Ziyan (Cecilia) Xia

4127584935 xiaziyan@cmu.edu
515 South Aiken Ave, Pittsburgh, PA 15232
https://www.linkedin.com/in/xiaziyan

SKILLS

- Programming Languages: SQL, Python, R, MATLAB, SAS, Javascript
- Tools: PostgreSQL, MySQL, Docker, Jupyter Notebook, Rstudio, Google Cloud Platform, MATLAB, SAS

EDUCATION

Carnegie Mellon University Aug 2021 - May 2022
Master of Statistical Practice | Department of Statistics and Data Science Pittsburgh, PA, US
GPA: 4.0/4.0 | **Spring 2022 Courses:** Cloud Computing, Statistical Consulting Capstone

Central China Normal University Sep 2017 - Jul 2021
Bachelor of Statistics | School of Mathematics and Statistics Wuhan, China
GPA: 87.69/100 (**Rank** 4/24)

APPLIED STATISTICAL EXPERIENCE

University of California, San Francisco (UCSF) Roland Henry Lab San Francisco, CA, US
Undergraduate Researcher in Data Science and Neuroscience (UC Berkeley URAP)
August 2020 - March 2021

- Built activity recognition models that identify different types of activity and can recognize relative intensity using methods including Peak Detection Algorithm, Dynamic Time Wrapping (DTW), and k-means clustering
- Developed scalable tools using Javascript for efficient annotations of activity graphs
- Designed R Shiny Web Apps to realize interactive visualization of data and analysis results

March 2020 - July 2020

- Modeled Fitbit-collected time series data of 120 Multiple Sclerosis patients' step counts over the past three years to develop remotely predictive health tests
- Tested multiple methods to handle missing Fitbit data (including interpolation, moving average, State Space Model, etc) using Cross-Validation
- Achieved time series data aggregation by hours, days, and weeks for original minute-by-minute data.
- Conducted predictive Regression-ARIMA hybrid modeling to predict future changes in motor outcomes for the patients

IMDB Recommendation Engine Wuhan, China
October 2020 - December 2020

- Designed a recommendation engine in Python and R to predict Users' ratings for different genres of movies by implementing the concept of User-Based Collaborative Filtering using Pearson Correlation
- Built an interactive R Shiny App for visualization of the recommendation results

Machine Learning Methods to Discover Interactions between Transcription Factors Berkeley, CA, US
February 2020 - June 2020

- Aimed to extract meanings from large complex gene data to discover pairwise and higher-order interactions between TFs of a model organism and detect important transcription factors for gene expression mostly through the Random Forest (RF) and iterative Random Forest (iRF) methods
- The results not only found the most frequent and stable pairwise and higher-order TF interactions but also found the TF pair ‘caup’ and ‘onecut’ was likely to have a similar function on gene expression and suggested future research on this

Mediocre Social Network Apps Inc. - Stock: Analysis and Forecast Berkeley, CA, US
March 2020 - May 2020

- Conducted SARIMA models and Regression-ARIMA hybrid models to achieve the stock price forecasting based on the stock price of Mediocre Social Network Apps Inc. across various time horizons
- Achieved model selection according to Information Criterion and MSE generated by Nested Cross-Validation and the performance of the final model ranked top 20% in class based on MSE

TEACHING EXPERIENCE

Carnegie Mellon University-Department of Statistics and Data Science Sep 2021 - Present
Teaching Assistant for Course: Methods for Statistics and Data Science Pittsburgh, PA, US

- Leading labs, holding office hours, grading exams and homework