Wu (Emma) Yiying

Website: https://wuyiying2018.github.io/index.html Phone: (702)637-5693 E-mail: yw3996@cumc.columbia.edu

EDUCATION

Columbia University Mailman School of Public Health New York, United States

Master of Public Health in Biostatistics 09/2022 - 06/2024 (Expected)

- Relevant Coursework: Applied Regression, Data Science, Analysis of Health Survey Data, Database(MySQL)

Shanghai University of Finance and Economics Shanghai, China

Bachelor of Science in Statistics 09/2018 - 07/2022

- Relevant Coursework: Real Analysis, Time Series, Survival Analysis, Machine Learning, Stochastic Processes
- Honor: First Prize for The 5th Mathematical Modeling Competition of SUFE (Top 1%, 04/2019)
- Certification: SUFE-Cambridge Leadership Online Programme (07/2020)

INTERNSHIPS

Data Analyst *Wellness Equity Alliance* California, United States, 07/2023 - Present

- Assisted in constructing the public health system for Vernon, California, and neighboring regions through demographic data analysis
- Used Google Earth (GIS) to display demographic data, mental health services, and food security resources

AI Analyst *Microsoft China, Microsoft AI and IoT Insider Lab* Shanghai, China, 06/2021 - 08/2021

- Tested the Microsoft Azure cloud product “deep-voice-conversion-master” by converting audio text data into target text with a Convolutional Neural Network on a Linux system
- Wrote code in Python to convert the target text into the specified format in batch

Investment Analyst *Hwabao Securities* Shanghai, China, 06/2020 - 08/2020

- Examined historical and current market supply and demand of the pharmaceutical industry
- Conducted statistical description/regression analysis to predict the risk of overseas investment and surplus capacity

RESEARCH

Research Assistant New York, United States, 06/2023 - present

Project: Predicting Methylation from Sequence and Gene Expression Using Deep Learning Methods

- Developed deep learning models that can accurately predict DNA methylation patterns based on genomic sequence and gene expression data
- Utilized Python and Columbia University’s research computing resources to test, evaluate, and refine deep learning models, ensuring optimal performance and reliability

Undergraduate Researcher Shanghai, China, 03/2022 - 05/2022

Project: Research on Cluster Analysis of E-commerce customers based on RFM model

- Drew word clouds about the most frequent words in the product description and brand column using the R package wordcloud2, to have a brief knowledge of the dataset
- Constructed a Recency, Frequency, and Monetary model (RFM) based on the data collected, and analyzed the model using the K-means algorithm
- Analyzed the groups of customer data after segmentation, and predicted their behavior in the future

Undergraduate Researcher Shanghai, China, 06/2021 - 08/2021

Project: Prospects and Challenges of "Internet+" for Community Healthcare in Shanghai

- Developed a questionnaire based on the Technology Acceptance Model (TAM) to investigate the acceptance of “Internet+” community healthcare by Shanghai residents
- Constructed a Structural Equation Model (SEM) of Shanghai residents' satisfaction with "Internet Community Clinic Services" using the TAM, and Analyzed the model using Partial Least Squares (PLS) written in R

SKILLS

- Software: R, MySQL, Python, SAS, Stata, LaTeX, and MS Office
- Language: English (Fluent), Chinese (Native)