

EDUCATION

- **University of Wisconsin Madison** Madison, WI
Master of Science in Data Science; GPA: 4.00 Aug. 2018 – Present
- **University of California, Los Angeles** Los Angeles, CA
Visiting Student (Summer session); GPA: 4.00 Aug. 2017 – Sep. 2017
- **Zhejiang University** Hangzhou, China
Bachelor of Science in Statistics; GPA: 3.71 Aug. 2015 – Jun. 2019

PROJECTS AND RESEARCH EXPERIENCE

- **Build and select models for hemopoiesis prediction** Dec. 2018
STAT 601: course project Madison, WI
 - **Modeling:** Use LASSO, PCA to reduce data dimension and fit linear model that is capable for high-dimensional data. Evaluate models using GMC (generalized measure of correlation), which evaluates non-linear relationship.
 - **Application:** Select the principal components or genes that contribute most to the response. Select best models using different standards. Interpret the findings.
- **Tweet sentiment analysis** Mar. 2019 - May. 2019
NLP project using deep learning methods Madison, WI
 - **Train neural network models:** Build 3-class sentiment classifiers for Sentiment 140 dataset, which contains 1.6 million tweets.
 - **Model performance comparison:** Compare different deep learning methods with traditional machine learning model using criteria of training speed and prediction accuracy. Reach 90% accuracy in 3-class classification.
- **Generating amino acid sequence for protein synthesis** Aug. 2019 - Present
collaborative research project(Prof. Raschka and Prof. Romero) Madison, WI
 - **Modeling:** Implement character-level recurrent neural network and transformer model to learn patterns of amino acid sequences. Model long-range dependencies in the sequence data.
 - **Evaluation:** Use reconstruction accuracy and secondary structure to serve as semantic meaning in text generation.
- **Yelp Business Analysis** Oct. 2019 - Present
STAT 628: course project Madison, WI
 - **Data cleaning and visualization:** Clean 1 million bar business review data and merge corresponding business, user, tip information. Visualize word importance based on TF-IDF.
 - **Topic Modeling:** Implement LDA and NMF in python based on review and user weight. Use TC-W2V coherence to evaluate topic quality. Provide specific advice to 8000 bar business owners.
 - **Interactive shiny app:** Build shiny app to present procedure of analysis and provide real-time suggestions for improvement.

SPECIALIZED SKILLS

- **Related Courses:** Mathematical statistics, Real analysis, Regression analysis, Mathematical software, Multi-variate statistical analysis, Deep learning
- **Software:** Proficient in R and Python, experienced in SQL, Latex and C
- **Statistics:** Experimental design, deep learning, machine learning, probability theory, linear regression, PCA, ANOVA, A/B testing, stochastic processes, time series
- **Standardized Tests:**
TOEFL iBT: 111 (R:29, L:30, W:26, S:26)
GRE: 324 (V:154, Q:170, AW:3)

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)