Varun Kulshrestha

+91-8218704244 — varunkulshrestha2137@gmail.com — LinkedIn — GitHub

Summary

Software Engineer with 4+ years of experience building data platforms and AI/ML-driven features, applying data science and LLMs to deliver scalable, real-world products.

Experience

HCL Technologies – Engineering R&D Team — Jan 2024 – Present

Project: Propel Data Solutions

- Developing an **Automated Data Archival Solution**, optimizing data lifecycle management and reducing active storage costs by **~35%**.
- Developing a **Data Quality** validation **POC** using **Great Expectations**, assessing dataset reliability and enabling downstream enrichment workflows for analytics and ML readiness.
- Built a **Dynamic Visualization platform** leveraging LLMs to convert natural-language prompts into structured queries and interactive visualizations, cutting report turnaround time by **~60%**.
- Built a **Test Data Generation backend to eliminate testing bottlenecks**, generating realistic datasets for files and **relational datasets with inter-table relationships** using **Mimesis**, reducing dependency on sensitive production data.

HCL Software – Observability Team — Nov 2021 – Dec 2023

Project: PepsiCo and Mashreq

- **Implemented Anomaly detection POCs on time-series observability data using Elasticsearch ML**, automating alerting workflows and reducing manual monitoring effort by **~50%**.
- **Automated ELK cluster provisioning for R&D POCs using Docker Compose**, deploying Elasticsearch (data & ML nodes) and Kibana to standardize DevOps workflows and remove manual infrastructure setup.
- **Developed interactive operational dashboards in Kibana**, enabling faster incident detection and reducing mean time to resolution (MTTR) by **~50%**.
- **Built end-to-end ETL pipelines using Logstash and Elasticsearch**, ingesting and transforming Okta API logs during Splunk-to-Elastic migration, improving log traceability and reducing debugging time by **~60%**.
- **Containerized and deployed observability services using Docker on AWS EC2**, improving environment consistency, scalability, and DevOps efficiency across multiple client POCs.

Projects

- Built a **Retrieval-Augmented Generation** (RAG) application using GPT-3.5 and Milvus Vector DB for domain-specific finance queries, improving answer relevance and reducing **latency** by **~40%.**
- **Insurance Premium Prediction** – Developed a **regression**-based machine learning model using demographic and behavioral features to estimate insurance premiums.
- **Loan Approval Risk Modeling** – Developed a binary **classification model** to predict loan default risk, generating probability-based scores to support data-driven approval decisions.

Skills

Programming: Python, C++.
Machine Learning & AI: LLMs, RAG, Regression, Classification, Time-Series Analysis, Anomaly Detection.
Data & Platforms: PostgreSQL, FastAPI, Airflow, ELK Stack.
Cloud & DevOps: AWS, Docker, Git, Linux.

Education

Abdul Kalam Technical University — Jul 2017 – Aug 2021

B. Tech in Computer Science and Engineering (Honors) | CGPA: 8.23 / 10