DHRUVKUMAR PATEL

Data Scientist

+91-9328132446 | iamdhruv1563@gmail.com | website | linkedin | github

EDUCATION

IIIT DELHI — Delhi, IN (Jul 2024 - Jun 2026)
M.Tech in Computer Science and Engineering, (GPA: 8.42/10.0)

LDRP ITR — Gandhinagar, IN (Aug 2020 - May 2024)
B.Tech in Information Technology, (GPA: 8.14/10.0)

WORK EXPERIENCE

Myntra — **Bangalore**, IN
Data Scientist Intern — Jan 2026 – Present

- Boosted production CTR **F1** by **~6%** via **overhauled data ingestion and feature pipelines** for the AdsRank team.
- **Halved training time** via a **modular automated pipeline** with **MLflow** and **continuous evaluation**.
- **Optimized the AdsRank re-architecture** by developing scheduled **Databricks** workflows using **PySpark** for **embedding generation** and crafting **idempotent, snapshot-based pipelines**, resulting in a **25% reduction** in IO latency.
- Owned **model quality validation** through **sanity tests**, **score-distribution analysis**, and **Grafana dashboards** surfacing alerts across 5 key model-health and run-artifact metrics..
- **Automated 80% of the end-to-end Ads ranking experimentation** by writing **custom agents** with **MCP integrations** across 5 internal platforms (*Databricks, Confluence, GitHub, Grafana, Kibana*).

MIDAS: Multimodal Digital Media Analysis Lab — **IIIT-Delhi**, IN
Graduate Researcher (Thesis/Research) — Dec 2024 – Present

- Developed a training pipeline for small **vision-LLMs** using **reward-driven feedback**, **LoRA**, and a **Mixture of Refinement Agents** on **5× H100-SXM cloud GPUs**, evaluating **Qwen-2.5-VL-72B-Instruct** for data anomaly detection.
- **Improved inference throughput by 100+ requests/second** by deploying a **GPU-backed containerized backend** with **FastAPI** and **Celery** for asynchronous batch image processing, **cutting end-to-end latency by 35%**.

Ishitva Robotics Systems Pvt. Ltd. — **Ahmedabad**, IN
Junior SDE Intern (Industrial) — Jan 2024 – Jun 2024

- Built a robust **synthetic data generation pipeline** with **OpenCV** (C++) featuring **10+ augmentation strategies** and real-time monitoring, enabling **5x class balance improvement**, **reducing validation error by 12%** on downstream tasks.
- Engineered a **modular Python utility** for efficient **batch data orchestration**, streamlining large image dataset organization through **metric learning and clustering**, significantly **reducing manual efforts by 80%**.

PROJECTS

DeathStarBench – Social Network — Code | Feb 2025 – Mar 2025

- Deployed and orchestrated a **microservices-based** social network using **Docker,** & **Kubernetes**, and migrated from local clusters to **Google Cloud Platform** for scalability and observability.
- Conducted **throughput benchmarking** & **observability** with **Pixie, Jaeger**, and custom load testing to optimize resource allocation, inter-service communication, and **improved latency by 5%**.

Advanced ANPR & FR (National Hackathon organized by MoE, India) — Code | Feb 2023 – Aug 2023

- Built an ANPR & FR pipeline using **YOLOv8-nano** and fine-tuned **ResNet-18**, achieving **92% precision** and **91% recall** on real-world traffic data in challenging conditions.
- Engineered a scalable ML backend using **FastAPI**, integrating **EasyOCR** and **Siamese ResNet** for efficient text extraction and face recognition at the edge.

PUBLICATIONS & RESEARCH

- "**Unlocking Enigmatic Pathways: Empowering Student Dropout Analysis with Machine Learning and Energizing Holistic Investigation**", *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*
- (Under Review) "**EduFace: Employing a Novel Unified Patch Spatial Channel Attention (UPSCA) Architecture for Enhanced Face Recognition in a Real-Time Classroom Attendance System**", *Springer Nature Computer Science (SNCS)*

ADDITIONAL

- **Certifications:** AWS MLA-C01 *(Machine Learning Associate)*, Microsoft Azure DP-100 *(Data Scientist Associate)*
- **Languages:** Python, C++, SQL, JavaScript
- **Cloud & Infra:** AWS, Azure, GCP, Docker, Kubernetes
- **Backend & Data:** FastAPI, Celery, PySpark, Databricks, MySQL
- **Observability and Monitoring:** MLflow, Grafana, Kibana
- **ML & APIs:** OpenCV, PyTorch, SciKit-Learn, TensorFlow, OpenAI API, GCP Vertex-AI API, Hugging-Face Transformers