

Kamesh Kotwani

kameshkotwani@gmail.com | +447932851567 | LinkedIn | GitHub | Portfolio

PROFILE

Azure Certified Data Scientist with over **2 years'** of experience applying **ML, NLP, LLMs**, and **big data analytics** to solve business problems. Skilled in building **end-to-end ML/LLM pipelines**, **optimising data workflows**, and deploying **scalable solutions** on **AWS/Azure** using modern **LLMOps** practices. Delivered measurable results, including reducing **pipeline runtimes by 33%**, cutting **cloud costs by 8%**, and improving overall **model accuracy**.

EXPERIENCE

University of Sheffield

Machine Learning Engineer

Sheffield, UK

Feb 2025 – Sep 2025

- Implemented a **Faster R-CNN with Swin Transformer** model, improving **malaria parasite detection accuracy by 12% over YOLO models** on 4,000 blood-smear images.
- Developed an **end-to-end ML pipeline** (data annotation, preprocessing, model training on **HPC**), cutting **development cycles by 50%**.
- Deployed MLflow to **standardise experiment tracking**, reducing experiment setup time by 40% and enabling reproducibility across projects.

IQVIA

Data Engineer

Bangalore, India

Jun 2022 – Aug 2023

- Developed **Spark based ETL pipelines** to process **10M+ healthcare records weekly**, achieving **99.8% data quality** for pharma analytics teams.
- Refactored high-load **Spark jobs** and optimised **SQL queries**, reducing pipeline runtimes by **33% (12h → 8h)** and reducing **Azure compute costs by 8% annually**.
- Built **Airflow automation scripts** that reduced **job failure rates by 40%**, improving **uptime** and incident response speed.
- Partnered with **Data Scientists** to deploy ML models into production via **MLOps (CI/CD, MLflow, Docker)**, ensuring **scalable model delivery** for business-critical applications.

CDAC

Machine Learning Engineer

Pune, India

Oct 2021 – Jun 2022

- Developed an **NLP analytics dashboard** (Django + SpaCy) that accelerated domain experts' **document review by 40%**, directly supporting regulatory projects.
- Delivered a **Streamlit-based validation platform** that halved tester effort and **doubled model release cycles**, streamlining deployment workflows.
- Provided **technical mentorship** to junior ML engineers, strengthening team capability in **NLP, Python best practices**, and **model evaluation**.

SKILLS

- Machine Learning & Analytics:** Python, SQL, PyTorch, TensorFlow, Scikit-learn, NLP (Transformers/LLMs), Power BI, Matplotlib, XGBoost/LightGBM
- Data Engineering & Cloud:** Spark, Airflow, Azure (Data Factory/ML/AutoML), AWS (EC2/S3, SageMaker), Docker, MLflow, LLMOps/MLOps
- Collaboration & Soft Skills:** Git, Confluence, JIRA, MySQL, MongoDB, Stakeholder Communication, Problem-Solving, Team Collaboration, Adaptability, Critical Thinking

PROJECTS

- AI-Powered RAG Chatbot (Streamlit + LangChain + FAISS + OpenAI)** (GitHub): Built a retrieval-augmented generation chatbot using **LLMOps**, cutting **document search time by 90%** with retrieval-based QA and cited responses.
- Multi-Modal Product Classification (BERT + ResNet + FastAPI)** (GitHub): Automated product tagging for e-commerce, processing 10K+ SKUs weekly, **reducing manual tagging by 80%**, deployed as an **API service**.
- Cassava Leaf Disease Detection (TensorFlow + Keras + Streamlit)** (MSc Dissertation): Achieved 0.87 F1 score on cassava disease dataset; deployed as an interactive app for **agricultural decision support**.

EDUCATION

- MSc Data Science**, University of Sheffield — Merit Sep 2023 – Sep 2024
- BEng Computer Science**, GGCT — Distinction Aug 2016 – Sep 2020

CERTIFICATIONS

- Data Analysis using Power BI**, CampusX (earned Jul 2025) Verify credential
- Microsoft Certified Azure Data Scientist Associate (DP100)**, Microsoft (earned Dec 2024) Verify credential
- Technical Diploma Big Data Analytics**, CDAC (earned 2021)