

# Nikhil Advani

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

## EXPERIENCE

### UB School of Pharmaceutical Sciences

AI Engineer

Jun. 2024 – Present

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 5 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

Data Scientist

Jan. 2022 – Apr. 2023

Mumbai, India

#### Customer Retention Modeling and Deployment for Scalable Business Impact

- Developed LSTM and XGBoost models with Tensorflow to predict customer churn & enable proactive retention strategies
- Improved customer engagement by conducting A/B tests and statistical hypothesis testing on marketing interventions
- 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 (S3, Sagemaker, Lambda) 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.

Data Engineer

Aug. 2021 – Dec. 2021

Gwalior, India

#### ETL Optimization and Spark-Based Data Integration

- Created Apache Spark jobs to clean and validate large operational datasets, ensuring high-quality input for reporting
- Designed automated RESTful services to sync inventory metrics across PostgreSQL and MongoDB
- Implemented schema transformation logic and JSON/XML parsing in Python, improving system throughput by 60%
- Documented dataflows, process maps, and ETL workflows to enable faster onboarding and operational efficiency

## 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

Masters in Data Science

Aug. 2023 – Dec. 2024

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

Bachelor of Technology, Electrical Engineering

Aug. 2018 – Jul. 2022

## SKILLS

**Programming Languages:** Python, R, SQL (MySQL, PostgreSQL, PySQL), SAS, bash, DAX, VBA

**Tools and Technologies:** Microsoft (Power BI, Office), Tableau, Docker, MongoDB, Postgres, Hadoop, Git, Kubernetes, Kafka, Retrieval Augmented Generation (RAG), Agentic AI, Large Language Models (LLama, GPT, Deepseek), Natural Language Processing, Machine Learning, Time Series Forecasting, Deep Learning, Data Structures, Algorithms, Quantization, MLOps, Causal Inference

**Cloud:** Amazon Web Services (Redshift, S3, Bedrock, Sagemaker), Google Cloud Platform (BigQuery, VertexAI, Dataproc), OpenAI

**Frameworks & Libraries:** Pytorch, Tensorflow, Flask, FastAPI, LoRA, XGBoost, GAN, Langchain, LlamaIndex, LangGraph, DeepEval, RAGAs, Rouge, Pandas, Numpy, VectorDBs (ChromaDB, Faiss, Qdrant), FastAPI, Teradata