Nirajan Bekoju

+977-9843-965-781 | nirajan.bekoju@gmail.com | nirajanbekoju.github.io
nirajan-bekoju | nirajanbekoju |

Experience

- **FuseMachines [🌐]** *March 2024 - Current*
  Machine Learning Engineer Kathmandu, Nepal
  - Designed and deployed an XGBoost model for effective anomaly detection and diagnostic purposes.
  - Implemented the Google MADI model for anomaly detection and diagnosis, significantly enhancing data cleaning processes and improving the performance of XGBoost model by approximately 9%.
  - Analyzed historical data to evaluate model results, boosting confidence in the model's reliability before its deployment.
  - Specialized Training: Big Data Analytics, Recommendation System, Time Series Analysis and Forecasting.
- **Khwopa College of Engineering [🌐]** *December 2024 - Current*
  AI Researcher | Project Supervisor Bhaktapur, Nepal
  - Assisted Ph.D. candidate in research related to Large Language Models (LLMs), machine translation and chatbot development.
  - Supervised and mentored two student groups in the development of AI-focused minor projects

Education

- **Pulchowk Campus, Tribhuwan University** *2019 - 2024*
  Bachelor in Computer Engineering Lalitpur, Nepal
- **Fusemachines** *2023-2024*
  Microdegree in AI Kathmandu, Nepal

Publications

[1]] N. Luitel, **N. Bekoju**, A. K. Sah and S. Shakya, **"Contextual Spelling Correction with Language Model for Low-Resource Setting,"** 2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2024, pp. 582-589, doi: 10.1109/ICICT60155.2024.10544712.

[2]] N. Luitel, **N. Bekoju**, A. K. Sah, and S. Shakya, **"Can Perplexity Predict Fine-Tuning Performance? An Investigation of Tokenization Effects on Sequential Language Models for Nepali,"** arXiv preprint arXiv:2404.18071, 2024.

Projects

- **Transformer Based Model for Nepali Language Generation and Spelling Correction** *[🐙]*
  Tools: Pytorch, Tensorboard, Transformer
  - Achieved 110 perplexity on test data of Oscar Nepali Corpus
  - Implemented Noisy Channel Model for Spelling Correction.
  - Prototype on huggingface
- **Energy and Price Forecasting - Hitachi** *[🐙]*
  Tools: RandomForestRegressor, SARIMA, TFT
  - Analyzed time series data of energy, weather, and price, and conducted experiments using RandomForestRegressor, SARIMA, and TFT models.
  - Achieved 1.651% Mean absolute percentage error. Best score in the Hitachi Technergy Hackathon 2024.
- **Other Projects**
  - **Wine Quality Classification:** Study of various physio chemical properties of wine to classify their quality | Handle Imbalanced Datasets | Machine Learning Experiments.
  - **Arxiv Paper Recommendation System:** Developed multi-class classifier model to classify the arxiv papers and built paper recommendation system.
  - **Fourier Transform Drawing:** Draw any 2D closed diagram using DFT | Technologies: OpenCV-Python3 for image processing to generate image coordinates and C++, SFML for image drawing.

Technical Skills

- **Programming Languages:** Python, C, C++
- **Data Science & Machine Learning:** Data analysis and visualization, Tensorflow, Keras, Pytorch
- **Experiment Tracking:** MLflow, Tensorboard
- **Big Data Analytics:** Pyspark
- **Cloud Technologies:** AWS, Vertex AI
- **Web Technologies:** Flask, Django, Django Rest Framework
- **DevOps & Version Control:** Docker, Git and Github
- **Specialized Area:** Anomaly Detection and Diagnosis, Recommendation System, Time series analysis and forecasting, NLP, RAG, Prompt Engineering
- **Mathematical & Statistical Tools:** numPy, pandas, matplotlib, Scienceplot, Scikit-learn, SciPy

Honors and Awards

- **Hitachi Technergy Hackathon 2024 - First Place** *Mar 2024*
  Locus, Pulchowk Campus [🌐]
  - Developed the best time series model for energy and price forecasting.
- **AI Competition - First Place** *Dec 2023*
  IT Meet 2023, Kathmandu University [🌐]
  - Developed the best model for the identification of tables and key-value pairs from the provided invoice samples
- **Docsumo Dataverse 2023 - Data Insights Category - First Place** *Jan 2023*
  Locus, Pulchowk Campus
  - Prepared the best data insight report by analyzing and classifying text data from Arxiv papers.

Certifications and Trainings

- **LangChain Chat with Your Data**
- **AWS Academy Graduate - AWS Academy Cloud Architecting**
- **Machine Learning by Stanford University on Coursera**
- **Bayesian Statistics: From Concept to Data Analysis**
- **Neural Network and Deep Learning by DeepLearning.AI**
- **Convolutional Neural Networks on Coursera**
- **Natural Language Processing on Coursera**