

Jay Dugad

Data Scientist | AI, Machine Learning & Data Engineering | Python, SQL, NLP, AWS | MSc Data Science

London | jaydugad10@gmail.com | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

SUMMARY

Data Scientist with experience building and **deploying ML and NLP systems** in production environments, with hands-on work across **Python, SQL, cloud platforms**, and scalable **data pipelines**. Strong foundation in **machine learning, distributed systems, analytics engineering**, and problem-solving, with experience translating complex data challenges into practical business outcomes. MSc in Data Science (Distinction) and a personal [recommendation](#) from Thomas Wood, Director, Fast Data Science.

SKILLS

- **ML & Data:** Machine Learning (Supervised & Unsupervised) | NLP & LLMs | Feature Engineering | Statistical Modelling | Time Series Forecasting | A/B Testing
- **Programming:** Python (Pandas, NumPy, Scikit-learn) | SQL | Bash Scripting | C/C++ (Foundational OOP)
- **Cloud & Infrastructure:** AWS | Azure | Docker | Git/GitHub | CI/CD | MLOps | Distributed System
- **Core Competencies:** Problem-Solving | Algorithms & Data Structures | Cross-functional Collaboration | Stakeholder Communication | Agile Delivery

WORK EXPERIENCE

Fast Data Science – Data Scientist

Dec 2024 – Present | London, UK

- Deployed and maintained a **Hugging Face** hosted **NLP/LLM pipeline** used by **20+ researchers**, enabling scalable unstructured text analysis on **cloud infrastructure**
- Engineered a molecular entity recognition module using **OMOP** and **SMILES data**, increasing **drug detection accuracy by 30%** in clinical **NLP pipelines**, involving **algorithm design, schema validation, and performance tuning**
- **Reviewed and merged 50+ GitHub pull requests**, contributing to **code quality** and accelerating team **delivery velocity by 25%**
- Validated LLM-generated outputs for **schema alignment** and consistency, ensuring suitability for downstream analytical use in **distributed data workflows**

getCredible – Data Analyst (Student Ambassador)

Feb 2024 – Jan 2025 | London, UK

- **Built SQL models** for **funnel and transaction trend analysis**, boosting campaign **ROI by 30%** through data driven visibility into **user journeys**
- Designed **automated forecasting** and **A/B test pipelines**, **reducing manual reporting effort by 25%** and improving experiment reliability
- Aligned fragmented **data schemas** across **marketing reports** using **SQL transformation logic**, ensuring consistency across **analytics dashboards**

Sun Micronet – Computer Software & Network Support Manager

Apr 2021 – Nov 2022 | India

- **Directed a team of 15** to optimise IT systems and recovery processes, **reducing system downtime by 30%** and ensuring **100% data continuity**
- **Automated ticketing** and **monitoring systems** with customised scripting workflows, **cutting resolution times by 25%** and improving SLA compliance
- **Analysed 100+ system logs** to detect anomalies, **optimise workflows**, and reduce **false alerts by 20%**, a direct analogue to **storage performance troubleshooting**

PROJECTS

[Intrusion Detection System – Python, Random Forest, XGBoost](#)

Designed and trained a high-performance intrusion detection model on 100,000+ records using Random Forest and XGBoost, achieving 97% classification accuracy. Applied advanced feature engineering and algorithm benchmarking, directly applicable to data anomaly detection in storage systems.

[Flight Punctuality Data Warehouse – SQL, Tableau, OLAP](#)

Built a data warehouse processing 1M+ flight record; used SQL for ETL pipelines, OLAP for multi-dimensional analysis, and Tableau for 10+ interactive dashboards. Improved reporting efficiency by 40%.

[Sales Forecasting – Python, LSTM, XGBoost, Random Forest](#)

Developed a predictive sales model using time series analysis and ensemble methods, improving forecast accuracy by 30% via feature engineering and cross-validation.

[Recommendation Systems – Python, ML, Streamlit, Spotify API](#)

Constructed movie, music, and medicine recommendation engines using vectorisation and real-time API integration. Built interactive front-end with Streamlit for live user-facing suggestions.

EDUCATION

MSc Data Science

Jan 2024 – Jan 2025

Kingston University London | **Grade:** Distinction

Core Modules: Machine Learning & AI, Data Programming, Database Management, Data Visualisation

Highlight: Developed an [Intrusion Detection System](#) using **Random Forest** and **XGBoost**, achieving **97% accuracy** on a dataset of over **100,000+ records**.

Bachelor's Degree in Computer Applications

Jun 2020 – Aug 2023

B.Y.K College

Core Modules: Python, Advanced SQL, Java, Cyber Security, DBMS, Algorithms & Data Structures, Software Testing

Highlight: Led a team of four to build a full-stack library management system with real-time updates, reducing manual tracking by 60% and improving efficiency by 40%.

AWARDS & Recognition

- Special [Letter of Recommendation](#) from a previous employer, 2021
- Special [Letter of Recommendation](#) from Vice Principal, 2021
- Member of [British Computer Society](#), 2024
- [Letter of Recommendation](#) from Director of Fast Data Science, 2025
- [Microsoft Badges](#) for Machine Learning Module Assessments, 2024
- [LinkedIn recommendation](#) from previous employers, 2024

CERTIFICATIONS & AWARDS

- [Data Science Job Simulation - British Airways](#), Aug 2024
- [Data Engineering with Databricks - Databricks](#), Jan 2025
- [Data Analysis Job Simulation - Quantum](#), Nov 2024
- [Power BI Job Simulation - PwC](#), Dec 2024
- [Data Warehousing with Amazon Redshift - AWS](#), Aug 2024