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 (3.8/4.0 GPA) *2025*

- Graduate (3.9/4.0 GPA, cross enrollment): Computer Vision, Recommender Systems, Deep Learning
- Undergraduate: Principles & Techniques of Data Science, Geospatial Data Science, AB Testing, Statistics, Databases, Operating 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, NumPy, PyTorch, Sci-Kit Learn, Tensorflow, Dash/Plotly, Cuda, Dask/Ray, SQL, PostgreSQL, D3, Streamlit
AWS S3, EC2, RDS, ECR, EKS, Docker, Kubernetes, Spark, Hadoop, Databricks, Distributed Systems, Github, Jira, ArcGIS

Experience

Bio-Rad June 2024 - Present
Data Science Intern - Bioinformatics, Clinical Diagnostics Group *Pleasanton*

- Built ETL pipeline w/ Pandas, RestAPI to ensure 100% data integrity & improved consistency from 91% to 99.9%
- Implemented unit tests & utilized Dask/NumPy to achieve 5.1x / 80.1% speedup on data validation across datasets
- Leveraged AWS ECR, Docker, PostgreSQL to deploy database, reduced storage usage by 35% through schema optimization
- Saved 61% in computing budget per month by utilizing spot instances based on stakeholder heuristics

San Diego Supercomputer Center June 2023 - September 2023
Machine Learning Engineer Intern *Remote*

- Designed Content-Based Filtering Recommender System utilizing Cosine and Jaccard metrics, supporting 5+ user constraints; Trained an RL agent using SB3, non greedy Q-Learning to improve recommendation quality by 100 iterations
- Utilized AWS EC2, AWS S3, & PostgreSQL for data queries and vectorized code to achieve 1.7x runtime speedup
- Deployed Recommender System on AWS EC2, Lambda, achieving a stateless design that scaled to process 200,000+ points

Deloitte February 2023 - June 2023
Data Science Apprentice *Remote*

- 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

Chan Zuckerburg Biohub June 2022 - January 2023
Data Science Intern - Infectious Disease Group *San Francisco*

- Built 9 interactive visualizations of CRISPR screen comparisons between 20000 features using Pandas/Dash/Plotly
- Improved data processing of a Nextflow data pipeline (16,000,000 data points) to minimize runtime by 10%
- Designed algorithms to compare across 30+ virus screens to yield insights in virus-host interactions using vectorized code
- Wrote documentation for 23 functions from scratch and improved 3K+ codebase readability using Readthedocs

University of California, San Diego September 2023 - March 2024
Instructional Assistant *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

Projects & Leadership

University of California, San Diego April 2024 - Present
Machine Learning Researcher *La Jolla*

- Investigating robustness of LLM's for Spatial Data Science - Spatial Information Systems Lab
- Researching & designing models to predict public transportation accessibility in New York City (RMSE of .1785)
- Utilizing machine learning to identify and predict crime hotspots in cities supervised by Prof. Wartell and Prof. Zaslavsky

Exploring CNN Architecture for Semantic Segmentation — PyTorch, OpenCV, HuggingFace February 2024

- 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%

Data Science Society October 2023 - Present
Projects Director

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