Joyce Hu

joh011@ucsd.edu | (510) 378-1127 | www.jjoycehu.github.io

EDUCATION

UNIVERSITY OF CALIFORNIA- SAN DIEGO
Bachelors of Science in Political Science: Data Analytics; GPA: 3.95 **Expected June 2024**

- Minors: Data Science and General Biology
- Relevant Coursework: Machine Learning for Political Scientists, Text as Data, Data Structures and Algorithms for Data Science, Data Science in Practice, Applied Data Analysis for Political Science, Multivariable Calculus

EXPERIENCE

Foreign Policy Data Analytics and Research | UC San Diego **January 2023- Present**
Research Assistant

- Assist Ph.D. candidate, Nicholas Smith, in coding data for machine learning models on U.S. foreign policy data
- Clean *American Public Opinion and U.S. Foreign Policy* survey data from 1974-2019 in R using tidyverse, dplyr, and other libraries
- Create data visualizations using ggplot

U.S. Immigration Policy Center | UC San Diego **June 2022- Present**
Research Fellow

- Gather original survey data on immigrant populations
- Lead project team to compile 100+ page documentation on DACA research
- Educate voters and support voter turnout during the 2022 Midterm elections

Prospect Journal of International Affairs | UC San Diego **March 2022- Present**
Director of Finance | Editor | Staff Writer

- Prospect Journal is a student-run journal on contemporary global issues and events
- Our website: prospect-journal.org/

PROJECTS

UN General Debate Speech Webscraper | *Beautiful Soup, Pandas, PdfReader*

- Wrote a webscraper to scrape and convert pdfs of 2022 UN General Debate Speeches into a CSV dataset
- Formatted final data frame to include relevant data (year, speaker, country ISO) and wrote a function to create a unique document ID for future indexing
- Cleaned text for page numbers and unnecessary headers

Environmental Performance Classifier | *Tensorflow, Keras, Scikit-learn*

- Constructed an environmental performance index using Principal Component Analysis and grouped countries by their environmental performance using KMeans Clustering
- Predicted environmental performance of countries using LASSO and achieved an R-square of 0.929 after cross-validation
- Created and optimized a deep learning model to predict country clusters and achieved a macro f1 score of 0.84

Box Office Sales Prediction | *Pandas, Vader, Textblob, NLTK, Seaborn, Scikit-learn*

- Collaborated with project team to predict box office sales from movie data scraped from Wikipedia
- Used a Decision Tree Regressor to find the variables with the highest MSE

TECHNICAL SKILLS

Programming Languages: Python, R, Java, Matlab, Stata
Languages: Mandarin (Advanced Proficiency), Spanish (Basic Proficiency)
Other: Pandas, Matplotlib, Seaborn, ggplot, Quanteda, Scikit-lean