**ZHIYU LIN**

1733 Newton St NW Unit B, Washington, D. C. 20010 • zhiyu.lin18@gmail.com • 651-202-5755

**EDUCATION**

**GEORGETOWN UNIVERSITY**, Graduate School of Arts and Sciences **Washington, DC, Aug 2019 – May 2021**

**Master of Science in Analytics - Data Science GPA: 3.78**

* Related courses: Massive Data Fundamentals, Probabilistic Modeling and Statistical Computing, Optimization, Advanced Python and R

**MACALESTER COLLEGE** **Saint Paul, MN, Aug 2015 – May 2019**

**Bachelor of Arts in Applied Math and Statistics, Concentration in Legal Studies**

* Related courses: Intro to Data Science, Object-Oriented Programming & Data Structures, Machine Learning, Numerical Analysis
* Award: Kofi Annan Scholarship for Academic Excellence and Global Citizenship
* Exchange: Institut d'études politiques de Paris, IPA (SciencesPo)

**RELEVANT SKILLS**

**Math/CS/Stats:** Python, R, Java, SQL, VBA, Git, APIs, Mathematcia, MatLab, LaTex**,** Tableau, D3, plotly, ggplot, ggmap, ArcGIS, Conda

**Data science:** neural nets, Tensorflow, anomaly detection, classification, regression, optimization, hypothesis testing, PCA, Bayesian models

**ANALYTICS PROJECTS**

**Survival Analysis on Predicting the U.S. Supreme Court’s Decisions Saint Paul, MN, Aug 2018 – May 2019**

***Thesis Author/ Statistics Researcher****, Macalester College MSCS and Legal Studies Departments*

* Constructed a 354 row time-series data set for survival analysis from 8,893 documented Supreme Court decisions (1947-2018) by performing data wrangling (join, merge, aggregate, filter) in R tidyverse, dplyr packages; documented for future research.
* Proposed 5 variables as statistical predictors for when the Court overrules a prior decision by reviewing 160+ law literatures.
* Assessed predictors strength by implementing and analyzing 35 probabilistic models (Cox PH, Weibull, log-normal) in R survival library; compared and selected models using AIC and Cox-Snell methods; visualized case durations with Kaplan-Meier curves
* Added robust up-to-date quantitative statistics evidence to the legal research field by contextualizing the results into legal doctrines; presented Court pattern discovery at Pi Sigma Alpha conference (political science honor society).

**Network Analysis on the Health Hazard Exposome Correlations Minneapolis, MN, Nov 2016 – May 2017**

***Statistics Research Assistant****, University of Minnesota Biostatistics*

* Created a globe visualization in R by calculating weighted connectivity using network analysis on the NHANES data.

**WORK EXPERIENCE**

**Enterprise Data Governance & Business Design and Optimization Washington, DC, Sep 2019 - Present**

***Analytics Consulting Intern****, Georgetown University Business Design and Optimization Group*

* Enterprise Data & Analytics Group: wrangle financial data in R; analyze 657 survey responses with heatmaps, linear correlation plots, word clouds, etc in R; design Qualtrics surveys by user cohorts; build, present and automate reporting dashboards in Smartsheet
* Optimization Group: assist with process mapping and stakeholder interviews; extract and report higher-ed news trends; build supplier diversity vis in ArcGIS; research procurement websites from peer institutions and redesign Georgetown websites, take meeting minutes

**Machine Learning & Natural Language Processing (NLP) on Fraud Detection Beijing, China, May 2018 – Aug 2018**

***Data Science Intern****, The Dragon Catcher L.L.C.*

* Created training and test NLP data sets with 4,000 variables by scraping 17,000 customer reviews about Vans sneakers on retail webpages and performing text tokenization in Python pandas and jieba dictionaries.
* Extracted more manageable datasets by cutting down the 4,000 variable to the essential 180 with Lasso selection method in Python.
* Achieved 87% correct classification on Vans trademark infringement detection by training both supervised and unsupervised classification models such as Decision Tree, Random Forest and K-means clustering in Python on top of explanatory analysis.
* Presented the project to the CEO and led the company sign with Amazon China by extracting business insights from the machine learning algorithms through effective visualizations (multidimensional dimensional density plots and histograms).

**Database Management & Automation Eden Prairie, MN, Jun 2017 – Aug 2017**

***Data Science Intern****, UnitedHealth Group – Optum Technology*

* Automated and optimized the technology consumption database research process through VBA; scaled the efficiency by 55 times.
* Queried and initiated and 1,000+ database decommission requests with 100% accuracy; saved clients $500,000 within 11 weeks.
* Designed and implemented a cost-savings dashboard in Tableau; presented to the senior leadership for optimal business operation

**LEADERSHIP & ENGAGEMENT**

***Program Coordinator****, Macalester College Career Development Center* **Saint Paul, MN, Aug 2018 – May 2019**

* Led 70+ undergrad career consultations; coordinated 8 recruiting events; designed, tailored and presented slides to targeted students.

***Philanthropy Coordinator****, Macalester College Annual Fund* **Saint Paul, MN, Jan 2017 – May 2017**

* Communicated with 10+ campus offices and for a fund-raising campaign; achieved a 75% donation rate among 2019 grads.

***Teaching Assistant****, Macalester College Math/Statistics/CS* **Saint Paul, MN, Jan 2016 – May 2016**

* Assisted professors and tutored statistical modeling; held 4-8 office hours/week; graded homework, quizzes and lab reports.

***Pianist****, Macalester College Chamber Ensemble* **Saint Paul, MN, Jan 2017 – May 2017**

* Performed Mozart Piano Trio in E major, K.542, movement 1 at the Macalester Janet Wallace Art Center.