

Suchit Bhayani

suchit.bhayani@gmail.com | (925) 875-8737 | linkedin.com/in/suchit-bhayani | suchitbhayani.github.io

EDUCATION

University of California San Diego

Bachelor of Science in Data Science, Minor in Mathematics

San Diego, CA

Expected: June 2027

- GPA: 3.9/4.0

EXPERIENCE

Nike

Data and Machine Learning Engineer Intern

June 2025 - Aug. 2025

- Designed governance frameworks for enterprise ML, generative AI, and BI infrastructure, enforcing compliance
- Implemented model card functionality in Databricks to improve transparency and governance of deployed models
- Engineered workflow using MLflow to automatically flag ethics violations in deployed models and notify relevant teams
- Scaled a recommender system using PySpark, speeding up product similarity recommendations by **20%**

UC San Diego

Teaching Assistant

Jan. 2025 - June 2025

- Automated grade calculations and imports using pandas and canvas API, reducing manual workload
- Facilitated engaging discussion sections, wrote testing scripts, and created exams to enhance learning outcomes
- Resolved **500+** tickets regarding data structures and algorithms during office hours and on the class forum
- Received a **100%** recommendation rate across multiple offerings of various courses, totaling **1000+** students

Li Lab

Machine Learning Researcher

Oct. 2024 - June 2025

- Engineered and cleaned datasets to build ML models analyzing key drivers of hematopoietic stem cell self-renewal
- Conducted regression and correlation analyses using scikit-learn, identifying **31** genes for further biological testing
- Performed time series differential gene expression analysis in R, identifying biologically significant gene clusters

Digital Prudentia

Data Science and Engineer Intern

June 2024 - Sep. 2024

- Utilized retrieval-augmented generation (RAG) with Azure SDK to develop image-based skin cancer detection model
- Linked documents with text and image data to the model using Azure AI Search, enabling contextual insights
- Used prompt engineering to reduce inference costs by **30%**, optimized model with hyperparameter tuning

PROJECTS

Music Recommender System

Full-Stack Development, Software Engineering, Agile Development, REST API, Containerization, MERN Stack

- Trained a LightFM-based music recommender system, deployed via FastAPI, and containerized using Docker
- Built full-stack web app using React, Express/Node.js, and MongoDB to serve and display music recommendations
- Implemented client-side features for real-time tracking and visualization of top tracks and listening trends

Personalized AI Health Insights

Big Data, Scalable Systems, Natural Language Processing (NLP), Healthcare AI, Data Engineering

- Developed ETL pipeline in dask to efficiently process and analyze millions of Apple Watch health data points
- Engineered health insight generation pipelines with fine-tuned Hugging Face LLMs, enabling health-specific inferences
- Built an interactive dashboard using plotly to visualize trends, insights, and actionable recommendations

Accelerating ML with Automated Feature Engineering

AutoML, Large Language Models, Statistical Feature Selection

- Automated an ETL pipeline with OpenRouter API for integrating LLM domain knowledge into the AutoML paradigm
- Validated performance of generated features using XGBoost and RandomForest models across 3 benchmark datasets

SKILLS

- **Languages:** Python, Java, JavaScript, SQL (PostgreSQL, NoSQL), R, HTML/CSS
- **Libraries:** pandas, PySpark, dask, scikit-learn, NumPy, scipy, plotly, seaborn, Matplotlib, BeautifulSoup
- **Frameworks:** PyTorch, TensorFlow, Keras, React, Express, Node.js, FastAPI
- **Tools:** AWS (S3, EC2), Azure (DevOps), Docker, Git, GitHub, MLflow, Jupyter, Linux/Unix
- **Interests:** Gym, nature, music, basketball, football