

Mengyuan Yu

New York, NY 10032 | +1 929 657 4130 | my2838@cumc.columbia.edu

EDUCATION

Columbia University

09/2023 - 05/2024

Master of Science in Biostatistics (Public Health Data Science Track)

Relevant Coursework: Data Science, Biostatistical Method

Beijing Jiaotong University

09/2019 - 06/2023

Bachelor of Science in Information and Computation Science

PROJECT

College Students Innovation and Entrepreneurship Competition Project

04/2021 - 06/2022

Team Leader

- Extracted the patient's transcriptomic data of various cancer species in The Cancer Genome Atlas Database (TCGA).
- Performed a model to calculate immune and stromal scores through Python and conducted survival and correlation analysis between different clinical factors (whether alcohol, hepatitis) to provide recommendation for personalized treatment.
- Constructed and visualized the protein-protein interaction network and immune pathway and analyzed the gene expression levels to obtain overlapping HUB genes related to prognosis.
- Decided the independent HUB genes through univariate and multivariate COX analysis using R and then verified.

Celebrity Popularity Ranking Based on Hierarchical Analysis

05/2021

Independent Researcher

- Qualified the indexes (search volume, social media followers, etc.) to evaluate the popularity of celebrities are determined according to the cognition of the public and market research.
- Quantified different types of data by Technique for Order Preference by Similarity (TOPSIS) and built a hierarchical structure model using MATLAB.
- Constructed a judgment matrix by multi-objective analysis of hierarchical planning (AHP) to synthesize the popularity of celebrities into a broad category ranking using MATLAB.

Time Series Complexity Analysis based on the Theory of Information Entropy

03/2023 - 05/2023

Independent Researcher

- Analyzes several mainstream entropy analysis models using simulated data (Gaussian white noise, 1/f pink noise, and logistics map systems) to explore the effects when parameters and features of data sequences changed by Python.
- Classify various pathology with ECG signals based on the multi-scale sample entropy model by Python.

WORK EXPERIENCE

Publicis (Beijing) Cultural Communication Co.,Ltd

11/2022 - 03/2023

Data Analyst Intern (of the market Nestle's professional nutrition online team)

- Organized and built the DataMart of 300k+ consumer consumptions records to increase in operational efficiency.
- Optimized data pipelines in SQL to automate in collecting and categorizing the customer satisfaction feedback data and provide decision-making suggestions.
- Aggregated and visualized daily and weekly sales data in dashboards of Excel and Power BI by different dimensions (locations, membership types, etc.) to examine and identify the sales risks.

Third People's Hospital of Jiujiang City (TPHJC)

06/2022 - 08/2023

Biomedical Engineering Intern

- Build an automatic patient data processing system with Python and docked the port of the system with the PACS system of the local hospitals to invoke relevant data from database directly.
- Designed, developed, and implemented the links to support the data processing system both front- and back-end usage.

SKILL

- Programming: **Python** (Pandas, numpy, matplotlib), **R** (dplyr, tidyr, ggplot2), **SQL**, **MATLAB**, **C++**
- Tools: Tableau, Power BI, MS Office, EViews, SPSS