

Rachel Warren

SKILLS

Fluent In: Scala, Java, Spark, Maven, sbt, Hadoop, Yarn, Python, Git, Unix, SQL

Can Hack in: JavaScript, HTML, CSS, Hive, PIG, R, Stata,

Paradigms: Object Oriented Programming, Functional Programming, Agile and Scrum methodology

PUBLICATIONS:

High Performance Spark, June 2017(Co-author with Holden Karau) O-Reilly.

EXPERIENCE

Software Engineer in Data Science, Alpine Data Spring 2015- Present

- Platform and machine learning engineer on Alpine's enterprise analytics platform.
- Working on the infrastructure to allow customers to integrate legacy python code
- Completed a project to automatically optimize configurations for Spark jobs.
- Added support for new algorithms on Hadoop -- Collaborative Filtering, NLP package (TFIDF, custom tokenizing, LDA) T-Tests, rank statistics calculator, custom aggregation for wide data. This includes writing or integrating an implementation of the algorithm to scale to a billion rows as well as implementing a graphical interface by which users will interact with it.
- Helped develop and continue to lead maintenance of our customer-facing extensibility framework, which allows users to add their own algorithms in Hadoop or on database. I added functionality to read and write from HDFS (including support for Avro and Parquet), date/time support, dirty data support, user facing examples and templates, and integration with D3 to support visualizations.

Accomplishments: Optimized rank statistics algorithm to be over 100x faster for wide data. Developed first iteration of collaborative filtering in two weeks to close a sales deal. Trained our data science team, who had no previous scala or Spark experience to be proficient in the extensibility framework.

Computer Science Faculty Assistant, Ashesi University in Ghana Fall 2014

- Three courses: software engineering (Angular Js, Javascript), intro (Java, Robot C), and Algorithms

Computer Science Senior Project, Wesleyan Media Project Fall 2013

- Developed a Java application that scrapes articles from the web, parses them, and uses machine learning (implemented with WEKA machine learning library) to categorize them for human review.

Accomplishments: Successfully reduced the time and associated costs of human readers by 20%.

Product Analytics Intern (paid), Pandora Summer 2013

- Used Hive on Hadoop cluster to mine consumer data
- Used Python, R, and Excel to produce and present accessible, quantitative insights to support marketing and sales designs and product features.
- Focused on analyzing differences across modes of usage such as subscription vs. free usage, and mobile vs. automotive users.

Accomplishments: At least one product feature implemented from my recommendations.

Research Assistant (paid), Wesleyan Media Project Summer 2012 - January 2013

- Selected for the paid Quantitative Analysis Summer Apprentice Program, included instruction in statistics and analysis for a Government Professor.
- Completed a statistical analysis in R, Stata, and SPSS using survival analysis and logistic regressions.

Accomplishments: At the end of the summer, my research partner and I received funding to continue our research, which we presented at an academic conference in 2013.

EDUCATION

Wesleyan University, CT, Bachelor of Arts in Computer Science 2014 GPA 3.6

ADVENTURE

Unsupported Bicycle Tour of Patagonia February 2017 - April 2017

My partner and I bicycled from Puerto Aysen, Chile to Ushuaia, Argentina (at the southern shore of Tierra del Fuego). We covered more than 1000 miles, most over mountainous dirt roads, carrying all our own supplies and doing all of our own bicycle maintenance.