Aidan Nguyen

aidannguyen27@ucla.edu
https://www.linkedin.com/in/nguyenaidan/
https://github.com/aidannguyen23

EDUCATION

University of California, Los Angeles (UCLA)   Expected June 2026
B.S. Statistics and Data Science   GPA: 3.97
Activities: Statistics Club (Social Chair), DataFest (Marketing Team), Bruin Sports Analytics (Consultant)
Relevant Coursework: Python with Applications, Statistical Programming with R, Data Structures and Algorithms, Linear Algebra, Multivariable Calculus, Probability, Data Justice & Society

EXPERIENCE

Signaling Systems Lab   January 2024 – Present
Undergraduate Research Student
- Analyzed Single Cell RNA sequencing data (6M+ rows) with R, caret, and Seurat to classify HSPC cells
- Deployed supervised learning models, including SVMs and coefficient classifiers, to evaluate data quality & scalability
- Fine-tuned SVM hyperparameters with R and e1071, improving model accuracy by 10% through cross-validation
- Secured grant funding for the lab by creating visual t-SNE plots that verified data from ongoing research initiatives

Dune Road Capital   June 2024 – September 2024
AI Research Intern
- Engineered SQL queries to analyze large financial datasets (10M+ rows) and calculate ETF performance metrics
- Utilized Python/pandas to extract ETF performance data and apply time-series analysis to forecast future returns
- Developed data visualizations with Tableau to present key ETF metrics such as expense ratio and asset allocation

PROJECTS

AWS Covid Analysis | *Amazon RDS, S3, EC2, QuickSight, MySQL*
- Designed a scalable data pipeline to process COVID-19 data using Amazon RDS & S3 for data storage and retrieval
- Applied EC2 for running MySQL queries, data processing, and transforming large datasets into manageable formats
- Connected AWS RDS to MySQL Workbench for database management and visualization of real-time data
- Developed an interactive dashboard in AWS QuickSight to visualize key pandemic metrics across regions

DataFest at UCLA | *R, dplyr, ggplot2, caret, tidyverse*
- Achieved top finalist out of 400+ participants in a 48-hour data hackathon, focused on analyzing complex datasets
- Analyzed 7 spreadsheets with 100K+ rows of user and site interaction data to improve a student textbook platform
- Utilized R and tidyverse to clean, transform, and merge datasets, identifying key patterns in user behavior
- Developed visualizations recommending data-driven improvements for greater engagement & educational outcomes

YouTube Success Metrics Analysis | *PyTorch, transformers, pandas, numpy, googleapiclient, nltk, matplotlib*
- Analyzed 600,000+ YouTube comments using a fine-tuned DistilBERT model for multilingual sentiment analysis
- Created predictive models, including a gradient boosting regressor, to assess how titles & thumbnails impact views
- Implemented OpenCV to implement face detection in video thumbnails, correlating findings with viewership metrics
- Visualized data insights using matplotlib to present findings on viewer engagement and content effectiveness

Galaxy Classification | *Python, numpy, pandas, tensorflow, scikit-learn, matplotlib*
- Applied K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression to classify galaxy data into 5+ categories
- Achieved 87% accuracy using KNN after fine-tuning hyperparameters and optimizing feature selection
- Developed a convolutional neural network (CNN) using TensorFlow's Sequential model, achieving 90% accuracy

Programming Languages & Tools: Java, C++, Python (pandas, numpy, scikit-learn, plotly, flask, tensorflow), JavaScript, Git, SQL, R, Excel, Tableau, AWS (RDS, S3, EC2)