

Pukar Rimal

+977 9866337295 | pukarrimal11@gmail.com | [Linkedin/PukarRimal](#) | [Github/PukarRimal](#) | [Portfolio](#) | Balaju, Kathmandu, Nepal

SUMMARY

Computer Science Graduate specializing in Artificial Intelligence, Machine Learning and Data Science, with hands-on experience in developing LLM powered systems, web crawling, machine learning models, recommendation systems, speech transcription systems, image processing and OCR applications. Passionate about building AI driven, data centric solutions while continuously enhancing practical research and engineering expertise.

EXPERIENCE

JulAiy Tech Labs Mid Level AI/ML Developer

July 2026 – Present
Bafal, KTM

Built multi-level sales forecasting models using Decision Trees and ensemble models including Random Forest, XGBoost, and LightGBM, applying feature engineering techniques to improve model accuracy. Developed an ERP chatbot using Retrieval Augmented Generation (RAG) to provide context-aware responses from enterprise data. Wrote and optimized multiple SQL queries for handling ERP databases, report generation, and data analysis. Created REST APIs using FastAPI and Django, developed backend services, and integrated the APIs into an ERP website using Next.js.

MarginTop Solutions PVT LTD Junior AI/ML Engineer

March 2025 – July 2026
Phulchowk, KTM

Built Retrieval Augmented Generation (RAG) systems integrating Large Language Models (LLMs) with external knowledge sources for accurate, context-aware outputs, and developed Artificial Intelligence (AI) driven recommendation systems for personalized content. Built and deployed Optical Character Recognition (OCR) applications for automated text extraction, and designed web crawling pipelines with automated pagination, workflow automation, and vector embeddings for semantic search. Applied prompt engineering to improve Large Language Model (LLM) response quality and reduce hallucinations, developed Representational State Transfer (RESTful) APIs using FastAPI, and containerized and deployed AI systems using Docker on cloud infrastructure.

AI/ML Engineering Intern

Researched and evaluated web crawling and scraping strategies for dynamic, large-scale websites, and designed scalable Crawl4AI pipelines for high-volume data extraction with pagination, session handling, and workflow automation. Cleaned and normalized unstructured web data using LLM-based workflows and NLP techniques, performed unit and end-to-end testing using pytest, and applied prompt engineering to improve Large Language Model (LLM) response quality and reduce hallucinations.

TECHNICAL SKILLS

Programming Languages: Python, JavaScript, C++

Database: Postgresql, NoSQL(Mongodb), Vector Database(Qdrant, Chromadb)

AI/ML: TensorFlow, Scikit-learn, Transformers, NLTK, PyTorch, NLP, OpenCV, RAG, LangChain, MCP, HuggingFace, Langgraph

Data Visualization: Pandas, Numpy, Matplotlib, Seaborn

Web and Automation: MERN stack, FastAPI, Streamlit, Rest APIs, Redux, JWT Token, Playwright, Crawl4ai

Libraries & Tools: Git, Docker, Postman, Anaconda, Hoppscotch, Coolify, Slack, Jira

PROJECTS

Documate- Helping your document (RAG APP)

[\[Github\]](#) [\[Live Demo\]](#)

- Built DocuMate, an AI-powered document assistant that analyzes user uploaded documents and generates accurate, context-aware responses using a Retrieval-Augmented Generation (RAG) approach.
- Implemented document parsing, intelligent text chunking, vector embeddings, and semantic similarity search to ensure precise information retrieval with clear references to source documents.

ApplyRush – A Smart Job Portal

[\[Github\]](#) [\[Live Demo\]](#)

- Developed a MERN-based job portal application enabling user authentication, job posting, smart job search, and CV upload with skill based job matching using TF-IDF and cosine similarity algorithms.
- Designed a scalable, modular architecture integrating React, Node.js, MongoDB, and Python-based recommendation services for skill-aware job matching, live update and career guidance.

City VS Village Image Classification

[Github](#)

- Developed an automatic image caption generation system that analyzes visual features using a CNN-based encoder and generates human-like captions through a sequence-based language model.
- Designed an efficient end-to-end pipeline combining deep learning, feature extraction, and natural language processing to produce context-aware and semantically meaningful image descriptions.

Scrappy – Google Map Scraper & Insta Automation

[Github](#)

- Developed an automated Instagram analysis system using Python, Playwright, and LLM integration, where users provide a business Instagram URL and the system extracts recent post comments to generate sentiment analysis and customer perception insights about the company.
- Implemented automated browser interaction and data extraction workflows using user behavior simulation techniques such as dynamic delays, scrolling, and interaction patterns to enhance scraper stability and minimize detection.

Agentic Ecommerce System

[\[Github\]](#) [\[Live Demo\]](#)

- Built an Agentic Ecommerce Assistant using RAG and tool-calling AI, combining Qdrant vector search, Supabase inventory management, and Groq LLM to handle product queries, live stock checks, and purchases through natural language.
- Integrated LangChain pipelines for vector embeddings, semantic PDF retrieval, and autonomous tool execution, with a Streamlit-based chat interface

EDUCATION

Bachelor's in Computer Science and Information Technology

National College of Computer Studies (NCCS) Final Percentage: 81.9%

10+2 Science

National School of Science (NIST) Final CGPA: 3.46

Secondary Education Examination

Srijana Gyan Sagar School Final CGPA: 3.70

CERTIFICATIONS

Agentic AI Foundation

[CERTIFICATE](#)

Course Provider: Oracle

Completed Oracle's Agentic AI Foundation course, gaining foundational knowledge of agentic AI concepts, AI agents, and autonomous task execution. Developed an understanding of how AI agents can reason, plan, and interact with tools to accomplish complex tasks.

Building with Claude API

[CERTIFICATE](#)

Course Provider: Anthropic

Completed a comprehensive course on integrating the Claude AI model into systems, covering RAG (Retrieval-Augmented Generation), Prompt Engineering, and Model Context Protocol (MCP).

REFERENCES

- Angel Ghimire – Senior AI Engineer at Autolab Technologies | +977 9843459037
- Rudal Kunwar – Senior Software Engineer at MarginTop Solutions | +977 9866115178