Vraj Dobariya

vrajdobariya.ict18@gmail.com https://www.linkedin.com/in/vraj-dobariya/ https://github.com/Vraj-Data-Scientist
San Francisco Bay Area Portfolio : https://vraj-dobariya.netlify.app/

EDUCATION

Master of Science - MS, Statistics Data Science

California State University, East Bay

Present Hayward, United States

GPA **3.85** / 4.0

- Engaging in hands-on projects that apply AI and statistical methods to solve real-world problems.

Bachelor of Technology - BTech, Information Technology

Adani University

06/2018 - 06/2022 Ahmedabad, India

GPA **3.9** / 4.0

- Received a **gold medal** for attaining a semester percentage index (SPI) of 9.79 out of 10.
- Maintained the highest academic standing in my Computer Science class for two years in a row.
- Ranked in the **Top 15%** in the GATE Computer Science exam, one of India's most competitive national exams.

SKILLS

Programming Languages: Python, R, SQL. Databases & Vector Stores: MySQL, MongoDB, ChromaDB. AI/ML Frameworks & Libraries: Scikit-learn, XGBoost, PyTorch, TensorFlow, Keras, spaCy, Transformers & Diffusers (Hugging Face), LangChain, FastAPI, Streamlit. MLOps & Deployment: Docker, Kubernetes, CI/CD, AWS, Azure, MLflow AI/ML Expertise: Machine Learning, Deep Learning, NLP, Generative AI, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), AI Agents, Hyperparameter Tuning.

PROJECTS

Intelligent RAG E-commerce Chatbot

https://gen-ai-powered-e-commerce-chatbot-vraj-dobariya.streamlit.app/

Developed a chatbot, enhancing user experience by ~60% and revenue potential by ~40% with a conversational interface. Integrated RAG with LLaMA 3.3 (Groq), semantic routing, and real-time SQL to replace inefficient filters and FAQs, reducing API costs by ~50% using HuggingFace embeddings, ChromaDB, and quantization.

RAG-Driven Real Estate Insights App

https://real-estate-assistant-using-rag-vraj-dobariya.streamlit.app/

Developed a web app, reducing LLM API costs by ~70% and research time by ~50% using Retrieval-Augmented Generation (RAG) for real estate insights. Enabled URL-based queries with precise answers via LangChain's UnstructuredURLLoader, HuggingFace all-MiniLM-L6-v2 embeddings, ChromaDB retrieval, and Llama3 (Groq) with references.

Finance Credit Risk Predictor

https://ml-project-credit-risk-model-vraj-dobariya.streamlit.app/

Developed a Logistic Regression model for credit risk, achieving 93% accuracy and 94% recall by addressing class imbalance with SMOTE Tomek and optimizing via Optuna. Predicts loan default probability, assigns credit scores (300-900), and rates creditworthiness for finance applications.

Damage Prediction CNN Model

https://dl-project-damage-prediction-vraj-dobariya.streamlit.app/

Built a deep learning model with CNNs and transfer learning, improving accuracy from 57.74% (baseline) to 80.87% using EfficientNet, ResNet fine-tuning, and hyperparameter optimization (dropout: 0.2, learning rate: 0.005). Applied regularization and computer vision techniques for robust damage prediction.

EXPERIENCE

Kaggle Participant & AI/Data Science Enthusiast

San Jose, United States

- Participated in Kaggle competitions to gain practical experience with real-world datasets, achieving top rankings in challenges across tabular data, computer vision, and NLP.
- Shared knowledge through Kaggle Notebooks, showcasing skills in EDA, feature engineering, and model development using Python, Scikit-learn, and TensorFlow.
- Engaged with the Kaggle community to collaborate and learn, applying a comprehensive skill set in machine learning, deep learning, and generative AI to deliver solutions.