

Stephen Yu

[Email](#) | [LinkedIn](#) | [Website](#) | [GitHub](#)

Education

University of Pennsylvania | Master's Computer and Information Technology

Philadelphia, PA

GPA: TBD | C, Python, Natural Language Processing, Deep Learning

Aug 2024 – May 2026

University of California, Los Angeles | B.S. Statistics and Data Science

Los Angeles, CA

GPA: 3.88/4.0 | C++, Python, R, Machine Learning, Data Mining, Mathematical Statistics

Sep 2020 – Jun 2024

Experience

UCLA Statistical Consulting | Statistical Consultant

Los Angeles, CA

Python: xgboost, scikit-learn, pandas, numpy | R: dplyr, rose, dmwr

Apr 2024 – May 2024

- Merged pitch and injury data sets by date and pitcher to form set containing 100k+ observations
- Utilized ROSE and SMOTE to balance out minority class leading to an increase testing accuracy by 30%
- Modelled with xgboost, logistic regression, and random forest achieving up to 65% accuracy on test set

Tsiang Statistics Lab | Undergraduate Researcher

Los Angeles, CA

R: rvest, tidyverse, ggplot2, mice, mclust class, mass, dmwr

Sep 2023 – Dec 2023

- Scripted an algorithm to scrape 25 predictors for 4577 observations from assorted websites
- Imputed data using MICE and leveraged SMOTE to create 3000+ synthetic data points for underrepresented class
- Utilized dimension reduction techniques such as Lasso Regression and Stepwise BIC to reduce to 9 significant predictors

Northrop Grumman | Software Engineer Intern

Woodland Hills, CA

MATLAB: Simulink

Jun 2022 – Aug 2022 / Jun 2023 – Aug 2023

- Performed testing on 11 hyper models for hybrid navigation integration
- Effectively updated JIRA to communicate task progress to my SCRUM team for project roadmap
- Developed a byte unpack algorithm for SCC buses, meeting IICD standards for IMU to System Processor bit stream
- Documented 100+ risks and opportunities using proprietary risk-management software (TETRA)

Project Kairos | Resale Strategist

Porter Ranch, CA

Microsoft Excel

Mar 2021 – Dec 2021

- Provided strategic guidance for the planning and execution of retail arbitrage strategies generating \$20k+ in returns
- Predicted optimal sell points based off seasonality, supply, and implied market volatility to increase profits by 40%

Electrum Homes | Data Analyst Intern

Irving, TX

Microsoft Excel

May 2021 – Sep 2021

- Scraped data from Zillow to expedite algorithm inputs which increased team productivity by 60%
- Examined 1k+ residences to gain experience in resale, rentals, renovation, financing, contracts, wholesaler, and property tax

Projects

Smart Sprite

Python: tensorflow, pygame, pandas, numpy

Aug 2023 – Apr 2024

- Constructed an unsupervised reinforcement learning algorithm on an agent to train and progress through a 2D platformer
- Utilized neural network mutation to tweak 400+ parameters that use present state stimuli to dictate decisions
- Agent learned to complete levels with enemies and obstacles on average 10% faster than human

K-Cluster Gaussian Hotspots (KCGH)

Python: scikit-learn, scipy, seaborn, pandas, numpy

Jan 2024 – Mar 2024

- Devised a novel supervised learning algorithm using gaussian clusters within each class
- Compared accuracy to K-Nearest Neighbor on real world datasets resulting in an average boost of 3.5%
- Advantages over KNN include increased interpretability, reduced prediction cost of $O(K)$, and a stronger probabilistic insight

IMDB Sentiment Mining

R: tidyverse, car, mclust, class, mass, proc

Sep 2023 – Dec 2023

- Preprocessed 50k+ IMDB reviews using K-Means Data Clustering and measured inertia to find optimal hyper parameter
- Applied Logistic Regression, KNN, LDA, QDA, and Random Forrest with K-Fold CV receiving a maximum accuracy of 88%

Titanic

C++

Oct 2023 – Dec 2023

- Developed a dynamic data management system using vectors and hash maps for a dataset with 1k+ observations
- Optimized memory usage by 30% through efficient data storage and retrieval techniques
- Leveraged inheritance and polymorphism for modular design and class flow

Additional Information

Awards / Honors: Guidance, Outstanding Intern, Team Project Showcase, Dean's List

Activities / Interests: Surfing, Downhill Skateboarding, Chess, Basketball, Weightlifting