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 deep learning model for predicting missing charges on medical insurance claims; deployed to US' largest medical payment network.
- Development of a N.N. architecture for revenue prediction on a per-procedure basis for submitted claims.
- 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 and semantic classification.
- Lead research, development, and production deployment on first machine learning based product at Nitro.
- Implemented first end-to-end automatic model training and deployment system for Nitro Cloud.
- Full stack production development and deployment in Scala and Javascript with Play!, Angular.js, Postgres, and Kafka.

Alpine Data Labs — *San Francisco, CA*

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.

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*: BASH, C, R, 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 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.