

Roni Kobrosly, Ph.D.

Austin, TX
1-512-965-0601
roni.kobrosly@gmail.com
www.linkedin.com/in/ronikobrosly
https://github.com/ronikobrosly
www.kobrosly.net

Experience

Senior Data Scientist

May 2016 – present

Charles Schwab

- Prototyped multi-input, LSTM, deep neural network with Keras to predict conversion of Schwab website visitors and to perform cross-channel attribution, based on their prospect journey. Productionizing on Google Cloud.
- Replaced \$150k/year Ipsos social media brand health reporting with an in-house tool created in Python. Obtained data through Twitter and Reddit APIs and custom-made news scraping pipeline. Employed topic modeling, sentiment analysis, and anomaly detection algorithms.
- Completed other high-impact projects with HR, customer support, and client experience teams. Employed causal inference modeling, regression analysis, and Bayesian inference.

Data Scientist and Program Director

Sept 2015 – May 2016

Insight Data Science

- Provided technical guidance and machine learning expertise to 66 unique Insight Fellow projects that included time series, supervised and unsupervised learning, and recommendation system components.
- Performed analyses that identified key factors that predicted Fellow success in program.
- Developed a Flask, D3, and Python-based internal dashboard to assess Fellow status and session KPIs.

Fellow

June 2015 – Sept 2015

Insight Data Science

- Created SciClarify.com, a tool to help social scientists improve their chances of publishing in a high impact journal.
- Conducted natural language processing to derive features related to structure and syntax. Classified texts using logistic regression, support vector machines, and random forest. Obtained data through PubMed API, stored data in MySQL, processed with Python pandas.
- Built front-end with Flask, Bootstrap, and D3 and hosted web app on Amazon AWS.

Research Scientist

Jan 2013 – Sept 2014

Pure Earth / The Blacksmith Institute

- Conducted the first-ever meta-analysis of Mexican environmental studies to estimate the true extent of lead exposure in children. Found that average blood lead levels were five times those of children in the US.
- Assisted staff with regression analysis and hypothesis testing for projects. Designed spatial survey methods to estimate the total number of industrial waste sites in Ghana.

Postdoctoral Fellow

Sept 2012 – May 2015

Icahn School of Medicine at Mount Sinai

- Applied machine learning techniques (support vector machines, random forest, multivariate adaptive regression splines, and multiple regression) to identify harmful chemical exposure among children, resulting in a total of 15 publications in top peer-reviewed journals.
- Created MySQL database to organize extensive laboratory data from 200,000+ patients from a large CDC study, which was used by multiple research groups at Mount Sinai, resulting in several publications.
- Predicted 3rd grade academic proficiency among 75,000 school children, with 78% cross-validated accuracy, in cooperation with the New York City Department of Health and Mental Hygiene.

Skills

Languages: Python, R, SQL, HiveQL, SAS, Stata (experienced). Javascript, Scala, C++ (prior exposure).

Tools: IPython, Spark, Docker, pandas, NumPy, SciPy, scikit-learn, Flask, Matplotlib, Rstudio, ggplot2, caret, git, UNIX, D3, Keras, TensorFlow, Amazon Web Services, Google Cloud, BayesiaLab.

Education

Ph.D. in Epidemiology, University of Rochester, 2012
M.P.H. in Epidemiology, University of Michigan, 2008
B.A. in Biological Sciences, Rice University, 2005