

YASH DUDEJA

Data Scientist & ML Engineer

I'm a 26 year old passionate and adaptive tech enthusiast with Bachelors degree in IT with Masters of Science in Data Science with that hands-on experience in building AI-powered applications, smart agents, and scalable backend systems. My work spans across multiple domains including Machine Learning, Deep Learning, Natural Language Processing, and Generative AI.

WORK EXPERIENCE

Capgemini

2021 - 2023

Cloud Developer

- Designation: Analyst/Software Engineer
- Role: Webmethods DevOps
- Administered WebMethods IS, Broker, MWS, and Optimization
- Worked with JDBC Adapter Notifications and Adapter services
- Successfully delivered two migration projects, moving systems from one cloud platform to another.

Snapwork

Jun 2024 - July 2024

Data Scientist

- Collaborated with the HDFC data science team
- Utilized Label Studio, RoBERTa, Streamlit, and Computer Vision for various tasks
- Developed a system for converting image-based cheque data into structured text.

Millipixels Interactive LLP

Feb 2025 - Present

Data Scientist

- Worked On LLMs
- Fine tuning Models
- Agentic AI

PROJECTS

Event Log Categorization System

- Developed a deep learning pipeline to classify system event logs into categories like Authentication, Database, Security, etc. Used **BERT, T5, and XGBoost with TF-IDF for experimentation**. Designed a synthetic log dataset and built custom evaluation tools for accuracy benchmarking.

Knowledge-Augmented Document with Smart QA (RAG System)

- Built an intelligent document processing pipeline to handle PDFs and scanned images. Extracted and chunked text, stored metadata in **PostgreSQL**, and modeled inter-document relationships using **Neo4j**. Leveraged **Qdrant** for vector-based retrieval and integrated **Large Language Models (LLMs)** for accurate and scalable question answering over documents.

Cross Browser testing platform

- Designed and implemented a cross-browser testing platform using **Selenium, FastAPI**, and **Lighthouse** to validate URLs across Chrome, Firefox, and Safari before production release. Automated extraction of detailed reports highlighting **CSS/JS errors, accessibility issues, and performance metrics**. Ensured high **reliability** and **responsiveness** of web applications by integrating the system into a CI/CD pipeline.

Agentic AI for IT Helpdesk Automation

- Designed an AI-driven helpdesk system that intelligently processes IT complaints, auto-resolves repetitive low-priority tickets, and routes critical issues to the right team with detailed summaries. The agent performs issue classification, impact assessment, and generates recommended action plans to assist IT teams in faster resolution. Also includes a knowledge retrieval module to suggest fixes based on past resolutions and internal SOPs. Built using modern **NLP** and agent orchestration frameworks with a focus on real-world enterprise applicability.



CONTACT

Phone

+91 9833732836

Email

dudejayash@yahoo.com

Github link

[Yash- Github](#)

Portfolio

[Yash- Portfolio](#)

EDUCATION

2017-2020

Mumbai University

- Bachelors of Science in Information Technology

2023-2025

NMIMS

- Master of Science in Data Science

SKILLS

- | | |
|--------------------|----------------|
| • Machine Learning | • LangChain |
| • NLP | • Docker |
| • Deep Learning | • Fast API |
| • Generative AI | • Pandas |
| • Computer Vision | • NumPy |
| • MongoDB | • Scikit-learn |
| • Neo4j | • PyTorch |
| • Postgres | • Playwright |
| • Python | • OpenCV |
| • SQL | |

LANGUAGE

Hindi

English