

# Stephen Yu

19655 Singing Hills Dr, Los Angeles, CA 91326

☎ (+1) 818-294-3292 | ✉ [spyu26@seas.upenn.edu](mailto:spyu26@seas.upenn.edu) | 🏠 [stephenyu2.netlify.app](https://stephenyu2.netlify.app) | 🐙 [stephenyu2](https://github.com/stephenyu2) | 📺 [stephenyu2](https://www.youtube.com/channel/UC8W33333333333333333333)

## Education

### University of Pennsylvania (UPenn)

#### MS IN COMPUTER & INFORMATION TECHNOLOGY

Aug 2024 - May 2026

- Planned Coursework: · Software Development (Python & Java) · Computer Systems (C) · Natural Language Processing
- Deep Learning · Computer Vision & Computational Photography · Big Data Analytics · Algorithms & Computation

### University of California Los Angeles (UCLA)

#### BS IN STATISTICS & DATA SCIENCE (GPA: 3.88)

Sep 2020 - Jun 2024

- Coursework: · Computational Statistics with R · Python with Applications · Computer Science (C++)
- Monte Carlo Methods · Mathematical Statistics · Probability · Computation and Optimization for Statistics

## Skills

- Language:** Python (PyTorch, TensorFlow, scikit-learn, Seaborn, pandas), R, SQL, C++, MATLAB (Simulink), Java
- Data Science:** Computer Vision, NLP, Feature Engineering, Reinforcement Learning, Web Scraping, Monte Carlo Simulation
- Other:** Git, JIRA, AGILE Framework, Microsoft Excel, Cross-Functional Solutions, Microsoft Teams, Google PageSpeed

## Work Experience

### Northrop Grumman | Software Engineer Intern (Received Full Time Offer)

Woodland Hills, CA

#### MATLAB: SIMULINK

Jun 2022 - Sep 2024 (Summers Only)

- Improved plane navigational accuracy by 5% by identifying 40+ deficiencies through testing on 17+ IMU hyper models.
- Reduced team workload by 20+ hours by developing a byte unpack algorithm for SCC buses meeting ICD standards.
- Generated \$50k in funding by developing a tool to locate weaknesses in the risk registry concerning airworthiness.

### Project Kairos | Data Analyst

Porter Ranch, CA

#### R: FORECAST, PROPHET, TSERIES, GGLOT2

Mar 2021 - Dec 2021

- Provided data-centric guidance for the planning and execution of retail arbitrage strategies, yielding \$20k+ in returns.
- Increased profits by 10% by quantitatively inferring optimal sell points based on seasonality, supply, and market volatility.

### Electrum Homes | Data Analyst Intern (Received Full Time Offer)

Irving, TX

#### R: RVEST, DPLYR, TIDYVERSE

May 2021 - Sep 2021

- Improved team efficiency by 15% by constructing a robust web scraping algorithm to expedite model inputs.
- Increased cash flow by 1.5% by identifying residential properties with features meeting the buy criteria.

## Research Experience

### Tsiang Statistics Lab | Undergraduate Researcher

Los Angeles, CA

#### R: RVEST, TIDYVERSE, GGLOT2, MICE, MCLUST CLASS, MASS, DMWR

Sep 2023 - Dec 2023

- Scripted an algorithm to scrape 25 features for 4500+ observations to assemble a full data set from 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.

## Projects

### UCLA Statistical Consulting

#### PYTHON: XGBOOST, SCIKIT-LEARN, PANDAS, NUMPY | R: DPLYR, ROSE, DMWR

Apr 2024 - May 2024

- Offered data-driven guidance to MLB Medical Services to potentially reduce pitcher UCL injury by 15%.
- Raised inference accuracy by 8.5% by utilizing ROSE and SMOTE to reduce majority class bias.
- Modeled with XGBoost, logistic regression, and random forest, achieving up to 65% accuracy on the test set.

### A Million Kirby's Fail at Walking

#### PYTHON: TENSORFLOW, PYGAME, IMAGEIO.V3, NUMPY

Mar 2024 - Jun 2024

- Deployed a reinforcement learning agent that completed 4 levels of a platformer video game.
- Utilized a genetic algorithm that improved the dense neural network reward by an average of 7.5% each generation.
- Improved completion time by 12.5% by implementing sightlines, allowing the agent to react precisely around obstacles.

### IMDB Sentiment Analysis

#### R: TIDYVERSE, CAR, MCLUST, CLASS, MASS, PROC

Sep 2023 - Dec 2023

- Designed a supervised learning voting ensemble with K-Fold CV, achieving accuracy 3.5% greater than any singular model.
- Increased testing accuracy by 5% over Bag of Words by using Term Frequency-Inverse Document Frequency (TF-IDF) method.
- Boosted model efficiency by 15% by implementing PCA for dimensionality reduction.