YUNQIU (LULU) YAO

100 Haven Avenue Apt 19G, New York, NY 10032
yy2827@columbia.edu | (646)-683-6625 | linkedin.com/in/yunqiu-yao

EDUCATION

Columbia University New York, NY
Master of Science in Biostatistics, GPA 3.97/4.0 *May 2019*

- Courses: Deep Learning, Database, Statistical Learning, Data Science, Probability, Inference, etc.
- Graduate Teaching Assistant (Spring 2018, Fall 2018)

Shanghai Jiao Tong University Shanghai, China
Bachelor of Science, Food Science and Engineering *June 2017*

- Academic Progress Scholarship; Academic Excellence Scholarship

SELECTED PROJECTS

Cancer Detection on Gigapixel Pathology Images *Sept. 2018 – Dec. 2018*

- Created samples by sliding a window across the gigapixel biopsy images to extract patches and labels
- Constructed the model based on convolutional neural network, made predictions on tissue patches from a new biopsy slide and output a heatmap showing the cancerous regions with an accuracy of 96.92%

Is Venmo Safe to Use? *Sept. 2018 – Dec. 2018*

- Developed Python code to web-scrape the recent public transaction records via Venmo public API
- Constructed a database with PostgreSQL on the Google Cloud Platform, calculated the risk score for each user to reminded them of the potential information leakage, and implemented the application with Flask

Study on the Readmission Rate for Diabetes *Mar. 2018 – May 2018*

- Analyzed 67,069 electronic medical records regarding the readmission status of patients with diabetes
- Constructed models with kNN, random forest, support vector machine, logistic regression algorithms to evaluate the efficacy of factors such as different treatments on readmission rate and make predictions

Why is a TED Talk Popular? *Nov. 2017 – Dec. 2017*

- Collaborated on GitHub and performed regression, sentiment analysis and data visualization in R Markdown with dplyr, ggplot2, shiny and plotly packages to determine the key factors influencing views
- Integrated findings in a scientific report on project website and recorded a screencast for presentation

Empirical Bayes (EB) Method for Haplotype-based GWAS *Jan. 2016 – Oct. 2016*

- Utilized R to collaborate on the construction of a linear mixed model (LMM) for GWAS with EB theory
- Spearheaded the initiative to apply the constructed EB-LMM model to haplotype-based GWAS, applied the model to the genome of 1092 subjects, and inferred 3 genes associated with the trait of interest
- Drafted a paper based on the study findings and published in the journal PLOS ONE

EXPERIENCES

Cepheid | Biostatistics Intern | Sunnyvale, CA *May 2018 – Aug. 2018*

- Implemented R/SAS to perform analyses for clinical trials (Cox PH models, contingency tables, etc.)
- Developed SAS Macros for multiple raw clinical datasets to streamline data cleaning and manipulation
- Summarized and presented the work to the Clinical Affairs group and suggested techniques to improve

Edenred–Accentiv' | Data Analyst Intern | Shanghai, China *Dec. 2016 – Feb. 2017*

- Wrote R, SAS and SQL scripts to automate the data cleaning and produce weekly/monthly sales report
- Collaborated on constructing a recommendation engine with clustering analysis, association rule mining
- Analyzed 1.86 million sales records, held discussions to troubleshoot the problems and contributed to future marketing strategies for different groups with different purchase habits

SKILLS

- Software: Python, R, SQL, SAS, SPSS, Latex, Perl, MS Office (Word, Excel, PowerPoint)
- Machine/Deep Learning: regression, kNN, random forest, SVM, clustering, neural network, NLP, CNN
- Python libraries: keras, tensorflow, nltk, numpy, pandas, sklearn, matplotlib, flask, etc.
- R packages: tidyverse (dplyr, ggplot2, tidyr, etc.), shiny, plotly, flexdashboard, etc.
- Language: English (Proficient), Chinese (Native)