Sunil Thapa

Toronto, M6S 3J4 | +1 (437) 970-0802 | sunil43thapa@gmail.com | linkedin.com/in/sunil-thapa99

TECHNICAL SKILLS

- **Languages**: Python, SQL, JavaScript, Shell Scripting, PySpark
- **NLP:** Langchain, Transformers, RAG, Cohere, OpenAI, Agentic AI, NLTK, Spacy, Gensim, Hugging Face, Stanza
- **Web Development Technologies**: HTML, CSS, Flask, Django, JavaScript, jQuery
- **Libraries:** Tensorflow, PyTorch, Keras, Sklearn, Numpy, OpenCV, Scipy, Pandas
- **Frameworks**: Rasa, Django, Flask
- **Analysis Tools**: Tableau Desktop, Power BI, MS Excel, Matplotlib, Seaborn, Pandas
- **Databases**: MySQL, MongoDB, Relational Databases, Oracle SQL, VectorDB
- **Cloud services skills**: AWS (S3, Lambda, CloudWatch), Google Cloud Platform (GCP), Microsoft Azure
- **Tools**: Git, JIRA
- **Operating systems**: Windows, LINUX, Mac OS
- **Soft Skills**: Communication, Teamwork, Leadership, Work Ethic, Time Management, Creativity

PROFESSIONAL EXPERIENCE

Agilepitch **July 2024 - December 2024**

AI Engineer

Responsibilities:

- Developed and deployed **LLM-based** systems utilizing retrieval-augmented generation (RAG) to enhance natural language understanding. Improved document retrieval **accuracy by 28%** through iterative testing and self-corrective algorithms.
- Designed pipelines to process and transform large datasets efficiently, implementing batch processing and distributed computing techniques, which enhanced data processing **efficiency by 40%**.
- Integrated scalable cloud infrastructure using **AWS S3, Lambda, and CloudWatch**, ensuring real-time monitoring and a **13% increase** in system reliability.
- Automated **CI/CD** pipelines with **Terraform** and **GitHub Actions**, reducing deployment time by **40%** and ensuring smooth integration of updates.
- Collaborated with cross-functional teams to fine-tune LLMs for specific applications, improving end-user satisfaction with tailored solutions.

MAGDOTNET **April 2022 - August 2022**

Software Engineer

Responsibilities:

- Built a recommendation system using collaborative filtering techniques, increasing **user engagement by 75%** and optimizing product suggestions for an online platform.
- Integrated Google Maps API to provide location-based services, enabling real-time updates and enhancing user accessibility for navigation services.
- Automated Docker container deployment on **Amazon ECS**, improving system scalability and **reducing response time by 16%** through efficient orchestration strategies.
- Conducted root cause analysis on datasets using **SQL**, uncovering insights that led to a **20% improvement** in operational efficiency.

InfoDevelopers Pvt. Ltd. **July 2019 - March 2022**

Software Engineer

Responsibilities:

- Designed and deployed an **inventory management system** powered by machine learning models to predict stock requirements, **reducing costs by 25%** and enhancing supply chain efficiency.
- Developed a **face-attendance system** with SVM and OpenCV, achieving **95% accuracy** by implementing robust image preprocessing and feature extraction methods.
- Utilized **NLP** techniques to perform sentiment analysis on user feedback, providing actionable insights that improved user satisfaction.
- Automated end-to-end testing with Selenium, streamlining quality assurance processes and reducing manual testing time by **30%**.

- Applied advanced computer vision techniques for real-time video analysis, successfully detecting and tracking objects with **35% higher efficiency**.

EDUCATION

Lambton College | Post Graduate in Artificial Intelligence and Machine Learning
Graduation Date: July 2024

University of Northampton | Bachelor of Science in Computing (Software Engineering)
Graduation Date: September 2019

NLP PROJECTS

Custom LLM Fine-Tuning
- Fine-tuned BERT for sentiment classification, achieving **96%** accuracy on a custom dataset.
- Developed a summarization tool using Hugging Face Transformers to process large documents effectively.

RAG-based Q&A System
- Designed a retrieval-augmented question-answering system focusing on document integrity and factual accuracy.
- Achieved a **30%** improvement in query response precision through self-corrective mechanisms.

Rental Rasa Chatbot
- Built a conversational interface for real estate rental property search using the **Rasa framework**.
- Integrated **natural language processing** capabilities to interpret user requests based on location, price, and property type.
- Leveraged **machine learning and rule-based approaches** for accurate and context-aware responses, achieving seamless user experience.

Hand Gesture Calculator
- Designed an interactive tool using **computer vision** to perform arithmetic calculations via real-time hand gesture recognition.
- Employed **image processing and machine learning** algorithms to interpret gestures, achieving high accuracy in mathematical operations.

Nepali Sentiment Analysis
- Developed a sentiment analysis model for Nepali text, enabling efficient classification of user sentiments in native language datasets.
- Created **word embeddings** for Nepali text using Word2Vec, GloVe, and BERT, addressing the lack of open-source solutions for the language.

RESEARCH PAPER

- "**NepaliBERT: Pre-training of Masked Language Model in Nepali Corpus**" (2023 7th International Conference on I-SMAC, Oct 2023): Developed word embeddings for the Nepali language using Word2Vec, Doc2Vec, and BERT architectures, laying a foundation for benchmarking results on diverse NLP tasks.
- "**Adult Income Prediction Using Various ML Algorithms**" (SSRN, Jan 2023): Conducted comparative analysis of machine learning models on an adult income dataset, achieving 86.3% training accuracy and 86% test accuracy with a Random Forest Classifier and explored feature engineering, selection, and EDA techniques.
- "**Clothes Identification Using Inception ResNet V2 and MobileNet V2**" (SSRN, Oct 2021): Designed a cloth identification model trained on nine categories of clothing, achieving 84.11% test accuracy. Utilized advanced deep learning architectures to improve classification performance.