Siddhartha Shakya

GitHub | Linkedin | siddhuushakyaa@gmail.com | +971 548889367 | Dubai, International City

Summary

Computer Science graduate with practical experience across the full data lifecycle, from building automated ETL pipelines to deploying NLP models in production. Gained real-world exposure through an AI internship and teaching assistantship, with a strong foundation in both data engineering and machine learning. Looking to bring these skills to an entry-level Data or AI engineering role.

Skills

Programming	Python, SQL, C/C++, Bash/Shell
Databases	PostgreSQL, MongoDB, MySQL
Cloud Platforms	AWS, Google Cloud Platform (GCP)
Data Engineering	Apache Airflow, Apache Spark, Hadoop, Pandas, NumPy
DevOps & MLOps	Docker, Docker Compose, Git, GitHub Actions (CI/CD), MLflow, DVC
AI & ML	PyTorch, HuggingFace Transformers, Scikit-learn

Work Experience

AI Intern | *Mach24 Orbitals* Oct 2025 – Feb 2026

- Built a Streamlit dashboard processing Landsat-8 satellite imagery across 5+ years of time series data, enabling real-time visualization of vegetation indices, land surface temperature, and urbanization patterns for environmental monitoring.
- Evaluated 3+ regression models (Random Forest, XGBoost, SVM) for Above Ground Biomass estimation using Sentinel-2 imagery via Google Earth Engine API, identifying XGBoost as the top performer with the lowest RMSE across test plots.

Graduate Assistant | *Herald College Kathmandu* 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.

Projects

Weather ETL Pipeline Repo Link

Automated end-to-end ETL pipeline ingesting live weather data from the Open-Meteo API, processing and loading into PostgreSQL with deduplication logic, error handling, and retry mechanisms. Containerized with Docker Compose and orchestrated via Apache Airflow with daily DAG scheduling.
Tech Stack: Apache Airflow, Python, PostgreSQL, PgAdmin, Docker, Docker Compose, Astronomer, Git, OPEN METEO API.

Movie Recommendation System Repo Link

ontent-based recommendation engine across 13,000+ movie titles using HuggingFace Transformer embeddings for semantic similarity, served via a FastAPI backend. Deployed with multi-stage Docker builds, achieving sub-200ms response times, with automated CI/CD testing via GitHub Actions and Pytest.
Tech Stack: Python, HuggingFace, Scikit-learn, NLTK, FastAPI, Jinja2, Docker, Docker Compose, GitHub Actions (CI), Pytest, logging.

Education

2026 - pres. PG Diploma in Data Science, **Manipal University**
2022 - 2024 BSc (Hons) Computer Science, **Herald College** (First Class Honors)

Last updated: February 27, 2026