MARJORIE SAYER DATA SCIENTIST

Palo Alto, CA • marjoriesayer@yahoo.com • http://marjoriesayer.com/about

PROFESSIONAL SUMMARY

- Mathematician + software engineer + published author = data scientist who asks pointed questions, solves complex problems, and conveys meaningful, understandable findings
- Machine learning: regression, trees, neural networks
- Data cleanup, dimension reduction, vectorization, ETL, SQL, R, Python
- Amazon Web Services, Google cloud services, BigQuery, Map Reduce
- Excellent writing, communication, networking skills

EDUCATION

Master of Information & Data Science, UC Berkeley School of Information, 2016 (anticipated)
PhD program in Mathematics & Master of Arts in Mathematics, University of California, Berkeley
Bachelor of Science in Honors Mathematics and Physics, University of British Columbia

EXPERIENCE

STARTUP.ML Fellow 2015

- Machine learning for agile development startup: classifying engineering text (check-ins, issues). Data mining via API query and BigQuery. Data preparation using NLTK and Google word2vec. Contrast/comparison of models: logistic regression, naïve Bayes, gradient-boosted trees, recurrent neural nets. Integrated model with Flask application.
- Financial technology modeling in peer-to-peer lending space. Set up ETLs from online lending APIs to mongoDB. Contrast/comparison of models: logistic regression, gradient-boosted trees. Created an anomaly detection model. Integrated models with ETL using Heroku framework.

UC BERKELEY Student Projects 2015

- Data storage and retrieval: using Amazon web services, establish weeks-long twitter API pipeline to S3 and mongodb; analysis of results using NLP and Amazon EMR; create inferred social-network graphs of public figures using data from Wikipedia, the New York Times Campaign Finance API, news, and high-profile blogs; created graph visualizations (see http://marjoriesayer.com/about/inferred-social-network-project).
- Machine learning with Python Scikit: classification of public data including MNIST, Iris, Mushroom, and SKLearn Newsgroups. Final project: kaggle.com San Francisco Crime Classification; reached top ten percent status (at time) with logistic regression and recurrent neural net classifiers.

NOMINUM Intern 2015

- Data retrieval from hadoop system, analysis and time series visualization using R and Python.
- Computed and graphed moving averages and most-frequent IPs using R and Python.

DEANZA COMMUNITY COLLEGE and COLLEGE OF SAN MATEO Adjunct Faculty, Mathematics 2011 – 2012

- Served as instructor for undergraduate algebra, trigonometry, calculus, statistics.
- Implemented a custom online homework system, trigonaut.com, which administered homework problems and automatically recorded participation and grades.

COMMUNICATIONS CONSULTANT 2002 – 2014 (UC Berkley, Gemini Mobile, Inktomi)

- Created a Python script that culled an error code database from source code, enabling the team to publish error code references with every release. Maintained intranet web pages.
- Authored guides for network infrastructure and search software. Regularly contributed to UI design.
- Wrote API documentation: wrote and debugged C++ code for SDK examples.

PUBLICATIONS AND AWARDS

- Co-author and editor of O'Reilly's *HTTP: The Definitive Guide*.
- D. G. Studer Memorial Award in Quantum Mechanics, University of British Columbia
- National Science and Engineering Research Council of Canada Fellowship