Siyu Wang

5002 Sheboygan Ave, Madison, WI, 53705 | (608) 381-5100 | iamwangsiyu@gmail.com | linkedin.com/in/siyu79/

EDUCATION

UNIVERSITY OF WISCONSIN-MADISON *Madison, WI*
Ph.D. student in Statistics *Expected Aug 2024*
Master of Science in Statistics - Data Science *May 2019*
Coursework: Experimental Design, Statistical Consulting, Causal Inference
Research area: Decision Tree Algorithms, under the supervision of Professor Wei-Yin Loh

UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA *Hefei, China*
Bachelor of Science in Statistics *May 2018*

PROFESSIONAL EXPERIENCE

Inari Medical *Irvine, CA*
Biostatistics Intern *May - Aug 2022*

- Participated in quality control for a data preprocessing pipeline, transforming data from the IBM database to the CDISC ADaM dataset using R and SAS. Executed comprehensive testing and validation procedures to guarantee accuracy and reliability. Instrumental in detecting and rectifying data issues during the integration of health records in CRF from real-world registries.
- Collaborated with the Clinical Insights team to evaluate the influence of biomarkers on forecasting patient severity. Implemented innovative survival models, not originally outlined in the Statistical Analysis Plan (SAP), for ad-hoc counterfactual analyses and identifying specific subgroups. This approach resulted in a 20% decrease in the Hazard Ratio for the treatment effect compared to traditional survival models.
- Engaged in methodological research focusing on the comparative nuances between SAS and R in handling repeated measurements. Conducted extensive literature review on marginal homogeneity. Used Python's manim package to develop explanatory animations, effectively illustrating complex concepts. Presented these research findings and insights at monthly department meetings, contributing to knowledge sharing and departmental learning.

RESEARCH

R Package for an Enhanced Decision Tree Method (LDATree) *Sep 2021 - Now*

- Authored a comprehensive decision tree package in R, currently available on CRAN. Identified and documented specific scenarios where this new decision tree method exhibits its strongest performance. The program supports missing values, generally outperforms traditional Linear Discriminant Analysis (LDA), and matches Random Forest's accuracy with better efficiency.
- Identified key areas within the decision tree framework where the integration of LDA could enhance performance. Conducted an in-depth review of existing literature on LDA development, comparing unique features of various models. Through extensive testing and analysis, developed a new stepwise LDA variant optimally tailored to our specific requirements.
- Carefully reviewed and refined each aspect of the decision tree, including variable selection (splits), model selection (pruning), and node modeling. Applied advanced coding techniques, algorithmic optimizations, and statistical tests to effectively balance speed and accuracy. Developed diverse visualization formats to facilitate clear communication of complex statistical concepts, enabling easier interpretation and insight extraction for professionals without a statistical background.