

Nikhil Advani

+1(716)939-6884 | nadvani@buffalo.edu | [linkedin.com/in/nikhil-advani-2b03aa201/](https://www.linkedin.com/in/nikhil-advani-2b03aa201/) | [Github](#) | [Portfolio](#)

EXPERIENCE

UB School of Pharmaceutical Sciences

Jun. 2024 – Present

Generative AI Engineer

NY, USA

RAG Chatbot for Drug Development Research

- Constructed a Retrieval Augmented Generation (RAG) System using Llama 3.1 to streamline drug development research
- Integrated cosine similarity-based chunking with an ensemble retriever combining BM25 and MMR, resulting in a robust RAG pipeline, improving information retrieval efficiency and reducing average query processing time by 50 seconds
- Optimized retrieval with FlashRank reranker to improve context generation, ensuring high relevance and domain-specific accuracy in the answers generated by the system
- Employed prompt engineering techniques including Chain-of-Thought and Iterative Prompting, boosting the quality of responses by 40%

EPL Limited

May. 2022 – May. 2023

Data Scientist/Analyst

Mumbai, India

Data Science Modeling and Visualization for Business Optimization

- Developed a logistic regression model to offer relevant product features to at-risk clients, leading to a boost in monthly customer retention by 0.92% and saving approximately \$120k annually
- Implemented time series forecasting models using ARIMA & SARIMAX for sales forecasting, enabling targeted marketing and boosting sales efforts by 22%
- Deployed ML models into production using AWS Lambda, S3, API Gateway, and SageMaker, automating real-time predictions and enabling scalable inference with Docker-containerized workflows, reducing manual intervention by 60%
- Delivered Power BI dashboards featuring KPI's, integrating manufacturing data from SAP, MongoDB, and Postgres. Collaborated with cross-functional teams and presented analysis to stakeholders, driving an YoY revenue gain of \$500K
- Designed, deployed, and monitored data pipelines using AWS Redshift, S3, Glue, and DMS, reducing data access delays by 32% and ensuring real-time integration

Numeric Infosystem Pvt. Ltd.

Aug. 2021 – Dec. 2021

Software Development Intern

Gwalior, India

Software Development for Operational Efficiency and Data Preparation

- Created an Inventory Management website integrating Django, python, SQL, and javascript, utilized by 5 clients, enhancing efficiency by 20% and saving 10 hours weekly
- Devised ETL workflows using Python, SQL and Excel to clean and prepare client data, enabling data scientists to build predictive models that lifted ROI by 7.3% annually

PROJECTS

AI Chatbot – Groq Inference Engine [\[demo\]](#) [\[code\]](#)

- Structured a RAG chatbot using Groq inference engine with the open-source Llama-3 model and Huggingface embeddings, reducing response time by 50%, combined tools like Wikipedia, PDF Search, and Arxiv for intelligent retrieval. By setting up Chroma DB for embeddings and using Langchain agents, boosted retrieval speed by 30% and improved reasoning efficiency.

Movie Recommendation System [\[demo\]](#) [\[code\]](#)

- Built a content-based movie recommendation system using BeautifulSoup for web scraping IMDb reviews and a NaiveBayes model for sentiment analysis, achieving 97.47% accuracy in classifying reviews. The system provides personalized movie recommendations, enabling users to discover relevant content based on sentiment and preferences.

Walmart eBay GCP Data Engineering Pipeline & Data Analytics [\[code\]](#)

- Orchestrated a data engineering solution to collect product data from Walmart's eBay API, process and store it in Google BigQuery, and create visualizations in Looker. Delivered e-commerce insights, including pricing trends, category performance, and product popularity analysis.

DeepFace Real-time Face Detection [\[code\]](#)

- Led the development of 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 50 users.

EDUCATION

University at Buffalo, Buffalo

Aug. 2023 – Dec. 2024

Masters in Data Science

NY, USA

SKILLS

Programming Languages: Python, R, SQL (MySQL, PostgreSQL, SSMS, PROC, PySQL), SAS, DAX, VBA, HTML, CSS

Tools: Microsoft (Power BI, Office), Tableau, Amazon Web Services (S3, Redshift, Glue, DMS, EC2, SageMaker, Lambda), Docker, MongoDB, Oracle, Databricks, Snowflake, Google Cloud Platform, Postman, JIRA, BigQuery, ServiceNow(CMDB, Discovery)

Data Science & Analytics: Python (Pandas, NumPy, Matplotlib, SciPy, Scikit-learn, ggplot2), R (dplyr, tidyr, table1, Shiny), ETL, Cleansing, Predictive Modeling, Feature Engineering, Data Analysis, Data Mining, Statistical Analysis, NLP, Spark, Data Visualization, Mathematics, Computer Science, Statistics, Finance, Business Analysis, Fraud Analysis, Git, Github, GitLab, A/B Testing

AI/ML: LLM (Open AI, Llama, Gemini, Claude), Langchain, Llama-Index, Prompt Engineering, Chunking, Hugging Face Transformers, LangGraph, DeepEval, RAGAS, Regression, Classification, Clustering, Cox proportional hazard, Kaplan Meier Curve, MLOps, Time Series Forecasting, CI/CD Pipeline, Causal Inference, Kubernetes, Kafka, REST API, Deep Learning (CNN, RNN, LSTM, GAN)