

Pedro Rodriguez | Resume

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

Experience

University of Colorado at Boulder **Machine Learning, Deep Learning, NLP, Knowledge Graphs**
Research Assistant to Jordan Boyd-Graber *August 2015 - Present*

- Primary research interests are large-scale machine learning, deep learning, knowledge graphs, and question answering.
- Creating a AI that plays a trivia game called Quiz Bowl which currently beats some human experts. This is an NSF funded project to investigate generative models and reinforcement learning for tasks with incrementally revealed information.
- Team Captain and founding member of the Colorado Data Science Team: leadership and technical role promoting data science education through applied experience and creating synergistic industry partnerships at CU Boulder.

Oracle Data Cloud/Datalogix **Apache Spark, Machine Learning, DevOps**
Data Science Research Intern *May 2016 - August 2016*

- Architected and implemented ETL pipeline with Apache Spark/Airflow which ingests 200GB/day of advertising data.
- Improved throughput from AWS S3 to Apache Spark by 40-60x which resulted in substantial cost savings.
- Prototyped data science workflows using Packer, Terraform, and Apache Zeppelin. This greatly increased productivity of two fellow interns, and provided valuable insights into requirements for a next-generation platform.

Trulia **Crowdsourcing, DevOps, Machine Learning**
Data Scientist *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. This code is now called on 100% of visits to trulia.com.
- Ran educational workshops within the company to facilitate the use and learning of Apache Spark and Docker.

UC Berkeley AMPLab **Distributed Machine Learning**
Undergraduate Research Assistant *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.

Boise State University Geophysics **Remote Sensing, Simulations**
Cryosphere Scientist, Field Researcher *May 2013 - October 2014*

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

UC Berkeley Center for Time Domain Informatics: Undergraduate Researcher in Astrophysics, 2012

Avalanche Safety Educator: Volunteer for National Ski Patrol, Professionally in South America, 2010-Present

Projects and Code

Qanta: An AI that plays quiz bowl. Uses deep learning on Tensorflow, Apache Spark, Vowpal Wabbit, Terraform, Packer, Amazon Web Services, XGBoost, and more to create running a robust state-of-the-art system

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

PyFunctional: Author of popular python package exposing a collections API borrowed from Scala/Spark

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

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

Education

University of Colorado at Boulder
PhD in Computer Science specializing in Machine Learning, Deep learning, NLP *Expected Spring 2020*

University of California at Berkeley
Bachelor of Arts in Computer Science with additional coursework in Astrophysics *2010 - 2014*