Dhruv Jitendra Limbani

(646) 281-9850, djl2204@columbia.edu, linkedin.com/in/dhruvlimbani, github.com/Dhruv-Limbani, dhruv-limbani-portfolio.streamlit.app

Education

Columbia University — New York City, US
MS in Computer Science, Machine Learning Track, GPA: 4.00 — Sep 2024 – Dec 2025

SRM Institute of Science and Technology — Chennai, IN
B.Tech in Computer Science and Engineering, GPA: 9.79/10 — Sep 2020 – Jun 2024

Relevant Coursework

Machine Learning, Data Science, Deep Learning for Computer Vision, Natural Language Processing, Data Structures and Algorithms, Database Management Systems

Skills

Programming: Python, SQL, C++, C
Machine Learning: PyTorch, TensorFlow, Scikit-Learn, OpenCV, NLTK, LangChain, LangGraph, MCP
Data Engineering & Analytics: Pandas, NumPy, Apache Airflow, Matplotlib, Seaborn
Databases: Snowflake, PostgreSQL, ClickHouse, Pinecone, MySQL

Experience

Chartmetric — New York, NY
Data Analytics Intern — May 2025 – Aug 2025
- Built analytics dashboards for music curation networks, enabling real-time trend insights, suspicious user detection (data scrapers) and product feature prioritization; used by C-suite for decision-making
- Optimized database schema by normalizing 50+ tables into relational structures, reducing system errors, storage overhead, and improving query performance for LLM-driven text-to-SQL pipelines
- Engineered a RAG proof-of-concept for text-to-SQL, reducing LLM token usage by 44% and operational costs by 35% while increasing query accuracy

Columbia University, Internet Real-Time Lab — New York, NY
Research Assistant (NLP/LLMs) — Jan 2025 – May 2025
- Developed a curated and annotated dataset of 450+ privacy policies with clauses on AI training, cross-border data transfers, and advertiser data sharing to conduct research on automated privacy policy analysis using LLMs
- Implemented benchmarking of multiple models (GPT-4o, DeepSeek, Mistral, BERT-family) under zero-shot, few-shot, and Retrieval-Augmented Generation (RAG) setups, achieving up to 99% recall in classification tasks

Samsung R&D Institute India — Bangalore, IN
Software Development Intern — May 2023 – Jul 2023
- Collaborated with On-Device AI Solutions team to develop a Recurrent Neural Network (RNN) based model for predicting user's next smartphone tasks, leveraging monthly data from 10+ apps, achieving an RMSE of 0.2

ML Research Intern — Jul 2022 – Feb 2023
- Designed and constructed a lightweight Multilayer Perceptron model deployed on-device (TFLite) for real-time mood prediction from sensor data, achieving 93.75% accuracy
- Published findings at the 2023 IEEE CONECCT conference; awarded Certificate of Excellence

Project Experience

Interactive Data Analytics Assistant using LLMs [Link] Python, LangChain, LangGraph, Multi-Server MCP
- Developed an intelligent assistant powered by Google Gemini-2.5-Flash, capable of performing end-to-end data analytics through conversational queries, including statistical analysis, insights generation, and visualizations
- Integrated multiple MCP-based tools (`csv_analyzer`, `local_python_executor`) to dynamically inspect datasets, run Python/NumPy/Pandas code, and generate visual outputs with Matplotlib and Seaborn
- Implemented system prompts and memory for context retention, enabling accurate and reliable multi-turn analysis

Renewable Energy Market and Risk Analysis [Link] Python, PyTorch, Airflow, PostgreSQL, Power BI
- Constructed a scalable ETL pipeline with Apache Airflow and PostgreSQL, automating data ingestion, transformation, and storage for energy and weather data across five Spanish cities
- Built a Long Short-Term Memory (LSTM) model in PyTorch, attaining 0.02 RMSE in energy price forecasting
- Computed 7-day VaR, CVaR, and volatility using SQL, integrating insights into a real-time Power BI dashboard for risk analysis

Pediatric Pneumonia Detection from Chest X-ray Images [Link] Python, TensorFlow, NumPy, Scikit-Learn
- Teamed up and created a CNN model to detect pneumonia from chest X-ray images, with 95.97% accuracy
- Designed and trained a DCGAN to generate 1.8K synthetic images for the minority class, addressing class imbalance
- Outperformed a fine-tuned pre-trained VGG16 model by achieving 2% higher accuracy, with a recall of 98% for Pneumonia class and 91% for Normal class on the benchmark dataset