KOSHAL S GOYAL

+91 7019263192 koshal.s.goyal0@gmail.com
linkedin.com/in/koshal-goyal-a63ab4255 github.com/koshalg07 leetcode.com/koshal_g30

TECHNICAL SKILLS

Java Core, Kotlin, Golang, Linux |Spring Boot, Spring Gateway, Spring Webflux, Quarkus, Hibernate, Redis, SQL, MongoDB, Couchbase, GraphQL |Python, Fast API, Machine Learning, Fine tuning-LLM,Embeddings-based search, Approximate Nearest Neighbors , Vector Databases, RAG Architecture|Docker, Kubernetes, AWS (EC2, S3, Lambda) |Reactjs, TypeScript, NextJs |NATS Message Queue, Kafka

EXPERIENCE

Rakuten **March 2024 — Present**

Intern → Associate Software Engineer *Bangalore, India*

- Built scalable microservices for campaign lifecycle management using **Java 17, Spring Gateway, and WebFlux**, delivering high-throughput, low-latency APIs for real-time personalization workflows.
- Engineered asynchronous communication between services with **NATS Message Queue** and **Kafka**, enabling reliable, decoupled event streaming for creation, targeting, and cancellation flows.
- Developed high-throughput Kafka consumers for processing large-scale user targeting data, integrating **user embeddings** and ANN-ready indexing in **Couchbase** for rapid retrieval and campaign automation.
- Built a location & behavior-based recommendation engine using **Java, Kotlin, Quarkus, Hibernate, and MySQL**, incorporating **embedding-based user profiling** to increase engagement by **25%**.
- Designed and optimized **geospatial + embedding queries (MySQL/PostGIS)** with spatial indexing, paving the way for **Approximate Nearest Neighbor** (ANN) retrieval and improving query performance by **30%**.
- Implemented a fraud detection model using **Isolation Forest**, achieving **90% detection accuracy** and reducing manual review overhead.
- Defined rule-based detection for **transactional, behavioral, and account anomalies**, strengthening the real-time fraud prevention and personalization pipeline.

Samsung Research– PRISM Program **March 2023 – September 2023**

AI/ML Intern– Conversational AI Capsule *Bangalore, India*

- Developed a voice-enabled product search capsule using a **decoder-based neural network, integrating text & audio embeddings** to achieve **95% accuracy** in multi-label intent detection and slot tagging.
- Built a multi-domain dialogue corpus to enhance data diversity, enabling robust context-aware recommendations and improving training efficiency by 20%.

PROJECTS

Smart Document Search Engine with ML Reranking | *Python, Flash, FAISS, sentence-transformer, SQLite* **Septemeber 2025**

- Engineered production RAG system processing 20+ safety documents (1.3GB corpus) with **40%** retrieval accuracy improvement using Python, Flask, FAISS, and custom ML reranker
- Developed end-to-end ML pipeline training logistic regression reranker on 50,000+ document chunks, achieving **85%** accuracy and **60%** reduction in false positives for safety-critical queries
- Architected scalable REST API with sub-200ms response times, hybrid vector-BM25 search, and confidence-based abstention mechanisms handling 1000+ concurrent industrial safety document queries.

Rufus – Scalable RAG-Ready Web Data Extraction Tool | *Python, FastAPI, LangChain, Google Gemini API* **November 2024**

- Engineered an AI-powered web crawler to scrape structured content from dynamic websites using user-defined prompts; designed for seamless integration with **RAG pipelines** and LLM-based systems.
- Built end-to-end backend (**FastAPI, Playwright, LangChain**) including dynamic content handling, intelligent chunking, and prompt-based routing with Gemini API.

EDUCATION, ACHIEVEMENTS, EXTRACURRICULAR

BMS College of Engineering (Bangalore,India) **B.Tech in Computer Science and Engineering (2020-2024) CGPA: 8.6**

Finalist in RevaHack 2022 **Two time Coding ninja top 5 in CodingNinja Contest**

Senior Core Member of CodeIO Club in BMSCE