# Shamindra Shrotriya

*PhD Candidate, Statistics and Data Science*

## Education

| | |
|---|---|
| 2017-pres | **PhD Statistics & Data Science**, *Carnegie Mellon University*, USA. |
| | ○ Expected June 2022 |
| 2017-2019 | **M.S. Statistics & Data Science**, *Carnegie Mellon University*, USA. |
| | ○ GPA of 3.92/4.0 |
| 2015-2016 | **M.A. Statistics**, *University of California at Berkeley*, USA. |
| | ○ GPA of 3.9/4.0 |
| 2003-2007 | **BCom (Actuarial/Finance)**, *University of New South Wales*, Australia. |
| | ○ Graduated with Distinction |
| | ○ UNSW Co-op Scholar in Actuarial Statistics |

## Publications

### Papers

1. Bong, H, W Li, S Shrotriya, and A Rinaldo (2020). Nonparametric Estimation in the Dynamic Bradley-Terry Model. In: *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 03-05 June 2020, Palermo, Sicily, Italy.*

2. Barter, R and S Shrotriya (2016). Integrated Data Analysis for Early Warning of Lung Failure. *ODBMS.org.*

### Workshops

1. Dalmasso, N, A Reinhart, and S Shrotriya (Dec. 2019). Predictive Inference of a Wildfire Risk Pipeline in the United States. In: *NeurIPS 2019 Workshop, Tackling Climate Change with Machine Learning.* Vancouver, Canada.

2. Bong, H, W Li, and S Shrotriya (Nov. 2019). Efficient Estimation of Distribution-free dynamics in the Bradley-Terry Model. In: *Carnegie Mellon Sports Analytics Conference (CMSAC).* Pittsburgh, United States.

## Presentations

### Workshops

| | |
|---|---|
| 2019-12-15 | **Predictive Inference of a Wildfire Risk Pipeline in the United States (Spotlight)**, *NeurIPS 2019 Climate Change Workshop*, Vancouver, BC. |
| 2019-11-02 | **Efficient Estimation of Distribution-free dynamics in the Bradley-Terry Model**, *CMSAC Reproducible Research Competition*, Pittsburgh, PA. |

### Posters

**Dec 2018** **Predictive Inference of a Wildfire Risk Pipeline in the United States**, *NeurIPS 2019 Climate Change Workshop*, Vancouver, BC.

**Dec 2018** **Efficient Convex Estimation of the Time Varying Bradley-Terry Model**, *COPTS conference*, Pittsburgh, PA.

### Talks

**Dec 2018** **Introduction to the Tidyverse**, *STAT 36-350*, Pittsburgh, PA.

**Dec 2018** **Functional Connectivity in iEEG Data**, *Advanced Data Analysis Presentation*, Pittsburgh, PA.

**Jul 2016** **Predicting COPD in pneumonia patients**, *Geisinger Collider Project*, Berkeley, CA.

## Research Experience

**2018-2019** **Advanced Data Analysis (ADA) Project**, *Carnegie Mellon University*, Pittsburgh, PA.
- Advised by: Prof. Max G'Sell and Prof. Avniel Singh Ghuman
- Investigated the dynamic functional connectivity in human epilepsy patients using iEEG data
- Successfully presented oral defense of research work

**2017** **Research Associate**, *University of California*, Berkeley, CA.
- Advised by: Prof. Bin Yu and Prof. Ben Brown
- Investigated the statistical properties of the iterative Random Forests (iRF) algorithm
- Co-developed the Python implementation of the iRF algorithm
- Helped complete a successful four-year NSF BIGDATA grant proposal for this project

**2016** **Geisinger Collider Project**, *University of California*, Berkeley, CA.
- Joint work with Rebecca Barter (UC Berkeley)
- Investigated using Electronic Medical Record data to determine whether a pneumonia patient will develop Chronic Obstructive Pulmonary Disease (COPD)
- Winner - Best paper award and overall competition

## Industry Experience

**2014-2015** **Data Science Infrastructure Team Lead**, *freelancer.com*, Sydney, Australia.
- Designed and implemented a prototype of the new A/B testing framework
- Co-designed and administered the entire Extract-Transform-Load (ETL) process written with Go and AWS Redshift
- Designed and improved the internal metrics monitoring dashboard

2012-2014 **Data Scientist**, *Quantium Consulting*, Sydney, Australia.
- ○ Led the end-to-end development of the behavioural 'lifestage' customer classifier for the entire 7 million Woolworths Supermarket customer base
- ○ Led the data-driven electronic marketing strategy for Woolworths Life Insurance which included developing scoring models (GLMs) and conducting A/B tests to optimise response rates
- ○ Co-designed and developed the National Australia Bank Online Retail Sales Index

2011-2012 **Microinsurance Fellow**, *UN - International Labor Organization*, Pune, India.
- ○ Wrote a report on the best actuarial pricing practices to be undertaken by microinsurance organisations

2007-2011 **Senior Actuarial Consultant**, *PricewaterhouseCoopers*, Sydney, Australia.
- ○ Built visualization dashboards for monitoring key risk metrics for Insurance Australia Group, Australias' largest private general insurer
- ○ Developed key reporting metrics used by Qantas airlines to assess key drivers and trends behind their Qantas Frequent Flyer Program (the largest customer loyalty program in Australia)

## Teaching Experience

### Head Teaching Assistant

2020 **STAT 36-350 (Statistical Computing)**, *Carnegie Mellon University*, Pittsburgh, PA.
- ○ Instructor: Prof. Peter Freeman
- ○ Developed R programming course materials
- ○ Managed 9 TAs and grading via Gradescope/Canvas, held office hours

2019 **STAT 36-350 (Statistical Computing)**, *Carnegie Mellon University*, Pittsburgh, PA.
- ○ Instructor: Prof. Peter Freeman
- ○ Developed R programming course materials
- ○ Managed 7 TAs and grading process, held office hours

2018 **STAT 36-700 (Intermediate Theoretical Statistics)**, *Carnegie Mellon University*, Pittsburgh, PA.
- ○ Instructor: Prof. Larry Wasserman
- ○ Wrote HW solutions, helped with HW/exam design
- ○ Managed other TAs and grading process, held office hours

2016 **STAT133 (Computing with Data)**, *University of California*, Berkeley, CA.
- ○ Instructor: Prof. Gaston Sanchez
- ○ Managed other TAs and grading process, Held weekly R tutorial sessions
- ○ Winner - Outstanding Graduate Student Instructor award

### Teaching Assistant

2019 **STAT 36-750 (Graduate Statistical Computing)**, *Carnegie Mellon University*, Pittsburgh, PA.
- ○ Instructor: Prof. Alex Reinhart
- ○ Wrote HW solutions, graded 300+ Github Pull Requests, held office hours

2018 **STAT 36-350 (Statistical Computing)**, *Carnegie Mellon University*, Pittsburgh, PA.
   ○ Instructor: Prof. Ryan Tibshirani
   ○ Reviewed course materials, held office hours

2017 **STAT 36-401 (Modern Regression)**, *Carnegie Mellon University*, Pittsburgh, PA.
   ○ Instructor: Prof. April Galyardt
   ○ Reviewed course materials, held office hours

## Awards and Honors

2020 **NGC Wildfire Research Scholar**, *American Australian Association*.

2020 **TA of the Year**, *Carnegie Mellon University*.

2019 **NeurIPS Climate Change Workshop Travel Award**.

2019 **CMSAC Best Paper Award**, *Carnegie Mellon University*.

2017 **Outstanding Graduate Student Instructor**, *University of California, Berkeley*.

2016 **Elizabeth Scott Memorial Award**, *University of California, Berkeley*.

2016 **Best Paper and Competition Winner**, *Geisinger Health Collider Project*.

2012 **Microinsurance Fellowship**, *UN - International Labor Organization*.

2007 **Associate of the Institute of the Actuaries Australia**.

2003 **Co-op Industrial Scholarship in Actuarial Studies**, *University of New South Wales, Australia*.

2003 **Council Tertiary Scholarship**, *Parramatta Council, Sydney, Australia*.

2002 **Entry Award Scholarship in Engineering (declined)**, *University of Sydney, Australia*.

2002 **Australian Students Prize for Academic Excellence**, *Australian Federal Government*.

2002 **Premier's Award for Academic Excellence**, *NSW Government, Australia*.

*Department of Statistics and Data Science, Carnegie Mellon University*
 *+1 415 937 3701* • *shamindra@cmu.edu* • *shamindras.com*
 *shamindraas* • *shamindras* *4/5*

2002 **University Admissions Index (UAI) 99.90 (top 0.1% in State)**.

## Service

### Reviewing

2020 **Program Committee**, *NeurIPS 2020 Workshop, Tackling Climate Change with Machine Learning*, (Held Virtually).
  ○ Peer-reviewed workshop papers

2020 **Program Committee**, *ICLR 2020 Workshop, Tackling Climate Change with Machine Learning*, Addis Ababa, Ethiopia.
  ○ Peer-reviewed workshop papers

## Software

1. Dalmasso, N, A Reinhart, and S Shrotriya (Feb. 2020). *backburner - An R Package for extracting and transforming open US wildfire and climate data.* (lifecycle: developing). `https://shamindras.github.io/backburner/`.

2. Holdgraf, C, K Kumbier, R Liu, S Shrotriya, S van der Walt, Y Wang, and E Xia (Dec. 2016). *iRF - A Python Package to implement the Iterative Random Forests algorithm.* (lifecycle: stable). `https://github.com/Yu-Group/iterative-Random-Forest`.

## Skills

Proficient `R`, `Python`, `SQL (Redshift/SQL Server/Teradata)`, `Git/Github`, LaTeX.

Competent `Bash`, `Make`, `SAS`.

*Department of Statistics and Data Science, Carnegie Mellon University*
📱 *+1 415 937 3701* • ✉ *shamindra@cmu.edu* • 🌐 *shamindras.com*
🐦 *shamindraas* • ⓖ *shamindras* *5/5*