

Email: datascisteven@gmail.com

Mobile: 646.991.4222

Connect:  @datascisteven  @datascisteven  @datascisteven

STEVEN YAN

Technical Skills

Machine Learning, SQL (MySQL), Python (NumPy, Pandas, SciPy, Scikit-learn, Tensorflow), Data Analysis, Data Visualization (Matplotlib, Seaborn, Tableau), Git, Microsoft Office, HTML/CSS

Technical Projects

Melanoma Image Classifier for Mobile App Deployment – tinyurl.com/mel-detector

- Developed CNN model from scratch alongside pretrained models to detect melanoma trained from 33,126 images in ISIC 2020 dataset ready for deployment for app accessible to all populations
- Evaluated performance on two holdout sets achieving 95%/80% accuracy with VGG19 after addressing class imbalance by bolstering with additional datasets and data augmentation with Albumentations
- Skills: Python, NumPy, Pandas, Scikit-learn, Tensorflow, Matplotlib, Seaborn

Automated Hate Tweet Detection – tinyurl.com/hate-detector

- Developed NLP binary classification model for detecting hate tweets some downloaded from Twitter API to address the growing concern of hate speech to incite recent acts of violence
- Implemented resampling techniques where best results achieved with SVM with minority class addition with the holdout accuracy of 92.0% considering ROC-AUC score of 0.965 with lowest % of FNs at 6.21%
- Skills: Python, NumPy, Pandas, Tensorflow, Matplotlib, Seaborn, NLTK, Tensorflow

Credit Card Default Prediction – tinyurl.com/cc-default

- Developed classification model to predict whether customer will default on next payment as well as identifying main factors for customer defaulting trained on dataset of 30000 customers
- Implemented resampling techniques from imbalance-learn and smote-variants across different models and engineered novel features for classifying CC default achieving 82% accuracy
- Skills: Python, NumPy, Pandas, SciPy, Scikit-learn, Plotly, Tableau

Work Experience

Educational Consultant (Freelance), Examsherpa, New York, NY

8/2009–10/2020

- Developed individualized study plans for premeds based on assessment and academic history
- Provided targeted content review, practice problems, experience and insight into each topic, essay writing and editing services
- Developed and delivered 20-session bootcamp for students in Mentoring In Medicine, a pipeline program for underrepresented premeds
- Collaborated with content writers led by Khan Academy and AAMC the testmaking org to develop open-source high-quality test prep materials in effort to equalize access to resources for all

Curriculum Developer/Teacher Trainer, Kaplan Test Prep, New York, NY

2/2006–5/2009

- Wrote and developed MCAT Tutoring Guide for new tutors based on my training/teaching experience
- 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
- Conducted regular prospective instructor auditions and training course for new hires

Education

Flatiron School, Data Science Online Immersive Program, NYC Campus
University of Chicago, BA Biology, 3.7 GPA, College Honor Scholarship

10/2020–4/2021
9/1996–6/2000