Vraj Dobariya

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

Key Skills

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

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

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)

AI/ML Frameworks & Libraries: Scikit-learn, XGBoost, PyTorch, TensorFlow, Keras, 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

Core CS: Data Structures & Algorithms, Graph Analysis, Software Engineering (System Design, Testing, OOP), API Integration (REST, GraphQL), Agile Methodologies

Professional Experience

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

AI/ML Intern (Internship) Dec 2024 - Mar 2025

- Led a team to develop two AI projects, driving communication with a cold storage firm and a healthcare company to define needs, completing a fruit freshness system and a Q&A tool using deep learning and generative AI.
- Spearheaded preprocessing of 16,000 fruit images and tuned a ResNet50 CNN, showcasing analytical skills to cut processing time by 30% and boost accuracy from 69.44% to 99.85%, earning recognition for impactful internship growth.
- Collaborated on a Streamlit-based healthcare AI tool, enhancing LLaMA 3 with RAG and retrieving PubMed data to deliver fast, evidence-based intermittent fasting insights for clinicians, improving decision-making efficiency.

Skills: PyTorch, ResNet, Hugging Face, Transformers, LLMs, RAG, Streamlit, AWS, Agile Development, Teamwork

Projects

Intelligent Chatbot Using RAG and LLM - **E-commerce** **Project Link**

[Transformers / LLaMA 3.3 / ChromaDB / SQL / Semantic Routing / REST API / Prompt Engineering] Oct 2024 - Jan 2025

- Engineered a conversational chatbot leveraging Retrieval-Augmented Generation (RAG) with LLaMA 3.3 (Groq), enhancing user experience by approximately 60% and boosting revenue potential by 40% through tailored, context-aware interactions.
- Integrated semantic routing and real-time SQL queries using HuggingFace embeddings and ChromaDB, streamlining data retrieval.

Damage Detection Using Deep Learning - **Automobile** **Project Link**

[PyTorch / CNN / CUDA / ResNet / Hyperparameter tuning / Data Augmentation / OpenCV / Matplotlib] Aug 2024 - Sep 2024

- Designed and implemented a CNN-based deep learning model leveraging transfer learning, boosting prediction accuracy from 57.74% (baseline) to 80.87% through EfficientNet and ResNet fine-tuning.
- Optimized model performance using hyperparameter tuning (dropout: 0.2, learning rate: 0.005) and regularization techniques, enhancing scalability while integrating OpenCV for advanced computer vision preprocessing, improving dataset quality by 15%.

Credit Risk Predictor Using Machine Learning - **Finance** **Project Link**

[Scikit-learn / XGBoost / SMOTE Tomek / Optuna / EDA / Feature Engineering / Data Preprocessing] Jun 2024 - Aug 2024

- Developed a Logistic Regression model for credit risk assessment, achieving 93% accuracy and 94% recall by mitigating class imbalance with SMOTE Tomek, reducing false negatives by 15% and enabling precise loan default predictions.
- Enhanced model performance through Optuna-based hyperparameter optimization, delivering scalable creditworthiness ratings (scores: 300-900), improving decision-making efficiency for financial stakeholders by 25%.

Credit Card A/B Testing Analytics - **Finance**

[EDA(Numpy, Pandas, Matplotlib, Seaborn) / Data Preprocessing / Hypothesis Testing / A/B Testing] Apr 2024 - May 2024

- Analyzed customer data with EDA to target 18-25-year-olds for a credit card launch in banking. Conducted A/B testing, rejecting the null hypothesis, proving higher engagement and uptake with the new card compared to the existing one.

Predictive Health Insurance Model - **Medical** **Project Link**

[XGBoost / Scikit-learn / Machine Learning / EDA / Feature Engineering / Data Preprocessing] Feb 2024 - Mar 2024

- Built a Health Insurance Premium Prediction Model using Linear Regression (98.87% accuracy, ages less than 25) and XGBoost (99.70% accuracy, ages > 25), applying StandardScaler and Random Search Tuning, reducing prediction errors by 15%.
- Deployed model via Streamlit, enabling real-time premium estimation from inputs, improving insurance provider efficiency by 20%.

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