

Yashasvi Shukla

977-35-48452 | yashasvi.shukla.me@gmail.com | [LinkedIn](#) | [Github](#) | [Portfolio](#)

SUMMARY

Full-stack engineer building scalable systems and production-ready AI/ML platforms, with hands-on experience in distributed backend architectures, applied NLP, and MLOps. Experienced in containerized services, ML pipelines, and model deployment workflows, with research exposure in LLM systems and retrieval-augmented generation.

EDUCATION

KIET Group of Institutions

Master of Technology in Computer Science and Engineering

2024 – 2026

Delhi-NCR, Ghaziabad

RKGIT, Ghaziabad

Bachelor of Technology in Computer Science and Engineering

2019 – 2023

Delhi-NCR, Ghaziabad

PROJECTS

CodeForge – Distributed Coding Platform | *React, Node.js, Express, MongoDB, Docker*

- Designed and built a distributed coding platform with authentication, problem management, and asynchronous submission evaluation
- Implemented secure, containerized code execution using Docker with language-agnostic evaluation via Judge0 integration
- Developed judging workflows including compilation, runtime execution, verdict mapping, and failure handling
- Architected backend services for scalable judging, test case validation, and submission lifecycle management

TalentForge - Resume Intelligence Platform | *Python, FastAPI, NLP, Scikit-learn*

- Built an AI-driven resume analysis system leveraging NLP techniques to match candidates with job descriptions
- Designed feature extraction pipelines for skills, experience, and semantic similarity scoring
- Implemented ATS-style ranking and keyword optimization logic exposed via RESTful ML inference APIs

ModelForge – AutoML & MLOps Platform | *Python, MLflow, Docker, AWS, FastAPI*

- Developed a production-oriented MLOps platform for training, tracking, and deploying machine learning models
- Integrated MLflow for experiment tracking, model versioning, and artifact management across training runs
- Designed automated pipelines supporting multiple ML algorithms and hyperparameter configurations
- Containerized model serving APIs and deployed inference services on AWS infrastructure

TECHNICAL SKILLS

Languages: Python, JavaScript, SQL

Backend Systems: FastAPI, Node.js, Express, RESTful APIs, JWT-based Authentication

Machine Learning & AI: Supervised and Unsupervised Learning, NLP, Deep Learning Fundamentals, Feature Engineering, Model Evaluation

LLMs & Retrieval Systems: Retrieval-Augmented Generation (RAG), Embedding-based Search, Prompt Engineering

MLOps & Observability: MLflow, Model Versioning, Experiment Tracking, Docker, Prometheus (metrics and monitoring concepts)

Data & Storage: PostgreSQL, MongoDB

Cloud: AWS (EC2, S3), Containerized Deployments

ACHIEVEMENTS & PROBLEM SOLVING

- Solved 100+ Data Structures and Algorithms problems on LeetCode and practiced consistently on Codeforces, focusing on optimized solutions and complexity analysis
- Authored research papers on *Reinforcement Learning for Memory Management in Language Model Agents* and *Retrieval-Augmented Generation Systems: Architectures, Techniques, and Empirical Analysis*
- Applied machine learning across multiple real-world datasets, gaining hands-on experience in model selection, evaluation, and deployment-ready pipelines