

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


 [github.com/caikitlearn](https://github.com/caikitlearn)

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


# – TIANYI “TAI” CAI, PhD –

Canadian Citizen with H-1B

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

Updated 10/2019 

[tcai@mail.harvard.edu](mailto:tcai@mail.harvard.edu) 

(857) 998-8619 

## SUMMARY

<b>Profile</b>	· Experienced Data Scientist and Statistician working at the intersection of applied statistics, machine learning, research science, software engineering, and product management
<b>Areas of Expertise</b>	· Data-driven software products and features put in production via cross-functional collaboration · Applied machine learning (data preparation, classification, regression, clustering, visualizations) · Bayesian statistics (causal inference, hierarchical models, model selection, nonparametric methods) · Classical statistics (experimentation, statistical learning, mixed models, time series)
<b>Current interests</b>	· Statistical properties of machine learning methods, causal inference, ProjectEuler+
<b>Frequently used</b>	· Python, R, SQL, Spark, Jupyter Notebook, Amazon S3, Amazon EC2, Vim, $\LaTeX$
<b>Used as necessary</b>	· Tensorflow, Keras, Java, C++, HTML, CSS, Javascript

## EXPERIENCE

<b>Senior Data Scientist, BitSight</b>	<i>Boston MA</i>	10/2017–Present
· Lead all data science projects for third-party cybersecurity risk management through designing metrics, researching methods, building prototypes, writing and reviewing production code, and communicating results to stakeholders and customers · Ship new software products and features (including 1 patent pending), working as the sole data scientist in collaboration with engineering, product, customer success, marketing, and sales teams · Design observational studies to evaluate the association and causality of relationships pertaining to efficacy of products, impact of extraneous events, and influence of internal interventions · Organize reading groups covering advanced topics in forecasting methods, prediction intervals, and model evaluation metrics · Supervised Master’s-level intern working on 1) validation of external financial data and 2) sales and marketing analytics · Promoted from Data Scientist on 02/2019 for exceeding expectations across all projects		
<b>Technical Advisor, Alumni Mentor, Data Science Fellow, Insight Data Science</b>	<i>Boston MA</i>	05/2017–09/2019
· As Technical Advisor: Mentored data science projects for 4 PhD graduates and postdoctoral fellows by providing weekly 1-on-1 feedback on project ideation/viability, data considerations, modeling techniques, and communication skills · As Alumni Mentor: Administered mock interviews covering computer science, machine learning, statistics, and case studies · 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 for Hubway bikes at 200 bike-sharing stations		
<b>PhD Student, Research Assistant, Harvard University</b>	<i>Cambridge MA</i>	08/2012–09/2017
· Conducted original statistical research on applied problems in genetics, health care policy, and end-of-life care · Developed novel R code for handling datasets with complexities such as sampling bias, misclassified outcomes, correlated outcomes, hierarchical structures, and confounding · Taught graduate-level classes with topics ranging from introductory statistics to seminar topics in Bayesian nonparametrics, decision theory, and sequential methods · Selected to tutor fellow PhD students for the biostatistics written qualifying exam		
<b>Full-time Co-op Work Semesters, University of Waterloo</b>	<i>Waterloo ON</i>	05/2008–12/2011
· Completed six semesters (two years) of full-time work at Munich Re, Manulife, Hewitt Associates, and Logitech · Applied statistical and actuarial methods to problems in biometric research, enterprise risk management, segregated fund valuation, group benefits pricing, pension administration, and database analysis		

## EDUCATION

<b>PhD, Biostatistics, Harvard University</b>	<i>Cambridge MA</i>	2017
· Thesis: Statistical Methods for the Analysis of Observational Data with Multiple Correlated Outcomes		
<b>Bachelor of Mathematics, University of Waterloo</b>	<i>Waterloo ON</i>	2012
· Honours Actuarial Science/Finance Option, Honours Statistics · Graduated with Distinction - Dean’s Honours List		
<b>Deep Learning Specialization, deeplearning.ai</b>	<i>Coursera</i>	2018
<b>Associate of the Society of Actuaries, Society of Actuaries</b>	<i>Schaumburg IL</i>	2012