

Sharon Muiruri

sharnmuiruri44@gmail.com • (254) 745728616 • [swariara.github.io](https://github.com/swariara) • linkedin.com/in/sharon-muiruri

EDUCATION

Massachusetts Institute of Technology

2022 - 2023

Data Science and Machine Learning Certificate

Jomo Kenyatta University of Agriculture and Technology

2018 - 2022

Bachelor of Business Information Technology

EXPERIENCE

Machine Learning Intern

Sep 2023 - present

Codsoft

- Engineered data pipelines and developed predictive machine learning models to address business challenges, resulting in a 15% improvement in decision-making efficiency.
- Built and deployed a customer churn prediction model, achieving a 20% increase in customer retention rates through targeted interventions.
- Developed a fraud detection classification model and enhanced its performance through feature engineering, leading to a 15% improvement in model accuracy.

Data Science Intern

Sep 2022 - Dec 2022

Flapmax AI Institute

- Implemented and customized a Commerce Marketplace SaaS Accelerator on Microsoft Azure, enhancing system scalability by 15% and improving key performance metrics such as response time and transaction throughput.
- Deployed two FastAPI applications using Kubernetes and Docker, improving containerization and orchestration efficiency by 30%.
- Collaborated with cross-functional teams to develop a car damage detection model using Mask R-CNN, achieving 90% accuracy in detecting and segmenting damaged areas.

PROJECTS

Stock Market Prediction

- Created a model to forecast stock market prices for three companies.
- Scraped stock data from the yfinance api library and conducted EDA to explore trends, volatility, and relationships between stock prices.
- Applied feature engineering with lag variables to capture dependencies in stock prices over different time periods.
- Built and evaluated three models: ARIMA, SARIMA, and LSTM, with LSTM achieving the best performance, showing the lowest RMSE score.

Credit Card Default Prediction

- Developed a predictive model to identify customers likely to default on credit card payments.
- Utilised Python libraries, including scikit-learn and pandas, for data cleaning, analysis, and preprocessing, and applied SMOTE to balance the dataset.
- Built and evaluated five classification models: Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and XGBoost, with Random Forest achieving the highest ROC AUC score.

TECHNICAL SKILLS

Programming & Databases: Python, SQL, MongoDB

Machine Learning: Regression, Classification, Clustering, NLP, Neural Networks

Libraries/ Frameworks: Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn, Tableau, TensorFlow, PyTorch, Keras, Apache Spark, Hadoop, Microsoft Azure, Docker, Kubernetes