



JOSHUA MUSIRA

Data Scientist & Full Stack Developer

✉ joshuamusira01@gmail.com
☎ +254111779815
🏠 10300 Nairobi
♂ Male
🌐 Portfolio: <http://joshmusira.netlify.app>
🐙 Github: <https://github.com/JoshMusira>

SKILLS

Gen AI

LangChain

Open AI

Fast API

Qdrant Vector store

Python

Machine Learning

Scikit-Learn

Natural Language
Processing

Exploratory Data Analysis

Data Analysis

Data Cleaning

Knowledge Graph

Algorithmic Trading

EDUCATION

Software Engineering: Undergraduate Sep 2021 – Present
Kirinyaga University, Kerugoya

Kenya Certificate of Secondary Education Jan 2017 – Apr 2021
Giraffe Education Centre, Machakos

Kenya Certificate of Primary Education Jan 2008 – Nov 2016
Swaminarayan Academy, Nairobi, Nairobi

EMPLOYMENT

Data Science Intern Mar 2024 – Sep 2024
Sankhyana Consultancy Services Limited, Ndia

I am working on numerous real-world projects focusing on business and market analysis. My portfolio highlights include advanced skills in Gen AI, OpenAI, LangChain, data mining, cleaning, visualization, and analysis, as well as machine learning and AI. I excel in communication, problem-solving, prescriptive analysis, process automation, and data security. My expertise also covers ; Credit Risk Analysis, customer analytics, conjoint analysis, financial modeling, geospatial analysis, topic modeling, and business intelligence, effectively supporting data-driven decision-making and model governance.

Data Science Intern May 2024 – Jun 2024
Intern InforTech, Ndia

I worked on various data analysis and machine learning projects during a virtual 1-month internship. My responsibilities included cleaning and preprocessing data, performing exploratory data analysis, and building predictive models using Python. I worked on four projects: car price prediction, shopper sentiment analysis, cinema tickets, and fake news detection.

Machine learning Intern Feb 2024 – Mar 2024
Menterness, Ndia

During my internship at Menterness, I worked on several machine learning projects aimed at developing and optimizing models for various applications. My tasks included data collection and

Deep Learning

Docker

Data mining

Statistics

Data Visualization

Git & Github

Power BI

Sentiment Analysis

Prescriptive Anlysis

Pandas, Numpy, Seaborn,
Matplotlib and Plotly

Figma

Feature Engineering

Microsoft Azure

MsSQL, MySQL & Mongo DB

Heroku, Netlify & Vercel

Natural Language
Processing

Reactjs(Nextjs)

Nodejs & Expressjs

Upstash Redis

PHP

preprocessing, feature engineering, and model training and evaluation using Python and popular ML library such as scikit-learn. Including deployment on Azure.

Data Science Trainee

Nov 2023 - Apr 2024

Sankhyana Consultancy Services Limited

Learned comprehensive data science skills, including data mining for pattern extraction, data cleaning for quality improvement, data visualization for insights, data analysis for interpretation, and developing machine learning and artificial intelligence models for predictive and intelligent decision-making.

Full Stack Software Developer Trainee

May 2023 - Jul 2023

TheJitu, Nyeri

Trained at TheJitu in SERN stack software development, Microsoft Azure for cloud computing, Docker for containerization, and enhancing communication skills.

COMPLETED PROJECTS

Loan Approval: I developed an extensive loan approval project, for feature selection, I employed Information Value (IV), Weight of Evidence (WOE), and Fisher score. I built a predictive model to minimize credit risk and created a Lift table to analyze performance. The model predicts credit scores and determines loan approval, review, or rejection based on a probability threshold. To complete the project, I built a user-friendly interface using ReactJS and developed a RESTful API with FastAPI for backend integration.

RAG Fullstack APP: I built a comprehensive and an innovative Retrieval-Augmented Generation (RAG) application using OpenAI, Qdrant vector store, embeddings, FastAPI, React for the UI, NewsAPI, Word Cloud, and Langchain. This project dynamically reads and performs Quation and Answer(QA) on over 500 news articles stored in our vector store, focusing on specific individuals and from different news sources.

Mdabujobs: Mdabujobs is a comprehensive job portal designed to connect job seekers with potential employers. Built as a full-stack project, Mdabujobs utilizes React for the front end, TanStack Query for data fetching and state management, Node.js and Express for the back end, and MySQL for the database.

Monte Carlo Simulation: The Monte Carlo Simulation project focuses on European option pricing using advanced probabilistic and statistical techniques. This project implements the Monte Carlo method to simulate the underlying asset paths and calculate the option price. Additionally, the Black-Scholes algorithm is used to provide a benchmark for the option pricing.

LSTM-neural-network-model: The LSTM-neural-network-model project focuses on stock price analysis and prediction using advanced machine learning techniques. The project involves pulling stock data from Yahoo Finance using the yfinance library, performing Long Short-Term Memory (LSTM) analysis to model the temporal dependencies in the data, and predicting future market prices.

Prisma ORM

Redux Toolkit

Stripe and Daraja APIs

Microservices

Topic Modeling

TypeScript

Problem-Solving

problem-solving skills

Business Intelligent

Collaboration

Kubernetes

LANGUAGES

English

Swahili

HOBBIES

■ Playing Chess

■ Swimming

■ Watching documentaries

■ Hiking

■ Football

QUALITIES

■ Versatility: Adaptability to embrace and quickly learn new technologies and methodologies.

■ Problem-Solving Aptitude: Strong problem-solving skills, capable of addressing

Company DashBoard : I began by collecting data from MySQL, performing ETL (Extract, Transform, Load) processes to cleanse and prepare the data. I then built a comprehensive schema to organize the data effectively. Using Power BI, I created an insightful dashboard that visualizes key metrics and trends. Finally, I published the dashboard for stakeholder access, enabling data-driven decision making.

Market movement on Sentiments : I conducted a unique project examining the correlation between stock market movements and social media sentiment. Using Yahoo Finance, I pulled stock data and collected tweets related to the stocks. I then merged this data to analyze how social media sentiment influences market movements, providing valuable insights into the impact of reviews on stock prices

ACHIEVEMENTS

Data Scientist Achievements:

- Conducted impactful data analysis, extracting actionable insights from large datasets.
- Developed predictive models for applications like customer churn and sales forecasting.
- Created compelling data visualizations, demonstrating business impact through data-driven decision support.
- Leveraged machine learning algorithms to improve data modeling results, resulting in a 25% increase in target user engagement.
- Redesigned existing data models in order to achieve a 10% increase in accuracy and a 5% cost savings.

Full-Stack Developer Achievements:

1. Developed and optimized dynamic, high-performance web applications using React and Next.js, improving user engagement and load times.
2. Implemented complex features in TypeScript and JavaScript, enhancing code maintainability and reducing runtime errors across projects.
3. Successfully deployed and managed scalable applications on Azure, leveraging Docker for containerization, resulting in streamlined deployment processes.
4. Engineered robust data storage solutions with MSSQL, MySQL, and MongoDB, ensuring high availability and efficient data retrieval in large-scale applications.
5. Automated application deployment and scaling using Kubernetes, leading to improved resource management and reduced downtime in production environments.

REFERENCES

Pavitra Biswas

Sankhyana Consultancies limited, Bengaluru
+919560450026, pabitra.stat@gmail.com

complex technical challenges effectively.

- Team Collaboration: Excellent collaboration abilities, fostering teamwork and communication in cross-functional environments.
- Coding Proficiency: Proficient coding skills, ensuring high-quality development in both front-end and back-end domains.
- Project Management: Successful project management, ensuring timely delivery and meeting high standards.

Tapaswani Nayak
Sankhyana Consultancies limited, Bengaluru
+9182174 26430, tapaswani.nayak@sankhyana.com

Mentorless, India
+916351567156,, mentorless.help@gmail.com

Kelvin Comba
TheJitu, Nyeri
0715364294, kelvin.gatimu@thejitu.com

Nairobi, September 2, 2024

Josh Musira