Thabang L. **Mashinini-Sekgoto**

Lead Data Scientist (Acting Head of Data Science) · AI & Analytics Leader

(+27) 79-5399-012 | lukhetobano@gmail.com | tmashininisekgoto.com | thabang-mashinini-0081b5b6 | Google Scholar

Summary

AI and data science leader with 10+ years operationalising enterprise ML in banking, insurance, and telecommunications. Currently Acting Head of Data Science at ABSA Insurance, leading a hyperpersonalisation intelligence platform processing 2M+ trips per day across 26,000+ telematics customers and geospatial flood-risk models spanning 230,000+ insured properties; previously led Vodacom's Smart Generator Optimisation platform, credited with approximately R1 billion in annual operational savings across 15,000 sites. NeurIPS-published climate ML researcher (IBM Research) whose doctoral work at Wits physics-informed self-supervised learning for flood extent mapping, extends a decade-long focus on climate and geospatial risk intelligence. Founder of Ubunye, an open-source Spark-native ML/ETL framework published to PyPI. Deep expertise in Databricks, streaming systems, LLM applications, and AI governance in regulated environments.

Education

University of the Witwatersrand — *Johannesburg, South Africa*

PhD in Computer Science (In Progress) — *2024 – Present*

- Research: Physics-informed self-supervised learning for SAR-based flood extent mapping, with applications to insurance risk and data-scarce regions.
- Intersection of remote sensing, self-supervised learning, and computational hydrology.

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.
- Proposed a novel spatiotemporal deep-learning formulation of variational level set segmentation, benchmarking five recurrent and convolutional architectures across four datasets.

University of the Witwatersrand — *Johannesburg, South Africa*

BSc Honours in Computer Science — *2017*

- Project: Wildfire Estimation Using Kernel Density Estimators.

University of the Witwatersrand — *Johannesburg, South Africa*

BSc in Computational & Applied Mathematics and Astronomy — *2014 – 2016*

- Focused on mathematical modelling, astrophysics, simulation, and numerical methods.

Experience

ABSA Insurance — *Johannesburg, South Africa*

Lead Data Scientist (Acting Head of Data Science) — *Mar. 2024 – Present*

- Lead a senior team of **three** (Lead Data Scientist, Specialist Data Scientist, Lead Data Engineer), owning analytical strategy and delivery standards for the Insurance data science function reported to the CDAIO.
- Architected a hyperpersonalisation intelligence platform translating **2M+** daily trips across **26,000+** telematics customers into production retention and growth interventions — churn prediction, driver behaviour deterioration, pre-claim risk signatures, CLV scoring, and cross-sell propensity.
- Modernised the telematics analytics platform on Databricks (Ubunye Engine), cutting data latency from approximately **2 months to under 24 hours** at scale.
- Delivered a production cashback rewards pipeline for the Activate programme — composite driving scoring, Mean Shift clustering, and portfolio-calibrated **5–30%** cashback allocation across **26,000+** policyholders.
- Built a telematics fraud detection suite covering **five** concurrent signal types, driver identity biometrics (autoencoder behavioural fingerprint), and LLM-powered claim narrative cross-referencing against recorded trip data.
- Developed geospatial flood and NatCat risk models (XGBoost, JBA ground truth) across **230,000+** insured properties, informing underwriting across homeowners' comprehensive and building portfolios.
- Standardised AI governance, CI/CD (Databricks Asset Bundle), and Unity Catalog table governance across a portfolio of **19** production analytical models spanning underwriting, retention, fraud, and growth.

Vodacom

Johannesburg, South Africa

Senior Data Scientist — *Nov. 2021 – Mar. 2024*

- Led a team of **10** (9 data scientists and 1 senior ML engineer) building real-time analytics and optimisation systems for national telecommunications infrastructure.
- Built streaming pipelines ingesting real-time telemetry from **15,000+** sites and mobile generators nationally via Kafka, processing **25M+** events daily with PySpark on Kubernetes.
- Led development of the Smart Generator Optimisation platform across **15,000** infrastructure sites nationally, contributing to approximately **R1 billion** in annual operational savings.
- Designed a custom PyFlink stream-processing framework for on-the-fly enrichment of Kafka topics with static reference data, integrating IoT telemetry, real-time alarms, and operational intelligence.
- Established scalable AI and analytics engineering standards adopted across the team.
- Received the Vodacom Star Award for innovation and engineering excellence.

IBM Research

Johannesburg, South Africa

Machine Learning Research Scientist — *Apr. 2020 – Nov. 2021*

- Developed machine learning and geospatial analytics solutions for environmental intelligence and climate-risk applications.
- Built predictive analytics systems using TensorFlow and distributed data platforms.
- Contributed to the Gauteng COVID-19 risk-index and prediction dashboard (IBM Research Africa, Wits University, and GCRO), used by the Gauteng Provincial Department of Health to identify infection hotspots and forecast hospital bed and ventilator demand for a province of **15M+** residents; deployed publicly via IBM Cloud.
- Deployed climate-forecasting models into the IBM PAIRS Geoscope platform, an enterprise geospatial-temporal system handling petabyte-scale data for clients globally.
- Co-authored peer-reviewed research and collaborated with international research and engineering teams across large-scale AI initiatives.

Business Intelligence Services – University of the Witwatersrand

Johannesburg, South Africa

Data Scientist — *Jun. 2018 – Apr. 2020*

- Developed recommendation and analytics systems supporting institutional planning and student success initiatives.
- Built reporting and analytics workflows using Python, SQL, and Power BI.
- Developed a clustering-based recommendation engine serving the Faculty of Humanities (**10,000+** students), generating over **R2 million** annually in government subsidy impact.
- Facilitated **7** analytics and machine learning workshops, training **76+** technical and non-technical staff and students.

Council for Scientific and Industrial Research (CSIR)

Pretoria, South Africa

Data Scientist and Software Engineer — *Nov. 2017 – Jan. 2018*

- Developed predictive analytics and operational intelligence systems supporting municipalities and public-sector decision-making initiatives.
- Built Django-based decision-support systems serving **17** municipalities — including the City of Cape Town and **16** across Gauteng — enabling operational visibility and real-time access to analytics for public-sector stakeholders.
- Applied machine learning and data engineering techniques to identify operational bottlenecks and improve service delivery planning processes.
- Collaborated within multidisciplinary teams combining software engineering, analytics, and public-sector innovation initiatives.
- Recognised for innovation in predictive modelling and enterprise solutions by Mail & Guardian , CSIR , and Department of Science and Technology (DST) .

ZAR X

Johannesburg, South Africa

System Analyst & Support — *Mar. 2017 – Nov. 2017*

- Provided systems analysis and operational support for a newly launched stock exchange serving **13,000+** clients, supporting trading platform stability during go-live and early operations.

Skills

Leadership	AI Strategy, Technical Leadership, Stakeholder Management, Team Mentorship, Agile Delivery
Programming	Python, SQL, R, JavaScript, C/C++, Shell Scripting
AI & Machine Learning	Machine Learning, Predictive Analytics, LLMs & RAG, Agentic AI Systems, NLP, Computer Vision
Data Platforms	Databricks, Apache Spark, Kafka, Flink, Hadoop, ETL/ELT Pipelines
Analytics	Power BI, Dashboarding, Data Visualisation, Business Intelligence, Statistical Analysis
MLOps & Engineering	MLflow, Docker, Kubernetes, CI/CD, GitHub Actions, Real-time ML Systems
Generative AI	LLM Application Development (Claude, Gemini, OpenAI APIs), RAG Architectures, Multi-Model Orchestration, Custom AI Agents & Gateways, AI-Assisted Engineering (Claude Code)
Cloud & Infrastructure	AWS, Kubernetes, Linux, Distributed Systems, Serverless Deployment (Vercel), Cloud-native AI Platforms

Publications

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

NeurIPS 2020 *2020*

- Published research focused on probabilistic forecasting and machine learning methods for climate intelligence systems.

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

EGU 2022 *2022*

- Research focused on machine learning approaches for probabilistic climate forecasting.

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

IFAC MMM 2019 *2019*

- Research focused on optimisation and recurrent neural networks in mining applications.

Selected Projects

AI Knowledge Platform & Virtual AI Mentor

AI Educator *2024 – Present*

- Built and lead a public AI education and technical leadership platform focused on practical AI adoption, career mentorship, and knowledge accessibility.
- Co-host **FabAcademic Unfiltered** (YouTube) with *Prof. Mamokgethi Phakeng* (profile), an AI-literacy series reaching a combined global audience of **300,000+** across platforms.
- Developed technical blogs attracting **12,000+** monthly visits, simplifying complex AI, distributed systems, MLOps, and machine learning engineering concepts for students, professionals, and business audiences.
- Architected and deployed "Thabang AI", a RAG-grounded assistant built on a multi-model architecture (Claude, Gemini, OpenAI) with a custom AI gateway and agent framework, deployed serverlessly to guide aspiring engineers and data scientists.
- Personal AI Platform

Ubunye AI Ecosystems

Founder & Architect *2024 – Present*

- Designed and lead development of an open-source, config-driven, Spark-native AI and ETL framework addressing fragmented ML operationalisation and data engineering workflows.
- Architected a modular distributed framework supporting Spark, Dask, and Pandas backends enabling scalable ML pipelines, reusable ETL workflows, and production-grade analytics systems.
- Introduced standardised engineering patterns for AI governance, reproducibility, model operationalisation, and scalable feature engineering.
- Published to PyPI and used as a dependency in enterprise data-engineering workstreams, reducing ML pipeline prototyping-to-deployment from approximately **2 weeks to under 8 hours**.
- Public documentation and working Databricks examples lowering adoption barriers for the data engineering community.
- Website Documentation

Tfilterspy

Creator & Principal Developer *2023 – Present*

- Developed a scalable Bayesian filtering and probabilistic state-estimation library designed for noisy real-world telemetry, IoT, and time-series environments.
- Implemented distributed Kalman and Particle filtering architectures using Dask to support large-scale streaming and sensor-data processing.
- Built to address operational reliability challenges common to production systems in insurance, telecommunications, and real-time analytics environments.
- Combined mathematical modelling, distributed computing, and production engineering principles into an open-source framework designed for practical enterprise AI applications.
- Website Documentation

Smart Generator Optimisation Platform

Senior Data Scientist *2021 – 2024*

- Led a team of **10** building a national-scale optimisation and operational intelligence platform supporting Vodacom's telecommunications infrastructure during severe load-shedding periods.
- Designed distributed real-time analytics pipelines integrating IoT telemetry, optimisation algorithms, streaming infrastructure, and operational intelligence systems.
- Built scalable AI and optimisation workflows using Apache Flink, Kafka, PySpark, Kubernetes, and graph-based optimisation techniques.
- Enabled intelligent generator allocation, operational prioritisation, and fuel optimisation across **15,000** infrastructure sites nationally.
- Platform contributed to approximately **R1 billion** in annual operational savings while improving infrastructure resilience and operational visibility.
- Vodacom Media Coverage

Kasi'Lam Digital Platforms

Co-Founder *2022 – Present*

- Co-founded digital platform initiatives focused on improving digital accessibility and online presence for township SMEs and local entrepreneurs.
- Designed lightweight, scalable React-based web platforms enabling low-cost digital transformation, with **17** township SMEs on-boarded.
- Focused on combining technology, accessibility, and practical business enablement to reduce barriers to digital adoption.
- Project website

Honors & Awards

2022	**Vodacom Star Award**, Recognised for innovation and engineering excellence	*Vodacom*
2019	**MSc with Distinction**, Outstanding academic and research performance	*University of the Witwatersrand*
2014 – 2017	**Co-Founder – Wits Astronomy Club**, STEM outreach and astronomy public engagement initiatives	*Wits University*

References

Dr.	**Etienne Vos (IBM Research Manager)**, Etienne.vos@ibm.com	*IBM Research*
Dr.	**Jacobus Du Toit (Machine Learning Lead)**, jacoDutoit@vodacom.co.za	*Vodacom*
Dr.	**Akram Zaytar (Senior Research Scientist)**, akramzaytar@microsoft.com	*IBM/Microsoft Research*
Mrs.	**Harshila Dulabh (Team Lead)**, Harshila.Dulabh@wits.ac.za	*Wits University (BIS)*