

Purv Kabaria

+91 8980374837 | purvkabaria@gmail.com | LinkedIn | GitHub | Portfolio | LeetCode | CodeChef | Codeforces

SUMMARY

Final-year Computer Science student at NIT Surat with hands-on experience across full-stack product development, distributed backend systems, and applied AI/ML, including production LLM infrastructure built as part of a startup's founding engineering team. Proficient in Python, JavaScript/TypeScript, and C++, with strong fundamentals in data structures, algorithms, and system design. Comfortable owning a feature end-to-end, from asynchronous backend pipelines to AI-powered product surfaces.

TECHNICAL SKILLS

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

AI/ML & Backend: FastAPI, Django, Node.js, NestJS, LangChain, Ollama, RAG, MCP, Agentic AI

Distributed Systems: Redis, RabbitMQ, BullMQ, Inngest, Microservices Architecture, Asynchronous Processing, GraphQL

Databases: PostgreSQL, MongoDB, Qdrant, Firestore, Supabase, Prisma

DevOps & Cloud: AWS (EC2, S3), GCP, Docker, Git, MinIO

CS Fundamentals: Data Structures & Algorithms, Operating Systems, Computer Networks, Database Management Systems, Object-Oriented Programming

EXPERIENCE

Stealth AI Startup

Remote

Software Developer Intern

May 2026 – July 2026

- Built an LLM-powered document parsing pipeline using NestJS, Inngest, and BullMQ-orchestrated Redis background workers, cutting p99 processing latency from 30s to ~12s
- Built an AI-powered candidate assessment engine, designing the backend workflows and LLM integrations behind automated evaluation
- Engineered third-party integrations with Google Forms, Workday, and Ashby using Firecrawl-based web scraping and a GraphQL API layer, auto-generating job descriptions and application forms in under 30 seconds

Mercurium Analytics

Remote

SDE Intern

June 2025 – August 2025

- Resolved Django ORM N+1 query bottlenecks and nested relationship overhead, improving response time of data-heavy dashboards by 35%
- Eliminated double-processing vulnerabilities across distributed background workers by designing atomic database state updates that let workers securely claim processing rights

PROJECTS

Distributed RAG | GitHub | *FastAPI, Redis, PostgreSQL, Qdrant, Docker*

2026

- Designed a distributed microservices architecture (gateway, ingestion worker, embedding worker, vector store, and query service), separating ingestion, embedding, and query paths for independent scaling
- Achieved sub-30ms p95 vector search latency and a full local query cycle under 800ms by decoupling embedding from the request path via Redis queues

FinBridge | GitHub | *BullMQ, Docker, Prisma, PostgreSQL, Redis, FastAPI*

2026

- Sustained over 100 records/sec write throughput while deferring up to 250MB of raw audio blobs per batch through an async BullMQ + Redis queue, running 10 concurrent audio-analysis jobs
- Optimized sentiment classification to under 50ms inference latency using a quantized RoBERTa model, integrated alongside Whisper/Wav2Vec2 transcription at 12% Word Error Rate

ConnectIQ | Live | GitHub | *Next.js, Supabase, FastAPI, Gemini*

2025

- Built a custom vector-graph similarity routing engine matching networking profiles at 93% alignment accuracy
- Enforced 100% strict cross-tenant data isolation using Row Level Security, validated to a 50+ tenant benchmark in a hackathon deployment

Prometheus | GitHub | *LightGBM, XGBoost, MCP, Ollama, Qwen*

2026

- Delivered quantile energy-demand forecasts (P10/P50/P90) at under 2.8% error, with a 98.2% capture rate on critical grid capacity limits
- Exposed the forecasting pipeline as a tool-calling Model Context Protocol (MCP) server backed by a Qwen 3.5B model served via Ollama, executing grid recommendations in under 1.8s

LEADERSHIP & ACHIEVEMENTS

Leadership: Chairperson, Cutting Edge Visionaries (Fintech Club), SVNIT; Developer, ACM SVNIT Student Chapter

Competitive Programming: 436 problems solved on LeetCode (Rating: 1597); 3-star CodeChef coder (Rating: 1625, max 1667); Codeforces Pupil (Rating: 1261, max 1359)

Hackathons: 12x Hackathon Finalist | 2x Hackathon Winner, including Adani Powermind Hackathon; Top 100 nationally in ZS Tech Challenge; Runner-up, AI/ML Hackathon hosted by WebOsmotic