

Pranjal Sapkota

Bhaktapur, Nepal • +977 9807099421 • pranjalsapkota25@gmail.com • [LinkedIn](#) • [Github](#) • [Kaggle](#) • [HuggingFace](#)

CS undergrad with hands-on experience building agentic AI systems, training LLMs from scratch, and deploying ML models to production. Interested in AI/ML engineering roles where I can work on real-world problems at the intersection of language models and applied machine learning.

EDUCATION

B.Tech in Computer Science & Engineering (Core)

Vellore Institute of Technology, Vellore • Expected May 2027 • CGPA: 8.48 / 10

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript

Data tools: Pandas, NumPy, Matplotlib, Seaborn

Agentic AI frameworks: LangChain, LangGraph, LangSmith, MCP

ML/DL: PyTorch, scikit-learn, XGBoost, LightGBM, FastAPI

Web development: Django, Streamlit, Tailwind, Next.js

Deployment: Railway, Render, Vercel

Developer Tools: Git, Docker, Linux, Google Colab, Kaggle, Google OAuth

Databases: MongoDB, PostgreSQL, SQLite, AWS S3

PROJECTS

FluxApply July 2026 | [Github](#) | [Website](#)

- AI-powered web app that tailors resumes and cover letters to job descriptions in seconds, built with FastAPI, LangGraph, PostgreSQL and Next.js.
- Designed an 8-node LLM pipeline with ID-based fact matching to rewrite only relevant resume content, integrated Google OAuth, AWS S3 storage, and Razorpay billing for a production-ready SaaS product.

MCP Agent Jun 2026 | [Github](#) | [Demo](#)

- Built a ReAct agent loop in LangGraph with conditional tool-calling edges across web search, filesystem, GitHub, and Gmail via MCP
- Persisted multi-turn conversation history via async SQLite checkpointer; Streamlit frontend with live streaming and tool-status indicators

SmallLLM – Pathology Language Model Jun 2026 | [Github](#) | [Dataset](#)

- Built a GPT-2 style decoder-only transformer in PyTorch (10.8M and 25M params) with self-attention, residual connections, and layer norm from scratch
- Curated 52M characters from 13 pathology textbooks, cleaned OCR artifacts, published dataset on HuggingFace (with 59 public downloads). achieved val loss 0.8794 (25M model)

Credit Card Fraud Detection Mar 2026 | [Github](#) | [API](#) | [Frontend](#)

- Built a stacked ensemble (RF, XGBoost, LightGBM, LR) on highly imbalanced data using class weighting and undersampling
- Achieved F1 of 0.85 (Precision: 0.92, Recall: 0.79) after threshold tuning; deployed as a FastAPI endpoint on Render

CERTIFICATIONS & ACHIEVEMENTS

- [Intermediate ML](#), [ML Explainability](#), [Data Cleaning](#), [Pandas](#), [Intro to ML](#)- Kaggle, 2026
- [Stanford University's Code in Place](#), 2025
- [Best freshers award](#) - InnovationX Hackathon organized by YANTRA, 2024
- [Response Web Design](#) – Free Code Camp, 2023