

Education

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

Courses: Web Programming, Introduction to Machine Learning, Introduction to Database Systems

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

Courses: Introduction to Computer Systems, Algorithms, Organization of Programming Languages

Work Experience

Software Engineering Intern at Verta

Summer-Fall 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*:
 - Designed a functional API with git-inspired immutable data types (e.g: Repository, Commit) and operations.
 - Works seamlessly with AWS S3 (data storage) and Spark MLLib (model development)
2. *Model Registry and Endpoint Deployment Functionality for Python Client and CLI*:
 - Allows user to register a model and expose it via an endpoint for prediction.
 - Allows for canary update, autoscaling and resource specification.
 - Available in Python script and command-line interface (implemented with Click).

Awards and Competitions

Winner of AIVIVN Sentiment Analysis Competition

Spring 2019

- Predicted accurately the sentiment of **over 90%** of product reviews, with an **F1 metric of 0.90012** on the test set.
- Model: word2vec embedding + ensemble of attentional recurrent neural network and residual convolutional neural network.
- GitHub Repository: https://github.com/petrpan26/Aivivn_1

8th Place in Emotion Recognition Competition 2019

Fall 2019

- Designed a convolutional neural network classifier of the speaker's emotion from raw audio data with nn-toolbox (see below).
- GitHub Repository: <https://github.com/nhatsmrt/erc>

Personal and Research Projects

Neural Network Toolbox (nn-toolbox)

Summer 2019

- Implemented deep learning procedures and models in PyTorch with a *composable and modular design*.
- Reduced the time to set up model development and prototyping for projects and competitions **from weeks to 1-2 days**.
- GitHub: <https://github.com/nhatsmrt/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.
- Implemented with PyTorch and my nn-toolbox (see above), based on Huang and Belongie's paper.
- GitHub Repository: <https://github.com/nhatsmrt/torch-styletransfer>

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
- Made API calls and geometric queries **3-6 times faster** with multiprocessing and spatial indexing (geopandas + rtree)

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/>

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. Client and command-line interface development.
 - 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. *Software Engineering Skills*: Agile project management via Jira. Versioning control with Git and GitHub.

Familiar With:

1. *Full Stack Web Development*: Server-side with Django. Client-side with HTML, CSS, JavaScript, ReactJS
2. *Other CS Topics*: Algorithms and data structures. Relational database and SQL.
3. *Other Technologies*: R (dplyr, tidyr, ggplot2, shiny). C, gdb, valgrind. Unix, emacs. AWS S3