Nathaniel del Rosario

natdosana@gmail.com | linkedin.com/in/natdosan | natdosan.github.io

EDUCATION

University of California, San Diego
La Jolla, CA

B.S. Data Science — 2025
- Graduate (cross enrollment): Computer Vision, Recommender Systems, Deep Learning
- Undergraduate: Principles & Techniques of Data Science, Geospatial Data Science, AB Testing, Statistics, Database Systems, Cloud Computing & Distributed Systems, NLP, Data Visualization, Computational Neuroscience

University of California, Berkeley
Berkeley, CA

Visiting, Computer Science — 2023
- Artificial Intelligence, Probabilistic Modeling & MDPs, Reinforcement Learning, Machine Learning, Algorithms

SKILLS

Python, Pandas, Dash/Plotly, Numpy, Scikit-learn, TensorFlow, PyTorch, Tensorflow, Cuda, Dask/Ray, SQL, PostgreSQL, D3
AWS, Google Cloud, Docker, Kubeflow, Spark, Databricks, OpenAI Gym, Distributed Systems, Github, Jira, ArcGIS

EXPERIENCE

San Diego Supercomputer Center
June 2023 - September 2023

Machine Learning Engineer Intern
- Utilized AWS EC2, AWS 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 over 200,000 items
- Implemented an RL agent using SB3, non greedy Q-Learning to improve recommendation quality

Chan Zuckerburg Biohub
August 2022 - January 2023

Data Scientist
- Designed algorithms to compare across 30+ virus screens to yield insights in virus-host interactions using Pandas methods
- Implemented 5+ UI/UX improvements to a database tool to promote intuitive workflow
- Wrote documentation for 23 functions from scratch and improved codebase readability using Readthedocs

Chan Zuckerburg Biohub
June 2022 - August 2023

Data Science Intern — San Francisco
- Built & deployed 3 interactive dashboard visualizations of CRISPR screen comparisons using Pandas/Dash/Plotly/Vercel
- Improved data processing of a Nextflow data pipeline (16,000,000 data points) to minimize runtime when traversing by 10%
- Created 6 new visualization and analysis methods for genome comparisons between 20,000 genes

Hacioglu Data Science Institute
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 methods to identify and predict crime hotspots in metropolitan areas supervised by Professor Wartell and Professor 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 over 500 students
- Updated deployment of course website using github pages & Docker supervised by under Suraj and Tiefenbruck
- Grading and hosting Office Hours for upper division data science course of over 700 students under Shannon Ellis

Deloitte
February 2023 - June 2023

Data Science Apprentice
- Cleaned data w/ 3000+ features, 1 billion observations 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, class weighting to improve F1 score from .35 to .70 and identify 10 highest risk demographics

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
- 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
- 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

Data Science Society
October 2023 - Present

Projects Director
- Sourcing and mentoring 12 data science projects during the academic year