Suchit Bhayani

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

EDUCATION

University of California San Diego San Diego, CA
Bachelor of Science in Data Science, Minor in Mathematics Expected: June 2027

- GPA: 3.9/4.0
- Relevant Coursework: Data Management, Scalable Analytics, Data Mining, Data Visualization, Probability, Statistics

SKILLS

- **Languages:** Python, Java, SQL (SQLite, PostgreSQL, NoSQL), HTML/CSS, JavaScript
- **Libraries:** pandas, PySpark, dask, scikit-learn, NumPy, scipy, plotly, seaborn, Matplotlib, BeautifulSoup, pytest
- **Frameworks:** PyTorch, TensorFlow, Keras, React, Express, Node.js, D3.js, FastAPI, JUnit
- **Tools:** AWS (S3, EC2), Azure (DevOps, Blob Storage, AI Search), Apache Spark, Databricks, Containerization (Docker), CI/CD (Github Actions), MLOps (MLflow), MongoDB, Linux/Unix, Bash, Git, Tableau, Excel, Word

EXPERIENCE

Nike June 2025 - Aug. 2025
Data and Machine Learning Engineer Intern

- Designed scalable governance frameworks to guide ethical use of BI, AI/ML, and GenAI across enterprise infrastructure
- Engineered `Databricks` workflow leveraging `MLflow` to automatically flag ethics violations in deployed models
- Scaled a `recommender system` using `PySpark`, enabling product similarity recommendations for consumers

UC San Diego | Data Science Department Jan. 2025 - Present
Teaching Assistant

- Tutor for DSC 20: Programming and Data Structures (`Python`), DSC 30: Data Structures and Algorithms (`Java`)
- Apply understanding of Python and Java via office hours and online question-answering platform

UC San Diego Health | Li Lab Oct. 2024 - Present
Machine Learning Researcher

- Conduct `time series` differential gene expression analysis in RNA sequencing data points using `PyDESeq2`
- Build ML models and use `causal inference` techniques to identify and analyze key factors of stem cell self-renewal
- Analyze correlations between genes and gene expression programs (GEPs) in progenitor and stem cells

WorldQuant June 2024 - Present
Quantitative Research Consultant

- Research, implement, and backtest 500+ equity trading strategies with `FastExpression` for potential portfolio integration
- Present research of high performing alpha strategies (2.83 Sharpe) to portfolio managers and executives

Digital Prudentia June 2024 - Sep. 2024
Data Science and Engineer Intern

- Utilized `retrieval-augmented generation` with `Azure OpenAI` to develop an image-based skin cancer detection model
- Created and stored multimodal embeddings in vector database with `Azure AI Search` and `Azure Blob Storage`
- Handled 700,000+ medical images and metadata, applying scalable practices for efficient model training and analysis

PROJECTS

Music Recommender System
Full-Stack Development, RESTful API, Containerization, MERN Stack, Software Engineering

- Built `LightFM` music recommender using `React`, `Express/Node.js`, and Python `FastAPI`, containerized with `Docker`
- Integrated `MongoDB Atlas` for data storage and managed API communication between frontend, backend, and ML service
- Implemented Spotify `OAuth 2.0` for secure user authentication and data access, enabling dynamic user preference retrieval

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

- Utilized `dask` to process and analyze millions of rows of Apple Watch health data, identifying key underperforming metrics
- Developed health insight generation pipelines leveraging fine-tuned `HuggingFace` LLMs, enabling health-specific inferences
- Developed 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