

Aayush Malla

📍 Kathmandu, Nepal ✉ mallaaayush59@gmail.com ☎ (+977) 9851203045 in Aayush Malla

Research Interests

AI, optimization modeling, human-centered system design, behavioral analytics, and geospatial intelligence, with a focus on applying machine learning and generative AI to study human decision making and animal behavior, advancing data-driven systems for real-world impact.

Education

Kathmandu University, Nepal *Bachelor of Computer Engineering* 2016 – 2020

- GPA: 3.60/4.00 (**Top 5% in the class**)
- **Relevant Coursework:** Machine Learning, Algorithms & Complexity, Data Structures, Database Management Systems, Artificial Intelligence, Computer Architecture & Organization, Computer Graphics, Microprocessor and Assembly Language, Compiler Design, Human Computer Interaction, Statistics & Probability
- **Capstone Projects:** Nepali License Plate Recognition — (**Best Final Year Project**)

Experience

Data Engineer

Fusemachines

Kathmandu, Nepal

March 2022 – Present

- **Quantum ERP Data Integration:** Built modular ETL pipelines in Azure ADF to integrate ERP data across HR, Finance, and Procurement systems. Developed PySpark and Spark SQL scripts in Databricks for data cleaning, transformation, and modeling, enabling structured ingestion into Azure Data Lake. Automated workflows and CI/CD using ADF, Databricks, and Azure DevOps, reducing manual effort by 50%, minimizing release errors, and ensuring over 95% data consistency.
- **Internal Analytics:** Designed ETL pipelines in Azure ADF to extract organizational data, including food operations, attendance, project metrics, and Jira logs. Integrated data into Azure Data Lake and built Power BI dashboards for leadership reporting. Leveraged Azure Synapse, Monitor, and Logic Apps to streamline scheduling, monitoring, and alerting.
- **Instructor & Mentor:** Delivered training sessions on SQL, PySpark, Spark SQL, Databricks, and Azure technologies. Mentored junior engineers in ETL development, CI/CD automation, and best practices for scalable data engineering.

Software Engineer

COTIVITI

Kathmandu, Nepal

Feb 2021 – 2022 Feb

- **ETL System for Health Care:** Developed PL/SQL ETL workflows for 20+ insurance clients, enabling custom data processing and automated claim validation. Built reusable ODI pipelines to normalize and centralize raw health insurance data from 20+ providers. Analyzed healthcare claims data and created Oracle Apex dashboards to support fraud detection and cost optimization. Automated routine ETL and data quality tasks with PL/SQL scripts, reducing manual effort and accelerating claim validation processes.
- **Data Hackathon:** Participated in a Data Hackathon and automated dynamic CRUD operations across multiple tables for onboarding new healthcare clients, improving efficiency and reducing manual effort.

Teaching Experience

Teaching Assistant – National Institute of Science and Technology, Nepal

May 2022 – March 2023

- Taught Object-Oriented Programming and Database Management Systems to undergraduate BCA students.
- Led lab sessions and tutoring, providing hands-on guidance in problem-solving and SQL programming.
- Contributed to assignment design and evaluation, fostering an interactive and collaborative learning environment.

Projects & Research Experience

ML Tool Suite (RAG, Summarizer, Keyphrases, Language→English, Sentiment) — MCP-enabled

2025

- Built an ML toolkit: BM25+MMR extractive Q&A, TextRank summarizer, TFIDF keyphrases, and language detection → English (MarianMT offline with online fallback).
- Implemented translation routing with Auto / Offline / Online modes and timeout-based fallback for on-device or cloud use.
- Designed a calibrated sentiment ensemble (custom ML + VADER + AFINN) with auto-learning (pseudo-labeling) for periodic self-retraining.
- Shipped a clean Streamlit UI and an MCP stdio server for tool interoperability (e.g., MCP Inspector/clients).
- **Technologies:** Python, scikit-learn, NLTK/VADER, AFINN, Hugging Face (MarianMT), rank-bm25, NetworkX, Streamlit, fastmcp (MCP)

Nepali License Plate Recognition

2020

- Developed a machine learning-based License Plate Recognition system to detect and extract Nepali license plate numbers from images with 95% accuracy.
- Created custom datasets by collecting and labeling thousands of Nepali character images for supervised model training.
- Built and trained an SVM model in OpenCV to improve recognition accuracy and feature extraction.
- **Technologies:** Python, OpenCV, SVM, Image Processing

Retail Data Analysis (US-based Company)

2022

- Analyzed multi-million row retail datasets using Apache Spark to uncover seasonal sales trends and product-level performance insights.
- Cleaned and transformed large retail datasets using Spark DataFrames to ensure high-quality input for BI reporting.
- Built data pipelines with Apache Spark DataFrames to transform and load processed retail data into SQL-based storage systems for downstream analytics.
- **Technologies:** Apache Spark, PySpark, SQL, DataFrames

Xplore Nepal

2019

- Developed an Android application using Java to serve as a tourism-focused map app for Nepal.
- Integrated features to display local areas, restaurants, and cultural heritage sites with detailed information.
- Added functionality for booking local guides, enhancing accessibility for tourists.
- **Technologies:** Java, Android Studio, Google Maps API

KU Digitized Wall

2018

- Developed a web application similar to Google Classroom for managing courses and class schedules.
- Enabled teachers to create virtual classrooms, post updates and assignments, and allowed students to submit assignments.
- Built using JavaScript with a local database for storage and retrieval of course-related data.
- **Technologies:** JavaScript, Node.js, Express.js, SQLite, REST APIs

Achievements and Awards

First Runner up – Data Hackathon:	Awarded for developing “Content Generator”, an AI-driven tool that automated dynamic content creation using NLP and generative models.
Data Engineering Trainer:	Taught PySpark, Python, Flask, SQL, and Azure to data engineering trainees through hands-on sessions.
Merit-Based Scholarship:	Awarded a merit-based scholarship in the 7th semester for outstanding GPA performance.
Outstanding Project Award:	Recognized for developing a machine learning-based Nepali License Plate Recognition system using SVM and OpenCV, achieving 95% accuracy.

Skills

Programming Languages:	SQL (5+ years), Python (4+ years)
Databases & Tools:	PostgreSQL, SQL Server, MSSQL, MySQL, Git, GitHub, Postman, MCP Inspector
Cloud Technologies:	AWS (S3, AWS Glue, Athena, MWAA), Azure (ADF, ADLS Gen2, Synapse, Delta Live Tables, Azure DevOps)
Orchestration Tools:	Apache Airflow, Databricks Workflows, Delta Live Tables, ADF Pipelines
Frameworks & Libraries:	PySpark, Pandas, NumPy, Matplotlib, Scikit-learn, OpenCV, Flask, Streamlit, Hugging Face Transformers (MarianMT), NetworkX, rank-bm25, NLTK/VADER, AFINN, PyPDF
NLP/IR Methods:	RAG (BM25), Top-K Retrieval, MMR (Maximal Marginal Relevance), TextRank, TF-IDF Keyphrases, Language Detection & Translation (MarianMT/LibreTranslate)
Machine Learning:	Supervised & Unsupervised Learning, Feature Engineering, Model Evaluation, Clustering (K-Means, DBSCAN), Random Forest, SVM, Sentiment (ensemble: ML + VADER + AFINN), calibration, pseudo-labeling/active learning
Agents/Protocols:	Model Context Protocol (MCP), fastmcp (studio tools)
Languages:	Nepali, Hindi, English