

# JOHN RUORO

Data Scientist Intern Singapore

+254706350181



ruorojohn@gmail.com

[linkedin.com/in/john-ruoro-3769b614b/](https://www.linkedin.com/in/john-ruoro-3769b614b/)



Nairobi, Kenya

---

## EXPERIENCE

---

### Data Scientist Intern Singapore

#### Data Glacier

02/2021-04/2022

Collaborated with several students to create visually impactful dashboards using Tableau, PowerBI and other visualization tools such as D3.js which helped in transforming raw data into meaningful information.

- Developed meaningful connections with other data professionals around the world which gave me perspective on what kind of data roles I should consider.
- Contributed to the Data Glacier's online community by providing datasets obtained through scraping the web to give future interns more data to practice on. Skills and technologies used: Data mining, Python scripting, SQL, R, AWS ecosystem.

### Data Scientist Intern Nairobi, Kenya

#### The Sparks Foundation

02/2021-04/2022

Contributed in statistical modelling projects to test students churn rate for an online course. Used Machine Learning Algorithms such as KNN, XG Boost, SVM, and Decision trees to determine the possibility of current students dropping out.

- Developed a personal project webapp with Streamlit to generate a word cloud from text which I have used to find the keywords in job posts to aid with the tedious job search process. The Streamlit app was deployed and shared with my friends giving them an edge when writing resumes.
- Collaborated with other data scientists in the company to work on shared company projects. The projects helped me learn collaboration and organization using Agile workflow in Jira. Skills and technologies used: Agile Lifecycle, Streamlit, Heroku, NLP, Kanban.

### Junior Data Analyst

#### Jomo Kenyatta University Nairobi

06/2020-11/2020

Developed visually impactful dashboards for the finance department using Excel and Tableau

- Collaborated with Software and Web Engineers to create data pipelines for the university's website reducing data loss by 40%

---

## SUMMARY

---

Dedicated data professional with 1.5 years of industry experience. \* Working experience and extensive knowledge in Python with libraries such as sklearn, TensorFlow, Numpy, Pandas, Matplotlib, cv2. \* Developed recommendation system for an e-commerce site using surprise library which boosted sales by 24% each month. \* Possesses experience in computer vision by creating an OCR model to copy text from YouTube coding videos. \* Can work within a team set-up owing to great communication and agile work experience

---

## SKILLS

---

Agile AWS Hadoop Computer

Vision Data Mining E-commerce

Github Heroku Kanban LinkedIn

NLP NoSQL Numpy Postgres

Python Scripting R Statistical

Modelling

---

## **PROJECTS**

---

### **DEEFAKE AUDIO RECOGNITION**

#### **Overview**

Deepfake audio recognition using machine learning involves training a model to distinguish between real and fake audio. Some of the techniques used for this purpose include analyzing the audio's frequency and spectral characteristics, as well as identifying artifacts or inconsistencies in the audio that may indicate that it has been tampered with.

**Technology and Libraries:** Python, Tensorflow, Hugging face, Librosa

### **E-COMMERCE RECOMMENDATION SYSTEM**

#### **Overview**

A recommendation system for an ecommerce store uses machine learning algorithms to suggest products to customers based on their browsing and purchasing history, as well as the behavior of similar customers. There are several types of recommendation systems, including collaborative filtering, content-based filtering, and hybrid recommendation.

**Technologies and Libraries:** Django, Surprise, LightFm, TensorRec, RecommenderLab

#### **REFERENCES.**

ASHA DEEN. Moringa School TM.  
+254791222153

BEATRICE KIPLAGAT. Jomo Kenyatta University (Data manager).  
+254724917773

FESTUS ORINA Ministry of Interior (Chief of Operations)  
+254797168356

#### **OTHER PROJECTS**

All these projects are available on my GitHub page [here](#).

- Computer vision, Paddle OCR for extracting text from Images.
- Data Visualization using R
- Tic-tac-toe with AI agent (Utilized reinforcement learning)
- Tik-Tok Analytics for Business Intelligence and Marketing Agencies.
- Fake news classifier and the bias of news (Left or Right Wing.)
- Projects on AWS, terraform, Hadoop, and Apache spark.