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 — July 2019 - June 2020 (Expected)

- Relevant Courses: 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 — Sept 2013 - Dec 2017

- Relevant Courses: Linear Algebra, Coding Theory, Probability and Statistics, Data Mining

Work Experience:

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

- Applying convolutional neural networks (CNNs) for object classification and segmentation tasks in medical image scans.
- Optimized code to query images from MongoDB by batch querying and multiprocessing, which led to a 5x decrease in execution time.
- Standardizing images by training a model to detect accidental rotation in images, using CNNs implemented in PyTorch.
- Led the migration of full-stack components from Google Cloud to IBM.

Associate at Infosys Limited — Feb 2018 - Apr 2019

- Greatly decreased the number of high and critical security vulnerabilities in production and non-production servers for the client.
- Coordinated activities between various application teams and patching teams so that server downtime is minimized during patching, 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 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 — Mar 2020 - Present

- A web application that recommends which streaming service a user should choose based on their preferences: genres, viewing habits, age, and so on.
- The application was selected to be pitched to four VCs on demo day
- Data was pulled from two APIs using python and standardized in order to join the data.

ML Algorithm Implementations — Oct 2019 - Present

- Recreated implementations of popular machine learning algorithms, such as Random Forests, Decision Trees, Naive Bayes, K-Means, and TFIDF.

Technical Skills/Libraries:

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