

# Rohan Posthumus

Data Scientist | Data Analyst | AI Specialist

Email • GitHub • LinkedIn • Website

---

## SUMMARY

---

With 14+ years of experience at the University of the Free State, I am a data scientist specializing in enhancing student success. I architect and build high-impact data solutions, from predictive ML models processing **150,000+ applications annually** to the core algorithms for a **R170 million** foundation partnership. I excel at translating complex data into institutional strategy and delivering robust, scalable systems.

## KEY PROJECTS

---

### NBT predictive algorithm & automation bot

**Problem:** During the pandemic, many students couldn't take the compulsory National Benchmark Test (NBT) for financial reasons, blocking them from registration.

**Solution:** I developed a Gradient Boosting Machines (GBM) algorithm to predict student academic literacy scores, allowing for exemptions. I then built a Python email bot to automate the generation and sending of 1,000+ personalized HTML notification emails per minute.

**Impact:** Processed over 150,000 applications per year for the last 5 years, achieved 85% out-of-sample accuracy, and reduced a task that took four people weeks to just eight minutes. The algorithm is institutionalized in the University's General Rules.

### Michael Susan Dell Foundation (MSDF) student success algorithms

**Problem:** The university needed a robust, scalable system to proactively identify at-risk students and track their progress toward career-readiness.

**Solution:** I authored all the core algorithms for a ZAR 170 million, 7-year partnership. This included the "Scaled Responsive Student Tracking" algorithm (an advanced early warning system) and the "Employability Tracking Algorithm" which pulls, integrates, and analyzes data to award badges as students hit milestones.

**Impact:** These algorithms form the foundation of the university's next-generation student success and retention strategy, directly improving graduation rates and securing a major long-term institutional partnership.

### COVID-19 Engagement Analytics (40,000+ Students)

**Problem:** At the start of the pandemic, top management had no visibility into whether teaching and learning were happening effectively online for over 40,000 students.

**Solution:** I led the tracking and analysis of all online engagement (LMS access, video sessions, assessment completion, etc.). I authored 65 detailed reports that were delivered directly to top management and our rector.

**Impact:** The reports became a national DHET reporting requirement, were presented to the Minister of Higher Education, significantly reduced the University's risk score, and led to multi-million Rand subsidies.

---

## **Learner Case Management (LCM) System**

**Problem:** The University needed an efficient, standardized, and scalable system to keep accurate records of advising support provided to students.

**Solution:** I led the conceptualization and development of the LCM system from the ground up. I wrote the initial prototype code in Python and C# and oversaw a two-year testing period to normalize, standardize, and validate all processes and data.

**Impact:** The LCM system is now the central, reliable platform for tracking all student support interactions and serves as the foundational data source that integrates with the R170 million MSDF student success algorithms.

## **"Baby AI" (Internal LLM Research & Development Project)**

**Problem:** Following the launch of ChatGPT, the University needed to urgently understand the technical capabilities, risks, and resource costs of Generative AI in our own context.

**Solution:** I built a small-scale transformer model ("Baby AI") from scratch. I trained it on institutional parallel corpora (an academic literacy study guide) to test its real-world performance on translation and question-answering tasks.

**Impact:** This R&D project provided critical, hands-on insights into training time, data requirements, accuracy, and hallucinations. The findings were presented at the 2023 SAAIR conference and directly informed the Center of Teaching and Learning's AI strategy.

## **NISS (Georgia State University) Data Collaboration**

**Problem:** The UFS is collaborating with the NISS to analyze 10 years of student learning pathways, a project requiring the extraction and processing of massive, complex historical datasets.

**Solution:** As the lead for institutional data provision, I developed Python scripts using multiprocessing to concurrently pull, process, and automate the handling of these large-scale data requests from our Microsoft SQL Server.

**Impact:** My automated scripts enabled this high-profile international research collaboration, accelerating the data-extraction-to-analysis pipeline and providing the foundation for new insights into student success.

## EXPERIENCE

---

### Data Scientist/Analyst

Nov 2018 - Present • 7 years

University of the Free State, Centre for Teaching and Learning

### Researcher

Oct 2012 - Nov 2018 • 6 years 2 months

University of the Free State, Metagenomics Platform

### Research Assistant

Jun 2010 - Aug 2011 • 1 year 3 months

University of the Free State, Metagenomics Platform

## FORMAL EDUCATION

---

### Level 1 exam passed in the CFA Program

2016

CFA Institute

### Management Development Programme Certificate

2014

University of the Free State

### Bachelor of Science (B.Sc.Hons)

2013

University of the Free State

### Bachelor of Science (B.Sc.)

2011

University of the Free State

## TOP 10 SKILLS

---

Data analytics and statistical modeling

Data science and machine learning

Generative AI development

Automation & process optimization

System architecture & development

Data manipulation and querying (SQL)

Applied science research

Technical communication & training

Programming (Python-first)

## RECENT ACHIEVEMENTS

---

- Authored the core algorithms for a **R170 million** student success partnership with the Michael Susan Dell Foundation.
- Developed and deployed a predictive ML model with **85% accuracy**, processing over **150,000+ applications** annually to inform university policy.
- Led COVID-19 analytics for **40,000+ students** annually, authoring 65 reports that became a **national DHET reporting best practice** and secured multi-million Rand subsidies.

- Architected and developed the **Learner Case Management (LCM) system** from initial prototype (Python/C#) to a production system central to student advising support.
- Conducted a comprehensive impact analysis on developmental modules, demonstrating their role in addressing the **national equity gap** and securing over **R20 million** in annual tuition.
- Built and trained an internal transformer model to **inform institutional AI strategy**, presenting the R&D findings at the 2023 SAAIR conference.
- Presented an invited 4-hour workshop on AI implementation and prompt engineering to the **Council for Higher Education (CHE)**.
- Co-authored research on **Generative AI for Assessment**, which was accepted for a special edition of the South African Journal of Higher Education.
- Automated a critical registration task using a Python bot, reducing a process that took four people **weeks down to just 8 minutes**.