

# Thabang L. Mashinini-Sekgoto

☎ (+27) 79-5399-012 | ✉ lukhetobano@gmail.com | 🏠 leparalamapara.github.io/myPortfolio/ | 📱 thabang-mashinini-0081b5b6

## Summary

Innovative data science leader with over 10 years experience in building and scaling AI organisations, setting technical standards, and guiding teams through modernisation, MLOps adoption, and cloud-native transformation specialising in distributed machine learning, probabilistic modelling, and large-scale data engineering. I design and operationalise end-to-end ML systems—from streaming architectures to model governance frameworks—ensuring reliability, scalability, and measurable business impact.

With experience across telecoms, banking, education, and research, I've built mission-critical platforms including Vodacom's Smart Generator Optimization engine, IBM's Environmental Intelligence Suite components, and ABSA's Databricks-centred ML transformation. My technical leadership includes introducing modern MLOps practices, architecting cross-cloud ML workflows, and stabilising production scoring engines used across the enterprise.

As founder of the Ubunye AI Ecosystem and author of Tfilterspy, I build tools that empower teams, standardise ML workflows, and accelerate innovation. I bring deep technical authority combined with the leadership needed to guide data science organisations through modernisation and scale

## Education

### University of the Witwatersrand

Johannesburg, South Africa

PHD IN COMPUTER SCIENCE (IN PROGRESS)

2024 – Present

- Research focus on advancing **self-supervised learning**, **probabilistic modelling**, and **distributed ML systems** to reduce dependency on labelled datasets and enable scalable AI adoption in real-world, low-resource environments.
- Application domains include **environmental risk modelling**, **sensor fusion**, and **real-time decisioning systems**.

### University of the Witwatersrand

Johannesburg, South Africa

MSC IN COMPUTER SCIENCE (DISTINCTION)

2018 – 2019

- 📄 **Thesis: Learning Level Set Method by Echo State Network for Image Segmentation**
- Developed a novel recurrent neural network approach using **Echo State Networks (ESNs)** for iterative image segmentation, enabling computationally efficient training and improved boundary detection.
- Research demonstrated potential applications in **medical imaging**, **remote sensing**, and **autonomous navigation** through enhanced segmentation performance.

### University of the Witwatersrand

Johannesburg, South Africa

BSC HONOURS IN COMPUTER SCIENCE

2017

- **Project: Wildfire Estimation Using Kernel Density Estimators**
- Applied Spatial KDE and unsupervised learning techniques to model wildfire occurrence patterns across South Africa, contributing insights for **disaster management** and **early-warning systems**.

### University of the Witwatersrand

Johannesburg, South Africa

BSC IN COMPUTATIONAL & APPLIED MATHEMATICS AND ASTRONOMY & ASTROPHYSICS

2014 – 2016

- Completed dual majors with emphasis on **mathematical modelling**, **numerical methods**, and **astrophysical simulation**, forming the foundation for later work in machine learning and probabilistic systems.

## Experience

## ABSA Insurance

Johannesburg, South Africa

LEAD DATA SCIENTIST (ACTING HEAD OF DATA SCIENCE)

Mar. 2024 – Present

- Built and led the **Data Science and Engineering function** from the ground up, setting strategy, operating models, and delivery standards for a team of **3 Lead Data Scientists, 1 Data Scientist and 1 Data Engineer** to advance data-driven risk mitigation across short-term insurance.
- Advise senior leadership on **AI strategy, roadmap, and capability scaling**, aligning technical delivery with business outcomes and long-term organisational goals.
- Architected and operationalized **enterprise ML pipelines on Databricks and AWS**, cutting model training/scoring latency by **80–90%** and enabling reliable real-time Telematics and Risk scoring.
- Designed and deployed **geospatial flood, lightning, and storm peril models** for property insurance significantly improving underwriting accuracy and geographic risk visibility by **80%**.
- Introduced **AI governance, monitoring, and model versioning**, strengthening reliability, regulatory compliance, and production ML performance.
- Stabilised and modernised the **Telematics driving-behaviour scoring ecosystem** used across vehicle finance and insurance products, eliminating multi-month backlogs, improving data quality, and enabling proactive digital engagement.
- Introduced enterprise **ETL/ML standards** through the [🔗 Ubunye-AI-Ecosystems \(UAIE\)](#), my open-source initiative, establishing consistent data pipelines and accelerating model development across the organisation by creating reusable patterns and supporting capability uplift.

## Vodacom

Johannesburg, South Africa

SENIOR DATA SCIENTIST (MLOPS, STREAMING AI, AND OPTIMISATION)

Nov. 2021 – Mar. 2024

- Developed and deployed **machine learning solutions** to optimise network and operational performance across Vodacom's national infrastructure, spanning predictive maintenance, anomaly detection, and capacity planning.
- Provided technical leadership across data science, engineering, and IoT teams—driving platform adoption, defining delivery standards, and mentoring **8 Data Scientists** and **2 ML Engineers**.
- Partnered with senior leadership to align AI initiatives with operational KPIs, embedding data-driven optimisation into Vodacom's national resilience strategy.
- Architected Vodacom's **real-time ML and optimisation platform** for network operations, integrating streaming analytics, optimisation models, and IoT telemetry to enhance maintenance decisions and strengthen network reliability during load-shedding.
- Developed a [🔗 smart-generator-monitoring optimization engine](#) for national smart-generator allocation, improving fuel efficiency, maintenance scheduling, and site-level prioritisation—contributing to an estimated **R1 billion in annual OPEX savings** through reduced generator expenditure and improved network uptime.
- Designed scalable **MLOps and streaming pipelines** using Apache Flink, PySpark, Kafka, and Kubernetes, enabling reliable real-time analytics and reducing model deployment cycles by **40%**.
- Recognized with the **Vodacom Star Award** for driving innovation and advancing the company's big data and AI capabilities.

## IBM Research

Johannesburg, South Africa

MACHINE LEARNING RESEARCH SCIENTIST

Apr. 2020 – Nov. 2021

- Collaborated with global IBM teams to design and operationalise **large-scale geospatial and meteorological analytics**, contributing to enterprise solutions for climate, environmental intelligence, and disaster resilience.
- Developed **state-of-the-art predictive models** using **TensorFlow**, improving seasonal climate forecast accuracy by **15%** and enabling more reliable decision-making for enterprise clients.
- Integrated advanced ML workflows into the [🔗 IBM PAIRS Geospatial Platform](#), enhancing real-time geospatial data access for strategic planning across multiple industries.
- Developed and deployed the [🔗 COVID-19 Analytics Dashboard](#) for Gauteng public health authorities, supporting operational response with real-time epidemiological insights.
- Led a cohort of **3 PhD and 2 MSc research interns**, fostering technical excellence, research rigor, and collaborative innovation across multiple machine learning projects.
- Served as team lead for an international cross-functional group, aligning research objectives, coordinating development across regions, and ensuring high-impact project delivery.
- Co-authored and published peer-reviewed research at [🔗 NeurIPS](#), advancing state-of-the-art machine learning methods and strengthening IBM's global research presence.

## Business Intelligence Service (University of the Witwatersrand)

JHB, South Africa

DATA SCIENTIST

Jun. 2018 - Apr. 2020

- Contributed to a central analytics division that integrated and operationalised institutional data, providing leadership with insights for strategic planning, performance management, and university-wide decision-making.
- Collaborated with BI Analysts and Engineers to design and operationalise **end-to-end machine learning pipelines**, improving model accuracy and streamlining analytics workflows that supported the university's student success strategy.
- Developed a **course recommendation system** using K-modes clustering, achieving a **90% silhouette score** and generating over **R2 million annually** of government subsidy through improved enrolment and pass rates alignment. [🔗 Git repository](#)
- Introduced and implemented feature engineering and machine learning model training pipelines.
- Led and facilitated over **50 workshops**, training staff and students in advanced data science and engineering practices, significantly enhancing institutional analytics capability and fostering a data-driven culture.

## CSIR

Pretoria, South Africa

### DATA SCIENTIST AND SOFTWARE ENGINEER INTERN

Nov. 2017 – Jan. 2018

- Collaborated in a team of **3** to design and develop **predictive profiling models using Python and Scikit-learn** for South African municipalities, achieving **75% accuracy** and identifying operational bottlenecks that improved service delivery planning.
- Developed and deployed a **Django-based decision-support web application** for municipal managers in Cape Town, enabling real-time data access and improving operational efficiency.
- Featured for innovation in predictive modelling and digital public-sector solutions by [Mail & Guardian](#), [CSIR](#), and [DST](#).

## ZAR X

Johannesburg, South Africa

### SYSTEMS ANALYST AND SUPPORT

Jun. 2017 – Oct. 2017

- Collaborated with the IT and engineering teams to enhance trading platform reliability, resolving critical system issues and implementing performance improvements that reduced downtime by **20%**.
- Improved reporting accuracy and operational visibility by refining data workflows and strengthening integration across trading and analytics systems.
- Provided technical support to internal stakeholders, ensuring rapid resolution of system-related issues and minimising disruptions to trading operations.

## Wits Astronomy Club (WAC)

Johannesburg, South Africa

### CO-FOUNDER AND EXECUTIVE

Jan. 2014 – Feb. 2017

- Co-founded the [Wits Astronomy Club](#) as a public engagement initiative aligned with the **Square Kilometre Array (SKA)** vision, promoting astronomy awareness among students and the broader community.
- Organised public lectures, stargazing events, and workshops that strengthened collaboration between students, researchers, and industry partners, contributing to science outreach in South Africa.
- Secured partnerships and funding to support club operations, enabling sustained growth, expanded outreach, and increased participation.
- Featured in media interviews with the **SKA** and **SABC NEWS, ETV**, highlighting the club's role in advancing astronomy outreach.  
[Crazy-TV](#) [SABC News](#)

## Skills

### Technical Leadership

AI/ML strategy, Data science capability building, Cross-functional leadership, Technical roadmaps, Mentorship, Agile delivery, Enterprise AI governance.

### Programming

Python, C/C++, MATLAB, R, CUDA, Shell Scripting, SQL, JavaScript.

### Data Engineering

Databricks, Apache Spark, Hadoop, Kafka, Flink, ETL/ELT pipelines, Feature stores, Data lakes, Serverless architectures.

### MLOps & Architecture

Model deployment, CI/CD (Jenkins, GitLab), Docker, Kubernetes, Infrastructure-as-Code (Terraform), MLflow, Monitoring, Versioning, Real-time ML systems.

### Machine Learning & AI

PyTorch, TensorFlow, Probabilistic ML, Distributed ML (Dask, Ray, Spark), Deep Learning, NLP, Computer Vision, Explainable AI, Optimization.

### Software Engineering

Microservices, REST APIs, High-performance computing, Systems design, Git-based workflows, Production-grade ML systems.

### Analytics & Visualization

Power BI, React, HTML/CSS, Dash, Interactive dashboards, Data storytelling.

### Domains & Expertise

Scalable ML Platforms, Real-time Decisioning Systems, AI Innovation, Telecom ML, Insurance Risk Models, Geospatial Intelligence, AI Security & Ethics.

### Languages

English (C2), Sesotho (C2), Setswana (C2), Sepedi (C2), isiZulu (C2)

## Publications

### ML-based Probabilistic Prediction of 2m Temperature and Total Precipitation

MOHAMED AKRAM ZAYTAR, BIANCA ZADROZNY, CAMPBELL WATSON, DANIEL SALLES CIVITARESE, ETIENNE EBEN

2022

VOS, THABANG MICHAEL MATHONSI, [THABANG LUKHETHO MASHININI](#)

- [paper](#)

### Long-Range Seasonal Forecasting of 2m-Temperature with Machine Learning

ETIENNE E. VOS, [THABANG MASHININI](#), ASHLEY GRITZMA, SIBUSISIWE MAKHANYA AND CAMPBELL WATSON

2020

- 34th Conference on Neural Information Processing Systems [NeurIPS 2020].
- [paper](#) [poster](#)

## Mine Worker Threshold Shift Estimation via Optimization Algorithms for Deep Recurrent Neural Network

MILKA C.I.MADAHANA, JOHN E.D.EKORU, [THABANG L.MASHININI](#), OTIS T.C.NYANDORO

2019

- 18th IFAC Symposium on Control, Optimization and Automation in Mining, Mineral and Metal Processing [MMM 2019]
- [🔗 paper](#)

## Noise Level Policy Advising System for Mine Workers

MILKA C.I.MADAHANA, JOHN E.D.EKORU, [THABANG L.MASHININI](#), OTIS T.C.NYANDORO

2019

- 18th IFAC Symposium on Control, Optimization and Automation in Mining, Mineral and Metal Processing [MMM 2019]
- [🔗 paper](#)

## Selected Projects

---

### Smart Generators Optimisation

*Vodacom, South Africa*

SENIOR DATA SCIENTIST

*Nov. 2021 – Mar. 2024*

- Addressed the need for resilient network operations during national loadshedding by developing a **constrained optimisation and graph-search framework** for intelligent generator prioritisation.
- Built real-time data pipelines using **Kafka** (stream ingestion), **PyFlink/Flink** (stream processing), and **PySpark** (batch analytics) to unify IoT telemetry, site metadata, and network alarms.
- Implemented optimisation models using **CVXPY**, operationalised with the **Analytics Engine**, and automated through **GitLab CI/CD** on **Docker + OpenShift**.
- Reduced generator fuel use by **10%**, cut operational costs by **30%**, improved network uptime by **5%**, and contributed to **R1 billion annual OPEX savings**.
- [🔗 smart-generator-monitoring](#)

### Tfilterspy — Bayesian Filtering Python Library

*South Africa*

CREATOR & PRINCIPAL DEVELOPER

*2023 – Present*

- Developed to address real-world challenges with **noisy IoT and sensor data** encountered at Vodacom (generator telemetry) and ABSA Insurance (telematics and risk signals), providing a reliable framework for state estimation and signal smoothing.
- Designed a scalable **Bayesian filtering library** supporting Kalman, Particle, and custom filters, with a **Dask** backend enabling distributed and large-scale time-series computation.
- Packaged and published on **PyPI** with full documentation, examples, and high-performance compute support for robotics, IoT analytics, and probabilistic modelling.
- [🔗 tfilterspy](#)

### Ubunye AI Ecosystem (UAIE)

*South Africa*

FOUNDER & ARCHITECT

*2024 – Present*

- Developed after joining ABSA to address the organisation's lack of **standardised ETL/ML processes**, fragmented capability, and limited engineering resources—creating a unified framework to uplift data science and engineering maturity.
- Designed modular components including **Ubunye Engines** for declarative ETL orchestration and **Ubunye ML** for probabilistic and distributed machine learning.
- Implemented **YAML-driven specifications**, pluggable compute backends (Pandas, Dask, Spark), and reusable transformation and model templates to ensure reproducible and scalable ML workflows.
- Adopted within ABSA to establish **organisation-wide ML standards**, datamarts, and feature stores, significantly improving consistency, reliability, and delivery efficiency across teams.
- [🔗 Ubunye-AI-Ecosystems](#)

### Kasi'Lam Digital Platforms

*South Africa*

CO-FOUNDER

*2022 – Present*

- Founded to enable township businesses to establish a **digital footprint at zero hosting cost**, addressing barriers faced by SMEs that lack access to affordable technology and online infrastructure.
- Developed scalable and user-friendly web platforms using **HTML, CSS, JavaScript**, and **React**, with deployment via **GitHub Pages** to ensure a completely cost-free hosting model.
- Empowered local entrepreneurs to expand their reach, improve customer engagement, and build stronger brand identities through accessible digital solutions.
- [🔗 View Projects](#)

### COVID-19 Visual Analytics Tool

*IBM Research, South Africa*

MACHINE LEARNING RESEARCH SCIENTIST

*Apr. 2020 – Nov. 2021*

- Built to support Gauteng public health decision-making during the pandemic.
- Developed a real-time visual analytics tool using **Python, Dash**, and **Plotly**, deployed on **IBM Cloud**.
- Enabled rapid monitoring of infections, hospital utilisation, and hotspot trends.
- [🔗 View Dashboard](#)

## Geospatial Analytics — IBM Environmental Intelligence Suite

IBM Research (USA, Brazil, South Africa)

MACHINE LEARNING RESEARCH SCIENTIST

Apr. 2020 – Nov. 2021

- Developed to meet the need for scalable environmental and climate-risk monitoring across global industries.
- Built **geospatial analytics workflows** using **IBM PAIRS**, **Python/Tensorflow**, **Hadoop**, and **Apache Airflow** to process and orchestrate multi-terabyte spatiotemporal datasets.
- Contributed to core analytical components powering the **IBM Environmental Intelligence Suite**, enhancing environmental risk prediction and enterprise-scale monitoring capabilities.
- [🔗 Environmental Intelligence Suite](#)

## Wits Course Recommendation System

University of the Witwatersrand, South Africa

JUNIOR DATA SCIENTIST

Jun. 2018 – Apr. 2020

- Addressed the need for personalised course guidance for first-year students.
- Developed a **K-modes recommendation system** using **Python**, **Scikit-learn**, and **Pandas**, achieving **90% silhouette score**.
- Integrated outputs into **PowerBI**, generating **R2M+ annual revenue** through improved enrolment alignment.
- [🔗 View Recommendation Engine](#)

## Honors & Awards

2022	<b>Vodacom Star Award</b> , Recognised for innovation, impact, and engineering excellence in developing the national Smart Generator Optimisation system.	South Africa
2019	<b>MSc Computer Science (Distinction)</b> , Awarded for outstanding academic performance and significant contribution to research in machine learning and big data systems.	University of the Witwatersrand
2014–2017	<b>Wits Astronomy Club — Co-Founder &amp; Executive</b> , Led public STEM engagement initiatives aligned with the SKA project, promoting astronomy outreach across the university and broader community.	South Africa

## References

PhD	<b>Dr. Etienne Vos — IBM Research Manager</b> , Etienne.vos@ibm.com	IBM Research
PhD	<b>Dr. Jacobus du Toit — Machine Learning Lead</b> , jacoDutoit@vodacom.co.za	Vodacom
PhD	<b>Dr. Akram Zaytar — Senior Research Scientist</b> , akramzaytar@microsoft.com	Microsoft
BSc	<b>Mr. Malachy Morrow — Principal Solution Architect</b> , mal@malmorrow.com	Altron & Vodacom
MSc	<b>Mrs Harshila Dulabh — Team Lead</b> , Harshila.Dulabh@wits.ac.za	Wits University (BIS)