# JEROEN VAN LIER

Data Scientist | ML Engineer

linkedin.com/in/jeroencvlier www.jeroencvlier.com jeroencvlier@gmail.com Dutch & South African

As an experienced Full Stack Data Scientist, I bring over four years of hands-on expertise in deploying Machine Learning solutions. At Flash, I led the implementation of MLOps, successfully operationalizing four key projects that significantly improved efficiency across model development, deployment, and monitoring. I have leveraged my seven-year background in geology to solve real-world problems.

## Skills

| | |
|---|---|
| **Programming & Data Analysis** | Python, SQL, SQLAlchemy, Pandas, NumPy, SciPy, Matplotlib, Seaborn |
| **Machine Learning & Big Data** | Scikit-Learn, PyTorch, Keras, XGBoost, PySpark, Ray[RLLIB], MLFlow |
| **Development & Deployment** | Docker, FastAPI, Flask, Streamlit, Taipy, GitHub, Git |
| **Cloud Platforms & Tools** | Azure ML, PythonAnywhere, Render, Weights & Biases, Gradient |

## Experience

### Flash Mobile Vending

Cape Town, South Africa

**Machine Learning Engineer** **Nov 2023 - Present**

- Developed a churn prediction model using XGBoost on Azure ML, improving retention rates by 27% with a balanced accuracy of 93% through effective hyperparameter tuning with MLFlow
- Automated the retraining and deployment processes of machine learning models through pipelines, including the setting up API endpoints for real-time model serving
- Conducted code reviews to uphold best coding practices and implemented comprehensive unit tests

**Data Scientist** **Jun 2022 - Nov 2023**

- Launched an SMS campaign project for customer retention, monitoring results by measuring statistical significance and achieving a €500,000 increase in monthly turnover
- Modeled and integrated clustering to group trader behaviours, significantly enhancing the accuracy of the fraud anomaly detection system by reducing false positives by 60%
- Established and optimized the data science workflow by automating continuous integration and deployment (CI/CD) pipelines using Docker and Azure, streamlining project production

**Fraud Analyst** **Jun 2021 - Jun 2022**

- Implemented a time-series analysis for anomaly detection, reducing fraudulent activity by €20,000 monthly
- Analyzed transactional data to identify and isolate trends, guiding the fraud team's rule-monitoring strategy
- Retrieved data upon request to aid in the investigation of fraudulent activity

### Quant Analytics

Remote

**Full Stack Data Scientist** **Dec 2020 - Feb 2024**

- Built a Proximal Policy Optimization Reinforcement Learning (RL) agent outperforming the SPY index by 46%
- Maintained the data ingestion pipeline into a data lake (AWS) for preprocessing and archiving historical data
- Deployed an RL model for 350+ stocks generating real-time, actionable insights and personalized trade alerts
- Developed an options strategy trading tool targeting earnings announcements, achieving a 93% success rate

## Omdena

Remote

**Machine Learning Engineer** **Sep 2023 - Nov 2023**

- Collaborated with an international team on a project to predict food prices in Nigeria, aiming to assist in famine prevention and enhance resource allocation efficiency
- Contributed to feature engineering and building of pipeline for XGBoost Regressor hyperparameter tuning
- Published in a hosted Streamlit app as an interactive tool for real-time analysis for stakeholder

## Ironhack

Amsterdam, The Netherlands

**Data Analytics Teaching Assistant** **Dec 2020 - Mar 2021**

- Mentored 30 students through boot camp challenges related to statistics, data analysis and machine learning
- Provided project guidance by offering feedback on coding practices and sharing insights into data analytics
- Assisted the lead instructor and facilitated online icebreakers and team events

## MapXact

Harfsen, The Netherlands

**Research Geophysicist** **Jun 2018 - Sep 2020**

- Advised the integration process of artificial intelligence to automate the interpretation of geophysical data
- Led the research department to investigate the limitations of ground-penetration radar (GPR)
- Mentored PhD students from two universities in a collaborated research project in GPR technology

## Professional Development & Certification

| | | | |
|---|---|---|---|
| **Machine Learning DevOps Engineer** | Udacity | Remote | **2024** |
| **Deep Reinforcement Learning** | Udacity | Remote | **2022** |
| **Data Analytics & Machine Learning** | Ironhack | Amsterdam, The Netherlands | **2020** |

## Education

| | | | |
|---|---|---|---|
| **MSc Geology** | Ludwig-Maximilian University | Munich, Germany | **2017** |
| **BSc Earth Science** | Stellenbosch University | Stellenbosch, South Africa | **2011** |

## Projects

- **Forecasting Dengue Fever (2023):** Developed an XGboost regressor model to forecast the spread of Dengue fever using environmental data, highlighting the impact of climate variables on disease transmission
- **ML Pipeline Rental Prediction (2023):** Rolled out an ML pipeline for predicting short-term rental prices leveraging MLflow and Hydra and hosted on W&B for continuous updates and performance tracking
- **Multi-Agent RL (2022):** Utilized a deep deterministic policy gradient (DDPG) algorithm to train two agents for collaborative and competitive play in a tennis environment, learning from shared experience
- **Continuous Control RL (2022):** Implemented a DDPG-based model to control a double-jointed arm for following a moving target
- **Navigation RL (2022):** Constructed a Deep Q-Learning network (DQN) to train an agent to differentiate and collect yellow bananas while avoiding blue ones in a Unity environment
- **Text Generation (2021):** Applied an LSTM network with Keras to replicate the writing style of a Dutch novella, generating contextually relevant sentences
- **Playlist Recommender (2020):** Developed a playlist recommender using Spotify's API and K-means clustering for unsupervised learning, performing nearest-neighbour search to generate playlists tailored to user's taste
- **Seattle Housing Price (2020):** Prediction of housing prices in Seattle, fine-tunning through experiment tracking and comparing multiple regression models and feature engineering techniques