

Zhengguang Wang

Email: zw4re@virginia.edu Telephone: (434)422-2399

Summary

- **Objective:** Seeking Summer 2023 Intern in Data Science or in related fields
- **Languages:** Bilingual in English and Chinese, Intermediate French
- **Programming Languages:** Proficient in Python and R, Intermediate Java and SQL
- **Tech Skills:** Object-Oriented Programming, Tableau, SAS, Linear Model, Applied Time Series, Neural Network, Scikit-Learn, Pytorch (Neural Network and Deep Learning) Bootstrapping, A/B Testing, Geocoding, ArcGIS, MS Office, D3
- **Natural Language Processing:** Pytorch, Scikit-Learn, NLTK, Seq2Seq, Word2Vec

Education

University of Virginia | Major in *Statistics and Computer Science* Graduation Date: Expected May 2024
Echols Scholar at the University of Virginia (top 5% at UVA), GPA 3.879/4.0, GRE 329/340

Professional Experience

Data Science Teaching Assistant | *Statistics Department, UVA* | Charlottesville, Virginia Aug 2022-Dec 2022

Teaching Machine Learning Algo, KNN classification, A/B Testing, Regression, and simulation in Python

- Held Office Hour twice per week to help students understand statistical concepts (sampling, unbiased estimator, type I error, etc); performed simulations to demonstrate these concepts
- Taught data wrangling and control-flow in Python; implemented KNN classifier and A/B Testing for example usage; wrote sample scripts for Bootstrapping and Monte-Carlo Simulation

Data Scientist Intern | *Local Energy Alliance Program* | Charlottesville, Virginia Jun 2022-Aug 2022

Web-Scraping with Python Selenium and Mapping with ArcGIS; Cleaning 7-year Solarize Program Data

- Used Python Pandas package to merge and tidy the 7-year Solarize Program Data; wrote Selenium scripts to retrieve county names and geographical coordinates from customer addresses using google map API
- Utilized ArcGIS and geocoding to generate four interactive maps of LEAP sites with features of amounts in Virginia; conducted geo-spatial analysis and present the finding to the Executive Director

Data Scientist Intern | *China Minmetals Corporation* | Beijing, China Jun 2021-Aug 2021

Identifying Merger & Acquisition (M&A) pattern by conducting Time-Series and Hypothesis testing

- Examined China's oversea mineral M&A pattern based on Big Data in the post-Financial Crisis era
- Accessed the S&P Data Base to retrieve M&A metadata such as case time, monetary amounts, minerals, locations; employed Python Pandas and Excel V-Lookup to clean data
- Conducted Time-Series Analysis on each year's M&A monetary amount using methods like differencing and detrending to remove time trend from data and employed Auto Regressive Moving Average model
- Applied statistical analysis techniques such as two-sample proportion t-test and Chi-Square test to compare the preferences of China's M&A with the rest of world
- Illustrated the global comparisons on a world map using Python Pyechart then compiled the findings and published a 16-page report on the Monthly publication of Economic Research Institute

Research

Linear Model in predicting the US birth rate by State | *University of Virginia* Sep 2021-Dec 2021

- Fitted a linear model predicting birth rate with higher order terms, interactive terms, and dummy variables with SAS resulted in an adjusted R-square of 0.67 with root MSE of 0.68; data was collected from Bureau of Economic Analysis and Statista and was wrangled by R
- Conducted the Exploratory Data Analysis and variable screening to select predictors then performed the three-stage analysis and used the regression diagnostics plots to validate the regression assumptions
- Resolved the Multicollinearity issues and high variable VIF values by substituting the Ordinary Least Squares method with the Ridge Regression with l_2 regularization; selected the ridge biasing constant manually based on criteria of stabilizing the parameter estimates while making the VIF values less than 1
- Utilized the Jackknife procedure and calculated studentized deleted residual and Cook's Distance; identified and removed one influential observation to further enhance the prediction force considering the significant increase of F test statistic and adjusted R-Square