

Siddhartha Shakya

[GitHub](#) | [Linkedin](#) | siddhuushakya@gmail.com | [+977 9810367988](#)

SUMMARY

AI & Data Engineering graduate with hands-on experience in ML development, ETL pipeline orchestration, and MLOps. Demonstrated ability to build end-to-end solutions through an AI internship and independent projects spanning data engineering, NLP, and model deployment. Eager to contribute and grow in a junior Data Engineer or ML Engineer role.

WORK EXPERIENCE

AI Intern <i>Mach24 Orbitals</i>	Oct 2025 - Feb 2026
- Built Streamlit dashboard for environmental monitoring using Landsat-8, performing time series analysis on vegetation indices, land surface temperature, and urbanization patterns.	
- Researched ML approaches for Above Ground Biomass estimation using Sentinel-2 imagery via Google Earth Engine API, evaluating regression models for biomass prediction from remote sensing data.	

Graduate Teaching Assistant <i>Herald College Kathmandu</i>	Aug 2024 - Sep 2025
- Led workshop sessions on parallel computing and performance optimization, covering parallel programming paradigms and GPU computing.	
- Conducted tutorials on Big Data technologies along with practical labs covering Hadoop ecosystem, MapReduce programming, distributed data processing frameworks, and real-time streaming. Guided students through implementation exercises involving large-scale data analysis and processing techniques.	

PROJECTS

Weather ETL Pipeline	Repo Link
Developed a fully automated ETL pipeline to ingest weather data. Engineered a containerized infrastructure with Docker Compose to orchestrate nested data transformations and loading into PostgreSQL, implementing robust error-handling and deduplication logic for 100% data reliability.	
Tech Stack: Apache Airflow, Python, PostgreSQL, PgAdmin, Docker, Docker Compose, Astronomer, Git, OPEN METEO API.	

Movie Recommendation System	Repo Link
Built a semantic movie recommendation engine using Transformer embeddings across 13,000+ titles. Implemented unified logging system tracking data ingestion, cleaning, and model training. Optimized deployment with multi-stage Docker setup and automated CI/CD testing using GitHub Actions and Pytest.	
Tech Stack: Python, HuggingFace, Scikit-learn, NLTK, FastAPI, Jinja2, Docker, Docker Compose, GitHub Actions (CI), Pytest, logging.	

Text Summarization MLOps Pipeline	Repo Link
Built an automated text summarization pipeline using NLP models to convert long-form text and dialogues into concise summaries through a multi-stage MLOps workflow.	
Tech Stack: Python, FastAPI, PyTorch, HuggingFace, NLTK, ROUGE, Docker, Docker Compose, Git, GitHub Actions (CI/CD), Pandas, PyYAML.	

EDUCATION

2026 - present PG Diploma in Data Science at **Manipal University**
2022 - 2024 BSc (Hons) in Computer Science at **Herald College** (Awards: First Class Honors)

CERTIFICATIONS

The Complete SQL Bootcamp: Zero to Hero Udemy 2026
Complete MLOps Bootcamp Udemy 2026

SKILLS

Programming Languages	Python, SQL, Java, C/C++, Bash/Shell	
Databases	PostgreSQL, MongoDB, MySQL, Redis	
Cloud Platforms	AWS, Google Cloud Platform (GCP)	
AI & ML Frameworks	PyTorch, HuggingFace Transformers, Keras, Scikit-learn	
Data Visualization & BI	Matplotlib, Seaborn, Plotly, Excel	
Data Engineering	Apache Airflow, Apache Spark, Hadoop, ETL Pipelines, Pandas	
DevOps & MLOps	Docker, Docker Compose, Git, GitHub Actions (CI/CD), MLflow, DVC, YAML	