

# Erik Schafer

My objective is to find consistent private sector data science or machine learning work, to leverage my experience as an engineer and my skills as a mathematician.

612 12th St. NE #201  
Washington, D.C. 20002  
562.833.5903  
[eSchafer90@gmail.com](mailto:eSchafer90@gmail.com)  
[erikSchafer.com](http://erikSchafer.com)

## EXPERIENCE

### **Pyramid Systems, Fairfax, VA — Data Scientist**

September 2019 - PRESENT

Led Data Model design, designed systems architecture, and supported core development on a critical modernization effort for the FHA. Worked directly with the client to understand business requirements, needs, and vision for the product. The success of this project resulted in additional contracts totaling more than \$42 Million ([hud.gov/catalyst](http://hud.gov/catalyst))

Developed a profile similarity metric using BERT latent word embeddings in a tech challenge to compete for work from federal agencies.

Demonstrated technical expertise for Pyramid Systems in a series of Proposals and Government Team interviews to compete for work.

### **AgencyQ, Washington, D.C. — Full Stack Developer**

August 2018 - February 2019

Supported development in a licensed CMS environment for commercial and government clients.

### **ICF International, Fairfax, VA — Data Scientist**

June 2016 - August 2018

Developed a technical approach and delivered a working prototype to track and report GHG data in accordance with the Paris Climate Accords for the government of Colombia.

Implemented gold-standard emissions and greenhouse gas accounting methodologies from the IPCC for USAID to track, report, and calculate greenhouse gas emissions reduced or avoided due to clean energy activity.

### **ICF International, Fairfax, VA — Full Stack Developer**

June 2013 - June 2016

Developed functionality, facilitated deployments, created UML documentation deliverables for a childcare subsidy management application for Navy childcare providers.

Supported O&M contracts for EPA and NPS

## SKILLS

Python, Numpy, PyTorch,  
Sklern, SciPy, OpenCV2

Data Science, Regression,  
Classification, Boosting,  
Bagging, Model Evaluation  
& Metrics, Hypothesis  
Testing, Statistics

Machine Learning,  
Supervised, Unsupervised,  
Reinforcement, Neural  
Nets, Convolution

C#, JAVA

PostgreSQL, T-SQL, PL/SQL  
(Oracle), MySQL

Mathematics, Probability,  
Statistics, Calculus

Computer Science,  
Algorithms, Data  
Structures, Complexity

Communication, Writing,  
Technical Writing, UML,  
Requirements Gathering,  
Client Facing

## LANGUAGES

German (Proficient)  
Dual Citizen

## EDUCATION

## **Georgia Institute of Technology, MSc Computer Science: Machine Learning**

Jan 2018 - Dec 2019

Pursued a Master's Degree from Georgia Tech with a chosen focus area in Machine Learning. Included coursework in Computational Photography, Cyber-Physical Systems, Computer Vision, Big Data For Healthcare, etc.

## **Hampden-Sydney College — B.S. Applied Mathematics, B.S. Computer Science, A.A. German Language and Literature**

Magna Cum Laude

Aug 2009 - May 2013

Achieved 3 concurrent degrees. Study in Mathematics focused in Statistics, Probability Theory, Linear Algebra, Computer Science studies focused in Object-Oriented programming, Graphics (OpenGL), and cryptography. Philosophy studies emphasizing Aristotelian Logic, Modern Symbolic Logic, Discrete Mathematics. Led a renaissance of the Math Club, held leadership roles in a Professional Fraternity. Lead student tutor for Mathematics and Computer Science curriculum. Performed TA duties for the Math Department, grading, etc.

### **SELECTED ACADEMIC PROJECT**

#### **Disease Recognition — PyTorch, CUDA, DenseNet121, ResNet152**

Spring 2019 - *Big Data for Healthcare*

Detected 14 distinct diseases in a multiclass-multilabel classification task using more than 500 GB of X-Ray image data. Compared performance of several different PyTorch architectures using the ROC AUC evaluation metric, and determined the accuracy for most diseases exceeded the accuracy of the average human doctor. Trained models without pre-trained weights locally and in a remote instance on GPU hardware with CUDA acceleration, and transferred trained and partially trained models between instances. Challenges faced included evaluating tradeoffs between image scale transformations and mini-batch size to allow limited GPU RAM resources to load the model weights and image data for training.