

## Education

**University of Washington, Data Science and Statistics, 2018 - 2019.** GPA: 3.97

Courses: CSE 154 Web Programming, CSE 416 Introduction to Machine Learning, CSE 414 Introduction to Database Systems

**University of Maryland, Computer Science: Data Science, 2020-2022.** GPA: 4.0

Courses: CMSC 132 Object Oriented Programming II, CMSC 250 Discrete Structures, CMSC 216 Intro to Computer Systems, CMSC 351 Algorithms

## Awards and Competitions

### AIVIVN Sentiment Analysis Competition

Spring 2019

- Design a text classification system for positive v.s negative product reviews (in Vietnamese).
- Models: word2vec + weighted average of (hierarchical) self-attention neural network, residual network.
- **Final Result: 1st Place in Public Leaderboard (F1: 0.90087) and Private Leaderboard (F1: 0.90012).**
- GitHub Repository: [https://github.com/petrpan26/Aivivn\\_1](https://github.com/petrpan26/Aivivn_1)

## Work Experience

### Software Engineering Intern at Verta.AI

Summer 2020

Top 5 contributor to ModelDB (<https://github.com/VertaAI/modeldb>), with **over 100 pull requests**. Main projects:

1. *Dataset and Metadata Versioning Functionality for Scala Client:*
  - Functional API with git-inspired immutable data types (Repository, Commit, Blob, etc.) and operations (merge, revert, etc.)
  - Works seamlessly with AWS S3 (data storage) and Spark MLLib (model development)
2. *Model Registry and Endpoint Deployment Functionality for Python Client and CLI:*
  - Register a model and expose it via an endpoint for prediction.
  - Allows for canary update, autoscaling and resource specification.
  - Supports a user-friendly command line interface (CLI), implemented with Click.

## Personal and Research Projects

### Neural Network Toolbox

Summer 2019

- Implement common deep learning procedures and papers using PyTorch for quick prototyping and model developing.
- GitHub Repository: <https://github.com/nhatsmrt/nn-toolbox>
- Documentation: <https://nhatsmrt.github.io/nn-toolbox/>

### Arbitrary Style Transfer

Summer 2019

- Given an arbitrary content photo and a piece of artwork, transfer the style of the artwork to the photo.
- Based on Huang and Belongie's paper "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization"
- Implemented using PyTorch and my toolbox (see above).
- GitHub Repository: <https://github.com/nhatsmrt/torch-styletransfer>

### Coding Platform for Solving Algorithmic Questions

Winter 2020

- Technology stack: Judge0 for evaluation. Django REST Framework (Backend); React (Frontend). Deployed on Heroku
- Prototype Website: <http://codingplatform-cp.herokuapp.com/>

### Data Collection for Food Desert Prediction Research Project at the University of Maryland

Spring 2020

- Web scraping (selenium, beautifulsoup4) for food data; Overpass and Nominatim API for geolocation data
- Speed up API calls, geometric queries, etc. **up to 3-6 times** with multiprocessing and spatial indexing (geopandas + rtree)

## Skills

### Proficient With:

1. *Java (4 years):* Object Oriented Programming Design Patterns, JUnit Unit Testing, Multithreading and Concurrency
2. *Python (4 years):* Data Science, Machine Learning, Deep Learning. Packages and tools proficient:
  - scipy, statsmodels, numpy, pandas, geopandas, matplotlib, scikit-learn, pytorch, keras, Google Colab; argparse, click; pytest
3. *Scala (1 year):* Functional Programming Principles and Design Patterns.
  - Behavior-Driven Testing and Property Based Testing with ScalaTest. Big Data with Apache Spark and MLLib
4. *Relational DMBS and SQL*

### Familiar With:

1. *Full Stack Web Development:* Server-side with Django; Client-side with HTML, CSS, JavaScript, ReactJS
2. *Other Technologies:* R (dplyr, tidyr, ggplot2, shiny); C, gdb, valgrind; Unix, emacs; AsteriskDB and SQL++; AWS S3
3. *Other CS and SWE Skills:* Agile Project Management via Jira; Software Versioning Control with Git and GitHub