EMAIL **DATASCISTEVEN@GMAIL.COM** LINKEDIN **@DATASCISTEVEN**
WEBSITE **DATASCISTEVEN.GITHUB.IO** GITHUB **@DATASCISTEVEN**

STEVEN YAN
(646) 991-4222

Former education management professional in NYC with passion for social good looking to apply data science skills to find impactful solutions while leveraging my interpersonal and collaborative skills

TECHNICAL SKILLS

Python, Machine Learning (Scikit-learn, Keras/Tensorflow**), Database Management** (SQL, PostgreSQL**), Data Analysis/Wrangling (**NumPy, Pandas, SciPy**), Data Viz (**Matplotlib, Seaborn, Tableau**), Version Control** (Git, Github)**, Deployment** (Flask, Heroku)**, Front-End** (HTML, CSS)**, Computer Vision** (OpenCV, Scikit-Image)**, NLP** (NLTK)

TECHNICAL PROJECTS

SMS Appointment Reminder Platform Using Twilio, Flask, Redis, and Celery (*Github*)
- Recognized by Essteem for individual contribution to developing SMS appointment reminder platform for Queens VA
- Implemented whole slurry of new skills for front-end development (Flask, Bootstrap, Twilio) for successful demo

Melanoma Detector Using Neural Network with Flask Implementation (*Github*)
- Trained CNN model from scratch or used transfer learning with various pre-trained models on 33,126 dermatoscopic images of skin lesions to detect melanoma
- Improved class imbalance from 1.8 to 31% by adding from other datasets and using data augmentation (Albumentations)
- Achieved accuracy of 95% and 80% on 2 different holdout sets using ensemble learning with VGG19 model

Hate Tweet Detector Using NLTK with Flask implementation (*Github*)
- Developed model for detecting hate tweets using 24,783 tweets collected from multiple labeled sources to address increasing anti-Asian hate crime and highlight the insidious potential of hate speech to incite people to violence
- Added hate tweets from other datasets to ensure good distribution of main categories: racist, sexist, homophobic
- Implemented SVM algorithm to achieve holdout set accuracy of 92.0%, while maximizing ROC-AUC score (0.965) and minimizing FNs (0.062), aka hate tweets that escape notice of detector

EXPERIENCE

Data Volunteer, *DataKind (Github)*   9/2021–10/2021
- Selected to collaborate post-hackathon to complete data analysis tasks for submission to original sponsor (CDAC)
- Code for pipeline for cleaning and joining datasets and Tableau visualizations for EDA highlighted in report

Educational Consultant (Freelance), *Examsherpa, New York, NY*   8/2009–10/2020
- Created individualized study plan based on assessment, delivered intensive targeted content review with providing insight based on experience into how topics will be tested, as well as offering essay writing and editing services
- Projects: creating and delivering 20-session bootcamp for students in Mentoring In Medicine, a pipeline program for underrepresented premeds; collaborating with content writers led by Khan Academy and AAMC (the test maker) to create open-source high-quality test prep materials to equalize access to resources

Curriculum Developer/Teacher Trainer, *Kaplan Test Prep, New York, NY*   2/2006–5/2009
- Created a training guide for new tutors to reflect changes to the new format and content in 2005
- Managed team of 15 instructors in Manhattan/Bronx region by assigning classes, conducting class observations, and providing feedback to increase student excellence ratings for region

EDUCATION

Flatiron School, *Data Science Online Immersive Bootcamp, NYC Campus*   10/2020–4/2021
University of Chicago, *BA Biology, 3.7 GPA, College Honor Scholarship*   9/1996–6/2000