

I Kadek Bagus Deva Diga Dana Putra

📍 Bali, Indonesia ✉ devadigadana@gmail.com 📧 in/bagus-deva 🌐 bagusdeva.com

SUMMARY

Data & AI Engineer with hands-on experience building production-grade pipelines, RAG systems, and cloud-based analytics. Proven track record in scaling ETL workflows (60k+ listings/month), optimizing dashboards, and deploying AI-powered applications for real estate and education sectors.

EXPERIENCE

Data Analyst & AI Automation Engineer

Investland Bali Properties • Full-time

July 2025 - Present

Building Bali's multi-source property market intelligence platform, from automated data acquisition across several listing websites to a production-grade dashboard and quarterly public market reports.

- Designed and operated an ETL pipeline aggregating 60,000+ listings from 9 property platforms into a unified Supabase database, with per-source field mapping and normalization across 100+ sub-area classifications.
- Built an end-to-end automated data pipeline that extracts Bali property market metrics (sales, supply, rental) from a third-party Looker Studio dashboard via reverse-engineered API calls, loads them into a Supabase/PostgreSQL warehouse, and runs scheduled SQL aggregation.
- Architected scraping infrastructure (VPS) using Docker Compose, cron/launchctl scheduling, and Cloudflare R2 for cross-machine coordination. Built Bronze (raw) and Silver (clean) data layers with staging review workflow and 31 data quality flags.
- Produced quarterly public market reports integrating proprietary data with 5 external sources (BPS Bali, Bank Indonesia, Colliers, Angkasa Pura, Global Property Guide), covering pricing, supply, leasehold analysis, and rental market across 8 regions.
- Developed backend for an AI-powered application, delivering REST APIs and integrating AI workflows into service endpoints. Implemented contract-first API design (OpenAPI/YAML) with consistent response schema.
- Built an interactive Next.js dashboard with region filtering, property type segmentation, and pricing analytics powered by Supabase.

Software Engineer & Co-Founder

Rayaska

July 2025 - Present

MariBelajar is an educational platform offering various courses and learning resources, including E-Certificates and video lessons.

- Co-founded a software house delivering projects across mobile, web, data science, IoT, and infrastructure.
- Built a Random Forest model predicting project delay levels for the Ministry of Public Works (PUPR Gianyar Regency) with 70%+ accuracy.
- Owned backend and inference pipeline end-to-end, ensuring scalable deployment.
- Led frontend development for company profile web build using Svelte and TypeScript.

Microsoft Data Engineer Cohort

MariBelajar

February 2024 - June 2024

MariBelajar is an educational platform offering various courses and learning resources, including E-Certificates and video lessons.

- Completed a 5-month intensive training in cloud-based data engineering with Microsoft Azure, covering key services like Azure Data Factory, Azure Databricks, and Azure Synapse Analytics.
- Designed and deployed scalable data pipelines, processing large datasets efficiently using ETL workflows in real-world project simulations.
- Collaborated in agile teams with peers and mentors to solve data-driven business problems, simulating professional data engineering environments.
- Gained hands-on experience with version control (Git), SQL-based data transformations, and pipeline automation, strengthening job-ready technical skills.

Machine Learning Cohort

Bangkit Academy led by Google, GoTo, & Traveloka

August 2023 - December 2023

Bangkit Academy is a tech career program by Google, GoTo, and Traveloka for Indonesian students.

- Completed Bangkit Academy's Machine Learning learning path, a 900+ hour training program led by Google, GoTo, and Traveloka under the Kampus Merdeka initiative.
- Gained in-depth knowledge of supervised and unsupervised learning, deep learning, computer vision, and TensorFlow-based model development.
- Collaborated with a multidisciplinary team on a capstone project to solve a social impact problem using ML, following end-to-end ML development practices.
- Strengthened soft skills through weekly mentoring sessions and training in agile teamwork, English for the workplace, and career readiness.

PROJECTS

Bali Properties Data Platform

Investland Bali Properties • Private repository

August 2025 - Present

- Architected a two-pipeline data infrastructure that collects 63,000+ property listings per monthly cycle from 10 sources, replacing manual market research that previously took the team several days.
- Engineered a quadtree-based scraper for Airbnb using spatial coordinate decomposition to bypass the platform's result-count limit, capturing 40,000+ listings per cycle versus ~1,000 with naive scraping.
- Built REID Pipeline that extracts statistics from Looker Studio's internal API across 41,000+ calls/month, applying weighted-average methodology to normalize data across heterogeneous property types.
- Implemented bronze/gold data architecture on Cloudflare R2 and Supabase with a human-in-the-loop staging step, ensuring data quality before promotion to production.
- Containerized all pipelines with Docker and automated execution via cron, eliminating manual data refresh entirely.

Bali Properties Intelligence Dashboard

Investland Bali Properties · *Private repository*

January 2026 – Present

- Built a full-stack analytics dashboard that transforms 63,000+ raw property listings into actionable market intelligence for analysts, agents, and management.
- Developed a property comparison tool enabling side-by-side evaluation of multiple listings across price, location, amenities, and ROI indicators.
- Implemented an automated report generation system that exports formatted market reports ready for client presentations, reducing report preparation time from hours to minutes.
- Designed interactive visualizations including market trend charts, price distribution maps, and occupancy rate breakdowns across Bali regions.
- Integrated directly with the Data Platform's Supabase gold layer so the dashboard reflects the latest monthly data cycle without manual imports.

Construction Delay Prediction System

Rayaska Project · S2 Thesis · *Private repository*

March 2026 – April 2026

- Built an end-to-end binary classifier predicting construction delays for public infrastructure projects, achieving 73% accuracy and 0.83 F1 score on a 484-project dataset.
- Engineered the full ML pipeline from feature engineering through Random Forest training to FastAPI inference service, with documented Pydantic schemas defining the contract for the frontend team.
- Containerized the inference service with Docker for reproducible deployment, delivering a production-ready REST API rather than a Jupyter notebook.
- Conducted comparative model evaluation across Random Forest, Logistic Regression, and Gradient Boosting, selecting Random Forest for its balance of accuracy and interpretability.

Chatbot Assistant with PDF

Personal Project · github.com/Bagusdevaa/Chat-with-PDF

March 2025 – April 2025

- Built a Retrieval-Augmented Generation (RAG) application for natural language Q&A over user-uploaded PDF documents, using LangChain, OpenAI embeddings, and a FAISS vector store.
- Engineered the document ingestion pipeline with recursive character chunking (1,000 tokens, 200-token overlap) and dense embedding generation for semantic retrieval.
- Implemented top-K context retrieval feeding the most relevant chunks to the LLM with grounded prompts to minimize hallucination on out-of-document queries.
- Iterated on chunk size, retrieval depth, and prompt structure through manual evaluation, balancing answer quality against context window constraints.

EDUCATION

Informatics Bachelor's Degree

Minor in Smart Computing · Udayana University · 2025 · GPA 3.95

CERTIFICATIONS

Microsoft Certified: Azure Fundamentals

Microsoft · 2024

- Gained foundational understanding of Microsoft Azure cloud services, including compute, storage, networking, and security solutions.
- Learned key concepts of cloud computing, such as cloud models (IaaS, PaaS, SaaS) and Azure pricing, helping in cost optimization and resource management.
- Developed a strong understanding of Azure governance and compliance, essential for building secure and scalable cloud applications.

Microsoft Certified: Azure Data Fundamentals

Microsoft · 2024

- Gained foundational knowledge of core Azure data services, including Azure SQL Database, Cosmos DB, and Azure Blob Storage.
- Learned key data concepts such as data warehousing, data modeling, and ETL processes, essential for managing and analyzing large datasets in the cloud.
- Gained practical skills in querying and managing data on Azure, providing a solid foundation for building and deploying cloud-based data solutions.

DeepLearning.AI TensorFlow Developer

Coursera · 2023

- Built, trained, and deployed deep learning models using TensorFlow and Keras, including CNNs for image classification and NLP models for text processing.
- Gained practical skills in handling real-world ML tasks such as time series forecasting, image recognition, and sentiment analysis using custom neural networks.
- Learned how to optimize model performance and generalization through techniques like regularization, callbacks, and hyperparameter tuning in TensorFlow workflows.

SKILLS

Languages: Python, SQL, JavaScript/TypeScript

AI/ML: TensorFlow, PyTorch, Scikit-learn, LangChain, OpenAI API, RAG, Transformers, MLflow, Embeddings, FAISS, ChromaDB

Data Engineering: PostgreSQL, Supabase, MongoDB, Apache Spark, Airflow, ETL/ELT pipelines, Medallion Architecture

Backend & Infra: FastAPI, Flask, Docker, Docker Compose, REST API design (OpenAPI/YAML), Git, CI/CD

Cloud & BI: Microsoft Azure (ADF, Databricks, Synapse), Power BI, Google Cloud Platform
