ZHIYU LIN, MS Candidate

zl281@georgetown.edu | zhiyu.lin18@gmail.com | 651-202-5755 |Washington, DC 20010

JUNIOR DATA SCIENTIST & BUSINESS ANALYST

Machine Learning | Statistical Modeling | Cloud Computing | Effective Visualizations

Highly motivated data scientist with 4 years of experience and a Master's degree in progress. Adept at collecting, analyzing and interpreting large datasets, designing and implementing innovative solutions for complex business and research problems, as well as extracting patterns from noisy data and communicating findings to a diverse audience. Intellectually curious and specializes in machine learning & natural language processing. Looking for summer 2020 internships and full-time positions post December 2020.

CORE COMPETENCIES

♦ Python & Java	♦ Big data: Spark & API	♦ Regression models	♦ PCA & NLP	♦ Data visualization
♦ R & Shiny	♦ MapReduce & Hive	♦ Classification models	♦ TensorFlow	♦ Tableau & ArcGIS
♦ SQL, Oracle	♦ Hadoop & Pig	♦ Bayesian models	♦ MatLab	♦ Data Wrangling
♦ HTML & LaTex	♦ Cloud: Azure & AWS	♦ Scikit-Learn	♦ Jupyter & Conda	♦ Numpy & Pandas

PROFESSIONAL EXPERIENCE

Enterprise Data Governance & Business Intelligence **Washington, DC, 2019 - Present**

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

- Lead a research project on DC homelessness by conducting spatial analyses (build & publish interactive maps with R, ArcGIS, HTML and JavaScript), identifying resource gaps, and delivering proposals for the University's current and future initiatives.
- Evaluate the University's data governance current state by pinpointing stakeholders and analyzing survey results: expose data structures with unsupervised learning (clusters & word clouds), build linear regression models, and troubleshoot with heatmaps.
- Innovate project management operations by migrating resources to Cloud and automating reporting dashboards in Smartsheet.

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

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

- Constructed a 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; 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).

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

***Data Science Intern**, The Dragon Catcher LLC*

- 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 manageable datasets by cutting down 4,000 variables 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, 2017**

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

- Automated 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 CEO for optimal business operation.

EDUCATION & CREDENTIALS

GEORGETOWN UNIVERSITY | M.S. Analytics | GPA: 3.78 **2019 - 2021**

- **Courses**: Massive Data Fundamentals, Statistical Computing, Optimization, High Dimension Streaming, Advanced Python & R

MACALESTER COLLEGE | B.A. Applied Mathematics and Statistics | Kofi Annan Scholar **2015 - 2019**

- **Courses**: Intro to Data Science, Object-Oriented Programming & Data Structures, Machine Learning, Numerical Analysis
- **Leadership**: Career Development Fellow, Philanthropy Fellow, TA for Statistical Modeling, Pianist for the Chamber Ensemble.

LinkedIn: www.linkedin.com/in/zhiyu-lin18 • **GitHub:** https://github.com/zlin18 • **Website:** http://zhiyulindata.com/