

Ritviksiddha Penchala

🇺🇸 U.S. Citizen

🌐 ritvikpen

in ritviksiddha-penchala

✉ rpenchala@ucsd.edu

☎ (408) 981-9355

EDUCATION

University of California, San Diego, La Jolla, CA

B.S. in Computer Science with Specialization in Bioinformatics

Minor in Biology

GPA: 3.7/4.0

Sep. 2021 – Dec. 2024

SKILLS & COURSES

Programming Languages: Python, C++, SQL, R, Java, JavaScript

Technologies/Libraries: PyTorch, Pandas, Sci-Kit Learn, SciPy, Seaborn, TensorFlow, Spark, jQuery, TypeScript, React, Node, Git, Kubernetes, Docker, Azure, AWS, Databricks

Coursework: Machine Learning, Statistical Natural Language Processing, Algorithms, Advanced Data Structures, Software Tools and Techniques, Biological Databases, Bioinformatics Algorithms

EXPERIENCE

University of California – San Francisco, San Francisco, CA

Data Science Specialist

Oct. 2025 – Present.

- Primarily tasked with investigating **large language models** in the context of public health and healthcare
- Assessed commercial models (ChatGPT, Gemini, Deepseek, Grok, etc.) for policy biases in a simulated pandemic.
- Implemented an **LLM-as-a-Judge** pipeline for the automated scoring of model reasoning in public health tasks.
- Utilized a custom machine learning pipeline consisting of **classifier, text generation, and prediction** models to extend survey data and simulate new personas for public health modeling.

University of California – San Diego, La Jolla, CA

Research Data Analyst

Jul. 2022 – Dec. 2024

- In the Sanchez-Roige Lab, analyzed data from **100,000 research participants** to investigate risk factors for Opium Use Disorder and utilized **machine learning classifiers** to identify novel predictors in high-dimensional datasets
- In the Saier Lab, utilized **R, Numpy, and Pandas** to analyze large, unannotated proteomic datasets for immunological significance, and presented findings at the **International Conference on Microbiome Engineering** hosted by AIChE.

Kaiser Permanente, Pleasanton, CA

Data Engineering Intern

Jun. 2023 – Sep. 2023

- Contributed to the modernization of an Electronic Health Record serving **over 8 million patients**
- Automated legacy database migration using **PySpark and SQL** on **Azure Databricks**
- Created **Python**-based testing scripts to ensure data integrity and accuracy during migration

Center for Applied Internet Data Analysis (CAIDA), La Jolla, CA

Web Development Intern

Jan. 2023 – Dec. 2024

- Managed and enhanced the CAIDA website, attracting over **10,000 monthly visitors**
- Applied **JavaScript, HTML/CSS, HUGO, jQuery**, and other web technologies to improve responsiveness and accessibility
- Improved and extended **REST API** integration for dynamic content delivery across previously static pages

NonExomics, Boston, MA

Data Engineering Intern

Mar. 2022 – Sep. 2022

- Developed automated analyses pipelines in **Python and Bash** for large-scale genomic datasets
- Leveraged **Amazon Web Services** cloud infrastructure to identify non-exomic genomic regions linked to carcinogenesis

PROJECTS

Infant and Maternal Health in Latinx Populations

Mentor: Dr. Janardhan Mydam, Humboldt Park Health, Chair of Pediatrics

Aug. 2024 – Present.

- Analyzed **CDC datasets** using **Python, R, and SAS** to uncover predictors of maternal and infant health outcomes
- Presented at the **Pediatric Academic Societies Scientific Conference** and **American Academy of Pediatrics Conference**

Melanoma Image Classification Model

Mentor: Dr. Edwin Solares, University of California San Diego, UC President's Postdoctoral Fellow

Jul. 2024 – Aug. 2024

- Built **image classification models** in Python with varying architecture, hyperparameters, pre-processing, etc. and compared to experimentally optimize for accuracy, efficiency, and complexity
- Achieved **over 97% classification accuracy**, surpassing the diagnostic performance of the median healthcare worker

NLP Models for Low-Resource Languages

Mentor: Dr. Ndapa Nakashole, University of California San Diego, Assistant Professor

Mar. 2024 – Jun. 2024

- Fine-tuned state-of-the-art NLP models (**mBERT and PEGASUS**) using custom tokenization, encoder adjustments, and targeted retraining to support low-resource languages such as Telugu, Pidgin, Swahili, etc.
- Achieved **over 70% accuracy** on multiple target languages, including those not seen during pretraining, demonstrating marked improvements in cross-lingual generalization