Zhiyang Li

zhiyangli.hi@gmail.com | (+1) 206.889.7265 | github.com/zhiyangli-hi | linkedin.com/in/zhiyang-li

Education

New York University, New York City, NY — Sept. 2021 – May 2023 (expected)
M.S. in *Data Science* — GPA: **3.78**/4.00

- *Coursework:* Machine Learning, Natural Language Processing, Big Data, Optimization and Linear Algebra

University of Washington, Seattle, WA — Sept. 2017 – June 2021
B.S. in *Applied and Computational Mathematical Sciences - Data Science and Statistics* — GPA: **3.82**/4.00

- *Coursework:* Statistics, Linear Algebra, Differential Equations, Database Systems, Data Structures and Algorithms
- *Honors:* cum laude, 11 times Dean's List, 3 years of Annual Dean's List

Professional experience

Research Assistant — Seattle, WA, USA
Department of Mathematics, University of Washington — Jan. 2021 – July 2021

- Worked with a math Ph.D. student in studying short-term stock option forecast with the Black-Scholes equation.
- Programmed models with Python and utilized UW Hyak supercomputer to approximate minimizers of 170,000 options.
- Used finite element method and Tikhonov regularization to solve the Black-Scholes equation in reversed time.
- Delivered a presentation about ill-posed problems and Tikhonov regularization in the WDRP department seminar.

Research Assistant — Seattle, WA, USA
School of Public Health, University of Washington — July 2020 – Oct. 2020

- Worked with Dr. Backonja and researcher Melinda Schultz on data collection for the SHARE-NW project.
- Coded scripts to crawl, search, and filter official datasets suitable for studying mental health in rural areas.
- Cleaned data and unified variable naming conventions of six datasets and merged them into one.
- Analyze and visualized the processed data using Tableau to illustrate counties with potential areas for future studies.

Reporting and Analytics Intern — Beijing, China
Business-intelligence of Oriental Nations Corporation Ltd. — July 2018 – Aug. 2018

- Developed an automatic service that deploys the company's big data package (BEH) on virtual machines, which includes a collection of important open-source software such as Hadoop, MapReduce, Zookeeper, Spark.
- Worked closely with colleagues to identify and revise outdated instructions in the installation manual.
- Tested the company's big data course kit for universities and provided feedback from a user's view.

Project

Spam detection in text message using Bidirectional LSTM (Python) — Feb. 2022

- Cleaned text data, and coded a custom word Vectorizer class, a light weight adaptation of the sklearn CountVectorizer.
- Built a Bidirectional LSTM model and increased accuracy to 98% compared to 91% by baseline logistic regression model.

Document clustering with k-Means Algorithm (Python) — Jan. 2022

- Constructed a normalized Term Frequency-Inverse Document Frequency (TF-IDF) matrix of 5000+ Wiki documents.
- Implemented Silhouette algorithm and found the optimal number of cluster, then labeled clusters with top words.

Topic Modeling on COVID-19 Tweets with Non-negative Matrix Factorization (NMF) (Python) — Dec. 2021

- Cleaned data by limiting the vocabulary and applied stemming and lemmatization to obtain root form of words.
- Extracted words with highest weights of five topics by NMF on the TF-IDF matrix.

Rain Forecast with Binary Classification (Python) — Aug. 2021

- Implemented hyperparameter tuning for Logistic Regression, k-Nearest Neighbors, and Random Forest models.
- Compared performances of all models by assessing the ROC Curve and AUC.

Award

***3rd Place* in East Coast Terminal by Correlation One (programming competition)** — Sept. 2021

Skill

Programming Languages: Python (Numpy, Pandas, scikit-learn, PyTorch, Keras), R, SQL, SQL++, Java
Tools: Spark, AsterixDB, Tableau, Microsoft Excel/Azure/PowerPoint, LaTeX, Git