Kai Zhang

kaizhang2019master@163.com/ (+86) 18062099816

Personal Website: https://kaizhang0624.github.io/

EDUCATION BACKGROUND

Cornell University   9/2019–9/2020
M.S. in Biostatistics and Data Science | Overall GPA: 3.98 / 4.0
Selected Courses: Statistical Learning, Generalized Linear Model, Survival Analysis, Causal Inference

The Chinese University of Hong Kong (Shenzhen)   9/2015–7/2019
B.S. in Statistics with the 1st Honor Degree | Overall GPA: 3.40 / 4.0 (**Ranked 7/71**) | Major GPA: 3.66/4.0 (**Ranked 3/71**)
Selected Courses: Data Mining, Data Management, Time Series, Stochastic Process, Financial Data Analysis

Online Courses: SQL Fundamentals (DataCamp), Deep Learning with Keras (DataCamp), ML in Python (Coursera)

WORK EXPERIENCE

Graduate Research Assistant   1/2020–Present
Weill Cornell Medicine   New York City, NY
➢ Collaborated with two professors and one PI in performing statistical and bioinformatic analysis for 2 research projects.
➢ Performed Single-cell transcriptomic data analysis implementing unsupervised learning algorithms, including K-means, DBSCAN and graph-based clustering, to detect and characterize subtype of mouse retinal cell.
➢ Performed mediation analysis for the longitudinal data, drafted planned abstract and research paper for peer-review submission.
➢ Managed and presented bi-weekly integrated progress to advisors and clients.

Research Data Analyst (Intern)   9/2018–6/2019
Shenzhen Research Institute of Big Data   Shenzhen, China
➢ Developed pipelines to convert raw Wi-Fi connection records to meaningful behavior data of university students.
➢ Predicted students' potential decline in academic performance using machine learning methods based on behavioral data.
➢ Architected and implemented analytics and visualization components for smart campus platform to analyze social and trajectory network using NetworkX toolkit, which supports multiple research subprojects.

Credit Risk Analyst (Intern)   7/2018–9/2019
Changjiang Securities Company Limited   Wuhan, China
➢ Participated the project of developing risk-rating scores for listed companies using logistic regression-based model.
➢ Delivered data summaries and financial reports to senior management on regular base.

PROJECTS

Single-cell RNA Data Analysis on Mouse Retina (Supervisor: Prof. Kathy Zhou)
➢ This project uses machine learning methods (clustering) to identify the cellular composition of mutant and control mouse retinal pigment epithelium (RPE), and further to detect the gene markers for different grades of injury RPE cells.
➢ Processed raw sequence data into a high-quality expression dataset following *Scater* workflow, including rigorous pre-processing, quality control, features selection, and dimensionality reduction.
➢ Performed and compared multiple clustering algorithm on cells, including K-means, DBSCAN and SNN graph-based clustering, and detected marker genes for clusters using pairwise t-test with adjustment for multiple comparison.

Causal Inference on Relationship Between Professions and Wage Using Data-adaptive Method (link)
➢ This project found causal effect of professions (information or industry) on male wage based on machine learning method.
➢ Followed roadmap for causal inference, and estimated average treatment effect using doubly robust estimation (AIPW and TMLE) based on Super Learner stacking multiple machine learning models (LR model, Random Forest, XGBoost, etc).

Predicting GPA Declines Based on Behavioral Data
➢ This project used machine learning method to predict the potential declines in academic performance of college students, based on the behavioral data collected and extracted from Wi-Fi connections.
➢ Developed pipelines to convert raw Wi-Fi connection records to meaningful behavior data of college students, including class attendance, online activities (web traffic), monthly spending time in multiple buildings in campus, etc.
➢ Developed GPA prediction model using XGBoost based on behavioral data and students' profiles, and attained an overall accuracy of 96.3% with sensitivity of 81.8%.

AWARD

Dean's List	2018-2019
Master's List Scholarship of Shaw College (Top 10% Academic Scholarship)	2018-2019
Master's Program - Academic Achievement (BDS track)	2019-2020

LANGUAGES AND SKILLS

GRE General: V: 154 / Q: 170 / AW: 3.5 / Total: 324
Programming Languages / Software: R (tidyverse, CARET, R shiny), Python (pandas, scikit-learn, Keras), SQL