Rafif Shafwan

+6281223638228 | rafifshaf@gmail.com | linkedin.com/in/rafif-shafwan | github.com/rafifshaf-fun | Garut, West Java, Indonesia

PROFESSIONAL SUMMARY

Data Scientist and ML Engineer with 6+ years of experience building production-grade systems across finance, computer vision, NLP, and full-stack development. Expert in Python, SQL, deep learning (PyTorch, TensorFlow), and MLOps - from prototyping to deployment and monitoring. Proven ability to own the full ML lifecycle: problem scoping, data pipeline engineering, model development, production deployment, and continuous improvement. Strong communicator who bridges technical and business stakeholders. Open to remote and on-site roles in ML Engineering, Data Science, MLOps, and Computer Vision.

CORE COMPETENCIES

Machine Learning & Data Science | Computer Vision (YOLO, OpenCV) | Deep Learning (PyTorch, TensorFlow) | NLP & LLM / RAG | MLOps & Production Deployment | Data Pipeline Engineering | Python & SQL | Full-Stack & Docker

WORK EXPERIENCE

Freelance Data Scientist & ML Engineer

2020 - Present

Data Scientist & Machine Learning Engineer

Garut, Indonesia

- Designed, prototyped, and deployed **production ML systems** across finance, computer vision, and NLP - owning the full lifecycle from problem scoping to deployment and monitoring
- Built the **Indonesian Stock MLOps Platform**: an end-to-end automated ML system generating BUY/SELL signals for 45 blue-chip stocks with auto-retraining, Grafana monitoring, and zero daily ops overhead
- Developed **multi-stage computer vision pipelines** combining YOLO/SSD detection with fine-grained classifiers, outperforming single-pass models on small/densely packed objects
- Built **NLP / RAG systems** using LangChain, FAISS, and Groq, evaluated with RAGAS metrics for measurable AI quality
- Fine-tuned **deep learning models** in PyTorch and TensorFlow using transfer learning for domain-specific tasks
- Applied **machine learning fundamentals** (scikit-learn, statistical analysis) across diverse domains
- Delivered **full-stack applications** with PostgreSQL, Docker, and professional reporting

Confidential Clients (Public Sector)

Jun 2021 - Jul 2022

Data Analyst

Jakarta, Indonesia

- Built dashboards and reports translating complex engagement data into actionable insights for non-technical decision-makers
- Conducted multi-dimensional analysis on cross-platform digital media performance, identifying trends and optimization opportunities
- Collaborated with stakeholders to define analytical requirements and present clear, decision-ready findings

Star Energy

Jan 2019 - Feb 2019

Data Engineer Intern

Jakarta, Indonesia

- Supported enterprise data migration, including metadata standardization and data consistency validation across legacy systems

PROJECTS

Multi-Stage Computer Vision Pipeline

CV | Deep Learning

Two-stage inference architecture combining **YOLO/SSD** for broad object detection with a fine-grained classifier on cropped detections. Achieves significantly higher accuracy on small or densely packed objects. Built with Python, OpenCV, PyTorch, and TensorFlow. Enterprise version delivered under NDA.

Python | YOLO | OpenCV | PyTorch | TensorFlow

Indonesian Stock MLOps Platform

MLOps | Finance

End-to-end automated ML system generating BUY/SELL signals for 45 IDXBLUE blue-chip stocks. Features **auto-retraining pipelines**, real-time Grafana monitoring, and zero daily ops overhead. Demonstrates full-stack data science: data pipeline, model development, deployment, and business impact.

Python | MLflow | Docker | Grafana | Scikit-learn | Pandas

CV RAG Chatbot

NLP | LLM

Retrieval-augmented generation system answering questions about a professional resume. Built with LangChain LCEL, FAISS vector storage, FastEmbed (ONNX), and Groq Llama 3.3. Evaluated with **RAGAS metrics** (faithfulness, answer relevancy, context precision) for measurable AI quality.

Python | LangChain | FAISS | FastEmbed | Groq | RAGAS | Streamlit

Sovereign Ledger

Full-Stack | Fintech

Comprehensive financial management system for a health-worker cooperative. Features double-entry accounting, automated batch processing, multi-type loan management, role-based access, and professional PDF/CSV reporting. Deployed on-premise via **Docker** (PHP-FPM + Nginx + PostgreSQL) with Supervisor-managed scheduling.

Laravel | Livewire | PostgreSQL | Docker | DOMPDF

EDUCATION

Bachelor of Computer Science GPA: 3.36/4.0

Universitas Terbuka

Completed through distance learning while building full-time professional experience

CERTIFICATIONS

Python - Data Science	SanberCode	2020
CISDM - Data Modelling	Cybertrend	2019
CIPMA - Project Management	Cybertrend	2019
MTA - Microsoft Technology Associate	Microsoft	2018

SKILLS

Languages: Python (primary), SQL, PHP, JavaScript **ML & Data Science:** Scikit-learn, Pandas, NumPy, Statistical Analysis, EDA

Deep Learning & CV: PyTorch, TensorFlow, YOLO, OpenCV, image processing **NLP / LLM:** LangChain, FAISS, FastEmbed, Groq, RAGAS

MLOps & Deployment: MLflow, Docker, Grafana, FastAPI, model monitoring

Data & Database: PostgreSQL, SQL, ETL Pipelines, Data Validation **Tools:** Git, Jupyter, VS Code, Linux