Ziyan (Cecilia) Xia

+1 4127584935 xiaziyan@cmu.edu
515 South Aiken Ave, Pittsburgh, PA 15232
https://www.linkedin.com/in/xiaziyan https://xiaziyan1999.github.io

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

- Conducted time series segmentation using Peak Detection Algorithm
- Built activity recognition models to cluster segmented time series into different intensity levels using Dynamic Time Wrapping 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

- Upgraded missing Fitbit data imputation after testing methods including interpolation, moving average, State Space Model, and Kalman Filter using Nested Cross-Validation
- Achieved time series data aggregation by hours, days, and weeks from original minute-by-minute data
- Modeled Fitbit-collected time series data of 120 Multiple Sclerosis patients' step counts over the past three years to develop remotely predictive health tests
- Accelerated research process by automating real-time analysis reports (prediction data, output model parameters, and associated graphs) generating process

Compare Machine Learning Classification Models' Performance on Imbalanced Genetic Data Pittsburgh, PA, US
October 2021 - December 2021

- Conducted classification performance comparison of XGBoost, Random Forest, and iterative Random Forest on both the original imbalanced data and the sampled data using criteria including ROC Curve, AUC and Precision

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

- Extracted meanings from large complex genetic data and discovered pairwise and higher-order interactions between TFs of a model organism using methods including Random Forest and iterative Random Forest.

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

- Achieved top 2% performance on stock price forecasting of Mediocre Social Network Apps Inc. across various time horizons using SARIMA models and Regression-ARIMA hybrid models

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