Jacob Marold, Ph.D.

Senior Data Scientist

New York, NY - Email me on Indeed: indeed.com/r/5823722739b54184

https://www.linkedin.com/in/jdmarold
http://www.jakemarold.com

Analysis of textual data using Natural Language Processing (NLP), document classification, parsing, part-of-speech tagging, and metadata extraction from documents, neural network and deep learning language techniques (Doc2Vec, Word2Vec), time series analysis (convolutional neural networks, recurrent neural networks LSTMs), anomaly and outlier detection development. Extensive experience with conference-level speaking, leadership and mentoring, statistics, error analysis, and linear and non-linear regression techniques with global (shared) parameter analyses.

Interested in finance, health care, pharmaceutics, machine learning, efficiency optimization, web-analytics, and creative multi-dimensional data visualizations.

Python, SQL, Web Scraping and Crawling, Network analysis, Machine Learning (supervised and unsupervised) using classical and neural network algorithms, Regression techniques, Time-series analysis, and data visualization. Major packages used: Keras, Tensorflow, Theano, Pandas, NumPy, SciPy, Scikit learn, BeautifulSoup, matplotlib, SymPy, lmfit
Willing to relocate to: New York, NY - New York, NY - Brooklyn, NY
Authorized to work in the US for any employer

WORK EXPERIENCE

Senior Data Scientist

J.P. Morgan Chase & Co. - New York, NY - November 2015 to Present

Analysis of textual data using Natural Language Processing (NLP)

Parsing, part-of-speech tagging, and metadata extraction from documents

Supervised, Unsupervised, Semi-Supervised classification and clustering of documents

Machine learning classification of documents - Neural Network and Deep Learning language techniques, K-neighbors, K-means, Random Forest, Logistic Regression, SVM

Anomaly and outlier detection algorithm development. Time series analysis (regression and classification), fraud detection.

Extensive experience with conference-level speaking, leadership and mentoring

Data Science Fellow

The Data Incubator - Washington, DC - June 2015 to August 2015

The Data Incubator (http://www.thedataincubator.com/) is a top-tier data science program

• Selected in top 2% from over three thousand PhD applicants (based on statistical and programming ability and data analysis skills) to participate in a rigorous data science program.

• Predicted Yelp business ratings using machine learning algorithms and natural language processing (NLP) (Scikit learn/sklearn: Random Forest, K-nearest neighbors KNN, multiple regression techniques).

• Developed predictive models for NYC restaurant inspection grades using machine learning algorithms and NLP

• Analyzed 8GB of Centers for Medicare and Medicaid (CMS) Provider Utilization and Payment Data by regional networks to view cost distribution for services and created D3 compatible visualizations (major packages: Pandas, Vincent).

• Performed social network analysis on NYC social diary unstructured data photo captions (Web Scraping, Beautiful Soup, Regex, NetworkX). Calculated PageRank based on social graph analysis. Who is the most popular person in NYC? Which couples are most likely to be seen together?

• Generated predictive models for weather data using time-series methodology. What is the chance of rainfall at 8pm in NYC on Mondays in July?

• Used Hadoop and Map Reduce to analyze >10GB of Wikipedia data. What are the top 100 words used in English Wikipedia pages?

• Developed financial models for Capital One customers to predict risk

Skills Used
Technical skills: Python (major packages: Pandas, NetworkX, NumPy, SciPy, SymPy, BeautifulSoup, Multiprocessing, Matplotlib), R, Hadoop Map Reduce.

Soft skills: Communication (written and oral), expertise with common office software packages (Microsoft Office - PowerPoint, Excel, Word), and graphics generation

Computational Biophysicist / Research Scientist

Johns Hopkins University - Baltimore, MD - August 2009 to June 2015

• Wrote software packages for biophysical data analysis and visualization

• Developed platform for global (shared parameter) non-linear and linear least squares regression of biophysical titration and time series data - resulted in 10-fold increase in data analysis efficiency

• Generated custom statistical mathematical models (based on probability theory) to be used in fitting routines

• Improved F-statistics and Bootstrap methodology for obtaining parameter confidence limits. My bootstrapping procedure led to a 12-fold speed enhancement through the use of multiprocessing.

• Platform speaker at the 59th Biophysical Society Meeting (the world's largest Biophysics conference with greater than 7000 attendees http://www.biophysics.org/)

• Organizing committee member and speaker at the NSF Protein Folding Consortium (meeting for the most advanced protein folding research and analytical techniques in the world http://proteinfoldingconsortium.com/)

• Co-founder - Johns Hopkins University Computational Graphics Club (Founder of a club to share visualizations and analysis code in multiple programming languages)

• Invited seminar speaker in the Department of Biochemistry and Molecular Biology, University of MN

Skills Used
Technical skills: Python (major packages: NumPy, SciPy, SymPy, lmfit, MultiProcessing, Matplotlib), Matlab, and Mathematica.

Soft skills: Communication (written and oral), expertise with common office software packages (Microsoft Office - PowerPoint, Excel, Word), and graphics generation

EDUCATION

Doctor of Philosophy in Computational and Molecular Biophysics
Johns Hopkins University - Baltimore, MD
2009 to 2015

Bachelor of Science in Biochemistry/Molecular Biology
University of Minnesota Duluth - Duluth, MN
2009

Bachelor of Arts in Chemistry
University Minnesota - Duluth, MN
2009

SKILLS

Python (5 years), SQL (2 years), Non-linear and Linear Regression (6 years), Data Science (4 years), Natural Language Processing (NLP) (2 years), Neural Networks (1 year), Javascript (Less than 1 year), Java (Less than 1 year)

LINKS

http://www.jakemarold.com

https://www.linkedin.com/in/jdmarold