Malcolm **Greaves**

Machine Learning Scientist and Engineer

U.S. Citizen, San Francisco Bay Area

(414) 671-9696 | greaves.malcolm@gmail.com | www.malcolmgreaves.io | malcolmgreaves | malcolm-greaves-49959919

Experience

Nitro Software, Inc.

San Francisco, CA USA

Research Engineer — *Mar. 2015 - Oct. 2016*

- Created novel machine learning based solution for automatic form field detection and semantic classification.
- Lead research, development, and production deployment on first machine learning based product within company history.
- Full stack production development and deployment in Scala and Javascript with Play!, Angular.js, Postgres, and Kafka.

Alpine Data Labs

San Francisco, CA USA

Software and Machine Learning Engineer — *Jun. 2014 - Mar. 2015*

- Implemented machine learning and feature transformation algorithms in Scala as a part of Alpine's analytics and algorithms platform.
- Algorithm optimization for distributed execution on customer's Hadoop and Spark clusters.

Read the Web, Worldly Knowledge Research Groups

Pittsburgh, PA USA

Graduate and Undergraduate Student Researcher — *Jan. 2010 - May 2014*

- Relationship and entity extraction from unstructured text. Includes experience with non-linear learning algorithms, clustering, baysian inference, and probabalistic graphical modeling.
- Computational linguistics: syntax parsing, part of speech tagging, named entity recognition, sentence segmentation.
- Large scale text classification, feature engineering, vector space modeling, and experience with processing TBs of text on a 175 node Hadoop cluster.

Education

Carnegie Mellon University

Pittsburgh, PA USA

B.Sc. and M.Sc. in Computer Science — *Aug. 2009 - May 2014*

- Graduated with School of Computer Science Honors
- Four years of research experience in large scale machine learning, natural language processing, and information extraction
- Master's Thesis in semantic relation extraction from unstructured text: `http://goo.gl/DzMr6c`

Work Portfolio

Programming	*Proficient*: Scala, Python, Go, Java; *Moderate*: C, BASH, MATLAB, R, SQL; *Familiar*: LaTeX, Javascript, Typescript
Technologies	Apache Spark, Kafka, Hadoop, git, Github, *nix, Jira, Play!, Angular.js 1.x, HTML 5, CSS, Postgres, Websockets
Libraries	Pandas, NumPy, NLTK, scikit-learn, TensorFlow, Caffe, Keras, BLAS/LAPACK, Breeze, Spire, Scalaz, Akka, Spark, CoreNLP, boof-cv, OpenCV
Data Science	machine learning, deep learning & neural network models, linear algebra, convex optimization, statistics, probability, probabalistic graphical models, combinatorics, algorithm design and analysis (including complexity), distributed systems, information retrevial and extraction, search and ranking, recommender systems
Software Engineering	functional programming, distributed and concurrent programming, server-side programming, SQL, technical communication (oral, presentation, and written), small team technical leadership, Agile software development

auto-gfqg: Automatic gap-fill question generation

https://github.com/malcolmgreaves/auto-gfqg/

- An unsupervised learning system that automatically creates multiple choice, fill-in-the-blank questions from a single text corpus.

smo-fun: Efficient SMO implementation for non-linear SVMs in Scala

https://github.com/malcolmgreaves/smo-fun/

- Full implementation of the sequential minimal optimization algorithm. Trains linear and non-linear support vector machines.

fp4ml: Functional programming for machine learning

https://github.com/malcolmgreaves/fp4ml/

- An ML library in Scala with clean, functional APIs and a strategic, referentially transparent use of mutability for performance.
- A novel type class that provides a Scala collections-like API for describing data manipulation. Includes evidence for Spark RDDs, Flink DataSets, and in-memory structures: write once, run everywhere!