Nikhil Advani

+1(716)939-6884 | nadvani@buffalo.edu | linkedin.com/in/nikhil-advani/ | Github | Portfolio

EXPERIENCE

UB School of Pharmaceutical Sciences Jun. 2024 – Present
AI Engineer NY, USA
RAG Chatbot for Biomedical Research

- Built a Retrieval Augmented Generation system using LLMs to extract insights from 1000+ clinical trial & drug label documents
- Resolved factual inaccuracies in context retrieval by implementing semantic chunking and combining BM25 & MMR retrievers
- Optimized context generation with FlashRank reranker, reducing average query processing time from 1 minute to 20 seconds
- Employed prompt engineering including Chain-of-Thought, hierarchical prompting, boosting the quality of responses by 40%
- Automated the manual research workflow, reducing 36 hours of manual review to a 5-minute autogenerated insights report

EPL Limited Jan. 2022 – Apr. 2023
Data Scientist | Machine Learning Team Mumbai, India
Customer Retention Modeling and Deployment for Scalable Business Impact

- Developed a BERT-based text classification model to predict customer churn by analyzing customer reviews and email complaints, enabling early identification of at-risk users
- Conducted A/B tests with hypothesis testing to evaluate marketing interventions' impact on customer engagement
- Delivered Power BI dashboards to stakeholders, enabling budget optimization and sharing knowledge across 6 internal teams
- Led end-to-end deployment of the model using AWS and Docker, automated CI/CD pipelines via GitHub Actions for model updates, and monitored performance using MLflow, improving deployment reliability
- Achieved a 15% increase in user retention and saved ~$320K annually through targeted interventions

Numeric Infosystem Pvt. Ltd. Aug. 2021 – Dec. 2021
Software Development Engineer | Data Processing Team Gwalior, India
Software Development for Operational Efficiency and Data Preparation

- Created a full-stack Inventory Management System using Django, JavaScript, HTML, and CSS to enhance inventory tracking
- Reduced excess stock and minimized waste by through improved visibility and real-time inventory updates
- Addressed data ingestion errors in ETL pipelines by developing automated error-checking & validation scripts in Apache Spark
- Enabled faster business reporting by automating inventory data flows, reducing manual processing time by 50%

PROJECTS

Agentic AI-Powered Research Assistant [demo] [code]

- Engineered an AI-powered multi-agent financial assistant using Phidata agentic architecture, automating stock research with 60% faster responses and real-time insights. Deployed via Streamlit using Deepseek R1, yFinance, and DuckDuckGo APIs, it achieved 95% accuracy and reduced research time by 40%.

Movie Recommendation System with Sentiment Analysis [demo] [code]

- Orchestrated a content-based movie recommendation system by scraping IMDb reviews with BeautifulSoup and applying Naive Bayes and XGBoost for sentiment analysis. Achieved 91.47% accuracy in segmenting users by preference and viewing patterns and delivered personalized recommendations, demonstrating scalable personalization.

DeepFace Real-time Face Detection [code]

- Structured a deep learning face detection model with 92% accuracy, utilizing a custom-labeled dataset of 5,000+ images processed with OpenCV and annotated using LabelMe. Leveraged TensorFlow with a pre-trained VGG16 base, the model achieved validation losses of 0.12 and 0.18, with 88% precision and 85% recall on real-world testing with 200+ users.

EDUCATION

University at Buffalo, Buffalo, NY Aug. 2023 – Dec. 2024
Masters in Data Science GPA: 3.90/4.00

- *Relevant Coursework: Machine Learning, Statistics, Data Mining, NLP, Data Visualization, Big Data Analytics*

Madhav Institute of Technology and Science, Gwalior Aug. 2018 – Jul. 2022
Bachelor of Technology, Electrical Engineering

SKILLS

Programming Languages: Python, R, SQL (MySQL, PostgreSQL, PySQL), DAX, VBA, Golang
Tools and Technologies: Microsoft (Power BI, Office), Tableau, Docker, MongoDB, Postgres, Hadoop, Git, GitLab, Kubernetes, MLflow, Streamlit, Kafka, Retrieval Augmented Generation (RAG), Agentic AI, Graph-RAG, Large Language Models (LLama, GPT, Deepseek), Natural Language Processing, Machine Learning, Deep Learning, Data Structures, Algorithms, Prompt Engineering, Quantization, A/B Testing, MLOps, Causal Inference, Hypothesis Testing, VectorDBs (ChromaDB, Faiss, Qdrant), SAP
Cloud: Amazon Web Services (Redshift, S3, DMS, Glue), Google Cloud Platform (BigQuery, VertexAI, Dataproc), OpenAI
Frameworks & Libraries: Pytorch, Tensorflow, Flask, FastAPI, LoRA, XGBoost, Langchain, LlamaIndex, LangGraph, DeepEval, Pandas, Numpy