

# Pedro Rodriguez | Resume

✉ p.rodriguez@colorado.edu • 🌐 pedrorodriguez.io

## Experience

### University of Colorado at Boulder

*Research Assistant advised by Jordan Boyd-Graber*

### Machine Learning, Distributed Systems

*August 2015 - Present*

- Primary research interest is in distributed systems for large scale machine learning
- Current project is scaling Quiz Bowl question answering system with Spark to reduce feature extraction and training time from over a week to less than a day. This will allow for faster iteration speed on changes. This is part of an NSF grant funded research project on embedding generative models in reinforcement learning for sequentially revealed data.
- Manage Spark and Docker clusters running on research OpenStack cluster.

### Trulia

*Data Scientist*

### Crowdsourcing, DevOps, Machine Learning

*January 2015 - July 2015*

- Designed and iteratively improved crowdsourcing experiments to produce rank order for machine learning generated photo albums. In doing so wrote a Amazon Mechanical Turk library which managed approximately 20,000 tasks.
- Architected and implemented infrastructure for internal data science APIs built on Docker, Swagger, Jenkins, and predominantly Django as a web framework.
- Created a click through rate (CTR) model using Apache Spark to gain insight into user interaction with web application, and begin serving customized recommendations.
- Ran workshops within the company to facilitate the use and learning of Apache Spark and Docker.

### UC Berkeley AMPLab

*Undergraduate Research Assistant*

### Distributed Machine Learning

*Fall 2014*

- Created an efficient and distributed implementation of Gibbs sampling based Latent Dirichlet Allocation to perform topic modeling. Implementation is based on state of the art optimization techniques, and uses Apache Spark.
- Teaching Assistant at the UC Berkeley AMPCamp 2015.
- Contributed to Berkeley crowdsourcing project AMPCrowd.

### Boise State University Geophysics

*Cryosphere Scientist, Software Engineer, Field Researcher*

### Remote Sensing, Simulations

*May 2013 - October 2014*

- Developed software for instrumentation and data acquisition in FMCW radar systems in hydrological research
- Created thermodynamic simulations of complex processes of snow metamorphism

### UC Berkeley Center for Time Domain Informatics

*Undergraduate Researcher*

### Astrophysics

*January 2012 - August 2012, January 2013*

- Analyzed group's novel research of machine learned catalog of variable stars to identify followup observation candidates.

## Projects and Code

**Apache Spark:** Contributed to implementation of topic modeling algorithm (LDA), and expansion of SparkSQL built in functions. Code is in MLlib, Intel Analytics Topic Model package, and Spark 1.5

**ScalaFunctional:** Author of python package exposing a collections API borrowed from Scala/Spark

**Docker and Docker-Compose:** In depth workshop of using docker tools for multi-service APIs

**SnowGeek.org:** Website/company I created for snow/avalanche science and realtime conditions retrieval

**FMCW Radar Controller:** Controller code to operate research radar with a Raspberry Pi

**Code:** [github.com/EntilZha](https://github.com/EntilZha), [bitbucket.org/pedrorodriguez](https://bitbucket.org/pedrorodriguez), [pedrorodriguez.io](https://pedrorodriguez.io)

**Languages:** Python, Scala, Go, C/C++, HTML/CSS/Javascript, Java, Ruby, SQL, Matlab, Tex

## Education

### University of Colorado at Boulder

*PhD in Computer Science, Distributed Machine Learning*

*August 2015 - Present*

### University of California at Berkeley

*Bachelor of Arts in Computer Science*

*2010 - 2014*