Kumar Abhinav

Sr. Software Engineer

(650) 695-4346 | [rushtoabhinavin@gmail.com](mailto:rushtoabhinavin@gmail.com) | [LinkedIn](https://www.linkedin.com/in/kumarabhinav05/) | [Github](https://github.com/Kumar0905/Csueb) | [Tableau](https://public.tableau.com/app/profile/kumar.abhinav6408/vizzes) | [Portfolio](https://abhinav0905.github.io/) | San Francisco Bay Area, CA |

|  |  |
| --- | --- |
| **Core Technologies & Skills**  Distributed System & API Development  FastAPI, REST APIs  Asynchronous Programming,  Performance Optimization  ETL Pipelines  Relational & NoSQL Databases,  **AI & Machine Learning:**  Agentic AI. RAG, Recommendation System,  Benchmarking | Senior Software Engineer experience in designing and benchmarking large-scale search & recommendation evaluation pipelines. Expert in Python, Java, SQL/Snowflake, and production GenAI integrations—driving 40%+ test-cycle reductions.  **Career Highlights**   * **Enterprise Ticket Automation (LinkedIn):** Engineered AI-powered APIs using Python/FastAPI to automate IT ticket resolution via ServiceNow-JIRA integration, migrated a fragmented Java chatbot backend to a unified Python architecture, and orchestrated fault-tolerant workflows with Temporal—cutting resolution times by up to 45% and slashing operational overhead. * **Bahrain Supply-Chain Analytics Suite:** Led the end-to-end custom product development of a cloud-hosted dashboard platform for the Kingdom of Bahrain’s Ministry of Health. Defined the product roadmap and feature set—including real-time metrics, drill-down views, and alerting—while building a secure SQL→ Elasticsearch pipeline that cut data latency 50%. Launched predictive modeling modules that boosted supplier on-time performance by 30% and delivered a 15% reduction in client operating costs, driving a 25% uplift in overall efficiency. |

**Work Experience**

**LinkedIn** | **Sr. Software Engineer (Contractor)**, Mountain View, CA. October 2024 to Present

In this role, I develop and deploy AI-driven automation solutions using Python and FastAPI to enhance internal operations. By modernizing legacy systems, integrating data sources, and improving process efficiency, I reduce manual intervention and improve system resilience, all of which contribute to a more streamlined and scalable infrastructure.

**Specific projects:**

* **Ticket Automator:** An AI-driven ServiceNow + Jira pipeline that auto-resolves and closes routine tickets by mining our internal knowledge base. Led pilot rollout to 500 users, then scaled to 28K. Customers enjoy a 40–45% reduction in MTTR, 30–35% lower support costs, and faster engineering response times.
* **Backend Migration & Scalability:** Led a complete backend migration of an internal chatbot from Java to Python, consolidating fragmented codebases into a single multi-tier system with asynchronous workflows. Improved operational efficiency by ~40%, lowering infrastructure costs and enabling 24x7 reliability.
* **GenAI Robustness Testing:** Integrated OpenAI’s API into a FastAPI microservice to generate synthetic user queries and edge-case prompts, then looped those through our recommendation engine to surface blind spots and guide quality improvements.
* **Asynchronous Daemon Orchestration:** Implemented Temporal for migrating legacy background daemons into fault-tolerant workflows. Enhanced issue detection through real-time logs and drastically minimized downtime via automatic retries and improved operational visibility.
* **Cross-Functional Evaluation Workshops:** Ran bi-weekly sessions with ML engineers and statisticians to co-define test cases, align on performance thresholds, and prioritize algorithm fixes—cutting validation cycles by 60%.

**RFXCEL (Part of Antares Vision)**. | **Software Engineer,** San Ramon, CA. May 2023 – Dec 2024

After leaving VVF India Limited to pursue an MS in Business Analytics, I joined RFXCEL as a summer intern. My contributions during the internship led to a full-time offer as a Tech Lead - Data Analyst and later transitioned it into Software Engineering role. I drive data-driven initiatives, design scalable ETL pipelines, and implement real-time analytics solutions, & enhancing decision-making.

**Specific projects:**

* **SAP → Luprolink API for AbbVie:** Partnered with AbbVie’s event creation team and IT teams to capture integration requirements, then designed and delivered a Java Spring Boot public API that automated order handoffs from SAP into Luprolink. Customers now enjoy real-time order synchronization—eliminating manual entry, speeding processing by 30%, and scaling seamlessly as volume grows.
* **Configurable Exception-Alert Framework for Walgreens:** Partnered with Walgreens’ Exception Management team to identify dynamic notification requirements, then delivered a Java Spring Boot + Angular/TypeScript email-configuration UI and REST API. Support agents and business users can now customize exception-alert rules in real time, boosting SLA compliance and cutting unplanned downtime notifications by 45%.
* **Integrated GenAI for Test Automation:** Embedded OpenAI GPT-4 prompts into evaluation workflows to auto-generate edge-case test scenarios, cutting manual test authoring time by 40 %.
* **Supply-Chain Analytics Platform:** Co-designed and delivered an end-to-end dashboard solution for Bahrain’s Ministry of Health. Conducted on-site workshops to define KPIs, then built a secure SQL Elasticsearch pipeline. The resulting product halved data-query latency and empowered operations teams to make real-time decisions, boosting process efficiency by 25%.
* **Predictive Scorecards for Cardinal & Walgreens**: Developed predictive supplier-scorecard for Cardinal and Walgreens using Python, SQL, Tableau, and GenAI and Typescript for front end. Iterated on deliverables through weekly review sessions—improving supplier on-time performance by 30% and slashing client data-processing effort by 40%.
* **Metric Dashboard for Algorithm Performance:** Developed a Grafana dashboard fed by Prometheus clients in our test harness, visualizing daily drift in recall/precision and triggering alerts when performance dipped > 2%.
* **ETL Data Pipeline & Data Processing:** Designed and implemented a custom ETL pipeline in Java, transforming transactional data from SQL to NoSQL (Elasticsearch), improving data latency and reducing operational costs. Automated data cleaning, schema matching, and deduplication, ensuring high data accuracy while minimizing SQL database load.
* **Snowflake Evaluation Pipeline:** Extended our existing Elasticsearch pipelines to feed Snowflake, enabling ad-hoc SQL queries over terabytes of event data and reducing ad-hoc analysis time from days to minutes**.**

**Vvf India Limited** | **Business &** **Data Analyst,** Mumbai, India. Dec 2017 to Aug 2022

As the Manager Supply Chain led several key projects that optimized the supply chain process, improved manufacturing efficiency, and successfully passed internal audits by McKinsey. Achieved global inventory reduction and led the SAP GST Implementation roll-out team.

**Specific projects:**

* **Global Inventory Analysis:** Conducted in-depth inventory data analysis by creating segmentation models using SQL and Python (Dictionary). Identified slow-moving, obsolete, and dead stock through data clustering, leading to inventory reductions of 25% and savings of $X million.
* **Manufacturing Cost Analysis:** Built linear regression models for cost and production analytics, detecting manufacturing bottlenecks and operational inefficiencies. Reduced manufacturing losses by 22% through data-driven Strategic Planning**,** Capacity Management andcorrective actions, resulting in annual savings of $X million.
* **Forecasting & Demand Planning Analytics:** Created demand forecasting models using ARIMA and SARIMA, SQL, and Tableau by integrating historical sales, market trends, and supplier data. Improved forecast accuracy by 20%, reduced stockouts, and increased on-time delivery rates by optimizing data-driven planning processes.

**Personal Projects**

**Toy Transformer LLM from Scratch** – Architected a mini-transformer (~1.2 M params) in PyTorch. Implemented full training loop on an Apple M2 GPU, tracked train/Val loss and perplexity over 10 epochs, and deployed a FastAPI Q&A demo. Documented parameter-count math and architectural trade-offs.

**Skills**

**Technical:** Python, Java, FastAPI, MySQL, NoSQL, Elasticsearch, Kibana, Tableau, Advanced MS Excel, Lang Chain, Azure, AWS,

**Functional:** Distributed Systems, API Development, Performance Optimization, Asynchronous Programming, Process Mapping, Delivery Presentation, Agile, SDLC, Jira, Git, Temporal, Unit Testing, Code Reviews, Agile, SDLC, CI/CD, Docker, Jenkins.

**Education and Academic Background**

**California State University East Bay**, Hayward, CA

**M.S. in Business Analytics**, Aug 2022 to Dec 2023

**Sathyabama Deemed University** Chennai, India

**B.Tech. in Chemical Engineering (Major)** | **Computer Science & Engineering (Elective)**