Timothy N. Rubin, PhD

Data Science Team Lead
(215) 990-4012
Tim.Rubin@gmail.com
Personal Website | Github Profile |Google Scholar

SUMMARY

I'm a data scientist and research scientist, with a PhD and over 10 years of experience in machine learning and related areas. I love leveraging data to build products that have a positive impact on the world, and to help organizations understand how to best serve their customers. In leadership roles, I strive to create a supportive environment in which people aren't afraid to ask questions, to disagree with me, or to experiment with new ideas. I am passionate about mentorship, and supporting people's skill development and career growth.

PROFESSIONAL EXPERIENCE

DATA SCIENCE TEAM LEAD: CHANGE HEALTHCARE — 2018-PRESENT

- Managed cross-functional team of 5+ data scientists and ML engineers, with duties including technical leadership, mentorship and career growth, and project selection + management
- Built numerous healthcare-related ML products that helped position Change as a leader in healthcare technology
- Helped develop DS project lifecycle standards that were adopted throughout the AI group
- Contributed to a framework for ML pipeline training and deployment, as well as an in-house active learning solution

SENIOR DATA SCIENTIST: SURVEYMONKEY — 2016-2018

- Developed all algorithms underlying SurveyMonkey's *SurveyMonkey Genius* platform—a user-facing ML product that provides customers with predictions about their survey (e.g., estimated completion times) as well as personalized recommendations for how to improve their survey. SurveyMonkey Genius has received industry press as part of SurveyMonkey's 2017 company rebrand, and had significant financial and brand impact.
- Developed a new SurveyMonkey use-case ontology and automated classification model. Applications for this product include sales assistance and driving a variety of personalization opportunities for users.

SENIOR RESEARCH SCIENTIST: INDIANA UNIVERSITY — 2013 – 2016

- Led and collaborated on research projects leading to numerous publications in top-tier journals and conferences, including: (a) automatically identifying functional brain regions using GC-LDA (b) empirically evaluating semantic models, and (c) improving prediction methods for Latent Dirichlet Allocation models

GRADUATE RESEARCH SCIENTIST: UNIVERSITY OF CALIFORNIA, IRVINE — 2006 - 2012

- Developed, implemented, and published probabilistic machine learning models for multi-label document classification, recommendation systems, and leveraging metadata for text clustering

RELEVANT SKILLS

Programming: Python; SQL; AWS; Spark; Docker; Hive; Java (some)

Statistical Analysis Tools:: Python scientific stack (Pandas, SKLearn, etc.) R; MATLAB; QuickSight; SPSS; Excel

Analytical Skills: Machine learning; Research; Experimental Design; Probability theory and statistics; NLP

Communication and Leadership Experience: Experience managing a team of 5+ people; Public speaking; Presentation of applied and theoretical research to both technical and non-technical audiences; Teaching

EDUCATION

University of California, Irvine — Irvine, CA
Ph.D., M.A., Department of Cognitive Sciences — 2012, 2009

Tufts University — Medford, MA
B.S. Cognitive Science — May 2004

SELECTED PUBLICATIONS

Papanikolaou, Y., **Rubin, T.N.**, Tsoumakas, G. (2017) Dense Distributions from Sparse Samples: Improved Gibbs Sampling Parameter Estimators for LDA. *Journal of Machine Learning Research (JMLR)*.

Rubin, T.N., Koyejo, O., Jones, M.N., Yarkoni, Y., (2016). Generalized Correspondence-LDA Models (GC-LDA) for Identifying Functional Regions in the Brain. *30th Annual Conference on Neural Information Processing Systems (NIPS)*.

Rubin, T.N., Kievit-Kylar, B., Willits, J.A., Jones, M.N., (2014). Organizing the Space and Behavior of Semantic Models, *36th Annual Conference of the Cognitive Science Society*.

Rubin, T.N., Chambers, A., Smyth, P., Steyvers, M., (2012). Statistical Topic Models for Multi-Label Document Classification, *Machine Learning: special issue on Learning from Multi-Label Data*.

Rubin, T.N., Steyvers, M., (2009). A Topic Model For Movie Choices and Ratings, *9th International Conference on Cognitive Modeling (ICCM)*, (Supplementary Material)