

RYAN R. ROSARIO

“I am passionate about combining code with data and turning them into actionable insights and/or smart and efficient systems for decision making and analysis. I am always looking for new ways to acquire, manage and present data and information. I crave learning new technologies, using my knowledge to help others, and solving challenging and exciting problems.”

Education

Ph.D., Statistics 2017

University of California, Los Angeles

Specialization: Data Mining, Machine Learning, Natural Language Processing, High Performance Computing

Dissertation Title: *A Data Augmentation Approach for Short Text Classification*

Coursework: probability theory, large sample theory, sampling and experimental design, regression analysis, statistical computing, Bayesian modeling, time series analysis, spatial statistics, multivariate analysis.

M.S., Computer Science 2010

University of California, Los Angeles

Coursework: algorithms and complexity, programming language theory, operating systems, natural language processing, neural networks, artificial life, Bayesian networks, web mining/information retrieval, machine learning theory.

M.S., Statistics 2008

University of California, Los Angeles

B.S., Mathematics of Computation, Statistics 2006

University of California, Los Angeles

Graduated with Latin Honors, College Honors, Departmental Honors in Mathematics and Statistics.

Major Coursework: linear algebra, real analysis, complex analysis, numerical methods, differential equations, combinatorics, scientific software design, networking, linear regression, nonlinear regression, SAS, Stata, psychometrics, full year of C++.

Relevant Work Experience

Facebook, Inc.

Quantitative Engineer (Machine Learning Engineer) May 2013-March 2015
Menlo Park, CA

- Developed the sentiment analysis algorithm and system used as the company standard, and presented this work at PyData 2014 (<https://www.youtube.com/watch?v=y3ZTKFZ-1QQ>).
- Helped retain a multimillion-dollar advertising deal by using sentiment insights to include/exclude user targeting.
- Developed an algorithm and distributed system for categorizing and summarizing groups of text using Latent Dirichlet Allocation.
- Managed and maintained the company PyData software stack (scipy, numpy, pandas, scikit-learn, IPython) educating users on use and upkeep.
- Helped develop new People Analytics team and mentored new lead:
 - Performed analysis on retention and promotion of underrepresented groups using logistic regression and mixed models.
 - Used text mining techniques and statistical techniques on interview feedback to reduce SWE interview load and make the process more efficient by moving certain interview types earlier in the process.
 - Helped construct, execute and improve a structured interview process for the new People Analytics team and serving as the psychometrics/IRT evaluator.

GumGum, Inc.

Chief Data Scientist/Research Engineer October 2011-October 2012
Santa Monica, CA

- Developed a system for extracting keywords from various forms of text using map-reduce and statistical techniques with Hadoop.
- Conceptualized, developed and evaluated a system for detecting inappropriate content in a variety of categories using techniques such as Naive Bayes, Decision Trees and topic models.
- Languages used (in order of frequency): Java, Python, R, SQL, Bash.

Data Scientist

West Los Angeles, CA

- Developed web traffic simulator in R and Python.
- Constructed a model and system based on model for predicting the behavior of a noisy ad server.
- Developed Hadoop jobs for data extraction and transformation, and assisted in troubleshooting cluster issues.
- Developed and implemented a new ranking metric for ad selection using Markov Chain Monte Carlo.

Mentoring, Teaching and Leadership Experience

Springboard

Data Science Intensive Mentor, and Career Track Mentor

February 2016-Present

San Francisco, CA

- Met weekly with students to discuss career goals, answer questions about their learning, and guiding them through the capstone project from inception to evaluation and graduation.
- Monitored learning progress through a 3-6 month intensive curriculum of learning Python coding, data management, data intuition, inferential statistics and machine learning.
- Developed interactive content in Machine Learning for a new Career Track.
- Developed an interview guide to standardize the training of mentors for mock interview preparation.
- Thoroughly reviewed the content, style and grammar of the final capstone report, as well as all Python code.
- Assisted in capstone project selection and maintaining high professional standards for all mentees ensuring all data intuition and data management skills were valid, all methods were executed properly, evaluated properly and substantially presented in context of the original problem.
- Alumni mentees now work as Data Scientists and Machine Learning Engineers at Amazon, WeWork, Walmart Labs, AirBnb and others.

Department of Statistics, University of California, Los Angeles

Teaching Fellow

Los Angeles, CA

2007-Current

- Served as a mentor for graduate students in the Professional Masters in Applied Statistics program by providing career advice and resources for learning important skills. (Statistics 404)
- Developed and presented a three hour lecture on Unix tools, the Python programming language provided an immersive learning experience through the use of a live demonstration of common workflows in machine learning. (Statistics 404)
- Taught weekly lab sessions for graduate statistical programming and data mining courses using Unix tools, Python, R, MySQL, Processing and introductory Hadoop. (Statistics 202A)
- Assisted faculty member with brand new geostatistics course using open-source GRASS GIS. (Statistics C173/273)
- Set up and managed a five node Hadoop cluster on Department owned Mac Pro systems and authored a user's and administrator's guide for statisticians.
- Lead lab sessions and discussion sessions for a wide diversity of undergraduate courses ranging from Introduction to Statistics (Statistics 10) to Modeling with Markov Chain Monte Carlo (Statistics 202C).

Technical Skills

Programming Languages:

Most familiar: Python, SQL. *Also code in:* Java, C/C++

Statistical, Mathematical, Machine Learning Packages:

R, SciPy/NumPy, pandas, scikit-learn.

Operating Systems and Infrastructure:

Linux, Mac OS X, Windows, Hadoop, Amazon Web Services

Publications and Presentations

N. Christou, R. R. Rosario, and I. Dinov. "Technology-Enhanced Probability and Statistics Education Using the Statistics Online Computational Resource," Emerging Technologies for Online Learning Symposium 2010. 20 July 2010.

R. R. Rosario. Review of [Text Mining with Perl](#), by Roger Bilisoly. In *Journal of Statistical Software, Book Reviews*. 14 January 2009. Vol. 29, Book Review 9. <http://www.jstatsoft.org/v29/b09>

E. Agapie, G. Chen, D. Houston, E. Howard, J. Kim, M. Y. Mun, A. Mondschein, S. Reddy, R. Rosario, J. Ryder, A. Steiner, J. Burke, E. Estrin, M. Hansen and M. Rahimi. *Seeing Our Signals: Combining location traces and web-based models for personal discovery*, in *Proceedings of Hot Mobile*, 2007.