

Yifan Li

444 Washington Blvd, Jersey City, NJ, 07310

☎ (+1) 608-216-5993 | ✉ ivanlee142857@gmail.com | 🌐 posquit0

Education

University of Connecticut

PHD IN STATISTICS

Connecticut, USA

Sep 2018 - Sep 2023

- **Coursework:** Analysis of Survival Data, Bayesian Data Analysis, Computational Method for Optimization, Financial Data Mining, Bayesian Decision, Applied Multivariate Analysis, Linear Statistical Model

University of Wisconsin - Madison

MASTER IN STATISTICS

Wisconsin, USA

Sep 2016 - May 2018

- **Major GPA:** 3.87/4; **Overall GPA:** 3.77/4
- **Coursework:** Survival Analysis, Stochastic Modeling, Classification and Regression Tree, Statistical Method, Mathematical Statistics, Machine Learning, Multilevel Models, Design of Experiments

Nanjing University

BACHELOR IN STATISTICS

Jiangsu, China

Sep 2013 - Jun 2017

- **Coursework:** Mathematical Analysis, Higher Algebra, Discrete Mathematics, Ordinary Differential Equation, Partial Differential Equation, Function of Complex Variable, Stochastic Process, Real Analysis
- **Award:** Awarded People Scholarship

Work Experience

Quantitative Trading Book in Ernst & Young U.S. LLP

SENIOR CONSULTANT

New York, USA

Oct 2023 - Present

- Derivatives Pricing Algorithm Project
 - Led the modular redesign of derivatives pricing algorithm by decomposing it into service class and analysis units.
 - This architecture ensured high decoupling of coding, enabling independent updates to each component without affecting the overall system, significantly reducing redundancy and enhancing maintainability.
 - Designed and implemented robust unit testing frameworks, improving system debug reliability by proactively identifying potential errors.
- Equity Derivatives Pricing Algorithm Enhancement
 - Improved the pricing algorithm of equity derivatives by shifting from a market-based risk model to an underlying location-based risk analysis, enabling a more accurate and interpretable pricing framework.
 - Combine advanced machine learning techniques, such as LSTM, random forest models with traditional MCMC methods to price derivatives, enabling the pricing of complex toxic options with more than three underlying.
- Counterparty Credit Risk Monitoring Using SFT VaR-Based Models
 - Employed SFT VaR-based models to calculate and monitor Counterparty Credit Risk.
 - Interpreted complex data and model results to deliver clear insights to stakeholders, including cross-disciplinary teams and non-technical audiences.
 - Continuously updated model parameters in line with evolving market data, ensuring the models reflect current market conditions and deliver accurate risk assessments.
- Optimization of American Options Pricing
 - Applied the American Monte Carlo (AMC) method to price American options, replacing the computationally intensive Monte Carlo over Monte Carlo method. This optimization reduced the computational complexity from $O(n^2)$ to $O(n)$, significantly improving pricing speed and saving considerable computational resources.

Bank of China International Holdings Limited

SECURITIES ANALYST ASSISTANT (INTERN)

Shanghai, China

Jun 2021-Sep 2021

- Focused on battery and new energy industry. Predicted the short- and long-term performance of stocks of related companies based on time series model with a spike-and-slab error.
- Adjusted the prediction under a multinomial model based on the performance of correlated companies and avoided making an over-optimistic forecast compared with previous model.

HUATAI SECURITIES CO., LTD. 华泰证券

DATA ANALYST (INTERN)

Jiangsu, China

Jul 2017-Sep 2017

- Unsupervised screened visitors with a strong desire to buy products based on their records on company's APP.
- Cleaned and reshaped the 17 million visitor records by summarizing operations from the same visitor.
- Extracted useful variables by PCA (principal component analysis) method.
- Divided visitors into five groups by K-means methods and assigned visitors labels by their group.
- Fitted a decision tree with labeled data which could tag new visitor within 20 seconds while the target is 1 min.

- Credit Card Approval with Unbalanced Data and Outliers
 - Decide who to approve or decline for credit based on historical repayment records.
 - Adding new missing indicator variables before applying imputing missing value after checking randomness.
 - Generate features based on the distribution of outliers and assign different weights on unbalanced responses.
 - Fit logistic regression, XGBoost, and Random Forest models separately and use the linear combination of three models as final model after cross validation.
- Yelp Reviews Rating Prediction
 - Predicted Yelp reviews' rating on 1 million unlabeled text reviews.
 - Cleaned 1.5 million Yelp reviews by removing un-English comments, abbreviations, and spelling mistakes.
 - Extracted positive/negative words based on their relative frequency in differently rated reviews to avoid placing too much weight on everyday words like "the", "a" which can be mistaken as positive words.
 - Transfer text reviews into vectors by Sentence-To-Vector and generate new features from positive/negative words.
 - Fitted pre-processed data by Long-Short-Term-Memory (LSTM) neural network and achieved 0.6 root-mean-square-error.

Thesis

Item-Response-Theory Model with Power Parameter Adjusted for Unbalanced Data

Connecticut, USA

- Estimated individual's ability and item's difficulty based on their performances on several.
- Adapted logistical regression model by Item-Response-Theory model with a power parameter which can control the skewness of link function.
- Combined Sliced sampling and Gibbs sampling method (MCMC) to get estimations of interested variables.
- Reduced the prediction error by half compared with normal logistical regression model.

Joint Model of Item Response and Response Time with Dirichlet Process Prior

Connecticut, USA

- Estimated individual's ability based on both item response (IR) and response time (RT).
- Fitted separate logistic and linear regression for IR and RT. Combined them with a nonparametric Dirichlet Process prior on individual's ability which get rid of normality assumption of variables.
- Estimated individual's ability by Hamiltonian Monte Carlo and clustered individuals by patterns from Dirichlet Process.

Joint Model of Longitudinal Item Response and Survival Time

Connecticut, USA

- Examined trend of individual's ability over time and their effects on response time.
- Individual's ability was taken as longitudinal and estimated by forward and backward forecasting method.
- Response time was fitted as a Cox proportional hazards model through partial likelihood method which is a semiparametric approach.
- All unknown parameters are estimated by stochastic gradient descent algorithm.

Skill

Language: Mandarin Chinese (Native), English

Coding/Database Languages: Master R, Python, GitHub, Latex, Nimble, JUGS and HPC, familiar with SQL, SAS, MATLAB, C++ and Julia.

Certificate: CFA level 1