

# 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*

San Diego, CA  
Sep. 2023 - June 2026

## EXPERIENCE

**UC San Diego Health**

*Data Science Intern*

Oct. 2025 - Present

- Develop NLP pipelines to predict patient risk from clinical notes using transformer embeddings and TF-IDF
- Model clinical risk from patient records using XGBoost, delivering interpretable insights into key predictors
- Performed EDA on **5,000+** patient records across nutrition, functional, and cultural risk factors
- Utilize AWS WorkSpaces to run cloud-based model training, data analysis, and data processing workflows

**UC San Diego**

*Teaching Assistant*

Jan. 2025 - Present

- 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

**Nike**

*Data and Machine Learning Engineer Intern*

June 2025 - Aug. 2025

- Built a MLflow pipeline to backfill model metrics, achieving a **98%** success rate with zero manual intervention
- Enhanced MLOps pipelines with Jenkins metric gates, automating model promotion via performance validation
- Designed and implemented model card functionality in Databricks to improve transparency of deployed models
- Scaled feature engineering in a recommender system using PySpark, speeding up recommendations by **20%**

**Li Lab**

*Machine Learning Researcher*

Oct. 2024 - June 2025

- Engineered predictive ML models using pandas and sklearn, analyzing key drivers of stem cell self-renewal
- Conducted correlation analyses to isolate **31** high-priority genes, reducing experimental search space by **99%**
- Developed R-based time series DGE workflows to identify **6** significant gene cluster using DESeq2

**Digital Prudentia**

*Data Science Intern*

June 2024 - Sep. 2024

- Architected a multi-modal retrieval-augmented generation (RAG) system on Azure for skin cancer detection
- Integrated Azure AI Search as a vector database to index and retrieve embeddings from text and imagery
- Reduced inference costs by **30%** through systematic hyperparameter tuning and strategic prompt engineering

## PROJECTS

**Music Recommender System**

- Trained a music recommender system, deployed via FastAPI and Render, and containerized using Docker
- Built full-stack web app using React, Express/Node.js, and MongoDB to serve recommendations

**Personalized AI Health Insights**

- 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 Large Language Models

## SKILLS

- **Languages:** Python, SQL (MySQL, PostgreSQL, NoSQL), R, Java, JavaScript, HTML/CSS
- **Libraries/Frameworks:** pandas, PySpark, dask, scikit-learn, NumPy, PyTorch, TensorFlow, Keras, FastAPI
- **Tools/Platforms:** AWS (S3, EC2), Azure (DevOps), Docker, Databricks, MLflow, Jenkins, Linux/Unix