

EDUCATION

---

- **University of Wisconsin-Madison** Madison, WI  
*Ph.D. in Statistics; GPA: 3.91* *Sep. 2020 – present*  
*Master of Science in Data Science; GPA: 4.00* *Aug. 2019 – May. 2020*  
*Visiting Student; GPA: 4.00* *Aug. 2018 – May. 2019*
- **Zhejiang University** Hangzhou, China  
*Bachelor of Science in Statistics; GPA: 3.71* *Aug. 2015 – Jun. 2019*

PROJECTS AND RESEARCH EXPERIENCE

---

- **Selecting the number of principal components** May. 2021 - current  
*Advisor: Karl Rohe, University of Wisconsin-Madison* *Madison, WI*
  - **Hypothesis test:** Developed a hypothesis testing procedure based on the bi-cross-validated eigenvalues that selects a reasonable number of principal components.
- **Generating amino acid sequence for protein synthesis** Aug. 2020 - current  
*Advisor: Sebastian Raschka, University of Wisconsin-Madison* *Madison, WI*
  - **Model structure:** Integrated a GPT-based generative model and a BERT-based discriminative model as a semi-adversarial network to facilitate directed protein evolution.
  - **Discriminative model:** Pre-trained BERT using unlabeled protein sequence dataset. Fine-tuned the pre-trained model and reached 68% accuracy on 8-class secondary structure classification task using Pytorch.
  - **Generative model:** Modified existed proteins using GPT-based model under the combined loss of protein activity score and secondary structure prediction.
- **Citation graph analysis on semantic scholar dataset** Sep. 2020 - Dec. 2020  
*STAT 992: course project* *Madison, WI*
  - **Journal clustering:** Aggregated the citation network of 220 millions papers by journals. Applied VSP (vintage sparse PCA) on the citation graph and obtained journal field membership based on citation patterns.
  - **Statistical paper analysis:** Focused on the Statistics-related journal cluster and zoomed into paper-wise citation graph. Applied VSP and bff (best feature function) to obtain clustered statistical methods.
  - **Trend analysis:** Selected statistical methods of interest and tracked their development over time based on the appearance in papers' abstract.
- **Yelp Business Analysis** Oct. 2019 - Dec. 2019  
*STAT 628: course project* *Madison, WI*
  - **Data cleaning and visualization:** Cleaned 1 million bar business review data and merged corresponding business, user, tip information. Visualize word importance based on TF-IDF.
  - **Topic Modeling:** Implemented LDA and NMF in python based on review and user weight. Used TC-W2V coherence to evaluate topic quality. Provided specific advice to 8000 bar business owners.
  - **Web app:** Built interactive shiny app to present procedure of analysis and provide real-time suggestions for improvements.

SKILLS

---

- **Software:** Proficient in R and Python, experienced in SQL, Matlab and C
- **Packages:** Pytorch, Transformers, Pandas, Numpy
- **Related Coursework:** Mathematical Statistics, Statistical Learning, Experimental Design, Regression and Analysis of Variance, Nonlinear Optimization, Theoretical Machine Learning, Algorithms

HONOR AND AWARDS

---

- **2015-2016:** Zhejiang University Academic Excellence Award
- **2017:** Winner of the 15th Statistical Modeling Contest at Zhejiang University
- **2018-2019:** Exchange & Visiting International Student Academic Excellence Award (twice)