

Kamesh Kotwani | UK-PSW

✉ kameshkotwani@gmail.com ☎ +44 7932851567 📍 United Kingdom 🌐 Kamesh 🏠 GitHub

PROFILE

Data Scientist with 2+ years of experience in Python, Scala, and Big Data tools like Hadoop, Spark, and Elasticsearch. Keen on optimizing ETL pipelines, automating tasks with Bash/Python, and working in Linux environments. Skilled in Agile and Test-Driven Development (TDD) methodologies, with strong expertise in data processing and pipeline automation for large-scale datasets.

EDUCATION

MSc, Data Science 09/2023 – 09/2024
University of Sheffield

B.E. Computer Science 08/2016 – 08/2020
Gyan Ganga College of Technology

SKILLS

Soft Skills

Effective Communication, Collaboration, Curiosity and Continuous Learning, Problem-Solving Mindset, Adaptability, Time Management, Attention to Detail, and Creativity.

Technical Skills

Python, Scala, R, PL/SQL, MongoDB, Spark, Hadoop, Elastic Search, Airflow, PyTorch, Keras, Tensorflow, Linux, Bash, Ansible, CI/CD Pipelines, Big Data Analytics, Bash scripting, SOA, predictive modelling, regression analysis, gradient boosting, XGBoost, Catboost.

API & Cloud Computing

MLflow, AWS, Azure, GCP, Data Version Control, GitLab, GitHub, Git.

Analytics Tools

Power BI, Tableau, KNIME Analytics, IBM SPSS.

EXPERIENCE

Data Engineer 06/2022 – 08/2023
IQVIA

- Collaborated with a team of 10+ members, optimizing ETL Spark and Airflow pipelines, reducing run time by 30%, and decreasing runtime from 10 to 7 hours.
- Enhanced efficiency in team workflow and automated redundant tasks using bash and Python, which increased operational boost by 10%.
- Program airflow pipelines for cluster cleanup automation to solve space deadlocks, which received high appreciation.
- Crafted Ansible scripts to deploy multiple-system configurations for frameworks and libraries.
- Strived in collaboration with 5+ doctors in the team to create and analyze patient data for machine learning pipelines.

NLP / ML Project Engineer 10/2021 – 06/2022
Centre for Development of Advanced Computing

- Facilitated the kanthasth project, involving the development of an AI chatbot using JS/NLP (<https://kanthasth-rajbhasha.gov.in/>), which reduced manual queries from users by 35%.
- Joined a 20+ team to create an NLP and text analytics dashboard for research purposes, reducing analytics time by 50%.
- Created Language Translation Models across Indic languages with an accuracy of 85% using transformers for languages such as Hindi and Marathi.
- Engineered and implemented frontend user interfaces for text analysis using Django and NLP, enhancing user accessibility and leading to a 25% increase in research analysis.

PUBLICATIONS

Unlocking the Power of the Keras Functional API: A Titanic Case Study [↗](#)

This article demonstrates Keras using the Titanic dataset to build a deep learning model using Keras Functional API for multi-output prediction, showcasing the ability to predict multiple target columns with a single model.

PROJECTS

Stock Price Prediction using LSTMs

- Utilized data augmentation to create stock portfolios and calculate features such as RSI, EMA, and SMA to predict the close price using features from the previous 60 days.
- Utilized MLflow for experimentation tracking stored on dagshub and AWS, DVC for data versioning, keras and tensorflow with GPU for training LSTM, achieved root mean squared error of 5 in test data.
- Worked on this project with the collaboration of a fellow classmate with knowledge of finance data.

Breast Cancer Classification

- Achieved 97% accuracy in cancer classification by implementing diverse ML algorithms on 31 features, enhancing predictive healthcare solutions in Kaggle competition.
- The accuracy of test data on Kaggle were Decision Trees (95%), Logistic Regression (84%), and Random Forest (97%).

Leaf Disease Detection using Deep Learning

- Experimented with CNN architectures to create and test models with accuracy, AlexNet (68%), ResNet (55%), and ImageNet (78%).
- Trained models in iterations using MLOps (experiment-tracking), DVC (Data Version Control), CI/CD, Keras and Tensorflow for GPU acceleration, reducing training time from 2-3 hours to 40-50 minutes on large datasets.

Recommender Systems [↗](#)

- Designed a book and movie recommender using content filtering, which recommends books and movies based on the vectors' cosine similarity score.
- Indulged in data cleaning, data analysis, kNN Imputation for missing values, feature engineering and feature scaling to increase models' performance and recommendations.

WhatsApp Chat Analyser [↗](#)

- Produced an NLP tool for sentiment analysis and text mining capabilities to derive valuable insights using WhatsApp chat data.
- Analyzes chats and gives exciting insights about user chats, such as the most active time, frequently recycled words, and contacts.

CERTIFICATES

- Machine Learning and Data Science with Python [↗](#)
- Databricks Associate Developer Certification [↗](#)
- Big Data Analytics [↗](#)
- Complete Hands On with Apache Airflow [↗](#)
- Apache Spark 3 [↗](#)
- Apache Kafka [↗](#)