

Faiz Naufal Huda

Bogor, Indonesia

faiznaufal2015@gmail.com · linkedin.com/in/faizhuda · github.com/faizhuda · faizhuda.github.io

Summary

Computer Science undergraduate at IPB University (GPA: 3.85/4.00), BNSP-certified Associate Data Scientist, with hands-on experience building complete, well-tested systems end-to-end, from a full-stack academic document management system (118 tests, Clean Architecture) to an award-winning hackathon e-commerce platform and ML pipelines with CI/CD, Docker, and transformer fine-tuning. Completed the Pijak × IBM SkillsBuild AI Engineer Cohort as a selected participant, graduating as Top Graduate. Fluent in English (TOEFL ITP 607). Actively seeking a Data Science, Data Analyst, or ML engineering internship.

Technical Skills

Languages:	Python, SQL, JavaScript, TypeScript
Frontend:	React, Tailwind CSS, Vite
Backend & DB:	FastAPI, SQLAlchemy 2.0, PostgreSQL, JWT + bcrypt, Alembic
ML / Data:	LightGBM, XGBoost, Scikit-learn, Pandas, NumPy
DevOps / Tools:	Docker, GitHub Actions (CI/CD), Git, pytest
Familiar:	Next.js, Supabase, PyTorch, Hugging Face Transformers, TensorFlow/Keras, CatBoost, Optuna, MLflow, Prometheus, Grafana, Weights & Biases, Bash, ReportLab, PyPDF

Projects

LTI Incident Risk Prediction (github.com/faizhuda/lti-incident-risk-prediction) Jun 2026

- BNSP Associate Data Scientist certification project: classified Lost Time Injury (LTI) risk from 2,699 SHE incident records at PT Banana Coal.
- Addressed severe class imbalance (7.82% positive rate) and target leakage; final model was a Logistic Regression optimized for recall via a self-contained notebook pipeline.

AQI Bangladesh: Air Quality Prediction (github.com/faizhuda/aqi-bangladesh-prediction) May – Jun 2026

- Independently built an end-to-end ML system predicting AQI across 30 cities in Bangladesh; trained LightGBM on 1M+ rows of multi-pollutant sensor data achieving MAE = 1.13, RMSE = 1.77 (AQI index points), $R^2 = 0.9986$.
- Engineered modular pipeline with centralized YAML config, lag features, rolling windows, cyclical time encoding, and walk-forward cross-validation (3-fold) for time-aware evaluation; experiment tracking via CSV logging.
- Implemented 40 unit tests across 6 modules (pytest + coverage); CI/CD via GitHub Actions with black/flake8 enforcement and pre-commit hooks; containerized with Docker for reproducible deployment.

AgriDesk: Academic Document Management System (github.com/faizhuda/ads-agridesk) Feb – Jun 2026

- Led full-stack development of a 4-person team project; solely responsible for the complete local implementation: backend in FastAPI + SQLAlchemy 2.0 + Alembic on PostgreSQL and frontend in React 19 + Vite + TailwindCSS v4, built on Clean Architecture, Domain-Driven Design (DDD), and Finite State Machine document workflows with drag-and-drop PDF upload.
- Implemented SHA-256 document hashing with QR code verification enabling third-party authentication without system login; PDF generation (ReportLab + PyPDF) with coordinate-based digital signature stamping.
- Achieved 100% test coverage on core services across 118 passing tests; JWT + bcrypt authentication; in-memory SQLite for isolated unit testing.

Fruit & Vegetable Freshness Classification (github.com/faizhuda/produce-freshness-classifier) May – Jun 2026

- Independently designed and executed 11 experiment scenarios comparing classical ML (SVM, Random Forest) against a MobileNetV2 CNN for binary fresh/rotten produce classification across 4,391 test images; best CNN scenario reached 98.70% accuracy, a statistically significant improvement over the best classical model (87.34% via SVM, $p < 0.001$).
- Used paired significance testing to catch a counterintuitive result: a CNN on raw unprocessed images significantly outperformed the same CNN behind a full image-restoration and segmentation pipeline (98.70% vs. 97.22%, $p < 0.001$).
- Built with 64 pytest unit tests, GitHub Actions CI, and ruff linting; pipeline logic kept side-effect-free and decoupled from Kaggle-specific orchestration for reproducibility.

- Built an Optuna-tuned ensemble (LightGBM + XGBoost + CatBoost, 135 total trials) for a binary loan-default Kaggle competition, rank-blending the three models to 0.9325 out-of-fold AUC (StratifiedKFold, 5-fold).
- Iterated on feature engineering empirically rather than by intuition: reverted class weighting and most derived features after measuring they reduced OOF AUC, keeping only the one interaction term that measurably helped.

- **Placed 2nd** as full-stack developer in a 3-person team; built a complete customer + admin e-commerce platform using Next.js, TypeScript, Tailwind CSS, and Supabase (PostgreSQL) within hackathon timeframe.
- Implemented PostgreSQL stored procedures for atomic order cancellation with inventory rollback, eliminating race conditions on concurrent stock updates.
- Delivered real-time admin analytics dashboard (Recharts), PDF invoice generation, and full order lifecycle management (Pending → Paid → Shipped → Delivered).

Experience

- Completed structured learning path covering Python, supervised/unsupervised ML, deep learning (CNN & RNN), experiment tracking (MLflow), and production ML deployment via IBM SkillsBuild and Dicoding's industry curriculum.
- Built **PriorMail** as capstone (scored 98/100): an email intelligence system with DistilBERT fine-tuning for 4-class priority classification; designed phishing detection module with security-focused adversarial evaluation protocol.
- Graduated Lulus Penuh (Full Pass) and recognized as Top Graduate ("Lulus Terbaik") of the cohort.

- Managed IDR 33 million F&B consumption budget covering meals, snacks, and catering for 500+ participants and 90+ committee members across the full IT TODAY event series.
- Secured IDR 5 million in internal operational funding through entrepreneurial methods to support committee activities.

- Managed annual divisional budget of ~IDR 15–20 million across 2–3 major programs with accurate planning and real-time expenditure tracking.
- Oversaw 20–30 financial transactions per program cycle; handled reimbursements and documentation with full accountability and transparent reporting.

- Completed 90 volunteer teaching hours (Grade A, 94.96/100) across 14 sessions to 25+ children in Leuwiliang, Bogor over 9 months, covering literacy, English, mathematics, environmental awareness, and introductory programming through game-based activities.

Education

Certifications

- **AI Engineer Cohort (Top Graduate)**, IBM SkillsBuild + Dicoding Indonesia Jul 2026
- **Associate Data Scientist**, BNSP (Badan Nasional Sertifikasi Profesi) Jul 2026
- **Belajar Fundamental Deep Learning** (Deep Learning Fundamentals), Dicoding Indonesia May 2026
- **Belajar Machine Learning untuk Pemula** (ML for Beginners), Dicoding Indonesia Mar 2026
- **Python for Data Science and ML Essential Training (Part 1 & 2)**, LinkedIn Learning Feb 2026
- **TOEFL ITP**, ETS (Score: 607) Nov 2023