Rudra Verma

Roorkee, Uttarakhand | +91 73028 83788 | rudraverma2612@gmail.com | LinkedIn | GitHub | LeetCode

PROFESSIONAL SUMMARY

AI/ML Engineer with 1+ year of industry experience building production LLM features, including a Retrieval-Augmented Generation (RAG) pipeline with pgvector and hybrid search, and an evaluation-driven finance assistant. Experienced in training and deploying computer-vision and NLP models (ViT, EfficientNet, BERT) as scalable APIs using FastAPI, Docker and GCP. Research background in AI-based healthcare screening and a 9.08 CGPA in B.Tech CSE. Seeking AI/ML Engineer roles (2027 batch).

EDUCATION

Bennett University – B.Tech, Computer Science & Engineering — Greater Noida, UP | Expected Jun 2027
CGPA: **9.08 / 10** (Semester 6 SGPA 9.37) | Coursework: Artificial Intelligence, Soft Computing, Data Analysis, Data Structures & Algorithms, Operating Systems

Ambition Public School – Class XII, 87.6% — Jhabrera, Uttarakhand | 2023

EXPERIENCE

Devslane – Escecion Technologies Pvt. Ltd. — Dehradun, Uttarakhand
Software Engineer Intern — Nov 2025 – Present
*Full-stack engineer on a client's **fintech SaaS** platform: **React 18 + TypeScript**, **Django REST Framework**, PostgreSQL, Celery, Redis, **Kubernetes (GKE)**.*

- Built a **RAG pipeline** for the product's LLM finance assistant: help-centre docs and 53 page guides chunked and embedded in **PostgreSQL + pgvector**, **hybrid vector + keyword retrieval** with page-aware filtering, relevance thresholds and cited answers; evaluated on a **120-question** English/Spanish set (recall@k).
- Grounded the **LLM assistant** (OpenAI / Anthropic APIs) in live ledger and construction data: figures computed in code, not by the model; native **structured outputs**; **21 golden evaluation cases** and prompt snapshots.
- Shipped **Stripe** billing (idempotency, webhook reconciliation) and multi-company books with safe PostgreSQL backfill migrations.
- Delivered **350+ commits** across 3 codebases with pytest / Vitest regression tests and QA guides, working in Agile sprints with code reviews.

Undergraduate Research Intern – Computational Intelligence Research Group, Bennett University (under Dr. Prashant) — May 2025 – Aug 2025

- Designed a modular 3-phase architecture (preprocessing, **ViT** inference, post-processing) for an AI-based Autism Detection system.
- Containerized the ML pipeline with **Docker** for a **FastAPI** backend serving **PyTorch** models; deployed the inference API to web and mobile apps.

Velcoris (Deep-Tech Startup) – *Senior Associate, Business Research & Analysis* — Sep 2025 – Nov 2025

- Researched **15+ graphene applications** across defense, aerospace, batteries and electronics, and ran market studies on **10+ competitors** to shape early-stage strategy.
- Mentored a team of **5 junior analysts** and worked directly with the founding team on experimentation and documentation, speeding up research cycles.

PROJECTS

Autism Detection System | GitHub · Live — *AI Autism Screening with Face & Age Detection*

- Implemented a **FastAPI** pipeline chaining 3 models (face detection, age estimation, ViT-based ASD classifier) with **<1.5 s** end-to-end inference.
- Deployed with **Docker** on **Google Cloud**, cutting model load time by **40%** and API latency by **30%**; prediction + annotation APIs serve web and mobile apps.

Tech: Python, FastAPI, PyTorch, OpenCV, Docker, GCP

Krishi Kavach | GitHub · Live — *AI Crop Disease Identification Platform with Marketplace*

- Built a CNN/**EfficientNet** model covering **40+ crop diseases** with **95%+ accuracy**, served through a Flask API.
- Developed and deployed a 3-module platform (static site + **React** rainfall app + ML backend) on Vercel with **<2 s** response time.

Tech: Python, Flask, TensorFlow/Keras, React.js, JavaScript, OpenCV, Vercel

NNN-CVE | GitHub — *Nmap/Nikto/Nessus CVE Classification & Vulnerability Prioritization*

- Built an ML-driven CVE classifier over **25,000+ real-world CVEs** using NLP pipelines (TF-IDF, Word2Vec, BERT embeddings).

Tech: Python, scikit-learn, PyTorch, Pandas, NVD/CVE datasets

TECHNICAL SKILLS

ML / DL: PyTorch, TensorFlow/Keras, scikit-learn, OpenCV, Vision Transformers, EfficientNet/CNNs
LLM & NLP: RAG, pgvector, hybrid search, OpenAI & Anthropic APIs, structured outputs, LLM evals, BERT, Word2Vec, TF-IDF
Deployment: FastAPI, Django REST Framework, Flask, PostgreSQL, Docker, Kubernetes (GKE), Google Cloud
Languages & Tools: Python, C++, Java, C, JavaScript, TypeScript, Bash, Git/GitHub, Jira

CERTIFICATIONS

Certified Ethical Hacker (CEH) – EC-Council, Feb 2026 (ECC0971648352) | Google Cybersecurity | Software Engineering – IIT Kharagpur (NPTEL) | Accelerated Computing – NVIDIA