Nathan Warren

Data Scientist · Machine Learning Engineer · Duke University

916-261-0695 | Nathan.A.Warren1@gmail.com | nathan2warren.github.io/ | Nathan2Warren | nathan-warren-ds

Skills

Programming Python · Bash · LaTeX
Analytics IPython · SQL · Tableau · Excel/VBA · R · Dataiku
Machine Learning Pytorch · Tensorflow · Scikit-learn · Spark
Cloud Docker · AWS · GCP

Education

Duke University — *Durham, NC*

M.S. in Data Science — *Aug. 2019 - May. 2021*

- Coursework: Machine Learning, Deep Learning Engineering, Statistical Modeling, Text analysis, Databases, Algorithms
- GPA: 3.92

Drexel University — *Philadelphia, PA*

B.S. in Biological Sciences — *Aug. 2012 - May. 2017*

Experience

Duke University — *Durham, NC*

Graduate Student Teaching Assistant — *Aug. 2020 - Present*

- Teaching assistant for Data Science with Python (Fall 2020), and Causal Inference (Spring 2021)
- Helped students learn essential python programming skills for data science (numpy, pandas, sklearn, dask)

Data+ — *Durham, NC*

Data Science Intern — *May. 2020 - Aug. 2020*

- Developed machine learning models (Random Forest, CNN-LSTM, etc) to classify human activity based on multi-modal sensor data
- Achieved an accuracy of 0.85 in classifying activity over 10 second segments of time
- Shared work as open-source on the Digital Biomarker Discovery Pipeline Repository

Children's Hospital of Philadelphia — *Philadelphia, PA*

Neurology Research Assistant — *May. 2017 - Jun. 2019*

- Served as Study Director for a proof of concept gene therapy pharmaceutical collaboration
- Oversaw project planning and initialization of drug trials
- Quantified data in ImageJ and conducted statistical analysis in R for publications

GlaxoSmithKline — *King of Prussia, PA*

Research Co-Op: Drug Metabolism and Pharmacokinetics — *Mar. 2016 - Sep. 2016*

- Characterized pharmacokinetic profiles of small molecules for treatment of COPD
- Analyzed and modeled toxicokinetics of pre-candidate drugs for safety profiling

Projects

Capstone Project with The Conversation Fund

- Combined remote sensing, raster, and manual surveying data to estimate biomass in forest properties in order to reduce the cost of manual surveying
- Built a pipeline to download and subset over 13 TB of GEDI (LiDAR) data for client-owned forestland
- Created visualizations and qualified metrics to communicate magnitude of climate impact on forest conservation to stakeholders

Neural Style Transfer

- Successfully manipulated VGG19 to extract semantic content information and style features from images to create hybrid images

Publications

2018 **Neuronal Signaling**, Role of frataxin protein deficiency and metabolic dysfunction in Friedreich ataxia, an autosomal recessive mitochondrial disease

2017 **Disease Models & Mechanisms**, Early VGLUT1-specific parallel fiber synaptic deficits and dysregulated cerebellar circuit in the KIKO mouse model of Friedreich ataxia