Malcolm **Greaves**

MACHINE LEARNING SCIENTIST AND ENGINEER

U.S. Citizen, San Francisco Bay Area

(414) 704-9696 | greaves.malcolm@gmail.com | www.malcolmgreaves.io | malcolmgreaves | malcolm-greaves-49959919

Experience

Change Healthcare   *Emeryville, CA*

SENIOR DATA SCIENTIST   *November 2018 - Present*

- Developed neural sequence model to populate medical claims from patient narrative text.
- Developed deep NN model for predicting missing charges & deployed to US's largest medical insurance payment network.
- Increased scientist productivity by 25x with internal model training & serving framework.

Volley Labs, Inc.   *San Francisco, CA*

NATURAL LANGUAGE UNDERSTANDING SCIENTIST   *December 2016 - October 2018*

- Implemented graph-LSTM deep learning model with PSL collective reasoning for automatic knowledge graph creation.
- End-to-end ownership & development of multiple-choice question generation system from unsupervised to supervised neural network approach. Product-differentating feature that helped Volley stand-out and make its first sale to JPMC.
- Technical lead of production AI team. Data pipelines in Airflow; deployed ML models using Keras, Spacy, Tensorflow; Python 3.

Nitro Software, Inc.   *San Francisco, CA*

RESEARCH ENGINEER   *Mar. 2015 - Oct. 2016*

- Created novel machine learning based solution for automatic form field detection (FFD) and semantic classification.
- Lead research, development, and automatic pipeline re-training deployment of Nitro Cloud ML features.

Alpine Data Labs   *San Francisco, CA*

SOFTWARE AND MACHINE LEARNING ENGINEER   *Jun. 2014 - Mar. 2015*

- Algorithm optimization for large-scale data processing & modeling in Spark: L-BFGS, random forest models, feature encoding, etc.

Read the Web, Worldly Knowledge Research Groups   *Pittsburgh, PA*

GRADUATE AND UNDERGRADUATE STUDENT RESEARCHER   *Jan. 2010 - May 2014*

- Large scale text classification, feature engineering, information extraction on TBs of web text with a 175 node Hadoop cluster.
- Relationship and entity extraction from unstructured text: ensemble learning, co-training, KG construction, PGM inference.

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
- Master's Thesis in semantic relation extraction from unstructured text with probabalistic logic & SVMs: `http://goo.gl/DzMr6c`

Work Portfolio

Languages	*Proficient*: Scala, Python, Go, Java; *Moderate*: BASH, C, C++11, SQL; *Familiar*: LaTeX, Typescript
Libraries	Pandas, TensorFlow, NumPy, spaCy, scikit-learn, Keras, PyTorch, Flask, Spark, CoreNLP, 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 retreival 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.