

# Vraj Dobariya

✉ [vrajdobariya.ict18@gmail.com](mailto:vrajdobariya.ict18@gmail.com)   <https://www.linkedin.com/in/vraj-dobariya/>   <https://github.com/Vraj-Data-Scientist>  
📍 San Francisco Bay Area

## EDUCATION

### Master of Science – MS, Statistics Data Science

California State University, East Bay

📅 Present   📍 Hayward, United States

- Demonstrated **leadership** by independently managing project timelines, troubleshooting deployment issues, and integrating feedback from peers to improve functionality.

GPA

**3.8** / 4.0

### Bachelor of Technology – BTech, Information Technology

Adani University

📅 06/2018 - 06/2022   📍 Ahmedabad, India

- Earned a **Gold Medal** for achieving a 9.79/10 SPI, maintaining the highest academic standing for two consecutive years.
- Ranked in the **Top 15%** in the GATE Computer Science exam, one of India's most competitive national exams.

GPA

**3.9** / 4.0

## 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.

### 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.

### 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.

### 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.

## EXPERIENCE

### Kaggle – Competitor & Innovator

📅 Present   📍 San Jose, United States

- Competed in 10+ **Kaggle** challenges across tabular data, computer vision, and NLP, achieving **top 5%** rankings by delivering high-performing solutions in machine learning and deep learning.
- Built and optimized models using Python, Scikit-learn, and TensorFlow, showcasing expertise in EDA, feature engineering, and generative AI to solve complex data challenges.