YANGHONG GUO

2 years working experience and 10 years of Math/Statistics background.

 682-298-2152   yxg190031@utdallas.edu   yg2485.github.io   linkedin.com/in/yanghongguo

EDUCATION

University of Texas at Dallas — **Aug 2021 – Present**
*PhD in **Statistics**; Main Focus: Bayesian and Deep Leanring Methods in Biostatistics; GPA 3.97/4* — *Dallas, TX*

Columbia University — **Sep 2016 – May 2018**
*MA in **Statistics**; Core Courses: Statistical Machine Learning, Statistical Computation* — *New York, NY*

East China University of Science & Technology — **Sep 2012 – Jun 2016**
*BS in **Mathematics & Applied Mathematics*** — *Shanghai, China*

EXPERIENCE

Industry Researcher — **Jul 2019 – Aug 2020**
Bank of China — *Beijing, China*
- Analyzed bank customer service data by the NLP Semantic analysis system, extracted effective information by keyword
- Detected and predicted frequently happened customer issues, and generated over 30 reports

Bitcoin and Crypto Data Analyst — **Jul 2018 – Mar 2019**
Beijing Micai Investment Co., Ltd. — *Beijing, China*
- Implemented automatic quantitative trend-tracing trading strategy with Support Vector Regression(SVR)
- Applied Web-Crawler with Python to collect online STO data then saved to the database by MYSQL
- Generated bilingual research reports on the in-depth study of new-launched STO cryptocurrencies

Quantitative Analysis Intern — **Jun – Aug 2017**
China Merchants Securities Co., Ltd. — *Shenzhen, China*
- Designed investment strategies by applying LSTM to historical data, with indicators such as RSI and MACD
- Detected listed companies with default risk with XGBoost and reached an 80% backtesting accuracy

PROJECTS

Bayesian Integrative Model for Bulk RNA-Seq Data and Spatial Transcriptomics | *R, C++* — **June 2023**
- Applied advanced Bayesian methodologies for precise region detection within ST data, achieving an accuracy of 90%
- Leveraged the spatially-resolved regions from ST as a reference to deconvolute Bulk RNA-Seq data, successfully mapping gene expression to specific spatial domains
- Integrated clinical metadata with spatially-deconvoluted Bulk RNA-Seq insights, enabling a richer understanding of gene expression patterns in relation to patient outcomes.

Bayesian Dirichlet-Multinomial Model for Integration of Clinical and Single Cell Data | *R, C++* — **Sep 2022**
- Proposed a hierarchical Bayesian framework for the integrative analysis of cell-type abundance count data and covariates analysis
- Implemented Markov Chain Monte Carlo(MCMC) methods with R and C++
- Obtained the discriminating cell types that are aligned with the findings in the original papers, with new discoveries of the relationship between the abundance of cells and covariates at the cell-type level that were not studied before
- Presented poster on the Conference ABGOD 2023 and SRCOS 2023
    GitHub: Bayesian DM Regression Model for Integrative Analysis of Clinical and Single Cell Data

Potential ETC Customer Identification | *Python, TensorFlow, Scikit-learn* — **Nov 2019**
- Processed and cleaned the original customer dataset of 9 million samples with de-noise analysis
- Utilized data discretization methods to further reduce batch size and model complexity
- Refined the data by Grid Search and Ensemble Generation and achieved an AUC around 0.85 when targeting a potential ETC Customer (rank 10/200)

SKILLS

- **Languages:** Python, MATLAB, C++, Latex, HTML
- **Data Analysis:** R, SAS, MySQL, SQL Server
- **Version Control:** GitHub, Git, Google Colab
- **Machine Learning:** Tensorflow, Scikit-learn, Pytorch