Ziyu Xiao

xiaoziyu1997@gmail.com | www.linkedin.com/in/ziyu-xiao | 206-468-4156

QUALIFICATIONS

- Solid academic background in biostatistics and data science. - SAS Certified Base Programmer. - Hands-on experience in data manipulation, modeling and visualization. - Self-motivated; capable of learning new techniques fast; years of project-oriented teamwork experience; working individually and cooperating with fast-paced team.

Skills: survival, longitudinal, categorical data analysis, generalized linear models, hypothesis testing, statistical inference, bootstrap, permutation, machine learning, sample size calculation, statistical analysis plan.

Programming Languages: R, SAS, Python, SQL, Bash

EXPERIENCE

Risk-based Assessment of Heterogeneity of Treatment Effect Using PATH Strategies in LESS Trial, Seattle, WA

Biostatistician **09/2020 – 03/2021**

Developed predictive algorithms that can provide a prognostic score for functional limitations related to back pain and apply them to stratify patients within the clinical trial (LESS) to examine risk-based variation in treatment effects.

- Provided a detailed statistical analysis plan (SAP) of the analyses.
- Conducted the descriptive statistics and exploratory analysis by using the longitudinal data of back-pain patients from different sources.
- Wrote documented code and results for whole project, including data cleaning, merging & predictive algorithm.
- Gave oral presentation of the project to the whole research group.

Yidu Cloud Technology Company Ltd., Beijing, China

Data Mining and AI Product Manager **06 – 09/2020**

Collaborated with cross-functional team to optimize existing product (a biomedical big-data intelligence platform that enables researchers to organize, link and structure real-time data from a number of clinical business systems, including HIS, RIS, and EMR).

- Investigated the main practice of artificial intelligence combining with big data in the medical field.
- Cleaned, merged and applied APRIORI algorithm for leukemia datasets with R and Python.
- Developed and validated several predictive models (DT, LR, SVM, Naive Bayesian) for the happen of post-operative VTE using the real-world gastric cancer data.
- Presented data and communicated the results of analyses with non-data professionals.
- Designed literature recommendation feature within the product, drew flowchart, prototype & interactive design, wrote the code for the construction of user & paper label through TF-IDF, recommended paper according to cosine similarity, and finally accomplished the product requirement document (PRD) for software engineers.

China Novartis Institutes for Biomedical Research Co., Ltd., Shanghai, China

Programmer **07 - 08/2019**

Participated in the iteration of an internal web-based platform in supporting scientists throughout the company to accelerate drug R&D through the combination of data science and biology.

- Cleaned, analyzed, interpreted the gene data using Python and Bash scripts.
- Established machine learning model pipelines to predict association between gene and diseases.
- Explained, visualized and presented the final results across the company, winning team award.

RESEARCH

Determining the Balance between Drug Efficacy and Safety by the Network and Biological System Profile of its Therapeutic Target, Hangzhou, China

Research Assistant **03 - 07/2018**

Explored the balance between efficacy & safety of drugs by analyzing drug targets' network & system biology features.

- Analyzed data of 40, 468 human protein-protein interaction network and biological system profiles on the balance between drug efficacy and its safety through feature selection Boruta.
- Drafted manuscripts for publication and published the paper as coauthor. (Li, X.X., etc. 2018. Determining the Balance between Drug Efficacy and Safety by the Network and Biological System Profile of Its Therapeutic Target.)

EDUCATION

MS Biostatistics, University of Washington, Seattle WA **Expected 3/2021**

Nonparametric Regression & Classification; Biomedical Data Science; Computational Skills for Biostatistics; Data Analysis & Reporting; Longitudinal & Multilevel Data Analysis; Categorical & Survival Data Analysis; Statistical Methods for Spatial Data