

ASHISH KUMAR

Chennai, India

+91-6307465659 ak503005@gmail.com Github Leetcode LinkedIn

EDUCATION

Motilal Nehru National Institute of Technology, Allahabad

Dec 2021 – Jun 2025

Bachelor of Technology — Branch Rank: 6/56

Allahabad, India

EXPERIENCE

MedicalMine India Private Limited (Zoho Subsidiary)

Jun 2025 – Present

Member of Technical Staff

Chennai, India

- Built and deployed distributed microservices using **Java Spring Boot** and **Redis Cluster** caching for the MCP system, cutting average API cost by **40%** and reducing P95 latency from **450ms to 150ms**.
- Improved scalability of **ChARM Assist** (GPT-4o + FAISS + Redis) for 1,000+ healthcare practices; sustained sub-**200ms** response times across **100K+ requests/hour**.
- Collaborated with backend and infra teams to design a fault-tolerant caching and retry mechanism, achieving **99.99%** uptime and lowering error rates by **65%**.
- Delivered features through 2-week agile sprints, conducted API performance profiling with Grafana/Prometheus, and automated deployment pipelines via CI/CD.

Interview Tayari

Jan 2025 – Apr 2025

Software Engineering Intern

Remote

- Developed a full-stack **MERN** platform with JWT + Google OAuth authentication and role-based access control, supporting **1,000+ users**.
- Implemented a secure Python/SQL code execution engine with Docker-based sandboxing, automated test evaluation, and Monaco Editor integration for live coding.
- Optimized MongoDB queries using compound indexes, reducing dashboard load time by **50%** and enabling real-time progress tracking.

PROJECTS

RAG-Based Multi-Source Q&A System | Python, FAISS, Sentence-Transformers, Streamlit, Ollama LLM Oct 2025

- Developed a **Retrieval-Augmented Generation (RAG)** pipeline using **FAISS vector DB** and **all-MiniLM-L6-v2** embeddings to process 100+ documents (Wikipedia, NASA, National Geographic).
- Integrated **Ollama LLM (Llama 3 8B)** with intelligent caching and logging, achieving **91.1% BERT F1** and **65.25% entailment accuracy** across 1,000+ test queries.
- Built Streamlit-based evaluation dashboard using BLEU, ROUGE, and METEOR metrics to analyze response quality.

Movie Recommendation System | Python, Pandas, NumPy, scikit-learn, Matplotlib

Apr 2023

- Designed a **user-based collaborative filtering model** handling **100K+ ratings** across 9.7K movies to generate personalized recommendations.
- Compared similarity metrics (Cosine, Pearson, Euclidean) and optimized performance using **NumPy vectorization**, enabling real-time recommendation generation.

TECHNICAL SKILLS

Languages: Java, Python, JavaScript, C++, SQL

Backend: Spring Boot, Node.js, Express.js, REST APIs, WebSocket, Microservices Architecture

Frontend: React.js, Redux, Tailwind CSS, HTML5, CSS3, Responsive Design

Databases: MongoDB, PostgreSQL, MySQL, Redis

AI/ML: OpenAI APIs, Scikit-learn, Pandas, NumPy, FAISS

Concepts: System Design, Data Structures, Algorithms, OOP, Design Patterns, Scalability, Distributed Systems

ACHIEVEMENTS

- Ranked **2076th** in **LeetCode Biweekly Contest 141 (2024)** with rating **1654**.
- Contributed to open-source projects in **GSSoC** and **Hacktoberfest**, enhancing performance and functionality.
- Solved **500+ LeetCode problems** in C++ and Java, strengthening algorithmic and data-structure skills.