

# BRIAN LIU

briliu@mit.edu | brianliu97.github.io

## EDUCATION

---

**Massachusetts Institute of Technology**  
Ph.D. Candidate | Operations Research Center  
Advised by Professor Rahul Mazumder

September 2021 - Present

**Cornell University**  
B.S. in Operations Research | GPA: 4.1  
Summa Cum Laude | Omega Rho Honor Society

May 2020

## INTERESTS

---

**Topics** | Ensemble learning, feature selection, model compression, interpretable machine learning, stability

**Applications** | Clinical machine learning, pandemic modeling, healthcare operations

## PUBLICATIONS

---

Refereed Conferences and Journals:

1. **B. Liu** and R. Mazumder. Fast: An Optimization Framework for Fast Additive Segmentation, In ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2024, 2402.12630
2. **B. Liu** and R. Mazumder. Fire: An optimization framework for fast interpretable rule extraction. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2023, 2107.00219
3. **B. Liu** and R. Mazumder. ForestPrune: Compact depth-pruned tree ensembles. In Proceedings of the 26<sup>th</sup> International Conference on Artificial Intelligence and Statistics (AISTATS) 2023,
4. **B. Liu**, M. Xie, and M. Udell. ControlBurn: Feature selection by sparse forests. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2021, 2107.00219
5. **B. Liu\***, Y. Zhang\*, S. Henderson, D. Shmoys, P. Frazier. Modeling the risk of in-person instruction during the COVID-19 pandemic, To Appear in INFORMS Journal of Applied Analytics, 2024
6. P. Frazier, J. M. Cashore, N. Duan, S. Henderson, A. Janmohamed, **B. Liu**, D. Shmoys, J. Wan, Y. Zhang. Modeling for COVID-19 College Reopening Decisions: Cornell, A Case Study. *Proceedings of the National Academy of Sciences*.

Under Review:

1. **B. Liu** and R. Mazumder. Randomization Can Reduce Both Bias and Variance: A Case Study in Random Forests, 2024, 2402.12668

Preprints and Technical Reports:

1. **B. Liu** and M. Udell. Impact of accuracy on model interpretations, 2020, 2011.09903

## TEACHING

---

**Teaching Assistant (MIT)**  
15.075 Statistical Thinking and Data Analysis (Undergraduate)

Spring 2024

**Teaching Assistant (MIT)**  
15.072 Advanced Analytics Edge (Graduate MBAn)

Fall 2023

**Teaching Assistant (MIT)**  
15.067 Engineering Statistics and Data Science (Graduate LGO)

Summer 2023

**Teaching Assistant (Cornell)**  
ORIE 3300: Optimization I and ORIE 4740: Introduction to Statistical Learning.

Fall 2018 & Spring 2020

## INDUSTRY EXPERIENCE

---

**Graduate Student Researcher****July 2022 – Present**

Takeda | Cambridge, MA

Clinical machine learning project to improve eosinophilic esophagitis (EoE) diagnosis. Built survival models to improve the timely diagnosis of EoE and machine learning models to improve differential diagnosis.

**Graduate Student Researcher****September 2021 – February 2022**

Hartford Healthcare | Hartford, CT

Analyzed patient and provider data to determine the impact of socio-economic factors on disparities in diabetes management.

**Data and Applied Scientist****July 2020 – July 2021**

Microsoft (Bing Ads) | Bellevue, WA

Built and maintained machine learning models to forecast advertising account revenue and churn. Analyzed Bing demographic and interest data to optimize audience targeting for online search ads.

**Data and Applied Science Intern****Summer 2019**

Microsoft (Devices) | Redmond, WA

Built machine learning models with >80% accuracy to forecast returns for Surface commercial customers. Designed an inventory management plan to optimize the safety stock of spare parts in the supply chain.

**Data Science Intern****Summer 2018 & Summer 2017**

Tesla (Global Service Operations) | Fremont, CA

Built machine learning models to predict customer satisfaction with vehicle repairs. Used interpretability tools such as LIME and SHAP to identify the key drivers of customer satisfaction.

**OTHER RESEARCH**

---

**Cornell Pandemic Modeling Group** | D. Shmoys, P. Frazier, S. Henderson**March 2020 – December 2022**

Modeled classroom transmission using Monte Carlo simulation to assess student and instructor risk. Analyzed the impact of university reopening, weather, and compliance to shutdowns on pandemic spread. Presented various data analyses and visualizations to Cornell senior leadership to support decisions on COVID-19 interventions. (<https://datasciencecenter.cornell.edu/covid-19-modeling/>)

**Cornell Pandemic Reopening Planning** | David Shmoys**March 2020 – September 2020**

Modeled student movement and behavior on campus using mobility and university internal data. Leveraged models to identify congested sidewalks and bus routes and predict potential hotspots for transmission. Tested various de-densification strategies using agent-based simulations and submitted the most effective methods to the provost. Assisted scheduling socially distanced classrooms. (<https://www.orie.cornell.edu/spotlights/unsung-engineering-behind-cornells-fall-2020-schedule>)

**Public Transit Route Optimization** | David Shmoys**August 2019 – March 2020**

Developed probabilistic models using farebox and vehicle location data to estimate rider demand on bus routes in Tompkins County. Used demand estimates to redesign the county's second busiest bus route to improve efficiency. Changes were implemented by Tompkins County Area Transit (TCAT) in Fall 2020.

**TECHNICAL SKILLS**

---

Python, Julia, R, SQL, Spark, Databricks, AMPL, JuMP, Gurobi, Tableau, LaTeX

**COURSEWORK**

---

**Machine Learning and Statistics** | Statistical Machine Learning, Statistical Learning Theory, Non  
**Operations Research** | Simulation Modeling and Analysis, Financial Engineering, Optimization  
**Math** | Nonlinear Optimization, Real Analysis, Linear Algebra, Stochastic Processes, Probability  
**Computer Science** | Database Systems, Object Oriented Programming, Software Development

**SERVICE**

---

**ORC IAP Seminar Coordinator****Spring 2024****ORC Student Seminar Coordinator****Fall 2023**