

Prasant Bagale

Kathmandu, Nepal

prasantbagale78@gmail.com | 977-9846863875 | Website | LinkedIn | GitHub

Professional Summary

Machine Learning Engineer with hands-on experience in NLP, Computer Vision, and Big Data systems.

Technical Skills

- **Programming Languages:** Python (Strong), Dart (Intermediate), familiar with C++, Java
- **Machine Learning & AI:** PyTorch, TensorFlow, Scikit-Learn, LangChain, spaCy, OpenCV
- **Big Data:** Pandas, NumPy, Dask, Spark (basic)
- **Development:** Flask, FastAPI, RESTful APIs
- **DevOps & Tools:** Docker, Git, MLflow, Weights & Biases
- **Cloud/Edge AI:** Familiar with AWS, on-device inference, Docker containers
- **Core Competencies:** NLP, CV, Reinforcement Learning (intro), Feature Engineering, Hyperparameter Tuning
- **MLOps:** Model versioning, monitoring, retraining, scalable deployment pipelines

Professional Experience

Freelance ML Engineer – Remote

Jan 2025 – Present

- Built MultiModal RAG pipelines integrating LangChain and Qwen 3 for semantic document search across PDFs, websites, and images
- Implemented NLP pipelines for email, and reports summarization, prioritization, and entity tagging using spaCy and BART models

ML Engineer, Omdena – Remote

Sep 2024 – Dec 2024

- Designed data pipeline to process 1TB+ of daily weather, flood, and disease data from multiple sources
- Applied data engineering using Pandas, Dask to handle missing values and optimize preprocessing
- Conducted EDA and correlation analysis to support predictive modeling for outbreak detection

AI/ML Developer, Icebrkr – Kathmandu, Nepal

Mar 2024 – Aug 2024

- Built end-to-end machine learning pipeline: data collection, feature extraction, training, and deployment
- Deployed Random Forest and SVM classifiers; fine-tuned BERT for text classification, Phi-3 for summarization
- Containerized models using Docker; served with FastAPI; monitored with MLflow
- Collaborated with frontend and product teams to integrate models into production environments

Notable Projects

Nepali Text Summarization with MBart

HuggingFace Demo

- Fine-tuned MBart model for Nepali language text summarization using Hugging Face Transformers
- Integrated model into REST API with FastAPI; supported multi-length summarization queries

Novel Recommendation System

HuggingFace Demo

- Developed scalable book recommendation system using Hugging Face embeddings and FAISS
- Served results via FastAPI; optimized response latency and memory usage

Noyaku

Playstore

- Built Flutter-based mobile app to translate Korean/Chinese comics using local LLMs

- Applied OpenCV and OCR to detect speech bubbles and extract/replace translated text in real time
- Enabled offline translation and edge deployment with low-latency LLM inference

Education

Bachelor of Science in Computer Science and Information Technology

Indreni College, Tribhuvan University – Kathmandu, Nepal

- Final Year Project: ShopVia – A price comparison app for e-commerce products using Python and Flutter

Certifications

- **Generative AI with Large Language Models** (Coursera) – Certificate
- **Natural Language Processing with Classification and Vector Spaces** (Coursera) – Certificate

Languages

- **Nepali** (Native)
- **English** (Proficient)