Vraj Dobariya

+1 3412409035 | vdobariya@horizon.csueastbay.edu | Union City , California | LinkedIn | Portfolio | GitHub

Gold Medalist in Computer Science | GATE Top 15% - Competitive National Exam in CS | AI/Data Enthusiast

Domain/Functional Areas

- Healthcare
- Recommender Systems
- Gen AI to Agentic AI
- Fraud Detection/Risk Management

Key Skills

Programming Languages: Python, R, SQL, C, C++

Databases & Vector Stores: MySQL, MongoDB, ChromaDB(Vector Search)

AI/ML Frameworks & Libraries: Scikit-learn, XGBoost, PyTorch, TensorFlow, Keras, Statsmodels, SciPy (stats), Prophet, pmdarima, Flask API, OpenCV, spaCy, NLTK, Transformers, Hugging Face, LangChain, FastAPI, Streamlit, CUDA

MLOps & Deployment: Docker, Kubernetes, CI/CD Pipelines, AWS, Azure, MLflow, Linux, Unix-based OS, Git

AI/ML Expertise: Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Generative AI, LLM, RAG, Agentic Workflows, Model Fine Tuning(Transfer Learning), Model Evaluation, Prompt Engineering, Reinforcement Learning

Data Science & Analytics: EDA(Numpy, Pandas, Matplotlib, Seaborn), Data Preprocessing, Feature Engineering, Hypothesis Testing, A/B Testing, Time Series Analysis, Forecasting, Big Data Technologies(Apache Spark, Hadoop), Data Visualization (Tableau, Power BI)

Professional Experience

AtliQ Technologies Pvt. Ltd. **Vadodara, India**

AI/Data Intern (Internship) Dec 2024 - May 2025

- Collaborated with a cross-functional team engaging stakeholders across healthcare and generative AI domains to align solutions with business goals, driving communication that expedited delivery and improved domain impact.
- Built a hybrid BioBERT + Logistic Regression model for medical transcription, achieving 79.5% accuracy via SMOTE and SynonymAug; reduced manual effort by 50%, cut misclassifications by 25%, and ensured reliable, scalable EHR support.
- Built an intelligent chatbot using RAG with LLaMA 3.3 and ChromaDB, increasing user engagement 60% and revenue potential 40%, integrating semantic routing and real-time SQL queries to optimize response times and experience.

Skills: BioBERT / Logistic Regression / SMOTE / RAG / LLaMA (LLMs) / ChromaDB / SQL / Flask API / Docker / AWS EC2 / Teamwork

Projects

Personalized Product Recommendation System - **E-commerce and Recommender Systems** **Project Link**

[Python / TensorFlow / Keras / Tokenization / LSTM / Bidirectional RNN / Streamlit] Nov 2024 - Jan 2025

- Developed an LSTM and Bidirectional LSTM recommender system on Retailrocket data, achieving 85.35% Precision@5 and Recall@5 with dual item-category inputs, improving personalization by 20% over baselines.
- Optimized model with robust preprocessing, IQR outlier detection, tokenization, and a Streamlit app with caching, enabling interactive visualizations and enhancing user engagement and recommendation efficiency.

RAG-Driven Research Tool - **Real Estate** **Project Link**

[Langchain / Transformers / ChromaDB / FastAPI / System Design / Communication / Agentic Workflows] Aug 2024 - Oct 2024

- Engineered a Streamlit-based web app with Retrieval-Augmented Generation (RAG), slashing LLM API costs by ~70% and research time by ~50% for real estate insights, delivering precise, URL-driven answers using LangChain's UnstructuredURLLoader tool.
- Spearheaded the integration of HuggingFace's all-MiniLM-L6-v2 embeddings and ChromaDB retrieval system, driving analytical efficiency and fostering teamwork to produce scalable, source-referenced outputs via Llama3 platform.

A/B Testing for Drug Marketing Campaigns - **Pharmaceutical Marketing** **Project Link**

[Python / SciPy / A/B Testing / Statistical Analysis / Snowflake / T-Test / ANOVA / Visualization] May 2024 - Jul 2024

- Built an A/B testing pipeline for elderly patient marketing campaigns, achieving a 1.52 odds ratio (95% CI: 1.11–2.08, P=0.0092) for prescription uptake via t-tests and ANOVA (P=0.2811) for consistent health condition effects, reducing decision uncertainty by 30%.
- Engineered age, health condition, and time slot features on a Kaggle dataset with IQR outlier removal, using Pandas/NumPy and Snowflake pseudo-code for scalable integration, with Seaborn visualizations boosting stakeholder interpretability by 20%.

Customer Segmentation for Auto Finance Marketing - **Auto Finance** **Project Link**

[Scikit-learn / K-Means Clustering / RFM Analysis / Feature Engineering / PCA / SAS / JMP / SQL] Feb 2024 - Apr 2024

- Developed a K-Means clustering model using RFM analysis on 1M+ transactions, identifying 5 customer segments (k=5 via elbow method), improving targeted campaign precision by 35% and aligning segmentation with auto finance benchmarks.
- Engineered 6 scaled features with IQR-based outlier capping and PCA visualization in Seaborn, reducing campaign design time by 40% and improving interpretability by 25% using SQL, SAS, and JMP for scalable integration.

Education

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

Master of Science - Statistics Data Science Jan 2024 - Present

Adani University **Ahmedabad, India**

Bachelor of Technology - Information Technology Jun 2018 - Jul 2022