

YUNING PAN

✉ pyuning@bu.edu ☎ (857)-272-3640 🏠 Open to Relocation

Education

Boston University

Ph.D. in Statistics, GPA: 3.9/4.0

Advised by Eric D. Kolaczyk with **research interest** in Network Analysis, Machine Learning, Deep Learning

Sep. 2019 – May 2025 (expected)

Boston, USA

Boston University

M.A. in Statistics, GPA: 3.9/4.0

Sep. 2017 – Jan. 2019

Boston, USA

Hefei University of Technology

B.S. in Mathematics, GPA: 3.6/4.0

Sep. 2012 – Jun. 2016

Hefei, China

Technical Skills

Programming Skills: Python, R, Matlab, MySQL, LaTeX

Machine Learning Toolkits: Pytorch, Tensorflow, DGL, Keras, Scikit-learn, H2O, Stan, Pandas, Numpy, SciPy

Professional Experiences

Discover Financial Services

PhD Data Scientist Intern in Risk Modeling Team

June 2024 – Aug 2024

Chicago, USA

- Quickly mastered domain knowledge and developed interpretable machine learning models in Python adhering to regulatory constraints to support the underwriting process for one of Discover's major products — Personal Loans.
- Redesigned backward feature elimination, reducing runtime from **6 hours** to **30 minutes** while enhancing generalization with out-of-bag AUC on the entire dataset. Also implemented Bayesian optimization for hyper-parameter tuning, improving accuracy and efficiency by optimizing cross-validated AUC over a single split random grid search.
- Implemented optimization with customized business metrics, enhancing model interpretability and decision-making. Developed an automated, object-oriented pipeline for data cleaning, modeling, and visualization, eliminating manual tasks and ensuring resilience against kernel interruptions.

Boston University

Lead Statistical Consultant in Boston University Consulting Center

Sep. 2021 – May 2022

Boston, USA

- Supervised 5 teams (17 master's students), completing 25+ consulting projects for university-affiliated researchers.
- Assisted clients with experimental and survey design, providing data analyses to inform decisions, e.g. help to make decision to reduce redundant checkups to reduce patient costs at Maine Medical Center.
- Built reproducible and repeatable data analysis pipelines in R for clients' future use post-consultation.

Deloitte Touche Tohmatsu Limited

Intern in Data Analytic

June 2018 - Aug. 2018

Beijing, China

- Extracted and analyzed insurance data from the People's Insurance Company of China (PICC) database, efficiently computing financial descriptors and constructing working tables in R.
- Supported PICC's IPO qualification assessment through fluctuation analysis, including variations in long-term prepaid claim amounts, using computed numerical descriptors.

Research Experiences

AI Guided Design and Synthesis of Semiconducting Molecules (NSF Award 2141384) **Feb. 2021 – May 2024**

- Collaborated with BU chemistry department to propose the candidates for the next generation semiconducting molecules: predicting the wavelength of molecules based on a small and heterogeneous experimental organic database.
- Designed ensemble learning frameworks embedded with XGBoost weak learners for accurate and stable predictions; feature engineering on molecular structures to reduce density functional theory calculation time from **1300** to **300** hours; utilized the accumulated local effects in the context of ensemble frameworks for interpretation of correlated features.
- The poster was accepted by *Molecular Machine Learning Conference(MoML2023)*, Montreal QC, May 2023.
- The paper was accepted by and published on *The Journal of Physical Chemistry*, July 2024.

Bayesian Modeling Framework for Network-based Data

May 2022 – Present

- Developed convolutional network that utilizes both node-level and graph-level features and implemented with Pytorch, and achieved higher prediction accuracy compared to simple MLP and GCN (decrease prediction errors by **20%**).
- Designed a Bayesian framework to model the network labels with graph-based Gaussian process architecture that significantly reduces the sampling time of posterior distributions and implemented with Stan in R and Python.
- Collaborated with University of North Carolina, Chapel Hill to implement NetOTC algorithm and develop research software package, decreased the running time by **250%**. Paper to be submitted to *Journal of Open Source Software*.