Miaoqing Yu

2930 Chestnut Street, Philadelphia, PA 19104
267-921-8980 mqyu@sas.upenn.edu LinkedIn

Education

University of Pennsylvania **Sep. 2021 – May. 2023 (Expected)**
M.A. in Applied Math and Computational Science | GPA: 3.83 / 4.0 *Philadelphia, PA, US*

- **Core courses:** Machine Learning, Algebraic Techiques, Probability Theory

Huazhong University of Science and Technology **Sep. 2016 – Jul. 2020**
B.S. in Economics | GPA: 3.84 / 4.0 *Wuhan, China*

University of California, Berkeley **Jan. 2019 – May 2019**
Undergraduate Exchange Student | GPA: 3.81 / 4.0 *Berkeley, CA, US*

- **Core courses:** Discrete Math, Optimization

Technical Skills

Programming & Database: Python, PrestoSQL, MySQL, Matlab, Shell, Git
Data Analysis & Visualization: Pandas, Numpy, Sklearn, Pytorch, Matlibplot
Techniques: A/B Testing, Hypothesis Testing, Machine Learning (Prediction, Classification)

Experience

Yidu Tech Inc. | ***Healthcare Industry*** **Mar. 2021 – Jul. 2021**
Data Science Department, Intern *Beijing, China*

- Processed over 300 million patients' clinical data in PrestoSQL with shell commands to help improve clinical diagnosis and treatment outcomes.
- Established exploratory data analysis pipeline via the Pandas package and conducted data visualization using Matplotlib.
- Built a real-world lung cancer model by applying logistic regression, which yields 99% accuracy rate and 98% recall, saving 8h of manual classification work.
- Organized weekly tech seminars and gave presentations about Reinforcement Learning in the healthcare industry to colleagues.

Southern University of Science and Technology **Apr. 2020 – Sept. 2020**
Research Assistant *Shenzhen, China*

- Utilized machine learning combined with econometric causal inference methods to estimate heterogeneous treatment effects of various systemic risk events in the internet information environment.
- Applied random forest to evaluate how the internet information environment affects capital flows on online financial platforms.
- Conducted a literature review overviewing machine learning approaches to previous internet financial platforms research.

Projects

Stroke Prediction | *supervised learning, classification* **Nov. 2021**

- Predicted patient stroke based on labelled demographic and health data via Python programming.
- Preprocessed the dataset to deal with imbalanced and missing data, categorical feature transformation, standardization, etc.
- Trained classification machine learning methods including Logistic Regression, Support Vector Machine, Random Forest, AutoML and applied regularization to avoid overfitting.
- Evaluated model performance, finding the tuned random forest model performed the best, achieving 99.34% accuracy and 97.22% F-1 score.

Principal Component Analysis on Face Reconstruction **Dec. 2021**

- Applied Principal Component Analysis (PCA) on face image to reconstruct image and save file storage.
- Saved 80% storage while maintaining 80% of image information through dimension reduction.
- Improved original PCA algorithm by using Rayleigh Quotient Iteration, reduced reconstruction time by over 10 times.