

CHAU NGUYEN

Arlington, VA 22201 | (626) 898-1408 | cn490@georgetown.edu
github.com/chav-nguyen | linkedin.com/in/chavnguyen | chav-nguyen.com

EXPERIENCE

Sep 2021 -
Present

Massive Data Institute Scholar at Georgetown University

Devised statistical method to ease documentation burden in COVID-19 assistance applications for renters; collaborated in cross-functional team of economists & engineers from 3 federal agencies

Constructing neural nets pipeline to detect banner-esque objects in paintings with TensorFlow.js

Jun 2021 -
Sep 2021

Data Science Intern at Fraym — *geospatial ML and data analytics startup* fraym.io

Developed log retrieval process using boto3 (Python) to extract error messages from 5,200 cloud computing jobs on AWS, slashing 10+ minutes of manual search to view each record

Utilized clustering algorithm to discover text similarities in crash logs from over 1,000 failed cloud computing jobs; investigated clustered errors and pinpointed 8 gaps in R codebase

Presented actionable insights to management to improve core product in company-wide meeting

Dec 2016 -
Aug 2020

Research Analyst at International Monetary Fund

Analyzed 140 million USD in annual overseas transfers to Samoa; estimated that high legal compliance costs caused transaction fees to be 6% higher than UN sustainable targets

Interviewed country officials during field research to estimate missing datapoints and solved significant data gaps in Tuvalu's fishing sector worth over 30 million USD (55% of country's GDP)

Developed excellent verbal and written communication skills from relaying analysis results, research findings, and policy recommendations in non-technical terms to foreign government officials

EDUCATION

Georgetown University | *M.S. Data Science for Public Policy*

May 2022

Data Science for Public Policy McCourt Scholarship (\$15,500 per year), Massive Data Institute Scholars program

University of California, Berkeley | *B.A. Economics*

May 2016

PROJECTS

[Scrollable Interactive Tutorial on Density-based Clustering Algorithms](#) (D3.js, JavaScript, HTML, scikit-learn)

Explained intuition behind HDBSCAN through interactive data visualizations of unlabeled synthetic & real data

[Doyle, Christie, or LeBlanc? - A Deep Learning Approach to NLP & Authorship Detection](#) (TensorFlow, NLTK)

Created NLP pipeline to train recurrent neural nets to detect penmanship; outperformed random guess by 98%

[Hyperparameter Tuning Convolutional Neural Network Layers to Predict Forest Fires](#) (keras, sklearn, seaborn)

Optimized hidden layers & dropout rate, conducted feature engineering; predicted fires within 0.1 km² margin

[The Crowd or the Stadium? Home-field Advantage in the NFL with ML](#) (pandas, ggplot2, beautifulsoup, requests)

Web-scraped data for 7,292 NFL matches & locations; found travel distance important feature in win prediction

[How to Set Up Jupyter Notebooks to Check Data Visualizations for Colorblind Accessibility](#) (UX, OpenCV)

TECHNOLOGIES USED

Python: ML, deep learning, NLP libraries (spaCy, NLTK); R: exploratory data analysis, tidyverse, statistical packages; Spark, SQL, AWS (SageMaker, S3); Agile development, shell script, Jira, Git, LaTeX, UNIX commands, Tableau, D3.js