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

Skills

Programming Languages and Databases: C, C++, Python, MySQL, PostgreSQL
Data Engineering and Analysis: Apache-Airflow, MS Excel, PowerBI, NumPy, Pandas, Matplotlib, Seaborn
Machine Learning, Computer Vision and NLP: Scikit-Learn, TensorFlow, PyTorch, OpenCV, NLTK
Additional tools: FastAPI, Streamlit, Git

Professional Experience

Samsung R&D Institute India-Bangalore — Bangalore, IN
Software Development Intern — May 2023 - Jul 2023

- Collaborated with the On-Device AI Solutions team to develop an RNN-based model for predicting smartphone tasks, using monthly data from 10+ apps
- Reconstructed a graph-based approach to log user activity patterns, utilizing previous 7 days' graphs as sequential adjacency matrices to predict next day's graph, testing its applicability to the current problem statement
- Identified and communicated inefficiency and incompatibility of graph-based approach due to high memory usage despite an RMSE of 0.2 in task prediction, informing the team's decision-making process

Samsung PRISM - ML Research Intern (On-Device AI Team) — Jul 2022 - Feb 2023

- Partnered on a Sensor-based Mood Profiling system to detect emotions in real-time, integrating accelerometer, gyroscope, and heart rate data from 78 volunteers for accurate mood prediction
- Developed two Android WearOS apps using Java and Android Studio for seamless data collection and mood tracking, with Firebase as the backend database to store and sync sensor data in real-time
- Engineered a lightweight TFLite model based on a Multi-Layer Perceptron architecture, achieving 93.75% accuracy by optimizing sensor data inputs (Accelerometer, Gyroscope, Heart Rate) for mood prediction
- Presented and published findings at the 2023 IEEE CONECCT, earning a Certificate of Excellence

Data Science and Analytics Project Experience

Renewable Energy Market and Risk Analysis [Link] Python, PyTorch, Apache-Airflow, PostgreSQL, PowerBI

- Designed 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 an LSTM (Long Short-Term Memory) model in PyTorch, achieving 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

German Credit Risk Analysis and Classification Model [Link] Python, NumPy, Pandas, Scikit-Learn

- Preprocessed data, performed EDA with visualizations and feature engineering on German Credit Data
- Selected 12 key attributes using statistical tests and performed hyperparameter tuning on classification algorithms
- Developed and deployed a full-stack web app with an interactive dashboard and SVM-based classification model (78% accuracy) on Streamlit Cloud for real-time predictions

Superstore Sales Data Analysis [Link] Excel, Power Query, Pivot tables, Pivot Charts

- Built an ETL pipeline using Power Query to clean and analyze sales data across four regions, time, and products
- Analyzed top-performing regions, low-profit areas, and top 10 best-selling products by volume and revenue, with insights on customer segments, shipping mode impact, and order trends
- Created interactive visualizations to track sales, profit, and profit margin trends for data-driven decision-making

Computer Vision Project Experience

Pediatric Pneumonia Detection from Chest X-ray Images [Link] Python, TensorFlow, NumPy, Scikit-Learn

- Teamed up on developing a CNN model to detect pneumonia from chest X-ray images, achieving 95.97% accuracy
- Designed and trained a DCGAN to generate synthetic images for the minority class, addressing class imbalance
- Outperformed a fine-tuned pre-trained VGG16 model by 2% accuracy, with a recall of 98% for Pneumonia class and 91% for Normal class on the benchmark dataset

CNN based American Sign Language Translator [Link] Python, OpenCV, Mediapipe, TensorFlow, NumPy

- Collected and labeled 5200 ASL hand image data for training model using Python, OpenCV, and Mediapipe
- Constructed a CNN architecture using TensorFlow to classify 26 ASL alphabets with an accuracy of 99.71%
- Fine-tuned MobileNetV2 model to enhance performance of system and improved accuracy to 99.81%

NLP Project Experience

Financial Sentiment Analysis and Categorization [Link] Python, NLTK, TensorFlow, PyTorch, Transformers

- Carried out text preprocessing with NLP techniques (tokenization, stopword removal, lemmatization, n-gram extraction) to improve model performance on financial datasets
- Evaluated various models (Logistic Regression, SVM, Random Forest, DNNs, LSTM, BiLSTM, BERT) to identify the best-performing approach
- Achieved up to 90.5% accuracy on test data using Bidirectional LSTM for financial sentiment classification

Recipe Recommender based on Ingredients [Link] Python, NLTK, Spacy, NumPy, Pandas, Scikit-Learn

- Trained a custom NER model on TASTEset (700 recipes, 13,000+ entities) to extract ingredients from raw text
- Preprocessed 6,000+ recipes followed by NER-based ingredient extraction and TF-IDF vectorization
- Developed a recipe recommendation model based on ingredient, cuisine, and dietary preferences using cosine similarity across 10 diets, 20 courses, and 50+ cuisines

Software and Database Project Experience

SaaSy-Events - Event Management Portal [Link] Python, FastAPI, AWS RDS

- Collaborated with a team of six to develop a microservices-based event management application, with an Angular frontend deployed on an S3 bucket, a Python FastAPI backend, and a MySQL database hosted on AWS RDS
- Designed and implemented a robust MySQL database to manage users, organizers, events, and tickets, optimizing query performance for seamless event operations
- Developed a user management microservice with CRUD operations, Google OAuth 2.0 authentication, and JWT-based authorization, ensuring secure access control and integrated third-party APIs for ticket QR code generation

NoCodeML: Simplifying the Data Science Workflow [Link] Python, Streamlit, Pandas, Scikit-Learn, Seaborn

- Built a no-code platform enabling users to perform end-to-end data science workflows, tested on 20+ datasets
- Implemented features for data cleaning, transformation and exploratory data analysis with visualizations
- Streamlined model training/testing and data preparation (train-test split, normalization, encoding), reducing data preparation time by 50%

Online Banking System [Link] Python, MySQL, Streamlit

- Designed an online banking system for banking operations like account management, fund transfers, credit/debit card management, and loan repayment
- Optimized MySQL databases and queries for transaction history, account balances, and secure fund transfers
- Integrated real-time data from MySQL server with an interactive front-end for an enhanced user experience