

 [linkedin.com/in/taicai](https://www.linkedin.com/in/taicai)


 github.com/caikitlearn

 [hackerrank.com/caikitlearn](https://www.hackerrank.com/caikitlearn)


– TAI CAI, PHD –

caikitlearn.com

Canadian Citizen | US Permanent Resident

Updated 05/2020 

tcai@mail.harvard.edu 

(857) 998-8619 

SUMMARY

Profile	· Quantitative scientist working at the intersection of statistics/machine learning, research science, product management, and data/software engineering
Areas of Expertise	· Cross-functional collaboration to productionize/engineer data-driven products and features · Bayesian statistics (causal inference, hierarchical models, model selection, nonparametric methods) · Classical statistics (statistical learning, experimental design, mixed models, time series) · Deep learning (artificial/recurrent/convolutional neural networks)
Frequently Used	· Python, Tensorflow, Keras, LightGBM, R, SQL, Spark, jupyter, Vim, \LaTeX
Sometimes Used	· Java, C++, HTML, CSS, Javascript

EXPERIENCE

Data Scientist II , Uber	<i>San Francisco, CA</i>	01/2020–Present
· Engineered an ensemble of predictive models using internal and third-party data to identify rare user groups for multi-million dollar experiments; improved legacy model performance by a factor of 2 · Analyzed the aforementioned experiments to identify statistically significant metrics, and dissected the limitations of the existing experimental design to improve future allocation of resources · Estimated point and interval values of competitive intelligence metrics used in executive decision making · Identified and corrected multiple errors and inefficiencies found in the team repository, ranging from production code to deliverables used by stakeholders		
Senior Data Scientist , BitSight	<i>Boston, MA</i>	10/2017–01/2020
· Led all data science projects for third-party cybersecurity risk management, including: · Tier Recommender, BitSight's first machine learning based product, which leverages unique network data · Enterprise Analytics, a product that computes and evaluates hypothetical risk scenarios for corporations · Fourth-Party Risk Management, a tool that provides visibility into business ecosystems · Designed observational studies to evaluate the association and causality of relationships pertaining to efficacy of products, impact of extraneous events, and influence of internal interventions · Supervised Master's-level intern working on 1) validation of external financial data and 2) sales and marketing analytics · Organized reading groups covering advanced topics in forecasting, prediction intervals, and model evaluation		
Technical Advisor, Data Science Fellow , Insight Data Science	<i>Boston, MA</i>	05/2017–09/2019
· As Technical Advisor: Mentored data science projects for PhD graduates and postdoctoral fellows by providing weekly 1-on-1 feedback on project ideation/viability, data considerations, modeling techniques, and communication skills · As Data Science Fellow: Consolidated and cleaned multiple data sources to tally millions of labeled data points to predict and visualize real-time supply and demand at 200 bike-sharing stations		

EDUCATION

PhD, Biostatistics , Harvard University	<i>Cambridge, MA</i>	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 sampling bias, misclassified outcomes, correlated outcomes, intractable cross-validation, hierarchical structures, and confounding · Taught graduate-level classes in introductory statistics/epidemiology, statistical methods, and Bayesian statistics		
BMATH, Statistics and Actuarial Science , University of Waterloo	<i>Waterloo, ON, Canada</i>	2012
· Graduated with Distinction – Dean's Honours List (Highest Honors) · Completed six work semesters at Logitech, Hewitt Associates (now Blackstone), Manulife Financial, and Munich Re		

PERSONAL PROJECTS

Stephen A.I. Smith	· Twitter bot trained on Stephen A. Smith's tweets, using transfer learning via Open AI's GPT-2 model, GloVe Twitter embeddings, and a Bidirectional LSTM model for scoring generated tweets
CRISPR-be-deep	· Exploration of gene editing data using convolutional and recurrent neural networks