# Ahmadreza Zamanvand

**Data & Machine Learning Developer**

ahmad.zmvd@gmail.com · +98 918 795 0256 · GitHub · Kaggle · Hugging Face · LinkedIn

## Profile Summary

Data and Machine Learning developer (M.Sc. in Computer Science, Data Mining) focused on building automated data pipelines, cleaning noisy real-world datasets, and training computer vision models using Python, SQL, PyTorch, FastAPI, and Docker. Hands-on experience collecting large volumes of unstructured text and image records, filtering out corrupted or duplicate data, and turning raw streams into clean relational tables and practical software tools.

## Projects

**FGVC-IR8: Automotive Data Pipeline & Fine-Grained Recognition System** [ GitHub · Kaggle · Hugging Face ]

*End-to-end data collection, multi-stage cleaning pipeline, and vehicle recognition system built from Divar.ir listings (18.8 GB public dataset).*

- Collected **84,083 car listings** and **360,700 raw images** into a **SQLite** database by partitioning search queries across production years to bypass API pagination limits, using an asynchronous downloader with automatic rate pacing.
- Built a multi-stage filtering pipeline combining **SQL** rules, Persian text parsing, and vision models (**YOLO**, **SigLIP**, and **MiniCPM-V**) to remove salvage cars and **58.9% interior or document noise**, producing **121,928 clean exterior photos** across **50,974 ads** (**9.38/10 Kaggle rating**).
- Grouped reposted dealer ads using **Union-Find graph clustering** over listing tokens, perceptual hashes, and visual similarity prior to **GroupKFold** splitting, keeping train-test data leakage at **0.00%**.
- Fine-tuned a multi-view **DINOv2** vision model with **LoRA** and **ArcFace** to distinguish visually similar cars across 8 classes (**98.27% accuracy**), and packaged the service with **FastAPI**, **Docker**, and automated tests.

**TelePulse: Real-Time Telegram Stream Collector & Lead Classifier** [ GitHub ]

*Automated Python data pipeline that monitors 18 public Telegram channels in real time, filters out spam, and stores verified project leads.*

- Built an asynchronous **Python** collector using persistent connection pooling and sliding-window rate limiting to poll 18 public channels continuously with **1.25s average latency** and zero rate-limit blocks.
- Designed a Persian text normalizer and **13-topic classifier** to separate real employer project posts from freelancer ads (**100% precision** on audited records), while deduplicating cross-channel messages in **SQLite** via **SHA-1** content hashing.

## Skill Set

**Data Engineering & Mining:** Python (NumPy, Pandas, Asyncio), SQL & SQLite (Indexing, Window Functions), ETL, Web Scraping.

**Data Cleaning & Search:** Union-Find Graph Clustering, Perceptual Hashing, Embedding Similarity Search, Persian NLP, GroupKFold.

**Machine Learning & Vision:** PyTorch, DINOv2 (ViT), YOLO, OpenCV, SigLIP, MiniCPM-V (VLM), LoRA, ArcFace, Scikit-learn.

**Backend & Tooling:** FastAPI (REST APIs), Docker, Connection Pooling & Rate Limiting, Unit Testing (unittest), Git, Linux.

## Education

**M.Sc. in Computer Science, Data Mining** • IAU Science and Research Branch (SRBIAU) · 2026 – Present (Active Student Exemption)

**B.Sc. in Computer Science** • Razi University · 2020 – 2026

## Languages

**English:** Upper-Intermediate Technical Reading & Listening (B2/C1); Intermediate Spoken Communication (B1) · **Persian:** Native