

EDUCATION

University of Washington

2020 - Spring 2026

PhD in Statistics, Advanced Data Science Track. (*Advisor: Elena Erosheva*)

Carnegie Mellon University

2016 - 2019

B.S in Statistics and Machine Learning, University Honors

RESEARCH EXPERIENCE

University of Washington

Dissertation Research

Winter 2022 - Present

- Developing statistical methods for causal structure learning and causal decomposition

Research Assistant in WA Notify project

Spring 2021-Fall 2021

- Conducted research on the impact of the privacy-protected exposure notification app (WA Notify) on COVID-19 transmission and identified factors influencing willingness to quarantine and get tested using statistical methods

Carnegie Mellon University

Undergraduate Research Assistant advised by Alexandra Chouldechova

Fall 2019

- Assessed the presence of age, race, or gender-based disparity in the utilization of fully or semi-automated decision-making processes for determining when a case worker should investigate specific abuse cases

Undergraduate Research Assistant advised by Peter Freeman

Spring 2019

- Developed a data pipeline to aid astronomers in understanding the evolution of galaxies based on their current appearances, employing techniques to address imbalanced data

Undergraduate Research Intern in Black & Veatch Corporate Capstone Project

Fall 2018 - Spring 2019

- Created an R Shiny app for analyzing historical company data, predicting injury and property damage cases, and generating prevention strategies through partial dependence plots (pdp).

Undergraduate Research Intern for the KONAM Foundation

Fall 2017

- Designed and implemented a machine learning algorithm that assesses the risk of planting certain crops for marginalized farmers in India

TALKS, PUBLICATIONS & MEDIA

1. **S. Prakash**, F. Xia, E. Erosheva, “A Diagnostic Tool for Functional Causal Discovery”, *Preprint*, 2024. Available at [arXiv:2406.07787](https://arxiv.org/abs/2406.07787).
 - R package: [causalDiagnose](#)
 - Presented at American Causal Inference Conference, 2024 and Western North American Region of The International Biometric Society, 2023
2. F. Samir, E. Ahn, **S. Prakash**, M. Soskuthy, V. Shwartz, J. Zhu, “Efficiently Identifying Low-Quality Language Subsets in Multilingual Datasets: A Case Study on a Large-Scale Multilingual Audio Dataset,” Accepted to Transactions of the Association for Computational Linguistics (TALC), 2025. Available at [arXiv:2410.04292](https://arxiv.org/abs/2410.04292).
3. **S. Prakash**, C. Cinelli, E. Erosheva, C. Lee, “A Causal Decomposition Analysis of Black-White Disparity in Selection into Discussion during NIH Proposal Review Process”, (2025), (in preparation)

4. **S. Prakash**, F. Xia, E. Erosheva, “A Hypothesis Testing-based Approach for Functional Causal Discovery”, (2025), (in preparation)
5. **S. Prakash**, I. Javed, S. Lu, A. Adler, B. LeRoy, R. Nugent, “Characterizing Incidents at Black & Veatch”, *Carnegie Mellon University Meeting of the Minds Undergraduate Research Symposium*, (2019)
6. **S. Prakash**, P. Freeman, “Linking Galaxies Across Time via Conditional Density Estimation”, *Carnegie Mellon University Meeting of the Minds Undergraduate Research Symposium*, (2019).
7. S. Konam, **S. Prakash**, S. Papp, S. Mishra, X. Liu, Z. Ma, K. Siripurapu, “New App for Indigenous Farmers”, *The Hans India*, (2017), [Link](#).

PROFESSIONAL EXPERIENCE

Applied Science Intern, Amazon

Summer 2024

- Developed and implemented methodology to improve fallback priors for Amazon’s primary experimentation platform: Weblab

Data Scientist, Marinus Analytics

Winter-Summer 2020, Summer 2021

- Applied machine learning and time series analysis for unstructured child welfare case records
- Launched spam filter and underage person detection algorithms for TraffickJam: an application that uses human trafficking advertisement data to aid law enforcement with finding trafficking victims and traffickers
- Productionalized Infoshield: a text clustering algorithm for large scale human trafficking advertisement datasets.

Data Science and Research Intern, 84.51°

Summer 2019

- Fixed issues and tested optimization algorithms for grocery promotion; recommended running promotion optimization for 52 weeks to increase category performance by 4%

Software Developer Intern, Optum Technologies

Summer 2018

- Built authentication, UI and containerized existing application for cloud deployment, generating millions in savings and revenue

Research Intern, Royal Caliber D3M Program DARPA

Summer 2017

- Worked on machine learning on graph datasets and implemented a significantly more efficient way to estimate the number of triangles in a graph, (from $O(V^3)$ to $O(V)$, where V is the number of graph vertices), using a wedge sampling algorithm

SKILLS

Programming: Proficient in R, Python, SQL. Familiar with C, Standard ML, Matlab, Java, Bash, Mathematica

Libraries/Software: numpy, pandas, scipy, sklearn, statsmodels, tensorflow, seaborn, matplotlib, rjags, tidyverse, parallel, Git, Docker, AWS S3 & EC2, Keras

AWARDS

- Rising Star in Computational and Data Sciences 2025
- Center for Statistics and the Social Sciences Travel Award 2023
- 3rd Place winner in Meeting of Minds Undergraduate Research Symposium Poster Presentation Competition 2019

RELEVANT COURSEWORK

- Statistical Learning, Advanced Theory for Statistical Inference, Advanced Regression Methods, Causal Modeling, Foundations of Fairness in Machine Learning, Computational Biology, Statistical Graphics and Visualization, Parallel and Sequential Data Structures and Algorithms (C/SML), Linear Algebra

TEACHING EXPERIENCE

Teaching Assistant, *University of Washington (2021 - Present), Carnegie Mellon University (2017-2019)*

- Introduction to Mathematical Statistics (STAT 509), Applied Regression (STAT 504), Quantitative Introductory Statistics for Data Science (STAT 391), Elements of Statistical Methods (STAT 311), Statistical Reasoning (STAT 220), Causal Modeling (STAT 566), Statistical Concepts and Methods for the Social Sciences (STAT 221), Introduction to Probability Theory (36-225), Introduction to Machine Learning (10-601), Methods for Statistics and Data Science (36-202)