

# THOMPHO RAMBUDA

Data Scientist

Data Scientist and Researcher with a demonstrated history of working in the research industry. Over the years, I have worked on projects involving the use of natural language processing, machine learning, mathematical optimisation, and mathematical modelling.

## CONTACT

✉ Thomphorams@gmail.com

☎ 0788 559 873

🔗 Thomphorams.github.io

## SKILLS

### Programming

Python

Matlab

Mathematica

### Software & Tools

Optimisation

(DEAP)

Data Visualisation

(Matplotlib, Seaborn, Kibana)

Data analysis

(Numpy, Scipy, Pandas, Elastic search)

Machine Learning

(Scikit-learn)

Data Cleaning Data Modelling

Natural Language Processing

(NLTK, TextBlob, SpaCy)

Data Extraction

(Tweepy API, Talkwalker API)

Data Curation

(structured and unstructured data)

### Soft skills

Presentation

Research

### Operating Systems

Linux

Windows

## AWARDS

Collaboration award, CSIR  
(2022)

Masters Studentship, CSIR  
(2016)

## WORK EXPERIENCE

### Data Scientist

Council for Scientific and Industrial Research (CSIR),  
Pretoria

Aug 2020 - Present

Responsibilities and work outputs:

- Analyse, design, and construct algorithms and predictive models using data modelling processes.
- Social media analytics for South African Local Government Elections (LGE) in 2021.
- Documentation of the Software Requirements Specification (SRS) and System design Document (SDD).
- Created a Natural Language Processing Toolkit for analysing various document formats from structured and unstructured data.
- Scraping and analysis of social media data.
- Developed a sentiment analyser on consumer.
- Write reports for various projects.
- Technical Mentor for Dirisa Datathon (2022) hosted by the CSIR and the Centre for High Performance Computing (CHPC)
- Testing and validation of scripts
- Built a machine translation model for South African Languages.

### Scientific Researcher

Council for Scientific and Industrial Research (CSIR),  
Pretoria

Feb 2016 - Jul 2020

Responsibilities and work outputs:

- Write Python scripts for data extraction, data processing, data visualisation and data cleansing
- Researcher in the field of Data science
- Mentoring and supporting vacation work students for the Data Science for Impact and Decision Enablement (DSIDE) programme.
- Work on various projects

Research Assistant for the project: Multilingualism  
in foundation phase Mathematics  
University of the Witwatersrand, Johannesburg

Feb 2017 - Jan 2019

Responsibilities and work outputs:

**NRF- Scarce Skills Honours Block Grant, NRF** (2015)

**Certificate of first class in Coding and Cryptography III,** (2014)

**University Entrance scholarship, University of the Witwatersrand** (2011)

**Eskom Merit Award, Eskom** (2011)

**Top Learner in Mathematics, 2010 Grade 12 Awards** (2010)

## LEADERSHIP ROLES

**Constitutional Committee, Maitazwitoma Students Society** (2015-2018)

**Mentor, Counselling and Careers Development Unit** (2014-2016)

**Entertainment officer, Maitazwitoma Students Society** (2014)

**Academic officer, Maitazwitoma Students Society** (2013)

**Sub house committee-academics, Reith Hall Residence** (2011)

**Learners Christian Movement Leader, Thengwe High School** (2007-2009)

Investigated the way in which teachers and learners use language in Foundation Phase Mathematics. The tasks involved collecting data through student surveys and teacher interviews, transcribed interviews, and material analysis (textbooks). Reports were compiled for the Department of Basic Education (South Africa), which then influenced the South African Language Policy on Language and Teaching. Research findings were presented at the 26th and 27th Annual Conference of the Southern African Association for Research in Mathematics, Science and Technology Education (SAARMTE) in 2018 and 2019.

**Vacation Student**  
**Council for Scientific and Industrial Research (CSIR), Pretoria**

Nov 2015-Jan 2016

Responsibilities and work outputs:

Produced an interactive web platform that estimates the total radiation energy from a selected region using solar and spectral data.

**Teaching Assistant**  
**University of the Witwatersrand, Johannesburg**

2015 - 2018

Responsibilities and work outputs:

Tutoring Computational and Applied Mathematics I, Statistics for Construction Management I, Computational and Applied Mathematics III

## EDUCATION

**MSc - Computational and Applied Mathematics**  
**University of the Witwatersrand - Johannesburg, Gauteng (South Africa)**

ongoing

**Thesis:** Expectation Maximisation Algorithm for System Identification in Nonlinear Dynamical Systems

My Masters is on developing an algorithm for system identification of a nonlinear dynamical system with unknown states. The algorithm combines four algorithms; the Expectation Maximisation algorithm, Extended Kalman Smoothing, Gaussian Radial Basis Function Networks, and global optimisation techniques (such as Genetic Algorithms, Simulated Annealing, and Particle Swarm Optimisation) for estimating a noisy time series data and the underlying model. The algorithm aims at integrating over the uncertain states of the unknown hidden states and optimising the resulting marginal likelihood of the parameters given the observed data while modelling out the nonlinearities and searching a large space to find an optimal solution.

**BSc (Hon) - Computational and Applied Mathematics**  
**University of the Witwatersrand - Johannesburg, Gauteng (South Africa)**

2015

**Thesis:** Hydraulic Fractures in an Impermeable Rock: Mathematical models, approximate solutions and asymptotic solutions

**Bachelor of Science**  
**University of the Witwatersrand - Johannesburg, Gauteng (South Africa)**

2011 - 2014

**Majors:** Mathematics, Computational and Applied Mathematics

**Matriculation**

**Thengwe High School - Tshandama, Limpopo (South Africa)**

**2010**

## **PUBLICATIONS**

**A Toolkit for Text Extraction and Analysis for Natural Language Processing Tasks**

**Publication Title: 2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)**

**2022**

Status: Published