

Raditya Alfarisi

Payakumbuh 26263 | radityaalfarisi6@gmail.com | +62 83869916635 | kaggle.com/radityaalfarisi | github.com/arufadesawa | arufa.dev

PROFESSIONAL SUMMARY

Highly driven Vocational High School (SMK) student and detail-oriented AI Engineer with hands-on experience in optimizing large language models (LLMs) and building machine learning systems from scratch. Proven track record in reducing cloud inference costs by 30% through advanced quantization and pruning techniques, and improving model accuracy via custom fine-tuning. Passionate about bridge-building between low-level systems (Rust/Python) and production-ready AI applications.

WORK EXPERIENCE

AI Engineer Intern | PT Mbot Technology Innovate

January 2026 - June 2026

- **Model Optimization:** Reduced cloud inference costs by **30%** by implementing post-training quantization and structural pruning on internal LLM architectures, while successfully maintaining a **95%** accuracy threshold.
- **Fine-Tuning:** Enhanced model F1-score from **0.78 to 0.92** by fine-tuning open-source LLMs using LoRA (Low-Rank Adaptation) on specialized internal domain documentation.

PROJECT

Aino Is Neural Operation

Python, Numpy | github.com/arufadesuwa/AINO

- Developed a lightweight Deep Learning micro-library from scratch using only Python and NumPy to understand the underlying mathematics of backpropagation and neural networks.
- Implemented custom forward/backward passes, activation functions, and basic optimization algorithms.

OchreML

Rust, Python, PyO3 | github.com/arufadesuwa/OchreML

- Created a high-performance classical Machine Learning library written in Rust with Python bindings, enabling seamless integration and faster execution.
- Implemented fundamental algorithms (including Linear Regression, Tree-based models, and KNN) from scratch to master the low-level mechanics of ML computation.

Fina

Next.js, Bun, Google Gen AI SDK, PostgreSQL | github.com/arufadesuwa/fina

- Engineered and built an AI-powered application utilizing Gemini models via the Google Gen AI SDK, featuring Retrieval-Augmented Generation (RAG) and intelligent planning.
- Integrated vector databases (Postgresql with pg_vector) for efficient semantic search.

TECHNICAL SKILLS

Bahasa Pemrograman: Python, Rust, SQL, JavaScript/TypeScript

AI/ML Frameworks: TensorFlow, NumPy, Pandas, Polars, LangChain, Google Gen AI SDK

Keahlian AI: LLM Fine-Tuning (LoRA), Quantization, Pruning, RAG, Neural Networks, Vector Search

Tools & Infra: GCP, Docker, Git, Bun.js, Node.js, PostgreSQL, Qdrant, N8N

EDUCATION

SMKN 4 Payakumbuh

Vocational High School in Software Engineering / IT | Graduation Year: 2027

- **ML Head Division** at TEFA (Teaching Factory): Managed and coordinated AI/ML-related school projects, directing development workflows.
- **1st Winner** of LKS (Lomba Kompetensi Siswa) Web Technology, Sumatera Barat Province (2026).

MTSN 5 Lima Puluh Kota

Junior High School | Graduation Year: 2024

- Active member of OSIM (Student Organization), developing leadership and teamwork skills.

INTERESTS & HOBBIES

- **Chess:** 1200 ELO at Bullet, Blitz and Rapid (Thinking and Blunder as fast as Hikaru).