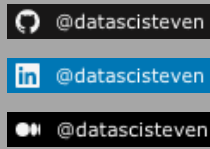


Email: [datascisteven@gmail.com](mailto:datascisteven@gmail.com)  
Portfolio: [datascisteven.github.io](https://datascisteven.github.io)  
Mobile: 646.991.4222  
Location: New York Metro



# STEVEN YAN

## WORK EXPERIENCE

### Data Scientist, StartOut, Remote, 6/2022-7/2023

- Developed algorithms and code for automating data preprocessing tasks: collecting from various sources and API, cleaning, etc.
- Implemented Sparse Principal Component Analysis and Difference in Differences models to produce a set of policy recommendations ranking highest in impact and statistical significance for each state and 120 selected metro areas for widget
- Analyzed data to identify trends and patterns, providing valuable insights to decision-making processes for partnering organizations and key stakeholders
- Delivered weekly report on data trends to assist colleagues in making informed decisions based on data-driven strategies
- Collaborated with manager and marketing team in authoring State of LGBTQ+ Entrepreneurship Report (*to be published*)
- Served as technical lead for small team of interns for researching data and performing data entry and cleaning tasks
- Website: <https://www.startout.org/index>

### Data Science Consultant and Intern, Startout, Remote, 1/2022- 6/2022

### Freelance Educational Consultant, NY Metro, 8/2009–10/2020

### Curriculum Developer, Kaplan Test Prep, NY Metro, 2/2006-5/2009

## VOLUNTEER

### Machine Learning Volunteer, Omdena, 1/2022-3/2022, Current

- Current: Detecting various diseases from lung X-ray images for web or app development for democratizing access
- Previous: Detecting school lots from satellite imagery in Southern Sudan using UNET segmentation model

### Data Analyst Volunteer, DataKind, 9/2021-10/2021

- Developed pipeline for cleaning and joining datasets and Tableau visualizations for EDA highlighted in report

## PROJECTS

### Melanoma Detector Using Neural Network with Flask ([Github](#), 6/2021)

- Implemented VGG19 CNN on dataset of 33126 images to achieve 96% accuracy, 95% precision, 92% ROC-AUC, and 88% recall

### Hate Tweet Detector Using NLTK with Flask ([Github](#), 5/2021)

- Implemented SVM algorithm on dataset of 24783 tweets to achieve 92% accuracy, 96.5% ROC-AUC, and 6.2% FN's

## EDUCATION

### Flatiron School

10/2020–4/2021

*Data Science Online Immersive Bootcamp*

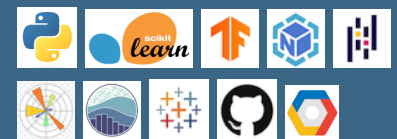
### University of Chicago

9/1996–6/2000

*BA Biology, 3.7 GPA*

*College Honor Scholarship*

## TECHNICAL SKILLS



### Programming Skills

*Python, SQL, HTML/CSS*

### Machine Learning, Predictive Modeling

*Scikit-learn, Tensorflow*

### Data Cleaning, Data Munging

*Numpy, Pandas*

### Statistics, Statistical Analysis

*Excel, SciPy*

### Data Mining, Data Interpretation

### Data Visualization

*Matplotlib, Seaborn, Tableau*

### Version Control

*Git, Github*

### Cloud Platform

*Google Cloud Platform*

## REFERENCES

Available Upon Request