Langat **Vincent**

DATA ANALYST ‖ DATA SCIENTIST ‖ ML/AI ENGINEER

☐ (+254) 712230453 | 🌐 Portfolio | ✉ vincentl@aims.ac.za | 💬 Vinylango25 | 💼 kipkemoi-vincent-19307a94/

Summary

A highly skilled Data Scientist with over 5 years experience in machine learning, data analytics, and predictive modeling. I specialize in anomaly & fraud detection, credit scoring, and developing scalable machine Learning models for capital markets. With a strong background in Python, SQL, and data visualization, I excel at transforming complex datasets into actionable insights across diverse sectors, including climate, fintech, healthcare, and agri-tech. Passionate about leveraging advanced analytics to optimize decision-making, minimize risks, and deliver data-driven solutions for business growth.

Expertise

- **Technical Stack:** Proficient in Python, R, SQL, LaTeX, Git, Jupyter Notebook, VS Code, SageMath, Gretl, DBeaver, PostgreSQL, and Python Django for data analysis, modeling, and software development.
- **Data Visualization & BI:** Expertise in designing automated dashboards, visualizations, and reports using Tableau, Preset, Matplotlib, Seaborn, Plotly, and Bokeh for effective communication of business insights.
- **Machine Learning & Predictive Analytics:** Experienced in building ML models (RF, XGBoost, CatBoost, LGBM, PyOD models, e.t.c) and time series forecasting (ARIMA, LSTM, fbprophet) using frameworks like Keras, Pytorch, TensorFlow and scikit-learn for classification, regression and forecasting.
- **Data Engineering & Cloud Computing:** Skilled in ETL processes, data pipeline automation, and cloud platforms (AWS, Azure, GCP) for scalable storage and processing using services like S3, EC2, and Google BigQuery.
- **MLOps, Deployment, & Project Management:** Experienced in managing projects, version control with Git, and CI/CD workflows for streamlined deployment. Proficient in productionizing ML models with Streamlit, Docker, Kubernetes, MLflow, and Databricks, ensuring scalability and monitoring.

Experience

Prospect 33 *New York, Remote*

DATA SCIENTIST (R&D) *Feb. 2025–Date*

- Responsible for researching, developing and deploying a Semi-Supervised Active Learning based Anomaly Detection ML models to accurately detect and remediate anomalies in financial and operational datasets.

Key Achievements:

- Developed and deployed an active-learning-based Isolation Forest model that incorporate domain-specific client feedback and expert insights to iteratively improve accuracy and reduce false positives.
- Co-developed DIVA and LEAP, flagship fraud and anomaly detection platforms at Prospect 33 for capital markets. Built scalable, adaptive models requiring minimal expert input, improving detection accuracy, reducing false positives, and accelerating time-to-value in high-stakes trading and compliance contexts.
- Strengthened backend and frontend capabilities within the LEAP application by enhancing API design and system architecture using Python, Django, and Django REST Framework—contributing to improved platform performance, scalability, and developer efficiency.
- Developed an explainability framework integrating SHAP, LIME, and gradient-based methods within the LEAP platform, enabling clear, localized insights into anomaly detection decisions—boosting stakeholder trust, model transparency, and regulatory alignment.
- Engaged closely with stakeholders to gather and prioritize product requirements, ensuring alignment between business needs and development efforts—driving targeted feature delivery and maximizing product impact.

JUNIOR DATA SCIENTIST (Global Data Lab) *Jun. 2024-jan. 2025*

- Researched and developed advanced anomaly detection techniques for financial and credit data, integrating explainability frameworks to enhance model transparency and interpretability. link

Twino (Moneza Kenya) *Westlands, Nairobi, Kenya*

DATA SCIENTIST *May. 2023 - Jan. 2025*

Responsible for developing and managing loan portfolio credit scoring models for both pay day loans (PDL) and buy now pay later (BNPL) products, analyzing trends and risks, generating actionable reports, implementing risk rules, and collaborating with teams to ensure data accuracy and minimize fraud.

Achievements:

- Developed and deployed credit scoring (XGBoost) and fraud detection (CatBoost) models using TransUnion and SEON data, significantly improving loan approval accuracy and strengthening fraud prevention measures.
- Created an automated loan default dashboard tracking daily, weekly, and monthly default rates, enabling proactive risk monitoring and delivering actionable insights that informed continuous improvements in credit risk models.
- Automated payment and debt collection reporting for back-office teams using PostgreSQL, streamlining operations and enabling data-driven decision-making to improve collection outcomes and reduce processing time.
- Gained in-depth expertise in SQL querying using DBeaver and PostgreSQL to extract and analyze data. Integrated insights into Preset for automated dashboards, enhancing reporting efficiency and enabling real-time, data-driven decision-making.
- Led and mentored a team of junior analysts, fostering skill development and driving the successful delivery of data-driven solutions in business intelligence, fraud detection, and credit risk analysis—enhancing team performance and accelerating project outcomes.
- Partnered cross-functionally to align technical solutions with business objectives, integrating fraud detection, credit risk, and compliance strategies to drive cohesive, high-impact outcomes.

Mananasi Fibre Ltd *Thika, Kenya*

Data Analyst and Reporting Specialist *July. 2022 − April. 2023*

Leading data analysis and reporting efforts for two climate focused pilot projects funded by the UK's SMEP and DEFRA programs: (1) fibre production from pineapple waste and (2) Accesing viability of carbon credits through biochar production from pineapple biomass. Responsibilities span technical, environmental, social, and financial domains to support data-driven decision-making and donor compliance.

Key Contributions:

- **Data Analysis & Reporting:** Collected, analyzed, and interpreted performance, efficiency, and environmental impact data; developed detailed analytical and narrative reports for stakeholders.
- **Scientific Reporting:** Produced technical documentation using statistical methods and data visualization; ensured adherence to scientific reporting standards and data accuracy.
- **Impact Reporting:** Compiled qualitative and quantitative data to demonstrate project impact; developed case studies and success stories highlighting socio-economic benefits.
- **Compliance & Documentation:** Ensured timely and accurate grant reporting in line with SMEP/DEFRA requirements; coordinated data gathering with project stakeholders.
- **Quality Assurance:** Implemented data validation processes to uphold data integrity and accuracy in all reporting deliverables.
- **Energy Consumption Analysis:** Analyzed energy usage to uncover efficiency opportunities and supported sustainability-focused decision-making.
- **Process Improvement:** Identified opportunities for automation and streamlined data workflows; kept abreast of best practices in data analysis and storytelling.

The Cyprus Institute *Nicosia, Cyprus*

Graduate research assistant- data science & air quality *Oct. 2020 - May. 2022*

- The project focused on developing machine learning algorithms to forecast air quality and improve the accuracy of data collected from low-cost air quality sensors as part of the **Horizon Europe funded EMME-CARE project**. The project aimed to enhance air quality monitoring capabilities in **Nicosia** by establishing a network of low-cost sensors. link

Key achievements:

- Designed and implemented a network of low-cost air quality sensors across Nicosia, including traffic, industrial, and background monitoring sites. This strategic setup enabled comprehensive spatial analysis and accurate source apportionment of air pollution, improving the understanding of pollution dynamics across the city.
- Designed and implemented automated data pipelines for real-time data collection, cleaning, and analysis, ensuring continuous monitoring of air quality in Nicosia. This helped in making real-time data available for decision-making and public health purposes.
- Created custom tools and dashboards for advanced real-time air quality data visualization, allowing stakeholders to monitor and analyze pollution levels dynamically, providing crucial insights for decision-making and public health responses.
- Developed and deployed advanced ML calibration techniques for low-cost air quality sensors, improving their measurement accuracy by reducing systematic errors and ensuring more reliable pollutant data in real-world urban environments.
- Conducted spatial and temporal analyses of air pollution in Nicosia, identifying key pollution hotspots and temporal patterns, which provided actionable insights for environmental policy-making and urban planning to mitigate pollution in high-risk areas.
- Developed a set of data quality indicators aligned with international air quality monitoring standards, including the US EPA and EU Directive 2008/50/EC, ensuring that the project complied with regulatory requirements and contributed to global best practices in environmental monitoring.

Finlays Kenya *Kericho, Kenya*

Research and data analyst *Nov. 2018 - Sep. 2020*

Responsible for analyzing sales, customer feedback, and production data to identify trends and optimize marketing strategies, while developing AI models to improve tea production efficiency and forecast sales. This enhanced inventory management and reduced production costs. Delivered actionable insights and reports to stakeholders, driving informed decision-making and business growth.

Education

The Cyprus Institute *Nicosia, Cyprus*

Mphil. in environmental sciences *Oct. 2020 May. 2022*

Thesis: Practicalities in Machine Learning Calibration of Measurements from Low Cost Gas Sensors. link

University of Western Cape | AIMS South Africa *Cape Town, South Africa*

Msc. in mathematical sciences (Data Science) *Aug. 2017 - June. 2018*

Thesis: Application of Machine Learning to High-Resolution Aerial Imagery for Vegetation Mapping. link

The University of Nairobi *Nairobi, Kenya*

Bsc. in mathematics (Statistics and Operations Research) *Sep. 2011 - Sep. 2015*

Grade: First Class Honours

Kenya Certificate of Secondary Education (KCSE)) *2006 - 2009*

Grade: A-

Licences & Certifications

1. Data Science Application in Financial Services: LINK   *Prospect 33 (Global Data Lab)*
2. MLOps with Databricks: LINK   *LinkedIn Learning*
3. Applied Data Science : LINK   *Prospect 33 (Global Data Lab)*
4. Data Science Application in Anti-Financial Crime : LINK   *Prospect 33 (Global Data Lab)*
5. Data & Databases : LINK   *Prospect 33 (Global Data Lab)*
6. ETL & ELT : LINK   *Prospect 33 (Global Data Lab)*

References

1 : Mark Clement: Global Head of AI and R&D, Prospect 33

+44 791 717 2754; mark.clement@prospect33.com

2 : Valtars Bondars • Head of Risk & Data Science, Twino(Moneza Kenya)

+37 120 290 947 ; valtars.bondars@twino.eu

3 : Gilbert Langat • Senior Data Scientist, EZRA,

+254 711 170 904 ; gilbemrlk@gmail.com