Yifan **Li**

444 Washington Blve, Jersey City, NJ, 07310

(+1) 608-216-5993 | ivanlee142857@gmail.com | yifan-li.com | yifan-lee

Education

University of Connecticut — *Connecticut, USA*
PhD in Statistics — *Sep 2018 - Sep 2023*

University of Wisconsin - Madison — *Wisconsin, USA*
Master in Statistics — *Sep 2016 - May 2018*

Nanjing University — *Jiangsu, China*
Bachelor in Statistics — *Sep 2013 - Jun 2017*

Work Experience

Quantitative Trading Book in Ernst & Young U.S. LLP — *New York, USA*
Senior Consultant — *Oct 2023 - Present*

- **Applied Machine Learning Projects in Generative AI Context**
  - **Assistant AI System**: Built a retrieval-augmented generation (**RAG**) system combining semantic text embeddings and file-based context retrieval to produce accurate and context-aware responses.
  - **Item Blurring Pipeline**: Built a two-stage object detection and classification framework with a Region Proposal Network and classifier to automatically blur specified items in images.
  - **Harmful Content Detection**: Designed a multi-task classification model with early fusion of text and image features to detect potentially harmful or policy-violating email content.
  - **Ad Click Prediction**: Constructed a personalized advertising model using Gradient Boosting Decision Trees (**GBDT**) and **DeepFM**, improving ad relevance and user engagement.
- Modular Redesign of Derivatives Pricing Algorithm
  - Led the architectural overhaul by decomposing the algorithm into service class and analysis units, archieving high **decoupling** of code.
  - Enabling independent updates to each component without affecting the overall system, significantly reducing redundancy and enhancing maintainability.
  - Designed robust unit testing frameworks, improving system **debug reliability** by proactively identifying potential errors.
- Optimization of American Options Pricing
  - Applied the American Monte Carlo (**AMC**) method to price American options, replacing the original Monte Carlo over Monte Carlo method.
  - Achieved a substantial reduction in computational complexity from $O(n^2)$ to **O(n)**, cutting pricing time and saving considerable resources.
- Equity Derivatives Pricing Algorithm Enhancement
  - Improved the pricing framework for equity derivatives by transitioning from a market-based risk model to an underlying location-based risk analysis, enhancing accuracy and **interpretablity**.
  - Intergrated 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
  - Employed SFT VaR-based models to calculate and monitor Counterparty Credit Risk.
  - **Interpreted** complex data and model results, and delivered clear insights to stakeholders, including cross-disciplinary teams and **non-technical** audiences.
  - Regularly updated model parameters in line with evolving market data, ensuring the models reflect current market conditions and deliver accurate risk assessments.

Bank of China International Holdings Limited — *Shanghai, China*
Securities Analyst Assistant (intern) — *Jun 2021-Sep 2021*

- Predicted the short- and long-term performance of new energy industry equity 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, avoiding an over-optimistic forecast.

HUATAI SECURITIES CO., LTD. (HTSC) — *Jiangsu, China*
Data Analyst (intern) — *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, and grouped them by **K-means**.
- Extracted useful variables by principal component analysis (**PCA**) method used in decision tree to tag visitor in 20s while the target is 1 min.

Thesis

Enhanceing Flexibility and Efficiency Item-Response-Theory Model — *Connecticut, USA*

- Developed innovative extensions to Item-Response-Theiry (IRT) models by incorporating flexible power link functions and random effects to address traditional limitations, such as skewness and parameter constraints.
- Integrated item response and response time into a joint model using a nonparametric Dirichlet Process prior, removing normality assumptions.
- Designed a joint model to examine the dynamics of individual abilities and their impact on response times using forward and backward forecasting methods. Modeled response times through a Cox proportional hazards framework with semiparametric partial likelihood estimation.

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