CHAU NGUYEN

Permanent US Resident / Green card holder
cn490@georgetown.edu | github.com/chav-ngvyen | linkedin.com/in/chavngvyen

EXPERIENCE

Dec 2021 – May 2022

Graduate Research Assistant at **Massive Data Institute, Georgetown University**

Guided research on neural networks' limitations in object detection in paintings due to training images biases; collaborated with Art History domain expert; presented findings in research showcase

Sep 2021 – Dec 2021

Graduate Research Assistant at **Office of Evaluation Sciences, GSA**

Created prototypes of statistical model to identify renters in low-income areas using R, collaborated with cross-functional team of economists & engineers to put model into production

Jun 2021 - Sep 2021

Data Scientist Intern at **Fraym**

Automated end-to-end log retrieval process to extract error messages from 5,200 AWS CloudWatch Logs using boto3 (Python), optimized efficiency by reducing record search and review time by 50%

Used clustering algorithm to detect text similarities in crash logs from over 1,000 failed cloud computing jobs in order to identify and group their root causes to inform engineering improvement

Created experimentation pipeline to help company evaluate performance of different clustering algorithms and calibrate model hyperparameters for future development of new data products

Dec 2016 – Aug 2020

Research Analyst at **International Monetary Fund**

Analyzed time-series data of money transfer costs for 8 countries to measure spillover impact of anti-money laundering regulations, found that prices in Pacific were 4 times the sustainable targets; findings led to regional collaboration efforts that reduce remittance costs by 75%

Conducted analysis on overseas direct investment by Chinese firms, built 7 internal dashboards using Tableau to help economists easily interpret results of analyses and provide policy recommendations

Developed excellent verbal and written communication skills from presenting modeling decisions, analytical findings, and policy recommendations in non-technical terms to foreign government officials

SKILLS

Programming Languages: Python (2 years), R (5 years), SQL (3 years), Stata (5 years)

Data Science: statistical modeling, causal inference, A/B testing, design of experiments, supervised and unsupervised machine learning, neural networks, NLP, web-scraping, data wrangling and manipulation, data visualization

Technologies Used: AWS, Hadoop, Spark, Github, Tableau, D3.js, UNIX command line

MACHINE LEARNING PROJECTS

Predicting Size of Forest Fires with Convolutional Neural Network (TensorFlow, pandas, sklearn)

Used CNN regression to predict the size of forest fires, performed cross-validation to prevent overfitting small dataset; conducted feature engineering and hyperparameters tuning, accurately predicted fire size within 0.1 km2 margin

Identifying Original Posters from Reddit Comments (AWS, Hadoop, PySpark, SparkML, SparkSQL)

Utilized cloud computing to manipulate big dataset of 8 million Reddit comments and build end-to-end gradient-boosted trees model to classify commenter's identity; achieved F-1 score of 0.92 for imbalanced class

EDUCATION

Georgetown University | *M.S. Data Science for Public Policy* May 2022

University of California, Berkeley | *B.A. Economics* May 2016