Wenxuan Zhu

wzhu@macalester.edu | Saint Paul, MN 55105 | (651) 529-4410
Personal Website: https://wenxuanzhu.netlify.app/

Education

Macalester College, Saint Paul, MN — Honors Bachelor of Arts Degree Expected May 2024
Major: Statistics *Minor*: Economics — *GPA*: 3.92 / 4.0
Honors: Recipient of the Charles J. Turck Honor Scholarship awarded to a limited number of international students who demonstrate special potential to excel academically (4-year scholarship); Macalester Dean's List (7/7 semesters)
Relevant Coursework: Correlated data, Real Analysis, Mathematical Statistics, Causal Inference, Computational Linear Algebra, Statistical Machine Learning, Linear Algebra, Probability, Applied Multivariable Calculus III

Publications and Presentations

- Zhao, B., **Zhu, W.**, Hao, S., Hua, M., Liao, Q., Jing, Y., Liu, L., & Gu, X. (2023). Prediction of heavy metals accumulation risk in rice using machine learning and mapping pollution risk. *Journal of Hazardous Materials*, *448*(15), 1–13. https://doi.org/10.1016/j.jhazmat.2023.130879
- **Zhu, W.**, Mills, L., Myint, L., Addona, V., Bosnakovski, D., "Investigating CIC-DUX4 Sarcoma Surfaceome." Masonic Cancer Center Research Symposium 2023, Minneapolis, MN. *Poster Presentation.* [View Poster]
- **Zhu, W.**, Mills, L., Myint, L., Bosnakovski, D., "Exploring Surfaceome of CIC-DUX4 Sarcoma." Joint Mathematics Meetings (JMM) 2024, San Francisco, CA. *Selected for Oral Presentation.*

Research Experience

Senior Honors Project, St. Paul, MN — Sep. 2023 - Present
Undergraduate Thesis at Macalester College
- In biological research, validating potential findings often involves tedious literature review, reading through a lot of journals. Is there an efficient "shortcut" for this process? This study seeks to develop such a shortcut, particularly in the context of CIC-DUX4 sarcoma, by employing network science and software packages to explore the comprehensive database SemMedBD, compiling texts extracted from all PubMed article abstracts.

University of Minnesota, Minneapolis, MN — Jun. 2023 - Aug. 2023
Research Intern at Bosnakovski Lab
- Investigated genes encoding surface proteins of CIC-DUX4 sarcoma, a pediatric cancer, across cell lines and tumor samples using RNA-Sequencing data in R
- Attained a robust understanding of genomics, searched public GEO bulk sequencing data, mapped the reads to reference genomes accordingly, and generated counts data using Linux
- Conducted differential expression analyses between control CIC-DUX4 cancer cell lines and iP300w 4-hrs treated cancer cell lines using `DESeq2` package; created related tables, heat maps, volcano plots, and PCA plots
- Developed Venn diagrams and identified a subset of 19 genes that showcased both a presence within the surface protein repertoire and differential expression analysis across cell lines subjected to iP300w treatment
- Performed pathway analysis for the identified genes and identified top 20 most significant pathways for each gene
- ***Collaborating with Prof. Darko Bosnakovski on drafting of a manuscript, with the intention of publishing***

Nanjing University, Nanjing, China — Oct. 2022 - Feb. 2023
Research Assistant at State Key Laboratory of Pollution Control and Resource Reuse
- Utilized machine learning models to map pollution and predict the risk of heavy metal accumulation (Cd, Hg, As, Pb, Zn, Ni, Cr, and Cu) in japonica rice in southern China
- Developed and implemented 10 modeling methods using 10-fold cross validation, including supervised models (Ridge, LASSO), and unsupervised models (SVR, ANN, RF, GBRT, ERT)
- Conducted thorough variable and model selection and comparison to identify ERT as the most effective approach to predict heavy metal accumulation in rice
- Uncovered the significant influence of temperature and precipitation on mercury accumulation in rice
- Identified soil-Cadmium level and pH value as the two most decisive features affecting rice-Cadmium uptake based on feature importance analysis (Zhao, et al. *Prediction of heavy metals accumulation*. 2023) [View Paper]

Selected Academic Experience

Spatial Analysis of Urban Heat Island (UHI) in Twin Cities & Phoenixs, St. Paul, US — Nov.2023 - Dec.2023
- [View Report]: Utilized Spatial Autoregressive (SAR) models and Conditional Autoregressive (CAR) models to conduct a spatial analysis of the UHI in two cities, highlighting its uneven distribution across American cities

- Compare model similarity and performance, confirming significant connections between race, income, and house value with UHI distribution, while each model identify city-specific patterns for policy implementation

Time Series Analysis of Macalester Main Campus Electricity Usage, St. Paul, US Sep. 2023 - Oct. 2023
- [View Report]: Modeled and removed trend using B-spline and seasonality using months & during-semester status
- Analyzed the ACF and PACF plots of remaining correlated noise, and constructed MA(1), AR(2), and ARMA models; used BIC index for model selection and performed two-year forecasting

Matrix Completion and Movie Recommendation Systems, St. Paul, US Mar. 2023 - May. 2023
- [View Report]: Analyzed MovieLens dataset with 9700+ films, and 99500+ rating scores, revealing real-world sparsity challenges in large user-item matrices
- Explored method advantages, limitations, and potential extensions to enhance prediction accuracy
- Researched and developed personalized recommendation system using Singular Value Thresholding (SVT), Nuclear Norm, and Cosine Similarity, surpassing baseline methods in RMSE

Investigating EM Algorithm, St. Paul, US Feb. 2023 - May. 2023
- [View Website]: Published interactive R bookdown website on EM Algorithm and its applications, featuring comprehensive content and examples
- Utilized R programming, R bookdown, Git, and collaborative tools for project development and maintenance
- Contributed to key sections of the bookdown (Introduction to EM, Image Segmentation, Gaussian Mixture Model, and Binomial Mixture Model) through extensive research and concept synthesis

Causal Inference: Self-Perception of Creativity & Participation in Arts, St. Paul, US Jan. 2023 - May. 2023
- [View Details]: Conducted in-depth assessment of the causal relationship between self-perception of creativity and arts participation using a dataset of 3,447 adults with diverse backgrounds
- Constructed a Directed Acyclic Graph (DAG) graph and employed Outcome Regression, Random Forest, and Inverse Probability Weighting (IPW) models, revealing a statistically significant positive relationship
- Validated the findings through sensitivity analysis, confirming the stability and validity of the estimates

DataFest 2023 & 2022, St. Paul, US Mar. 2023 & Mar. 2022
Team Leader
- [View Details]: Analyzed and presented findings within 48 hours, including data from an online pro bono legal service by the American Bar Association (2023) and a dataset focused on the Yale Play2Prevent game (2022)
- Collaborated with teammates to clean and join data with 2,100,000+ rows and 130+ variables in 15+ spreadsheet
- Conducted thorough statistical analysis, such as K-means clustering, to identify patterns and associations

Data Analysis Research on Reviews for Luxury Hotel in Europe, St. Paul, US Sep. 2021 - Dec. 2021
- [View Slides]: Cleaned dataset with 515,000+ cases by filtering, combining, and mutating variables to predict the reviewer's score for luxury hotels in Europe and led the final presentation
- Trained statistical models, including LASSO, GAMs, Logistic Regression, Random Forests, PCA, and K-Means Clustering, to predict the review score for Hotel and visualized analytical results and variable importance in R

Teaching

Macalester College MSCS Department, St. Paul, US Jan. 2023 - May. 2023
Teaching Assistant
- TA for *Statistical Machine Learning* and *Mathematical Statistic*s (Capstone level)
- Hold office hours to review their code, identify errors, and suggest effective methods for further study
- Explained machine learning concepts, pros and cons for each model, and the implication in real data to students
- Assessed and graded weekly assignments for 15+ students to provide valuable feedback

Additional Experience

MacProject Corps — Ramsey County, St. Paul, MN Jan. 2022
Team Leader
- Researched in available datasets, news, and citizens' reviews to identify the goal in community wellbeing
- Designed strategies for community-based responses to 911 via threat scale, post-call follow up, etc.

Skills

Language:	Fluent in Chinese and English
Programming:	R, LaTeX, GitHub, Mathematica, Linux, Python, SQL, and Matlab