 • emmanuelnkossi@outlook.com

LinkedIn • GitHub • Portfolio

💡 Professional Summary

Versatile and driven **Data Scientist** with a strong foundation in IT System Development and practical experience delivering real-world solutions in **AI**, **machine learning**, **NLP**, **data visualization**, and **full-stack development**. Demonstrated ability to build and deploy **end-to-end analytical platforms**, extract insights from complex datasets, and collaborate across multi-disciplinary teams. Experienced in Telecoms environments.

🔧 Technical Skills

Languages & Frameworks: Python, SQL, JavaScript, HTML/CSS, Flask, FastAPI, React, Streamlit
Libraries & Tools: Pandas, Matplotlib, Plotly, Scikit-learn, FAISS, UMAP, HDBSCAN, BERTopic, LLaMA2 & LLaMA3, SentenceTransformers, Power BI, Git
Cloud & Data: Azure SQL, AWS Foundations, REST APIs, Data Warehousing, Chatbot Development
Soft Skills: Critical Thinking, Communication, Team Collaboration.

🎓 Education

Explore AI Academy
IT System Development – NQF Level 5 Jan 2023 – Dec 2023

- Python, Power BI, SQL, Supervised/Unsupervised ML, Regression/Classification, AWS Cloud Foundations

Vukuzame High School
Matric (2019)

- Computer Applications Technology, Mathematics, Physical Sciences, Life Science, Home Language, English.

Achievements/Certification

Data Science Certificate

Awarded - 07 December 2023

📅 Experience

Data Science Analyst — University of the Witwatersrand (TISL)
Dec 2024 – Present

- Built a full-stack **Telkom Executive Dashboard** using Flask, HTML/CSS/JS for displaying business KPIs.

- Integrated **Google Gemini API** chatbot with dynamic prompt injection and response configuration.

- Developed a **CSV insights platform** for real-time upload, parsing, and Chart.js visualizations.

- Built a **PDF Research Assistant Web App** (FastAPI + FAISS + React + LLaMA3) for intelligent document QA.

- Designed and deployed **BERTopic NLP pipeline** with UMAP, HDBSCAN, KeyBERT, and LLaMA2 for social media topic clustering and labeling.

- Created a **Streamlit-powered dashboard** to visualize post frequencies (daily/monthly) and trends using Azure-hosted SQL data.

- Performed advanced **Matplotlib-based exploration** for tracking engagement spikes, post profiling, and temporal sentiment patterns.

Junior Data Scientist — Telkom
Jan 2024 – Dec 2024

- Built real-time monitoring dashboards for internal operational metrics.

- Developed machine learning models for trend prediction and classification.

- Conducted data wrangling, preprocessing, and deployed pipelines for large-scale usage.

- Actively contributed to innovation workshops and analytical strategy planning.

Data Scientist Intern — BCX (Business Connexon)
June 2023 – Dec 2023

- Developed and fine-tuned **deep learning NLP models** for speech sentiment analysis.

- Handled large-scale preprocessing and model optimization in a team environment.

- Delivered insights via reports and executive-level visualizations.

🏆 Projects

EmBot - AI Assistant

EmBot is a powerful AI chatbot that allows users to upload PDF documents and ask questions about their content. Built using FastAPI, FAISS, SentenceTransformers, and LLaMA3, EmBot is designed for intelligent document Q&A, with a clean React frontend and modern UX features like dark mode, chat bubbles, PDF preview, and file uploads.

Try EmBot Live: https://embot-ycjn.onrender.com/

🛍️ References

Available on request.