

# MARCELL HERMAWAN KRISTIANTO

[marcellkristianto.ai@gmail.com](mailto:marcellkristianto.ai@gmail.com) | East Jakarta, Indonesia | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

[www.linkedin.com/in/marcell-hermawan-kristianto](https://www.linkedin.com/in/marcell-hermawan-kristianto) | <https://github.com/marknshoot> | <https://marcell-kristianto.vercel.app/>

## SUMMARY

5th-semester Data Science student at Binus University (GPA 3.85/4.00) targeting AI Engineer intern roles. Co-author of an in-press Springer Nature chapter benchmarking 6 deep CNN architectures on biological cellular microscopy with TFLite FP16 quantization. Hands-on experience building production LLM applications — hybrid vector retrieval (pgvector + BM25), LangChain multi-agent copilots with RAG, NVIDIA NIM serving — and fine-tuned CLIP image–text retrieval (+47.6% Recall@1 gain). Teaching assistant for the 9-hour NVIDIA Rapid Application Development with LLMs workshop at BINUS, and a daily practitioner of AI agent harnesses (Hermes Agent, Pi Agent) since April 2026.

## EDUCATION

### Bachelor in Data Science, Bina Nusantara (BINUS) University

2024 - Present

- Current GPA: 3.85/4.00 (5th Semester)
- Relevant Coursework: Machine Learning, Deep Learning, Text Mining, Data Mining (R), Bayesian Statistics

## WORK EXPERIENCES

### Student Advisory and Support Center (SASC) BINUS University

*Mentor Scholarship: Junior Mentor (Even 25/26) → Senior Mentor (Odd 26/27)*

Feb 2026 - Present

- Selected on a merit tuition grant to deliver weekly academic tutoring in computer science and data science fundamentals, designing personalized study modules mapped to each mentee's corresponding semester coursework (Semesters 2-5).
- Selected as Junior Mentor (Feb 2026), then promoted to Senior Mentor for the following term (Aug 2026).

## ORGANIZATIONAL EXPERIENCES

### Data Science Club (DSC) BINUS University

*Teaching Assistant — NVIDIA DLI Workshop: Rapid Application Development with LLMs*

Oct 2026

- Selected as one of three teaching assistants for the 9-hour NVIDIA Deep Learning Institute workshop delivered at BINUS by an NVIDIA DLI Certified Instructor (BINUS AIRDC × Data Science Club); completed the course and earned the NVIDIA Certificate of Competency.
- Ran event operations end-to-end — designed the announcement poster, booked the venue (Room 427), built the registration form and screened applicants — and debugged participants' code during the hands-on LLM application labs. Evidence: [Data Science Club post listing the teaching assistants](#).

*PIC, Data Seeker (Competition Team)*

Oct 2025 - Present

- Directed 15 competitive data science delegates across 5 squads; engineered and judged an onboarding Kaggle competition evaluating notebook architectures, validation splits, and feature engineering pipelines.
- Managed competition registrations, squad budgets, and reimbursement workflows for national hackathons and data competitions, and moved them onto Data Seeker Web, a portal I built (Next.js, Supabase, \$0 hosting) that tracks each claim through review and compiles the university's 7-section reimbursement into one printable PDF.
- Conducted weekly technical mentorship sessions covering deep learning architectures, cross-validation discipline, and ensemble methods.
- Established a shared solution vault in the same portal, archiving winning codebases from past squads alongside 10 curated competition guides (Kaggle 101 to Bayesian ML) to standardize competitive playbooks and speed up onboarding.

*Research Activist*

Feb 2025 - Oct 2025

- Joined the club's research track with the BDSRC lab; the output was the ICOBAR 2025 paper listed under Publications (2nd of 5 authors, in press with Springer Nature).

## PUBLICATIONS & BIOLOGICAL AI RESEARCH

### Model Comparison for Toxigenic Cyanobacteria Image Classification using Transfer Learning and Post-Quantization

Aug 2025

*Springer Nature (LNEE 1693, in press) · Co-Author (2nd of 5) · ICOBAR 2025*

- Evaluated 6 pretrained CNN architectures on biological microscopy cellular imagery in TensorFlow/Keras in collaboration with BDSRC; EfficientNetB1 achieved top test accuracy of 96.72% (+6.68 pts Macro F1 over prior MobileNet benchmark).
- Implemented FP16 post-training quantization (TFLite), cutting storage footprint by 50% (25.5 MB → 12.8 MB) with 0.00% accuracy loss and 45.3ms CPU latency, evaluated for low-power analytical devices.
- Conducted confusion matrix failure analysis across morphological ambiguities in a 10-class dataset; paper accepted at the 7th International Conference on Biospheric Harmony (ICOBAR 2025).
- Link: [ResearchGate Early Access](#)

## CERTIFICATES

- Rapid Application Development with Large Language Models (LLMs), NVIDIA (Certificate of Competency) Issued Oct 2026
- PyTorch for Deep Learning (Professional Certificate, 87h), DeepLearning.AI Issued Mar 2026

|                                                                     |                 |
|---------------------------------------------------------------------|-----------------|
| • Generative AI with Large Language Models, DeepLearning.AI         | Issued Mar 2026 |
| • Foundation: Introduction to LangGraph (Python), LangChain Academy | Issued Sep 2026 |
| • Foundation: Introduction to LangChain (Python), LangChain Academy | Issued Aug 2026 |
| • Fundamentals of Deep Learning, NVIDIA                             | Issued Aug 2024 |
| • AI RAG Systems & Agentic AI Workshops, ZenerA                     | Issued Aug 2026 |

## ACADEMIC AND INDEPENDENT PROJECTS

---

### Toko Marcell — Multimodal Hybrid Retrieval & Agentic AI Engine (Live)

*PyTorch, OpenAI CLIP, PostgreSQL, pgvector, LangChain, FastAPI, Next.js 16, Docker*

- Fine-tuned OpenAI CLIP ViT-B/32 on 5,378 multimodal pairs in PyTorch with hard negative mining; champion Decoupled LR model achieved Recall@1: 39.15% (+12.6 pts / +47.6% relative gain over baseline) and MRR: 0.5592, exported to ONNX runtime for CPU/GPU serving.
- Engineered 3-way hybrid retrieval engine combining lexical BM25, 384-dim dense text vectors (all-MiniLM-L6-v2), and 512-dim visual embeddings using Reciprocal Rank Fusion (RRF, k=60) in PostgreSQL/pgvector HNSW indexes.
- Built an autonomous Tri-Modal AI sales copilot using LangChain with 4-node architecture (Fast-Path Guardrail → Planner → Concurrent Tool Execution → Persona Synthesizer) that searches the catalog for shoppers, recommends and sells products, and answers store-policy (RAG) and body-metric sizing (TB/BB) questions.
- Architected multi-tier caching (Next.js 16 Edge ISR <40ms TTFB, in-memory TTL facet cache dropping Neon queries from ~60ms to 2ms) and defensive Pydantic bounds; containerized fullstack with Docker Compose local auto-seeding.
- Links: [Live Storefront](#) | [GitHub Documentation](#)

### Telco Churn — Cost-Based Retention Targeting

*Python, scikit-learn, imbalanced-learn, XGBoost, LightGBM, pandas*

- Costed 6 retention strategies in dollars on 986 held-out customers; logistic regression with a cost-chosen threshold (0.06 vs. default 0.5) lost \$5,259/month vs. \$6,439 for discounting everyone (−18.3%), catching 257 of 263 churners.
- Benchmarked 7 classifiers with and without NCR undersampling on 5-fold PR-AUC, and reported that a one-line contract rule captures most of the gain (−14%) and wins once offer success falls below ~75%.
- Link: [GitHub Notebook](#)

## SKILLS

---

**Programming Languages:** Python, SQL, TypeScript, R, JavaScript, C, HTML, CSS

**Languages:** Native in Indonesian, Professional Fluency in English

**ML & Deep Learning:** PyTorch, Hugging Face Transformers, CLIP (VLM), TensorFlow / Keras, transfer learning, TFLite FP16 & ONNX Runtime, scikit-learn, XGBoost, LightGBM, imbalanced-learn

**Evaluation:** Recall@k, MRR, nDCG, PR-AUC, stratified CV, cost-based thresholds, baselines

**LLM & Agents:** LangChain, LangGraph, Gemini API, NVIDIA NIM, RAG, tool calling, cross-encoder reranking, agent harnesses (Hermes, Pi)

**Data & Retrieval:** SQL, PostgreSQL, pgvector (HNSW), BM25 hybrid search, PySpark, Apache Airflow, pandas, Supabase

**Backend, Web & DevOps:** FastAPI, Next.js, React, Tailwind, Docker Compose, GitHub Actions, pytest, Vitest, Git, Linux, Vercel