

Siddhartha Shakya

[GitHub](#) | [LinkedIn](#) | siddhuushakya@gmail.com | [+971 55 183 0660](#) | [UAE, Dubai](#)

PROFILE SUMMARY

Computer Science graduate with 1-2 years of hands-on experience building data pipelines, AI applications, and analytics solutions. Skilled in Python, SQL, and cloud platforms. Passionate about delivering real-world AI and data engineering solutions that drive business impact.

TECHNICAL SKILLS

Languages	Python, SQL, C/C++
Data Engineering & Pipelines	Databricks, Apache Spark, Apache Airflow, ETL/ELT, Data Pipelines
AI & Data Applications	PyTorch, Scikit-learn, Pandas, NumPy
Cloud Platforms	GCP (BigQuery, Earth Engine, Cloud Storage), AWS
Analytics & Visualization	Plotly, Seaborn, Matplotlib, Jupyter Notebook, Google Colab, Dashboards
Databases	PostgreSQL, MySQL, MongoDB
DevOps & Tools	Docker, Git, GitHub, GitHub Actions, Streamlit

WORK EXPERIENCE

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

- Developed and maintained Python applications and data pipelines processing 1,000+ imagery tiles monthly; automated ETL processes and API integrations using GCP to ensure data quality and reliability.
- Built and evaluated machine learning models for forecasting, evaluating performance across 450+ test cases to achieve a 23% lower RMSE (14.2 vs. 18.5 Mg/ha) than the Random Forest baseline.
- Supported decision-making by constructing interactive Streamlit dashboards and translating complex data into actionable insights for cross-functional stakeholders.

Graduate Teaching Assistant | *Herald College Kathmandu* **Aug 2024 – Sep 2025**

- Led technical workshops for 120+ students on programming paradigms, SQL, and big data technologies including Spark and real-time streaming pipelines.
- Built and documented data applications, ETL workflows, and pipeline design examples, improving student assessment scores by 15% through clear communication.

PROJECTS

EduFin Data Investigation & Analytics | 2026 [Project Link](#)

- Built and optimized data pipelines in Databricks using PySpark and SQL to ingest and analyze a 15,000+ student loan portfolio, resolving a Rs. 12 Crore data discrepancy.
- Translated ambiguous business requirements into practical solutions, segmenting risk factors and developing interactive dashboards to support stakeholder decisions.
- **Tech Stack:** Python, Databricks, PySpark SQL, Delta Tables, Data Pipelines, Data Visualization

Geospatial Environmental Monitoring Dashboard | 2025 [Project Link](#)

- Developed a Python application and data pipelines utilizing Streamlit and APIs to process USGS satellite datasets with a 200ms average load time.
- Designed interactive dashboards to visualize key performance indicators (KPIs), delivering real-time reporting interfaces to simplify complex data.
- **Tech Stack:** Python, Pandas, Streamlit, Folium, APIs, Docker, GCP

EDUCATION

PG Diploma in Data Science **Feb 2026 – present**

- *Manipal Academy of Higher Education*

BSc (Hons) Computer Science **Feb 2022 – Jun 2024**

- *University of Wolverhampton*
- Honors: **First Class**

CERTIFICATIONS

PySpark: Building Data Pipelines	LinkedIn Learning	Certificate
SQL Bootcamp	Udemy	Certificate
MLOps Bootcamp	Udemy	Certificate