

# AMIT SHAVIT, PHD

<http://shavitamit.github.io> • Boston, MA • (781) 696-8282 • [shavitamit@gmail.com](mailto:shavitamit@gmail.com)

---

## OBJECTIVE

Experienced data scientist with strong communication skills. Passionate about leadership roles, client-facing roles, and technical roles involving machine learning, natural language processing, and deep learning.

## EDUCATION

- Ph.D. in Chemical and Biomolecular Engineering** *University of Pennsylvania* (Philadelphia, PA) June 2015
- Focused on high performance computing and modeling complex systems using Python and C
- B.S. in Chemical Engineering** *University of Massachusetts Amherst* (Amherst, MA) May 2010
- Graduated *summa cum laude* in Chemical Engineering with a minor in Chemistry

## RELEVANT EXPERIENCE

- Thomson Reuters – Data Scientist at Thomson Reuters Labs**, Boston, MA 2016 – Present
- Develop code and solutions across the entire data science spectrum (ETL, analytics, learning, & inference)
  - Collaborate with the business units to define specifications and scope for high-impact data science projects
  - Explore ideas for feature development with partners through user testing and brainstorming workshops
  - Establish an expertise on applying data science methods to natural language and unstructured content
  - Tech stack: Python pandas & SQL for ETL, Python sklearn for ML, PyTorch for deep learning, Vowpal Wabbit for online ML, Matplotlib for data visualization, Python Flask and HTML/CSS/JS for app development
- Thomson Reuters – Data Scientist**, New York, NY 2015 – 2016
- Researched and developed quantitative finance algorithms to analyze the social network “StockTwits”
  - Utilized big data technologies such as Hive and Python pandas to facilitate analysis of 20 GB of tweet data
  - Published & presented manuscript detailing the results of this study in *DSAA 2016* (acceptance rate 21%)
- University of Pennsylvania – PhD Candidate**, Philadelphia, PA 2010 – 2015
- Developed algorithms and code to analyze particles using high performance computing and big data technologies (*e.g.*, C & C++, Python, Linux, CUDA, parallelization, supercomputers/clusters)
  - Published five first-author papers in reputable journals; presented research in 17 local and national venues

## SELECTED DATA SCIENCE PROJECTS

### Deep Learning for Text Summarization

Used recurrent neural networks to create generative summaries of legal documents. Trained word-level and character-level networks using PyTorch on AWS GPUs. Worked with the Legal editorial team to quantify model accuracy.

### Predicting News Significance

Developed methods to correctly predict news relevance to a stock trader analyst with high accuracy using a combination of cross-feature-interactions, word embeddings, and ensemble methods. Models are currently in production.

### Global Energy Leaders Ranking Program

Worked with the business unit throughout the development and launch of a new program to rank energy companies across 25+ metrics using disparate datasets. Utilized Bayesian methods to account for missing-and-correlated data.

### Kaggle Competitions (Currently Ranked Top 5% of Kaggle Users)

Completed collaborative global data science competitions focused on ML, NLP, ensemble methods, and image classification using deep neural networks; achieved honors for finishing in the top 5.5% and top 7% of participants.

## SELECTED LEADERSHIP ACTIVITIES

- CoderDojo – Organizer – promoting coding and programming to the younger generation** 2016, 2017
- Co-organized two CoderDojos (NYC, Boston) each with 20+ participants aged 7-17
- Graduate Student Symposium – Co-President**, University of Pennsylvania 2014
- Organized a symposium to connect industry leaders with graduate students; increased attendance by 50%

## SKILLS

Python, SQL, Linux, Python flask/celery/airflow, C & C++, git, HTML/CSS/JS

*Citizen of the United States of America*