

# SHANE VAN HEERDEN

MACHINE LEARNING ENGINEER



+27 76 812 4903



shaneavh8@gmail.com



[shaneandrewvanheerden](#)



Remote / Cape Town, South Africa



[shanevanheerden.github.io](#)

## Profile

Experienced Machine Learning Engineer driven by a strong passion for developing AI solutions that tackle real-world problems. Extensive expertise in Deep Learning, MLOps, and Natural Language Processing, with a proven track record of extracting valuable insights from data. Holds a PhD in Machine Learning and possesses significant industry experience in leading successful data science projects and delivering practical AI solutions with tangible business **impact**.

## Experience

### Luno

Sep 2021 – Present

Senior Data Scientist (promoted in July 2023)

- Led key company project to develop real-time customer Anti-Money Laundering risk scores. Responsibilities included operationalising risk score models in Databricks using Py-Spark and MLFlow, engaging with Compliance stakeholders to construct risk indicators and advising on model parameter choices. **Impact: Prevented potential regulatory fines up to R100 million.**
- Built declarative feature engineering infrastructure to service all ML models in Databricks. **Impact: 10x reduction in feature creation and operationalisation time as well as improved scalability and data lineage alerting.**
- Led key company project for defining and tracking how customers move through various lifecycle statuses. Responsibilities included designing PySpark-based reverse ETL pipeline and monitoring reliability of upstream data dependencies. **Impact: Critical component of a wider initiative to protecting \$30 million in revenue and generating \$1 million more.**
- Data Lead for CRM migration from Kustomer to Zendesk. Responsibilities included identifying critical data sources, coordinating with Engineers, and advising on new data architecture while considering cost and API rate limit constraints. **Impact: Part of initiative to scale Customer Success operations to handle a customer-base of 100 million customers.**
- Led project to produce accurate inbound customer message forecasts for workforce hiring and scheduling optimisation. Responsibilities included exploring correlation between Bitcoin spending behaviour and messages volumes, and benchmarked predictive performance of ARIMA and Prophet forecasting models. **Impact: 62% improvement in forecast accuracy.**
- Won 2023 Lunaut of the Year in a company of 600+ people.

### Cape AI

Mar 2020 – Aug 2021

Machine Learning Engineer

- Led project for South African job recruitment firm, utilising advanced NLP to match resumes to job specs. Developed production-ready preprocessing pipelines on AWS EC2, stored/query results on Elasticsearch, fine-tuned BERT models, and served results via Streamlit. **Impact: Client repositioned business offering around new AI-based recruitment matching system.**
- Data Science Lead for an internal Cape AI venture aimed at connecting employees for knowledge-sharing opportunities by analysing unstructured text communications. Leveraged state-of-the-art NER techniques, built production-ready data pipelines interfacing with Neo4j database, and deployed and monitored ML model recommending possible connection opportunities. **Impact: Venture secured external funding and remains operational.**
- Led project for South African insurance provider to decrease administrative burden in handling customer queries. Analysed text communications and developed proof-of-concept chatbot using DialogFlow. **Impact: Potential 60% containment rate in support ticket creation.**
- Led project for client's data exchange platform, developing C# and Python-based ML solution to identify sensitive client data fields. Used Logistic Regression classifier and established active learning feedback loop between client's application and Azure cloud. **Impact: 20x speed improvement in sensitive data field categorisation.**

## Experience

### DataProphet Data Science Intern

Jan 2020 – Mar 2020

- Projects involved discovering good operating regions for clients' manufacturing machinery by employing fast Fourier transforms together with variational autoencoders using the Scikit-learn and Tensorflow Python packages.
- 

## Education

### PhD (Data Science) Stellenbosch University

2017 – 2019

- Research focused on the design and development of a Data Mining framework for quantifying and characterising road accident risk using machine learning and implemented as a Python-based solution.

### BEng (Industrial Engineering) Stellenbosch University

2013 – 2016

- Invited to the Golden Key International Honour Society.
  - Won the prize for the best computer-based decision support system in the final year project.
- 

## Skills

### Data Science

Natural Language Processing, Data Analysis & Visualisation, Supervised/Unsupervised/Reinforcement Learning and Computer Vision

### Programming languages

Python, SQL, R, Bash, MATLAB/Octave, C, C#, HTML, VBA, LaTeX and Cypher

### Tools & Databases

Git, Docker, GIS, Neo4j, Elasticsearch and PostgreSQL

### Libraries & Frameworks

Numpy, Pandas, Matplotlib, Plotly, SpaCy, Tensorflow, Pytorch, PySpark, MLFlow, Scikit-learn, Transformers, SHAP, BeautifulSoup, Selenium, Flask, Streamlit, Logging and Pytest

### Platforms

Databricks, Looker, GitLab, AWS, Azure, Jira, Slack, Miro, Google Suite

---

## Academic exposure

### European Conference on Operational Research (EURO) University College Dublin, Ireland

Jun 2019

- Presented research findings at the 30th EURO conference in Dublin, exposed to new ideas in the fields of Operations Research and Machine Learning.

### Deep Learning Indaba

Sep 2018

Stellenbosch University, South Africa

- Attended a week-long meeting of the African Machine Learning community, exposed to teaching, research, exchange, and debate around the state-of-the-art in Machine Learning and Artificial Intelligence.

### Operations Research Society of South Africa Conference South Africa

2016, 2017, 2018 & 2019

- Presented research findings at the 2016, 2017, 2018 & 2019 ORSSA annual conferences.
- 

## References

Contact details available on request.