

Pedro Rodriguez | Resume

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

Experience

University of Colorado at Boulder

Research Assistant

Machine Learning, Distributed Systems

August 2015 - Present

I am currently a research assistant at CU Boulder advised by Jordan Boyd-Graber. My primary research area is in large scale, distributed, machine learning. Problems that require using advanced machine learning with large scale data greatly interest me. I also manage the lab's Spark and Docker clusters running on OpenStack.

Trulia

Data Scientist

Crowdsourcing, DevOps, Machine Learning

January 2015 - July 2015

I architected and implemented three projects. In the first project I ran crowdsourcing experiments to survey user opinion of machine created photo albums. I wrote a Mechanical Turk library, designed experiments, ran, interpreted approximately 20,000 AMT tasks. In the second project I created infrastructure built on Docker, Swagger, Jenkins, and various web frameworks to power data science team APIs. Lastly I implemented a click through rate prediction model using Apache Spark to drive user traffic. These projects supported a company initiative to drive growth through discovery. I also ran workshops on learning Apache Spark and Docker.

UC Berkeley AMPLab

Undergraduate Research Assistant

Distributed Machine Learning

Fall 2014

I worked in the Berkeley AMPLab as an undergraduate research assistant. My project was centered around creating an efficient implementation of Latent Dirichlet Allocation to perform topic modeling using Spark.

Boise State University Geophysics

Cryosphere Scientist, Software Engineer, Field Researcher

Remote Sensing, Simulations

May 2013 - October 2014

Worked in instrumentation development, data acquisition, instrumental control software and thermodynamic simulations. These supported research for hydrological and snow applications of FMCW radar systems.

UC Berkeley Center for Time Domain Informatics

Undergraduate Researcher

Astrophysics

January 2012 - August 2012, January 2013

My project was focused on leveraging the group's recently published machine learned catalog of variable stars and cross match with an x-ray catalog, followed by filtering for observation followup candidates.

Projects and Code

Apache Spark: Latent Dirichlet Allocation, two PRs in Spark SQL, ran workshop at Trulia

ScalaFunctional: Python package for pipeline functional programming in scala style

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

Quiz Bowl: AI to play quiz bowl. Current PhD project is to refactor to run in cluster

SnowGeek.org: Website/company I created for snow/avalanche science at snowgeek.org

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

Code: github.com/EntilZha, bitbucket.org/pedrorodriguez, pedrorodriguez.io

Languages: Python, Ruby, Scala, Go, C/C++, HTML/CSS/Javascript, Java, 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