Yiming Li

(857)763-9407 | yiming77@bu.edu | Boston, MA

EDUCATION

Peking University, School of Mathematical Science Beijing, China
Visiting Scholar Sep, 2023 – Present

Boston University Boston, USA
Master of Science in Applied Business Analytics Sep, 2021 – Jan, 2023

Northwest A&F University Xi'an, China
Bachelor of Science in Information and Computation Science Sep, 2015 – Jun, 2019

RESEARCH EXPERIENCE

Programmatic Calculations of Enumerative Geometry Sep, 2023 – Present

- Explored the asymptotic behavior of Gromov-Witten invariants.
- Built machine learning models, such as artificial neural networks, support vector machines, etc., to establish strict asymptotic theory.

Bank Customer Loan Risk Analysis Sep, 2022 – Dec, 2022

- Prepared bank loan application data set by data cleaning, transformation, validation, and visualization.
- Built confusion matrix to reduce the correlations between features.
- Developed 7 models, including logistic, decision tree, random forest, linear SVM, and etc. to establish the index system that contribute individual repayment ability the most.
- Measured and accuracy of different models, and selected the best performing logistic regression model to build customer loan risk analysis system.

Airbnb Rental Price Sep, 2022 – Dec, 2022

- Analyzed and Predicted Airbnb rental price by developing multiple linear regression and clustering models to improve customer experience, increase revenue and provide marketing strategy.
- Successfully increased customer conversion rates by 13.7% and boosted revenue by 17.6%.

New York City Real Estate Price Analysis Jan, 2022 – May, 2022

- Collected and manipulated 10,000+ transaction data to gain characteristics and clarify area distribution.
- Developed 3 time series models to forecast NYC real estate price in the next 2 years.
- Developed optimization model to adjust the commission and office area to gain the maximum profit.

Book Recommendation System Based on PageRank Algorithm Jun, 2018 – Jun, 2019

- Designed and executed PageRank algorithm to figure influence for Book Search Recommendation.
- Calculated PageRank score for every book and category of book using eigenvalue calculations in Python.
- Evaluated accuracy of new Book Recommendation Rank.
- Identified student requirement to bring technologies to current Book Recommendation System in library.
- Performed statistical analysis on 100,000+ book borrowing data using SQL and Tableau.

Statistical Research on Regional Poverty Alleviation Work May, 2017 – Apr, 2018

Team Leader

- Developed scale analysis using reliability coefficient and correlation analysis to ensure questionnaire more representative.
- Developed factor analysis to extract 5 common factors to build an independent index system to evaluate the performance of poverty alleviation work.
- Developed clustering analysis to divide surveyed counties into 6 groups.
- Researched and compared the indicators of the 6 groups of counties, and gave insights for poverty alleviation.

WORK EXPERIENCE

Margik, Inc Boston, USA

Data Analyst Feb, 2023 – Aug, 2023

- Conducted in-depth data analysis of the drone industry, lighting distribution and manufacture, and economic development councils to identify potential customers for the company's cutting-edge LED technology.
- implemented predictive models, including time series, logistic regression, and clustering models to forecast market trends and assess potential risks.
- Developed algorithm for market trend analysis, increased 20% the accuracy of market predictions.

Shenzhen Skyense Intelligent Co., Ltd Shenzhen, China

Data Scientist Aug, 2020 – Jun, 2021

- Collaborated with Business Operations and the leadership team to analyze 7000+ data and conduct strategic business analysis. Identified potential business opportunities and trends, and supported marketing strategy making decision.
- Established metrics and analytics to track and report all key performance indicators against annual, quarterly and monthly goals.
- Developed and maintained management dashboards, regional reports, and executive review documents that will drive business outcomes.
- Monitored existing KPI metrics, developing new metrics, identity correlations and root causes.
- Design and build visualization dashboards to accelerate information-to-action at scale.
- Effectively communicated findings and solutions for stakeholders and leadership. Incorporated on-going feedback to existing analysis to deliver highly reliable and accurate reports.

Industrial and Commercial Bank of China Hebei, China

Software Development Engineer Aug, 2019 – Jul, 2020

- Collaborated with cross-functional teams to gain in-depth understanding of credit risk prediction and management, and business strategy and service.
- Conducted data integration and data filling to establish baseline metrics and data mart with key important indicators.
- Developed 3 logistic regression models based on 4 datasets to analyze credit risk for corporate customers.
- Measured model quality in accuracy, discriminatory power and stability using out-of-sample data. Determined the best model, which can identify 70% of customers with high credit risk.
- Applied model to identify high-risk customers, adjusted lending strategies and thus reduced 40% lending risk.

SCIENTIFIC COMPETITIONS

Chinese Undergraduate Mathematical Contest in Modeling, Second Prize Sep, 2017

- Group leader, responsible for problem-solving ideas, R and MATLAB modeling of the pricing issue of photo tasks in a labor crowdsourcing app.
- Implemented principal component analysis (PCA) to identify critical variables affecting task completion.
- Built logistic regression model to clarify the correlation between task price and completion.
- Incorporated a categorical indicator, member credit rating system to optimize pricing strategies and increase task completion rate by 37.2%.

HONORS & AWARDS

Outstanding Bachelor's Thesis (Top 5%) Jun, 2019

First Class Academic Scholarship (Top 5%) Jun, 2016

SKILLS & TOOLS

- Computer skills: Python, R, SQL, Tableau, SPSS, MATLAB, SageMath
- Languages: Chinese (native), English (fluent)

Statement of Purpose

My fascination with **computer science** stems from its logical rigor and transformative potential in shaping the future of technology. Through my academic journey and professional experiences, I have consistently gravitated toward solving computational challenges, particularly in the realm of **machine learning**. My desire to pursue a **PhD in computer science at Columbia University** is driven by my deep interest in the emerging area of **Explainability and Transparency in Machine Learning**. I am eager to advance my theoretical understanding and contribute to cutting-edge research that addresses these vital aspects of AI.

I earned my **Bachelor's degree in Information and Computational Science** from **Northwest A&F University**, where I explored the application of computational techniques to real-world problems. Leading a team, I utilized **statistical and machine learning methods** to assess poverty alleviation efforts in economically underdeveloped areas. I also introduced the **PageRank algorithm** to the university's library system, predicting user preferences based on borrowing patterns. These experiences sparked my interest in **machine learning's societal applications** and revealed the challenges of interpreting model outputs, solidifying my long-term goal of focusing on model interpretability.

My academic journey continued at **Boston University**, where I obtained a **Master's degree in Applied Business Analytics**. This program provided me with a solid theoretical foundation in **statistical modeling, machine learning**, and **data analysis**. Through research projects like

predicting **Airbnb rental prices** and analyzing **bank customer lending risks**, I experienced firsthand the tension between **predictive accuracy and model interpretability**. This deepened my understanding of **explainable AI**, reinforcing my interest in ensuring that ML models not only perform well but also generate understandable explanations for their decisions.

Professionally, I refined my technical skills as a **data scientist at Shenzhen Skyense Intelligent Co. Ltd.** and later as a **data analyst at Margik Inc.** In both roles, I frequently built **predictive models for strategic decision-making**. However, I also faced challenges where stakeholders questioned the reasoning behind model predictions, underscoring the importance of **transparency in machine learning**. These experiences further confirmed my commitment to addressing the need for interpretable AI, especially in **high-stakes environments** where decisions must be transparent and trustworthy.

At **Peking University**, my research delved deeper into the intersection of **machine learning models and their interpretability**. I explored **algebraic geometry** and used ML models to investigate the **asymptotic behavior of Gromov-Witten invariants**, expanding the scope of existing theoretical derivations. This research helped me realize the significant **trade-offs between model complexity and interpretability**, a challenge I aim to address during my PhD studies. I hope to design novel techniques that enhance explainability without compromising performance, particularly in **sensitive domains like healthcare, finance, and autonomous systems**.

Columbia University's PhD program in Computer Science offers the perfect environment for me to pursue my research aspirations. I am especially drawn to the **world-class faculty** and the university's focus on **interdisciplinary research**, which aligns with my interest in applying ML models across various fields. The opportunity to collaborate with researchers who are equally passionate about **explainability in AI** excites me, and I am confident that Columbia will provide the tools and support I need to develop transparent and trustworthy machine learning systems.

In addition, I would like to address a key decision in my academic trajectory. After working as a research assistant in mathematics for a year, I realized that a PhD in computer science would better align with my passion for solving real-world computational challenges. This realization strengthened my commitment to pursuing a PhD in CS, and I believe **Columbia's program** will provide the perfect platform to channel my expertise and ambitions.

In conclusion, my **academic background, professional experience, and research** have prepared me to tackle the critical challenges of **explainability in machine learning**. I am excited about the possibility of contributing to **Columbia University's vibrant research community** and pushing the boundaries of AI to make it more interpretable, accountable, and trustworthy.