

Nthimbo Gift Tembo

Nationality: Zambian  Phone number: (+260) 765653492  Email address: nthimbo.tembo@gmail.com

 LinkedIn: <https://www.linkedin.com/in/nthimbo-gift-tembo-93728279/>

 Website: <https://github.com/nthimbo>

ABOUT ME

Senior Data Scientist & Big Data Engineer with 8+ years of experience delivering **predictive models, large-scale distributed data processing, and statistical intelligence solutions**. Expert in **Python, SQL, Apache Spark, Pandas, NumPy, SciPy**, and end-to-end ML development.

Proven track record designing scalable ML systems, processing **millions of records** in Spark, and applying advanced **statistical inference** to extract high-value KPIs from complex datasets. Strong collaborator, skilled at partnering with engineering and product teams to turn raw data into operational intelligence.

CORE SKILLS

Machine Learning & Statistics

- Predictive learning · classification · regression
- Statistical inference · hypothesis testing · Bayesian modelling
- Feature engineering · model selection · large-scale training
- scikit-learn · PySpark MLlib · XGBoost · SciPy · NumPy

Big Data & Distributed Systems

- Apache Spark (advanced) · Spark SQL
- Hadoop · Hive · Dask
- Distributed ETL/ELT pipelines

Programming & Data Engineering

- Python (Advanced), SQL (Advanced), R, JavaScript
- Apache Airflow · Apache Kafka · Databricks
- Data quality automation
- Reusable ETL pipelines · schema design

Deep Learning & NLP

- Transformers: BERT, GPT (Hugging Face)
- TensorFlow · PyTorch
- Address/text normalization · sequence modelling
- Prompt engineering & LLM optimization

Visualization & Tools

- Power BI · Tableau · Superset · Streamlit
- Git/GitHub · Docker · Kubernetes
- Jupyter · Google Colab

WORK EXPERIENCE

Ministry of Health - Zambia – Lusaka, Zambia

City: Lusaka | Country: Zambia

Senior Data Scientist & AI Engineer

[2021 – 09/2025]

- **Predictive Modelling & Statistical Intelligence**

- Built ML models for **patient retention**, **HIV viral load forecasting**, and risk assessment using linear, sequential, and Bayesian frameworks.
- Used statistical inference to extract actionable KPIs supporting national health policy.

- **Big Data Processing (Spark)**

- Designed high-volume processing pipelines handling patients using PySpark & Spark SQL to feed into Apache Superset for Data Visualisations.
- Automated ingestion, transformation, validation, and feature extraction across distributed datasets.

- **Data Visualization & Reporting**

- Built dashboards for surveillance intelligence using Superset, Power BI, and R Shiny.
- Automated reporting for disease indicators, risk maps, and EPR response summaries.
- Produced visually compelling, policy-ready figures and charts for national dissemination.

- **Team Collaboration & Leadership**

- Worked with EPR, SDI, CIO teams, clinicians, laboratories, and informatics experts to enable multi-source data integration.

AskData Italia – Rome, Italy

City: Rome | Country: Italy

Data Science Research Intern

[2018 – 2019]

- Developed ML solutions for **postal address segmentation and normalization** using HMMs and RNNs.
- Contributed to deep learning models that improved address standardization accuracy by **25%**, reducing delivery times and operational costs.
- Created analytical dashboards validating model behaviour and translation consistency.

Huawei Technologies - Zambia – Lusaka, Zambia

City: Lusaka | Country: Zambia

Intelligent Networks Engineer

[2013 – 2016]

- Built analytics dashboards for billing anomalies, reducing revenue leakage by **15%**.
- Performed churn analysis using Python, boosting customer retention by **10%**.
- Automated ETL pipelines for CDR ingestion, cutting manual workload by **30%**.

EDUCATION AND TRAINING

Master of Science

University of Milan Biccoca [2025 – Current]

Field(s) of study: Artificial Intelligence for Science & Technology

Master of Science

Sapienza University of Rome [2016 – 2019]

City: Rome | Country: Italy | Field(s) of study: Data Science | Final grade: 98/110 | Thesis: Application of Sequence Machine Learning Models to Address Segmentation and Standardization.

- Thesis: *Application of Sequence Machine Learning Models to Address Segmentation and Standardization*

- Focused on developing Hidden Markov Models (HMMs) and Recurrent Neural Networks (RNNs) for noisy dataset segmentation, introducing theoretical advances in referential learning and statistical corrector algorithms.

- Key Courses: Algorithmic Methods for Data Mining, Stochastic Processes for Machine Learning, Statistical Methods and Models, Security and Data Privacy.

Bachelor of Science

The University of Zambia - UNZA [2009 – 2013]

City: Lusaka | Country: Zambia | Field(s) of study: Computer Science | Final grade: 28/40 | Thesis: Online Bus Ticket Reservation System

Key Courses: Advanced Database Management Systems, Data Structures and Algorithms, Introduction to Artificial Intelligence.

CERTIFICATIONS

Online Certifications

- **IBM Data Engineering Professional Certificate:** Gained hands-on experience with tools like Apache Hadoop, Spark, Kafka, and Airflow, focusing on data pipeline creation and management.
- **IBM Data Analyst Professional Certificate:** Advanced skills in statistical analysis and data visualization using Python and R.
- **Data Visualization with Tableau Certification (University of California, Davis):** Mastered interactive visualizations and dashboards for decision-making.

FUTURE ASPIRATIONS

Goals and Contributions

- **Research Goals:** Advance research in Artificial Intelligence by exploring scalable machine learning, sequential modeling, and data-driven decision systems for complex real-world problems. Build collaborations with researchers and institutions to drive innovation and ensure ethical, impactful AI applications. My goal is to develop AI solutions that bridge theoretical advances with practical deployment across diverse domains.
- **Academic Community Contribution:** Foster diversity and collaboration through mentorship initiatives, hackathons, and workshops. Promote equitable AI practices and train ethical, responsible data scientists for global challenges.

RECOMMENDATIONS

Available upon request