CHAU NGUYEN

Washington, D.C. Metro Area | Email: cn490@georgetown.edu | Phone: (626) 898-1408
Github: chav-ngvyen | LinkedIn: chavngvyen | Website: https://www.chav-ngvyen.com

EXPERIENCE

Massive Data Institute – *Georgetown University's McCourt School of Public Policy* **Washington, D.C. (Hybrid)**
MDI Scholar *September 2021 - Present*
- Constructing custom TensorFlow.js pipeline to detect banner-esque objects in museum painting archives *(ongoing)*
- Devised statistical method to identify and ease documentation burden in COVID-19 assistance applications for low-income renters in collaboration with cross-functional team of economists & engineers from 3 federal agencies

Fraym – *Geospatial and data analytics company* – *fraym.io* **Arlington, VA (Remote)**
Data Science Intern *June 2021 - September 2021*
- Developed log retrieval process in Python using boto3 to connect to AWS and extract log messages from over 5,200 cloud computing jobs, slashing over 10 minutes of manual search and match to view any single 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 where spatial interpolation process can fail
- Implemented workarounds to avoid crashes during feature selection process, eliminating 1/8 of identified issues
- Presented to management actionable insights to further improve core product in company-wide meeting

International Monetary Fund **Washington, D.C.**
Research Analyst *December 2016 - August 2020*
- Analyzed 140 millions USD in annual overseas transfers to Samoa; estimated that high legal compliance costs caused transaction fees to be 6% higher than sustainable development goals; findings highlighted in IMF report
- Conducted field research, interviewed country officials to solve significant data gaps in Tuvalu's fishing sector which accounts for over 50% of GDP; authored first formal industry analysis published in 2018 IMF country report
- Developed strong verbal communication skills from relaying analysis results, forecasts, impact evaluations, and policy recommendations in non-technical terms to foreign government officials

EDUCATION

M.S. Data Science for Public Policy | **Georgetown University** | GPA: 3.82/4.0 *Expected: May 2022*
- Honors: DSPP McCourt Scholarship ($15,500 per academic year), Massive Data Institute Scholars program

B.A. Economics | **University of California, Berkeley** | GPA: 3.43 *Graduation: May 2016*

SKILLS

- **Programing Languages:** Python (preferred), R (exploratory data analysis, statistical packages, tidyverse), Stata
- **Quantitative Analysis:** A/B testing, causal inference, logistic regression, RCTs, experiments, hypothesis testing
- **Machine Learning**: Supervised (classifiers, decision trees, ensembles); Unsupervised (topic modeling, sentiment analysis); Deep Learning (anomaly detection, time series, PyTorch); NLP (embeddings, encoders, transformers)
- **Data Engineering & Big Data:** distributed computing, Hadoop, Spark, relational databases, SQL, APIs, AWS
- **Others:** Agile development, unit testing, shell scripting, Git, LaTeX, Unix commands, Tableau dashboards, ArcGIS

PROJECTS

- Scrollable Interactive Tutorial on Density-based Clustering Algorithms *(D3, JavaScript, HTML, SVG, 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 and Authorship Detection *(TensorFlow, NLTK)* Created NLP pipeline to train recurrent neural nets to detect penmanship; outperformed random guesses by 98%
- Hyperparameter Tuning Convolutional Neural Network Layers to Predict Forest Fires *(keras, sklearn, matplotlib)* Optimized layers & dropout rate, conducted feature engineering; predicted fire spreads with $0.1km^2$ error margin
- The Crowd or the Stadium? Home-field Advantage in the NFL with ML *(pandas, ggplot2, beautifulsoup, requests)* Webscraped data for 7,292 NFL matchups & locations; found travel distance important feature in win prediction
- How to Set Up Jupyter Notebooks to Check Data Visualizations for Colorblind Accessibility *(UX design, OpenCV)*

PUBLICATIONS

- Report on the Pacific Roundtables: Actions to Address Correspondent Banking and Remittances Pressures, *2019* ⬈
- Macroeconomic Dynamics and the IS puzzle, *Economics: The Open-Access, Open-Assessment E-Journal, 2018* ⬈
- The Economic Impact of Natural Disasters in Pacific Island Countries, *IMF Working Papers, 2018* ⬈
- Tuvalu: Staff Report for the 2018 Article IV Consultation, *IMF Staff Country Reports, 2018* ⬈
- Samoa: Staff Report for the 2018 Article IV Consultation, *IMF Staff Country Reports, 2018* ⬈