

# Kamesh Kotwani | Data Scientist

Sheffield, UK | kameshkotwani@gmail.com | +447932851567 | LinkedIn | GitHub

## PROFESSIONAL STATEMENT

---

Microsoft Certified Azure Data Scientist Associate with experience in building and deploying machine learning models, MLOps pipelines, and working on large-scale AI solutions. Proficient in Python, Spark, and cloud platforms (AWS, Azure), with expertise in tools such as MLflow and DVC for model tracking and versioning. Skilled in solving classification problems, natural language processing (NLP), reinforcement learning, and generative AI. Experienced in delivering scalable ML solutions using FastAPI, Django, and big data tools like Hadoop and PySpark.

## TECHINCAL SKILLS

---

**Programming:** Python, Scala, R, SQL, Bash

**Cloud & DevOps:** AWS (SageMaker, EC2, S3), Azure (ML, Fabric, Synapse, Databricks), Docker, Kubernetes, CI/CD, Ansible, Airflow

**Machine Learning Tools:** PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, Gradient Boosting, MLflow, DVC, FastAPI, GenAI, LLMs

**Data Visualisation:** Power BI, Tableau, KNIME, IBM SPSS

**Big Data Tools:** Spark, PySpark, Hadoop, MongoDB, Elasticsearch, Kafka

## EXPERIENCE

---

### MACHINE LEARNING ENGINEER | CDAC

October 2021– June 2022

- Led the **Kanthasth project**, developing an AI chatbot using JavaScript and NLP, reducing user queries by **35%**.
- Built an NLP dashboard with a team of 10+, cutting analytics time by **50%**.
- Developed translation models for Hindi and Marathi with **85% accuracy** using transformers.
- Implemented **Topic Modelling** via Django and NLP, boosting research efficiency by **25%**.

### BIG DATA ENGINEER | IQVIA

June 2022– August 2023

- Ensured data security and compliance with organisational data policies while handling large datasets.
- Built **Airflow pipelines** with Spark and Elasticsearch, improving platform reliability.
- Automated Spark SQL tasks using Bash, increasing productivity by **10%**.
- Developed a Hadoop cleanup pipeline, preventing storage disruptions.
- Optimised Spark runtimes by **30%** in collaboration with Spark and Hadoop teams.
- Enhanced CI/CD pipelines using Kubernetes, Docker, and GitLab, ensuring rapid deployments.

### TA – BIG DATA ANALYTICS | University of Sheffield

February 2025– Present

- Assisted in teaching **Big Data Analytics**, guiding students on Spark, Hadoop, and PySpark.
- Conducted lab sessions and tutorials on real-world big data solutions.
- Provided feedback on projects involving data mining and distributed computing.
- Developed course materials and assessments in line with industry standards.

## CERTIFICATIONS

---

**Microsoft Certified: Azure Data Scientist Associate (DP-100):** Verify Certification

## EDUCATION

---

MSC in Data Science | University of Sheffield | Grade: 2:1

Sept 2023 – Sept 2024

- Data Analysis, Data Visualisation, Data Mining, Database Design, Data Society, Business Intelligence.

## MACHINE LEARNING PROJECTS

---

### Sentiment Analysis with (FastAPI, Sklearn, DVC & MLflow) on AWS

- Engineered and deployed a sentiment analysis model using **Logistic Regression** with **Bag of Words & Count Vectorizer**, delivering real-time predictions via **FastAPI**.
- Designed a fully automated ML pipeline with **DVC** for version control and **MLflow** for experiment tracking, ensuring **reproducibility and seamless model iterations**.
- Deployed on **AWS** for scalable, production-ready inference, optimising performance and reliability in a **cloud environment**.

### Comprehensive WhatsApp Chat Analytics

- Created an **NLP-based tool** for sentiment analysis and behavioural trends, extracting insights like active chat times and frequently used words.
- Delivered actionable user behaviour insights, demonstrating **analytical thinking** and effective use of **text mining techniques**.

### Dynamic Sales Performance Dashboard (Power BI)

- Designed an interactive **Power BI** dashboard with heatmaps and drill-throughs to track revenue, profit margins, and growth trends.
- Leveraged **Power Query** and **DAX** for real-time analysis, enhancing **stakeholder decision-making** and operational efficiency.

### High-Performance Stock Price Predictor (LSTM Models)

- Predicted stock prices with a **Root Mean Squared Error (RMSE) of 5**, analysing 5 years of historical data with features like **RSI, EMA, and SMA**.
- Reduced training time by **40%** using **MLflow** for tracking, **DVC** for data versioning, and **TensorFlow GPU** acceleration.

### AI-Driven Breast Cancer Classification

- Achieved **97% accuracy** using **XGBoost, Decision Trees (95%), and Logistic Regression (84%)** on a dataset of 10,000 records.
- Analysed **31 clinical features**, showcasing **strong statistical analysis** and machine learning expertise in predictive healthcare.

### Rapid Cassava Leaf Disease Diagnosis (Deep Learning)

- Developed models like **AlexNet (68%), ResNet (55%), and ImageNet (78%)**, showcasing expertise in **model development and evaluation**.
- Reduced training time by **75%** using **MLOps tools** such as **DVC, CI/CD, and TensorFlow GPU**.

### Smart Recommender Systems for Books and Movies

- Built a content-based recommender system using **cosine similarity** and **feature engineering** to improve recommendation accuracy.
- Improved performance through **data cleaning** and preprocessing, providing a better user experience.