linkedin.com/in/taicai
github.com/caikitlearn
hackerrank.com/caikitlearn

– Tai Cai, PhD –

caikitlearn.github.io
Canadian Citizen | US Permanent Resident

Updated 05/2024
tcai@mail.harvard.edu
(857) 998-8619

Summary

Profile · Quantitative scientist with PhD + 7 years of tech industry experience in leveraging applied statistics, machine learning, and software engineering to influence high-level decision-making

Expertise · Bayesian statistics (probabilistic programming, causal inference, hierarchical models, nonparametric methods)
· Classical statistics (statistical learning, experimental design, mixed models, time series)
· Machine learning (tree-based models, deep learning, theory)
· Cross-functional collaboration to develop and productionize data-driven products and insights

Toolkit · Python, SQL, R, jupyter, Vim

Experience

Data Scientist, Planner Evaluation — ***Waymo**, New York, NY* 11/2022–Present

· Halved the family-wise error rate of experiments used to evaluate onboard software by successfully launching confidence interval adjustments in collaboration with front-end and back-end engineering teams
· Pioneered the use of statistical power and minimum detectable effect throughout the broader Planner Evaluation team by authoring the API and writing supporting documentation
· Provided statistical mentorship for teams across Waymo on topics such as resampling methods, type I/II error, and quantifying uncertainty; promoted better statistical fluency across Planner Evaluation by teaching a short course

Research Data Scientist, Quantitative Engineering — ***Meta**, New York, NY* 02/2021–11/2022

· Collaborated (with engineering, product, design, and UX research) as the scientific point of contact in a 100 person team to build, improve, and migrate users to infrastructure tools, doubling user engagement in the process
· Served as the technical lead for the data team through multiple product and feature launches by providing mentorship, planning, and stakeholder management
· Guided engineers to optimize the design of experiments intended to reduce latency in Core Infra services, yielding more principled decision-making

Applied Scientist, Global Intelligence — ***Uber**, San Francisco, CA* 01/2020–02/2021

· Estimated point and interval values of competitive intelligence metrics used in executive decision-making
· Engineered an ensemble of machine learning models using internal and third-party data to identify rare user groups for multi-million dollar experiments; improved the legacy model by doubling precision

Data Scientist, Data and Research — ***BitSight**, Boston, MA* 10/2017–01/2020

· Led all data science projects for third-party cybersecurity risk management, including BitSight's inaugural machine learning-based product, multiple analytical tools, and observational studies to evaluate internal processes
· Supervised Master's-level intern working on validation of external financial data and sales/marketing analytics

Technical Advisor — ***Insight Data Science**, Boston, MA* 06/2019–09/2019

· Mentored data science projects for PhD graduates and postdoctoral fellows by providing weekly 1-on-1 feedback

Education

PhD, Biostatistics — ***Harvard University**, Cambridge, MA* 08/2012–09/2017

· Conducted and published statistical research on applied problems in genetics, health care policy, and end-of-life care
· Developed novel methods and R code for handling confounding, sampling bias, misclassified and/or correlated outcomes, intractable cross-validation, and hierarchical structures
· Taught graduate-level classes in Bayesian statistics, statistical methods, and introductory statistics/epidemiology

BMath, Statistics and Actuarial Science — ***University of Waterloo**, Waterloo, ON, Canada* 09/2007–06/2012

· René Descartes National Scholar; Graduated with Distinction - Dean's Honours List (Highest Honours)
· Associate of the Society of Actuaries designation completed 03/2012
· Six work semesters at Logitech, Hewitt Associates, Manulife Financial, and Munich Re