

Nathaniel del Rosario

natdosana@gmail.com | [linkedin.com/in/natdosan](https://www.linkedin.com/in/natdosan) | github.com/natdosan | natdosan.github.io

Education

University of California, San Diego

2025

Bachelor of Science, Data Science

La Jolla

- Graduate: Computer Vision, Recommender Systems, Deep Learning
- Undergraduate: Principles & Techniques of Data Science, Geospatial Data Science, AB Testing, Statistics, Database Systems, Cloud Computing & Scalable Analytics, NLP, Data Visualization, Computational Neuroscience

University of California, Berkeley

2023

Visiting, Computer Science

Berkeley

- Artificial Intelligence, Probabilistic Modeling & MDPs, Reinforcement Learning, Machine Learning, Algorithms

Skills / Tools / Interests

- Python, Pandas, Dash/Plotly, Numpy, Sklearn, TensorFlow, PyTorch, Kera, Cuda, Dask/Ray, SQL, PostgreSQL, D3
- AWS, Google Cloud, Docker, Kubeflow, Spark, Databricks, OpenAI Gym, Distributed Systems, Github, ArcGIS, Jira
- Sports Analytics, Creative Problem Solving, Music Composition, French, Japanese, Jazz Trumpet, Impressionist Art

Experience

Hacioglu Data Science Institute, San Diego

April 2024 - Present

Undergraduate Researcher

La Jolla

- Researching & designing models to predict [public transportation accessibility](#) in New York City (RMSE of .1785)
- Utilizing machine learning to predict crime hotspots in metropolitan areas, advised by [Prof. Wartell](#) and [Prof. Zaslavsky](#)

University of California, San Diego

September 2023 - March 2024

Undergraduate Student Instructor

La Jolla

- Beta Testing assignment and exam questions, hosting Office Hours for a [data science course](#) of >500 students
- Updated deployment of course website using github pages & Docker working under [Suraj](#) and [Tiefenbruck](#)
- Grading and hosting Office Hours for upper division [data science course](#) of >700 students under [Shannon Ellis](#)

San Diego Supercomputer Center

June 2023 - September 2023

Machine Learning Engineer Intern

La Jolla

- Utilized **AWS EC2, S3/PostgreSQL** to allow user data queries and vectorized code to prune computations by **7%**
- Designed **Content Filtering Recommender System** supporting **5+** user constraints for **~200,000** items
- Implemented an RL agent using **SB3**, non greedy **Q-Learning** to improve recommendations by **50%**

Deloitte

February 2023 - June 2023

Data Science Apprentice

Remote

- Cleaned data w/ **3000+** features, **1 billion** rows using **Dask/** vectorized **Pandas** to decrease cleaning runtime by **20%**
- Leveraged **XGBoost, Lasso** to identify **850** significant features, predict drug use in young adults with **81%** accuracy
- Tuned Hyperparameters to improve **F1 score** by **from .35 to .70** and identify **10** unique demographics most at risk

Chan Zuckerberg Biohub

August 2022 - January 2023

Data Science Intern

Remote

- Designed **3** gene-virus pair querying algorithms in Python to query comparisons of specific Host Factors
- Implemented **5+** UI/UX improvements to a database tool to promote intuitive workflow
- Wrote documentation for **23** functions from scratch and improved codebase readability

Chan Zuckerberg Biohub

June 2022 - August 2022

Data Science Intern

San Francisco

- Built & deployed **3** interactive dashboard visualizations of CRISPR screen comparisons using **Dash/Plotly/Vercel**
- Improved data processing of a Nextflow data pipeline (**16,000,000** data points) to minimize runtime by **10%**
- Created **6** new visualization algorithms for genome comparisons between **20,000** genes across **30+** virus screens

Projects & Leadership

Seq2Seq Language Translation | *PyTorch, Google T5 Model*

- Fine tuned Google T5 on English to Mandarin text translation, increasing robustness (BLEU score of **15**)
- Developed Custom training pipeline from HuggingFace dataset retrieval for English to Mandarin translation task

Exploring CNN Architecture for Semantic Segmentation | *PyTorch, OpenCV* | [report](#)

- Implemented different UNet architectures with AdamW Optimization, Data Augmentation, weighted cross entropy loss, learning rate scheduling to improve IoU score from **.055** to **.071** and pixel accuracy from **73.4%** to **75.1%**
- Utilized FCN ResNet-101 for transfer learning further improving IoU score to **.33** and validation accuracy to **87.3%**

Spotify User Persona Clustering | *SpotiPy, Scikit-Learn* | [code](#)

- Wrote an automated pipeline using **SpotiPy, Spotify API** to scrape, preprocess, feature engineer the data
- Performed PCA and K-Means to identify **6** unique listening personas for identifying target audiences

ACM UCSD

April 2023 - Present

Projects Director

La Jolla