

Ehud Vaknin

[LinkedIn](#) | udivak9@gmail.com | 053-422-4700 | [GitHub](#) | [Portfolio](#)

Software Engineer (B.Sc.) with hands-on experience across the full machine-learning lifecycle - from raw data and feature engineering through model training, evaluation, and production serving. I enjoy building end-to-end data pipelines and shipping reliable, reproducible ML systems. Seeking an entry-level Data Science / ML Engineer role.

EDUCATION

2022 - 2026 B.Sc. in Software Engineering, Sami Shamoon College of Engineering · **GPA: 92**

EXPERIENCE & PROJECTS

Bird vs. Drone Detection (YOLO11) - B.Sc. Final Project

Computer Vision / Deep Learning · Ultralytics YOLO11s, PyTorch, Python

- Trained a custom **Ultralytics YOLO11s** object detector to detect and classify two classes - **birds vs. drones** - for airspace monitoring.
- Curated and labeled a **~170K-image dataset** (53% drone / 47% bird) with an 80/10/10 train/val/test split, and generated **12K synthetic drone images** in **Unreal Engine 5.4** to expand coverage of rare viewpoints.
- Achieved **mAP@0.5 of 0.90** ($mAP@0.5:0.95 = 0.66$) on the held-out validation set, with avg precision 0.95 and recall 0.82 across classes.
- Ran **real-time inference** on video streams, targeting low-latency / edge deployment.

Israel Housing Price Platform - End-to-End Data & ML Pipeline

Academic Data-Science Project · Python, pandas, FastAPI, MongoDB, Docker

- Built an end-to-end real-estate **data pipeline**: **Playwright** scrapers ingest five public sources (Govmap, Tax Authority, CBS, odata.org.il, Madlan) into **MongoDB**, feeding an ETL stage that normalizes and enriches **~280K property transactions**.
- Engineered **~88 features** per record - property attributes, macro-economic signals (CPI, prime rate, USD/ILS), and **50+ OSM geospatial features** (distances, POI counts, land-use ratios) - with multi-resolution **H3** geo-indexing for zoom-aware map clustering.
- Trained and served **seven price-prediction models** (LightGBM, CatBoost, stacked ensembles) behind a **FastAPI** service with LRU model caching, champion selection, and multi-model comparison (R^2 / MAE); microservices stack containerized with **Docker Compose**.

Machine Learning Pipeline - Acoustic Emission Research

Independent / Academic Research Project · Python, scikit-learn, pandas

- Built an **end-to-end ML pipeline** on raw acoustic-emission sensor data: a custom parser converts proprietary instrument .TXT files into structured datasets (handling multiple header formats and merged-cell edge cases) and engineers a **19-feature** signal representation, scaled with **RobustScaler**.
- Clustered signals with **K-Means** (k selected via **silhouette analysis**, PCA-inspected) and built an **anomaly-detection** stage benchmarking **Isolation Forest**, **LOF**, and an **AutoEncoder** with ROC-AUC.
- Standardized a **reproducible workflow**: hyperparameter grid search, serialized artifacts and scalers (**joblib**), JSON run metadata, and a reusable notebook template for new algorithms.

Software Engineer - Mind Artists Journey

2025 - 2026 · Production B2B SaaS platform

- Designed and shipped an AI **lead-nurturing system** built on a custom **"skills" framework** - modular, composable capabilities the AI selects from to engage and nurture leads - **orchestrating OpenAI, Anthropic, and Gemini** across content and transcript workflows, with memory-based automation for context retention.
- Built an **embeddings-based RAG pipeline** powering semantic search, retrieval, and an in-product AI coach with **OpenAI streaming** responses.
- Owned **MLOps & infrastructure** - CI/CD with parallelized E2E tests (**cut pipeline time 40 -> 12 min**), Dockerized services, and multi-environment deployments on Vercel/Supabase - and managed an offshore developer (scoping, PR review, releases).

SKILLS

Languages: Python, SQL, TypeScript, C, C++, C#, Java

ML & Deep Learning: scikit-learn, PyTorch, Ultralytics YOLO, LightGBM, CatBoost, XGBoost, K-Means, Isolation Forest, LOF, AutoEncoders, PCA, ensembles / stacking

Data & Features: pandas, NumPy, geopandas, shapely, H3, ETL pipelines, web scraping (Playwright), feature engineering, model evaluation (mAP, ROC-AUC, R^2 / MAE)

MLOps & Infrastructure: Docker / Docker Compose, CI/CD (GitHub Actions), FastAPI model serving, model serialization (joblib), MongoDB, Vercel, Supabase, Git

AI / LLM: Retrieval-Augmented Generation (RAG), vector embeddings, OpenAI / Anthropic / Gemini APIs, prompt engineering, streaming inference

LANGUAGES

Hebrew - Native | English - Fluent