# Virginia Ngami Paul

*Data Scientist*

ngamivirginia2@gmail.com +254 768645387

Results-driven Data Scientist with a passion for leveraging machine learning (ML) algorithms, data analytics, and AI-driven solutions to drive business outcomes. Experienced in building scalable recommendation systems, fraud detection models, and predictive analytics for various industries, including e-commerce and finance. Adept at using advanced tools such as Python, TensorFlow, Scikit-learn, and SQL to extract meaningful insights from complex datasets and create impactful data solutions. Strong communicator with a track record of collaborating with cross-functional teams to deliver data-driven solutions.

## Experience

### Part-Time Lecturer (AI and Machine Learning)

**Murang'a University of Technology** *2024 - Present*

Delivered lectures and workshops on **machine learning**, **AI algorithms**, **deep learning**, and **NLP** to undergrad and students.

Developed **curriculum** and **practical lab exercises** focused on AI applications in business and data science.

Mentored students on **AI project development**, guiding model design, training, and optimization.

Led seminars on **AI trends**, including **reinforcement learning**, **AI ethics**, and industry advancements.

Coordinated student projects, preparing them for careers in AI and data science with industry-relevant skills.

Assessed student performance through assignments, exams, and hands-on projects.

### Freelance Data Scientist (Remote) – Remotasks *2023 – 2024*

Designed and developed a **product recommendation system** for an e-commerce company using collaborative filtering, enhancing product sales by 15% within six months.

Built **fraud detection models** for a financial institution using Random Forest and XGBoost, achieving an accuracy rate of 95%.

Created **predictive models for sales forecasting** using time series analysis and regression algorithms, improving forecasting accuracy by 20%.

Led a project to **segment customers for targeted marketing** using K-means clustering, enabling the marketing team to optimize campaigns and improve conversion rates.

Utilized Python and SQL to preprocess, clean, and analyze large datasets, providing actionable insights to key stakeholders.

### AI Model Specialist (Remote) - Outlier *2022 - 2023*

Led the **end-to-end development and training** of AI models for NLP-based applications, including building and fine-tuning **chatbots** and **AI virtual assistants** to improve customer interaction, reducing resolution time by 30%.

Developed **predictive maintenance AI models** using sensor data for manufacturing environments, enhancing system reliability and reducing downtime by 25%.

Collaborated with cross-functional teams to implement **automated AI training pipelines** that improved model training efficiency and consistency.

Designed and executed **model optimization strategies**, including hyper parameter tuning, model ensembling, and feature engineering, to ensure high-quality outputs.

## Education

**Masters Computer Science** *Ongoing*

**Bachelor's Degree in Computer Science** *2018 - 2022*

- Second Class Honors

**Coursework**: Machine Learning, Data Structures, Algorithms, AI, Data Visualization, Database

## Projects

*Product Recommendation System*

- **Technologies Used**: Python, Pandas, Scikit-learn, Flask
- Developed a personalized product recommendation engine for an e-commerce platform using collaborative filtering.
- Utilized customer purchase history and product details to suggest relevant products, resulting in a 10% increase in sales.

*Customer Segmentation for Marketing Optimization*

- **Technologies Used**: Python, K-Means Clustering, Matplotlib, Seaborn
- Built a customer segmentation model using clustering algorithms to help businesses target marketing efforts more effectively.

- Visualized customer segments and provided actionable insights, leading to a 12% increase in marketing campaign ROI.

*Fraud Detection Model*

- **Technologies Used**: Python, Scikit-learn, XGBoost
- Developed a machine learning model to identify fraudulent transactions in real-time, achieving an 85% precision rate.
- The model was deployed in a production environment, helping the company mitigate financial losses from fraudulent activities.

*Sales Forecasting System*

- **Technologies Used**: Python, Time Series Analysis, ARIMA, Prophet
- Developed a sales forecasting system to predict future sales trends using historical data and time series models.
- The system improved inventory management and reduced overstock by 18%.

## Skills

**Programming Languages**: Python, R, SQL, JavaScript

**Data Science & Machine Learning**: Pandas, NumPy, Scikit-learn, TensorFlow, Keras, XGBoost

**Data Visualization**: Matplotlib, Seaborn, Plotly

**Web Development**: Flask, Django, and HTML/CSS (for data-driven web apps)

**Tools & Databases**: Git, Jupyter Notebooks, MySQL, PostgreSQL, AWS

**Big Data Technologies**: Hadoop, Spark (Optional)

**Soft Skills**: Problem-solving, Critical thinking, Communication, Teamwork, Leadership

## INTERESTS

AI Research and Innovation

Data-Driven Business Solutions

Data Ethics and Privacy

Open-Source Contributions