Shrikar Thodla

734-389-5134 • thodla.shrikar@gmail.com • San Francisco, CA • sthodla.github.io

Education:

University of San Francisco: M.S. in Data Science — June 2020

- Relevant Courses: Python, Machine Learning, SQL, Time Series Analysis, Distributed Computing, Data Acquisition, Data Ethics, Product Analytics, A/B Testing & Experimental Design, Data Structure & Algorithms

University of Michigan: B.S. in Informatics, Minor in Mathematics — Dec 2017

- Relevant Courses: Linear Algebra, Coding Theory, Probability and Statistics, Data Mining, Computation Biology/Bioinformatics, Animal Physiology, Biochemistry

Technical Skills/Libraries:

Python (scikit-learn, NumPy, pandas, SciPy, spaCy), PyTorch, PostgreSQL, AWS, Github, R, Spark (PySpark), ggplot2

Work Experience:

Data Scientist - Intern at Retrace Labs — Oct 2019 - Present

- Standardized images by training and deploying a neural network to detect accidental rotation in images.
- Segmented objects in medical image scans by using convolutional neural networks (CNNs), like UNet, resulting in a 0.91 dice score.
- Optimized code to query images from MongoDB by batch querying and multiprocessing, which led to a 5x decrease in execution time.
- Led the migration of full-stack components from Google Cloud to IBM by setting up virtual server instances and user access management rules.

Associate at Infosys Limited — Feb 2018 - Apr 2019

- Decreased the number of security vulnerabilities in production and non-production servers by coordinating activities between the application and patching teams, while also giving frequent updates to stakeholders.

Lab Assistant at Young Lab — Dec 2014 - Oct 2016

- Designed an experiment to estimate the effect of different growth hormones on bacterial strains to show that the growth hormone was not a confounding variable.
- Showed organoids are a viable alternative to human tissue cultures for experiments by collecting pertinent data.

Projects:

Bengali Character Classification (Kaggle Competition) — Jan 2020 - Mar 2020

- Top 10% in the competition out of 2059 total teams (Bronze Medal).
- Predicted 3 different labels for 200k+ Bengali characters by training a CNN with a custom classification layer and weighted cross-entropy loss function. Code can be found here: https://github.com/ash-jha/kaggle_bengali.
- Improved model performance by using findings from research papers, including advanced data augmentation techniques such as Cutmix and Mixup.

StreamHopper (streamhopper.video) — Mar 2020 - May 2020

- Developed and pitched a web application that recommends which streaming service a user should choose based on their preferences in order to stave off subscription fatigue.
- Data for all the shows and movies on each streaming platform was pulled from two APIs using Python and standardized in order to join the data. Code here: https://github.com/matthewcking312/StreamHopperPublic

ML Algorithm Implementations — Oct 2019 - Feb 2020

- Recreated implementations of popular machine learning algorithms in Python, such as Random Forests, Decision Trees, Naive Bayes, K-Means, TF-IDF, and multiple feature importance algorithms.