

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.
- Co-developed deep NN model for predicting missing charges & deployed to US's largest medical insurance payment network.
- Implemented re-usable model serving framework to decrease time to model deployment by 25x from existing process.

Volley Labs, Inc.

San Francisco, CA

NATURAL LANGUAGE UNDERSTANDING SCIENTIST

May 2017 - October 2018

- Implemented graph-LSTM deep learning model with PSL collective reasoning for automatic knowledge graph creation.
- Research, development, and production deploy of automatic cloze-style, multiple-choice question generation system using a mixture of supervised, unsupervised learning and manually identified lexico-syntactic patterns.
- 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 probabilistic logic & SVMs: <http://goo.gl/DzMr6c>

Work Portfolio

Languages Libraries

Proficient: Scala, Python, Go, Java; Moderate: BASH, C, R, SQL; Familiar: LaTeX, Typescript

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, probabilistic graphical models, combinatorics, algorithm design and analysis (including complexity), distributed systems, information retrieval 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/](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.

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

[HTTPS://GITHUB.COM/MALCOLMGREAVES/SNO-FUN/](https://github.com/malcolmgreaves/sno-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/](https://github.com/malcolmgreaves/fp4ml/)

- An ML library in Scala with clean, functional APIs and a strategic, referentially transparent use of mutability for performance.