

# YASH DUDEJA

## Data Scientist & ML Engineer

I'm a 26 year old Passionate AI & Data Consultant with hands-on experience in designing and deploying AI-driven solutions for real-world business challenges. Skilled in data-driven problem solving, automation, and scalable system design. Adept at translating complex technical requirements into actionable insights that drive measurable business outcomes. Experienced in client collaboration, proof-of-concept development, and end-to-end data solution delivery across industries.

## WORK EXPERIENCE

### Capgemini

2021 - 2023

Cloud Developer

- Designation: Analyst/Software Engineer
- Delivered two successful migration projects, administering WebMethods IS, Broker, and MWS. Enhanced cloud application performance and deployment efficiency across teams

### Snapwork

Jun 2024 - July 2024

Data Scientist

- Collaborated with HDFC's data science team to automate cheque data extraction using computer vision and NLP. Built models with RoBERTa, Label Studio, and Streamlit for text structuring.

### Millipixels Interactive LLP

Feb 2025 - Present

Data Scientist

- Worked on fine-tuning LLMs and developing agentic AI systems. Designed end-to-end intelligent automation pipelines integrating modern NLP and orchestration frameworks for enterprise use cases.

## PROJECTS

### Event Log Categorization System

- Developed a scalable AI pipeline to classify 5,000+ system logs across Authentication, Security, and Database categories. Improved anomaly detection accuracy by 28% using a BERT-based classifier. Designed end-to-end data generation, preprocessing, and evaluation modules simulating enterprise monitoring systems.

### Intelligent Document Summarization & Compliance Checker

- Designed a hybrid AI-powered system for processing and validating enterprise policy and regulatory documents. Automated extraction and chunking of text from PDFs and scanned files, storing metadata in PostgreSQL and modeling inter-document relationships in Neo4j. Integrated Large Language Models (LLMs) with Qdrant-based retrieval to perform context-aware summarization and compliance validation against predefined checklists. Reduced manual review effort by 60% and enhanced regulatory adherence accuracy across document workflows.

### Cross Browser testing platform

- Developed a CI/CD-integrated testing platform using Selenium and Lighthouse for automated web compatibility testing across browsers. Reduced pre-production testing time by 30% by detecting CSS/JS and accessibility issues early

### Agentic AI for IT Helpdesk Automation

- Built an intelligent helpdesk system that auto-resolves repetitive IT tickets and routes critical ones to the right team. Reduced low-priority workload by 40% using NLP-driven classification, impact analysis, and RAG-based resolution suggestions with LangChain and FastAPI.



## CONTACT

### Phone

+91 9833732836

### Email

[dudejayash@yahoo.com](mailto: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

#### • AI & Data :

- Machine Learning
- NLP
- Deep Learning
- Generative AI
- LLMs
- RAG
- Computer Vision

#### • Libraries & Frameworks:

- Pandas
- NumPy
- Scikit-learn
- PyTorch

#### • DevOps & Deployment:

- Docker
- CI/CD
- Selenium
- Playwright

#### • Data & Backend:

- Python
- FastAPI
- LangChain
- SQL
- MongoDB
- PostgreSQL
- Neo4j

## LANGUAGE

Hindi

English