SHANE VAN HEERDEN

DATA SCIENTIST
INDUSTRIAL ENGINEER

PROFILE

As a curious Industrial Engineer turned Data Scientist, I thrive on work that challenges me to think both logically and creatively when solving complex problems. I have a strong interest in Deep Learning, Operations Research and object-orientated programming, and excel when tasked with discovering insights in structured and unstructured data.

Inspired by practitioners like Jeff Dean and David Silver, my true passion lies in leveraging techniques and algorithms on the forefront of Machine Learning to develop productisable AI that can be a positive force for change in the world.

INFORMATION

Gender	Male
Nationality	South African
ID	9410085066084
Languages	English and Afrikaans

CONTACT DETAILS

- +27 76 812 4903
- shaneavh8@gmail.com
- shaneandrewvanheerden
- Cape Town, South Africa

WORK EXPERIENCE

Luno
Data Scientist (Sep 2021 – Present)

- Led projects for Luno's Customer Service Workforce Team which focused on automating various productivity reports. This involved developing a **PySpark**-based data pipeline within Luno's **Databricks** platform, and exposing the productivity metrics to the relevant Teams using Google's **Looker** BI tool.
- Completed an advanced Data Science training program offered by Google's ML@CapitalG Team where I was exposed to deep learning topics in **Sequence Modelling** and **Reinforcement Learning**.
- Led a project tasked with identifying when the number of inbound customer conversations exceeded the available Customer Service capacity. This involved developing a set of **indicator metrics** and constructing a backlog monitoring dashboard using **Looker**.

Cape AI
Machine Learning Engineer (Mar 2020 – Aug 2021)

Consulting Projects

- Involved in a project for a South African-based job recruitment firm which focused on matching resumes to job specifications by employing **advanced NLP techniques**. This involved developing and hosting multiple production-ready preprocessing pipelines to ingest ~70 000 documents on AWS EC2, storing and querying results from an **Elasticsearch** database, fine-tuning multiple **BERT-based** models for **text classification and NER**, and serving the final results to the client via an interactive **Streamlit dashboard**.
- Led a project for a Netherlands-based electronic supply client which focused on using **customer segmentation** to discover high-value customers, together with establishing customer look-alike profiles. This involved segmenting and clustering customer buying patterns by employing **UMAP and HDBSCAN**, as well as presenting results to the client by means of a **Streamlit dashboard**.
- Led a project for a South African client's product review website advising how to decrease webpage creation time by employing automatic content generation techniques. This involved understanding how to optimize for **SEO objectives** and developing a proof-of-concept pipeline using **Python and Huggingface Transformers**.
- Led a project for a leading South African insurance provider advising on how to decrease the administrative burden in handling customer queries and complaints. This involved analysing **unstructured text communications** and developing a proof-of-concept **chatbot using DialogFlow**.
- Led a project for a South African client's data exchange platform which entailed developing a **C# and Python-based machine learning solution** for automatically identifying sensitive client data fields. This involved employing a **Logistic Regression classifier** and establishing an **active learning feedback loop** between the client's local application and **Azure cloud platform**.

Knowledgemarker

Knowledgemarker is a Cape AI venture which takes the form of a simple browser add-on deployed within an organisation and aims to connect employees for knowledge-sharing opportunities.

- Lead Data Scientist (Jun 2020 – Apr 2021)
- Led the development of Knowledgemarker's Python-based "NLP Engine" which ingests **unstructured text communications** between colleagues and outputs possible opportunities for collaboration on technical topics.
- Responsibilities include breaking down high-level user stories into actionable development tickets, leveraging state-of-the-art **Natural Language Processing research**, and ensuring data pipelines are production-ready.

Data Prophet
Data Science Intern (Jan – 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, Industrial Engineering)

Stellenbosch University (2017 – 2019)

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

BEng (Industrial Engineering)

Stellenbosch University (2013 – 2016)

- Undergraduate average: 73%.
- Invited to the Golden Key International Honour Society.
- Won the prize for the best computer-based decision support system in the final year project.

National Senior Certificate

Crawford College North Coast (2008 — 2012)

- National Senior Certificate average: 88% with 6 distinctions.

PROFESSIONAL SKILLS

Data Science	Data Analysis & Visualisation, Supervised/Unsupervised/Reinforcement Learning, Natural Language Processing and Computer Vision
Programming languages	Python, R, Bash, MATLAB/Octave, C, C#, HTML, VBA, LaTeX, Cypher and SQL
Tools & Databases	Git, Docker, GIS, Neo4j and PostgreSQL
Libraries & Frameworks	Numpy, Pandas, Matplotlib, Plotly, SpaCy, Tensorflow, Pytorch, PySpark, Scikit-learn, Transformers, SHAP, BeautifulSoup, Selenium, Dash, Flask, Streamlit, Logging and Pytest
Platforms	Google Suite, Microsoft Office, Github, Gitlab, Jira, Slack, AWS and Azure
Methodologies	Agile, Scrum, Operations Research and Simulation

ACADEMIC EXPOSURE

European Conference on Operational Research (EURO)

University College Dublin (Jun 2019)

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

Deep Learning Indaba

Stellenbosch University (Sep 2018)

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

South African Journal of Industrial Engineering

2018 Article Publication

- **van Heerden, S.A., & van Vuuren, J.H.** (2018). *Optimisation of stock keeping unit placement in a retail distribution centre.* South African Journal of Industrial Engineering. 29(2), 74-91.

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

Mr Adit Mehta

Cape AI

- Lead Data Scientist

Mr Wasim Lorgot

Data Prophet

- Senior Data Scientist

Dr Stephan Nel

Stellenbosch University

- Character reference