Birat Poudel

Data Scientist | AI/ML Research Engineer

Phone: +977 9861281912 | **Email:** poudel.birat25@gmail.com

Research Portfolio: https:biratpoudel.com.np

About Me

I'm Birat Poudel, an AI/ML Engineer with a strong background in machine learning research, algorithm design, and scalable system development. My work bridges theoretical advancements in artificial intelligence with practical, real-world applications, aiming to create systems that are both intelligent and impactful. I'm motivated by a passion for advancing the frontiers of machine learning through rigorous experimentation, innovative methodologies, and interdisciplinary collaboration.

Research Papers

1. Nepali Sign Language Characters Recognition: Dataset Development and Deep Learning Approaches

- Developed the benchmark dataset for Nepali Sign Language (NSL) with 36 gesture classes and 1,500 samples per class
- Fine-tuned MobileNetV2 and ResNet50 architectures achieving classification accuracies of 90.45% and 88.78% respectively
- Demonstrated effectiveness of transfer learning and fine-tuning for underexplored sign languages in low-resource settings
- Contributed to accessibility research by creating systematic dataset and deep learning approaches for NSL recognition

  Status: Completed (Awaiting Publication)

2. Fine-Tuning DialoGPT on Common Diseases in Rural Nepal for Medical Conversations

- Fine-tuned Microsoft DialoGPT-medium on a synthetically generated dataset of 1,000 doctor–patient dialogues covering ten common diseases prevalent in rural Nepal to create an offline-capable medical conversational AI system
- Designed a two-stage data generation and validation pipeline using Gemini 2.5 Pro and Claude 4 Sonnet, followed by expert medical review, ensuring accuracy, empathy, and contextual relevance of dialogues
- Conducted quantitative (perplexity, loss, token-level F1) and qualitative evaluations by healthcare professionals, demonstrating the model's ability to produce coherent, medically appropriate, and empathetic responses for low-resource healthcare settings

  Status: Completed (Awaiting Publication)

Technical Skills

Programming	C/C++ · Java · Python · JavaScript · TypeScript
Python Libraries	NumPy · Pandas · Scikit-Learn · Matplotlib · Seaborn
ML Frameworks	Tensorflow · PyTorch
Backend	Spring Boot · Flask · Django · FastAPI
Frontend	HTML/CSS · ReactJS · NextJS · Redux
Database	MySQL · PostgreSQL · MongoDB · Redis
Others	Docker · Kubernetes · Firebase · AWS

Work Experience

Leapfrog Technology, Inc. *Kathmandu, Bagmati, Nepal*

AI/ML Research Engineer *June 2024 - Present*

- Built AI systems, including **IVR (Interactive Voice Response) and Conversational Voice AI** for medical patient follow-ups and referrals.
- Designed a **Foundational Models Evaluation Pipeline** that generates automated PDF reports for large-scale model benchmarking.
- Engineered and deployed **AI Agents and MCP Servers** to enable scalable, modular, and autonomous agent orchestration across workflows.

Jobsflow.ai *Kathmandu, Bagmati, Nepal*

ML Engineer 6 months | *Dec. 2024 - May 2024*

- Built AI systems, including an **AI Voice Interviewer** and an intelligent **Chatbot** capable of making tool calls to over ten plus services like Google Calendar, Meet, Gmail, Zoom, etc.
- Developed algorithms for calculating **match score** of a particular applicant for a job based on job descriptions and applicant's resume and answers for the job related questions.
- Implemented **contextual searching, filtering and sorting** using embeddings to enhance candidate selection accuracy.

Fusemachines *Kathmandu, Bagmati, Nepal*

ML Research Engineer *3 months | Sep. 2024 - Nov. 2024*

- Preprocessed and transformed datasets using **NumPy** and **Pandas**, applying advanced **feature engineering** techniques for time series forecasting and machine learning applications.
- Designed and implemented ML models, including **SARIMA**, **LSTM**, **Prophet**, and **XGBoost**, for time series forecasting and predictive analytics, achieving a 15% improvement over previous models.
- Enhanced **RAG-based** systems by optimizing vector storage and retrieval.

Maven Solutions Pvt. Ltd. *Kathmandu, Bagmati, Nepal*

ML Engineer *2 years | August. 2022 - August. 2024*

- Worked on data preprocessing and feature engineering using libraries like **Numpy** and **Pandas** to prepare datasets for model training.
- Developed and implemented machine learning algorithms using libraries such as **Scikit-Learn** for tasks like classification, regression, and clustering achieving an accuracy improvement of 15% over previous models.
- Employed advanced **automation scripts** and conducted precise **web scraping** operations to streamline workflows and gather mission-critical data efficiently.
- Orchestrated the development and seamless integration of **backend APIs**, collaborating closely with cross-functional teams to enhance application functionality and performance.

Certifications

Machine Learning Specialization
Certificate Link:
coursera.org/account/accomplishments/specialization/2HMG68ZZHEVA

Web Development Specialization
Certificate Link:
coursera.org/verify/specialization/G4HWQNSPNDTU

Education

Thapathali Engineering Campus *Thapathali, Kathmandu*

Bachelor of Electronics, Communication and Information Engineering *2019 – 2023*

Relevant Courses: AI, Probability & Statistics, Discrete Structures, Big Data and Web Development.

Research Projects

Automobile License Plate Detection and Recognition | OpenCV, Convolutional Neural Network (CNN), Inception-ResNet-v2, YOLOv8, Google Tesseract and Flask

GitHub Link: https://github.com/Birat-Poudel/Automobile-License-Plate-Detection-and-Recognition

- Developed a comprehensive two-stage system: initial implementation with Inception-ResNet-v2, later enhanced with YOLOv8.
- Integrated state-of-the-art object detection with Tesseract OCR for robust license plate localization and text extraction.
- Achieved robust performance across diverse environmental conditions including varying lighting, angles, and image quality.
- Demonstrated practical applicability for real-time traffic monitoring and automated vehicle identification systems.

Amazon Bedrock Foundational Models Evaluation Pipeline | LLM-as-a-Judge, Quality Metrics, Performance Metrics, Responsible AI Metrics

GitHub Link: https://github.com/Birat-Poudel/AWS-Bedrock-Models-Evaluation

- Implemented an end-to-end evaluation pipeline for Amazon Bedrock models (multi-region support) with performance (latency, throughput, time-to-first-token) and quality metrics (helpfulness, faithfulness, completeness, coherence, etc.).
- Used LangChain orchestration and an LLM-as-a-judge approach for automated, multi-dimensional quality and responsible-AI assessments (harmfulness, bias, refusal appropriateness), plus async processing for concurrent benchmarking.
- Produced structured evaluation outputs and human-readable reports to compare models, tune prompts, and inform safe deployment decisions.

Vector Search, Databases and Retrieval Augmented Generation (RAG)

GitHub Link: https://github.com/Birat-Poudel/Vector-Search-RAG-Projects

1. **Semantic Search for Movie Database**
- Implemented a semantic search feature to find movies using natural language queries. Utilized **Hugging Face sentence-transformers model** and **Atlas Vector Search.**

2. **Gemma Model Document Q&A**
- Developed a Document Q&A project using Gemma Model, Langchain and Streamlit. Utilized **Google Generative AI Embeddings** and **FAIS Vector Store**.

Machine Learning Projects

GitHub Link: https://github.com/Birat-Poudel/Machine-Learning

1. **Titanic Survival Prediction**
- Performed **Exploratory Data Analysis** on Titanic Dataset and obtained valuable insights from the dataset. Implemented a **Decision Tree Classifier** model to predict survival.

2. **Breast Cancer Classification**
- Developed a **Logistic Regression** model to classify breast cancer using 30 input features, achieving a training accuracy of 94.72% and testing accuracy of 92.98%.

3. **Automobile Price Prediction**
- Developed and evaluated automobile price prediction models using **Linear, Ridge, and**

4. **Supermarket Sales Prediction**
- Implemented an **XGBoost Regressor** model to predict supermarket sales. Fine-tuned the

Lasso Regression, with the Lasso model achieving the best performance on testing data (R-squared = 0.8709).

hyperparameters using **GridSearchCV** and increased the test accuracy by about 10%.

Natural Language Processing (NLP) and Computer Vision (CV) Projects

1. **SMS Spam Classifier**

- Implemented Word Cloud generation, Text featurization and built model using **Multinomial Naive Bayes.**
- *GitHub Link:* *https://github.com/Birat-Poudel/Natural-Language-Processing-Projects*

2. **Pokemon Image Classification**

- A multi-class classification project. No. of classes = 3 (Pikachu, Bulbasaur, Charmander)
- *GitHub Link:* *https://github.com/Birat-Poudel/Pokemon-Image-Classification*

IPL Dashboard | Cricket match and team analytics dashboard | **ReactJS, Spring Boot, MySQL**

- Read, processed and wrote data from CSV file to **MySQL** database using **Spring Batch.**
- Implemented business logic, API endpoints and User Interface (Home, Team and Match page).
- *GitHub Link: https://github.com/Birat-Poudel/IPL-Dashboard*