

# *Machine learning for science and society*

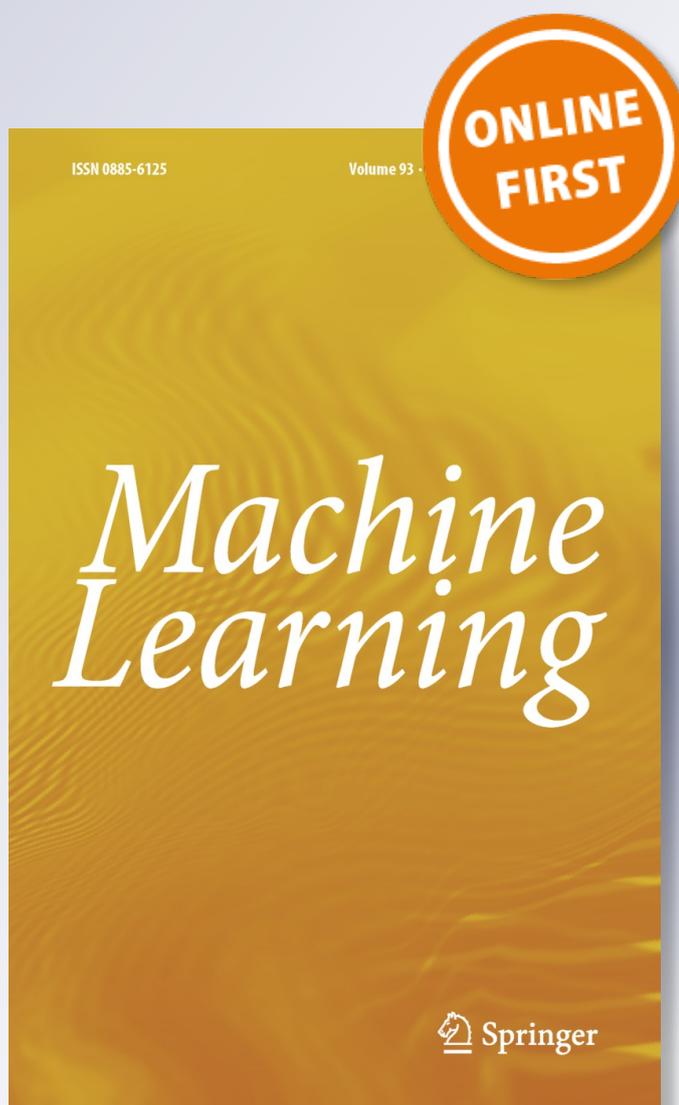
**Cynthia Rudin & Kiri L. Wagstaff**

**Machine Learning**

ISSN 0885-6125

Mach Learn

DOI 10.1007/s10994-013-5425-9



**Your article is protected by copyright and all rights are held exclusively by The Author(s). This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Machine learning for science and society

Cynthia Rudin · Kiri L. Wagstaff

Received: 17 October 2013 / Accepted: 15 November 2013  
© The Author(s) 2013

**Abstract** The special issue on “Machine Learning for Science and Society” showcases machine learning work with influence on our current and future society. These papers address several key problems such as how we perform repairs on critical infrastructure, how we predict severe weather and aviation turbulence, how we conduct tax audits, whether we can detect privacy breaches in access to healthcare data, and how we link individuals across census data sets for new insights into population changes. In this introduction, we discuss the need for such a special issue within the context of our field and its relationship to the broader world. In the era of “big data,” there is a need for machine learning to address important large-scale applied problems, yet it is difficult to find top venues in machine learning where such work is encouraged. We discuss the ramifications of this contradictory situation and encourage further discussion on the best strategy that we as a field may adopt. We also summarize key lessons learned from individual papers in the special issue so that the community as a whole can benefit.

## 1 Why a special issue on impact to science and society?

In this special issue, we showcase machine learning work that addresses problems of importance to science and society. Machine learning (ML) and data mining have been used, and will continue to be used, in many important domains that affect people’s lives every day; however, it is not common in many mainstream machine learning venues to publish work whose primary goal is to have impact on a new real-world problem. The collection of papers

---

C. Rudin (✉)  
Sloan School of Management, Massachusetts Institute of Technology, 100 Main Street, Cambridge,  
MA 02142, USA  
e-mail: [rudin@mit.edu](mailto:rudin@mit.edu)

K.L. Wagstaff  
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena,  
CA 91109, USA  
e-mail: [wkiri@jpl.nasa.gov](mailto:wkiri@jpl.nasa.gov)

in this special issue provides an updated answer to “what is machine learning good for?” in which impact is the guiding principle.

Because machine learning is primarily influencing the broader world through its implementation in a wide range of applications, rather than through its novel specialized algorithms or theory, aspects beyond algorithms and theory can be (and often are) the most important for knowledge discovery. This has been recognized for many years, by many different scientists, and is captured in the knowledge discovery frameworks of KDD and CRISP-DM (Frawley et al. 1992; Chapman et al. 2000) and by others who view the machine learning component as one part of a much larger formalized system for discovering knowledge (Hand 1994; Brodley and Smyth 1997).

Once we graduate from sandbox studies of benchmark data sets, many ML researchers are surprised to discover that differences in performance between individual ML algorithms, which are featured prominently in typical ML publications, diminish in importance. When ML is used in a real application, its success is instead primarily determined by how effectively we understand the unique aspects of the domain and how well we tailor the ML solution and evaluation measures to the domain (Hand 1994; Fayyad et al. 1996; Brodley and Smyth 1997; Saitta and Neri 1998). This special issue presents work that has successfully forged the necessary connections and incorporated domain expertise into the development, evaluation, and deployment of a machine learning system to benefit the larger world.

Some ML researchers may feel that applied work should be published in an applied venue (e.g., a journal devoted to the real-world application being solved by ML). This is a highly effective way to reach domain experts who stand to benefit directly from the ML contribution and who may not otherwise be exposed to ML innovations. However, relegating all applied work to non-ML venues would hurt our ability to learn from colleagues’ advances, create barriers to career advancement, and ultimately discourage ML researchers from investing the time and effort needed to use ML to solve society’s problems. We explore these and other problems with this approach to recognizing, and disseminating, applied ML work in Sect. 4. In the best case, applied work would be published in both venues, appropriately recast for the different audiences.

## 2 Lessons from the review process

For this special issue, we solicited papers describing successful ML efforts that have led to measurable benefits to science and society. In contrast to typical Machine Learning journal expectations, we encouraged but did not require that the work feature a new learning algorithm or theoretical advance. Instead, work would aim to tackle new problems, demonstrate their importance to science and society, and report on lessons learned from the deployment effort. We also encouraged authors to include a relevant domain expert as a co-author.

We received many excellent submissions, and we summarize the seven accepted papers below. These papers went through three revisions on average, and some papers went through five revisions. However, the collection of 27 submissions we received also taught us something about the current state of the machine learning field. While many ML researchers find applications motivating, they also face many barriers to achieving the end goal of measurable impact from their work. Collaboration with domain experts can sometimes be very difficult. Challenging tasks include: forming the right project, stimulating interest from domain experts and funding sources, communicating with domain experts, and attaining sufficient recognition for completing the project. Yet as above, intelligent application of domain knowledge is vital for achieving impact. ML experts cannot solve the world’s problems in isolation.

This special issue required ML researchers to think outside typical ML paper expectations. While we were aware that our field does not primarily reward work based on its impact to science or society, we were surprised to find that some authors were at a loss as to how to write a paper that focuses on measuring impact. We received some submissions that only reported results on benchmark data sets, used generic evaluation metrics, and/or proceeded without any input from domain experts. Some papers provided no discussion of a plan for or evidence of implementation and use of the ML system by the target domain. While acceptable in other ML publishing venues, these papers were not suitable for this special issue.

We also found that success by standard measures of performance (e.g., accuracy, recall, precision) did not always correlate with domain experts' assessment of the results. For this special issue, we obtained reviews from domain experts outside of the author teams to provide a domain-relevant assessment of the work's actual (or potential) impact. In some cases, they validated the importance of the work. In others, the ML metrics were deemed unrelated to functional performance that mattered to the experts. As a field, we should be aware of the dangers of convincing ourselves that we have solved a particular problem based on evidence provided by generic metrics that, while persuasive to an ML colleague, is insufficient for a domain expert.

We were fortunate to be able to recruit many open-minded reviewers for this special issue, and we very much appreciate their time and effort in ensuring the quality of the articles. We provided the reviewers with instructions stating that papers should have scientific novelty, but this novelty need not necessarily manifest in algorithmic or theoretical development. We provided these instructions to address a possible *cultural reviewing bias*: the assumption that a paper that does not contain a novel algorithm, statistical model, or mathematical proof is not substantial enough to merit publication in mainstream machine learning venues such as Machine Learning. In fact, there may be a large class of papers being submitted to machine learning venues that are rejected simply because they are not perceived to be novel, regardless of their potential scientific contribution to the greater good.

We hope that this issue will inspire the reviewers of ML journals and conference proceedings to adopt a modified outlook that is re-awakened to the importance of doing machine learning that matters, an outlook in which reviewers prioritize impact to the world along with algorithmic and theoretical novelty.

### 3 Machine Learning's impact on science and society

We accepted a set of papers that addresses several critical aspects of society: these papers may influence or are influencing the way we perform repairs on fundamental infrastructure (Li et al. 2013), the way we predict severe weather and aviation turbulence (McGovern et al. 2013; Williams 2013), how we conduct tax audits (Kong and Saar-Tsechansky 2013), whether we can detect privacy breaches in access to healthcare data (Menon et al. 2013), how we target advertisements (Perlich et al. 2013), and how we link census datasets to track an individual's career trajectory (Antonie et al. 2013).

We provide a brief summary of each paper and then report key lessons learned from these studies about how best to achieve impact.

- *Water Pipe Condition Assessment: A Hierarchical Beta Process Approach for Sparse Incident Data* by Zhidong Li, Bang Zhang, Yang Wang, Fang Chen, Ronnie Taib, Vicky Whiffin, and Yi Wang

This paper improves risk management of water distribution systems. The authors used machine learning to predict water pipe failures throughout the city of Sydney Australia. They were able to find pipes at high risk of failure before they fail and result in significant service disruption to the community.

- *Enhancing Understanding and Improving Prediction of Severe Weather Through Spatiotemporal Relational Learning* by Amy McGovern, David J. Gagne II, John K. Williams, Rodger A. Brown, and Jeffrey B. Basara

The goal of this work is to predict severe weather events. The authors use spatiotemporal machine learning to understand and predict phenomena such as tornadoes, hail, wind, and aircraft turbulence which annually cause loss of lives and property.

- *Using Random Forests to Diagnose Aviation Turbulence* by John K. Williams

Atmospheric turbulence poses a significant hazard to aviation, with severe encounters costing airlines millions of dollars per year in compensation, aircraft damage, and delays due to required post-event inspections and repairs. The work discussed in this paper is on a path to operations as part of the next update to the FAA's Graphical Turbulence Guidance system.

- *Collaborative Information Acquisition for Data-Driven Decisions* by Danxia Kong and Maytal Saar-Tsechansky

This work creates an active learning method to assist with tax audit decisions. Tax avoidance is a significant problem: a recent IRS estimate suggests that underreporting of income accounts for \$376 billion in lost revenues at the federal level alone. The active learning method of the paper allows multiple learners with different goals to reason collaboratively. This way, learners can acquire informative learning experiences that cost-effectively improve the decisions the models inform. The results of this work show that significantly higher profits can result from effective auditing decisions.

- *Detecting Inappropriate Access to Electronic Health Records Using Collaborative Filtering* by Aditya Krishna Menon, Xiaoqian Jiang, Jihoon Kim, Jaideep Vaidya, and Lucila Ohno-Machado

This paper's focus is on ensuring privacy of healthcare data. The authors use machine learning to detect inappropriate, unauthorized, or illegal access to (and use of) personal information in healthcare data from a hospital. Through their work, investigations based on the learning model's predictions were conducted by the hospital, which imposed sanctions on offending users. In fact, in one case a user faced termination of employment.

- *Machine Learning for Targeted Display Advertising: Transfer Learning in Action* by Claudia Perlich, Brian Dalessandro, Troy Raeder, Ori Stitelman, and Foster Provost

This paper is about personalizing online advertising. Given the huge portion of the US economy devoted to advertising (>2% of US GDP), targeted display advertising is an important domain for machine learning. This paper presents a fully deployed multistage transfer learning system that has been in continual use for years for thousands of advertising campaigns.

- *Tracking People over Time in 19th Century Canada for Longitudinal Analysis* by Luiza Antonie, Kris Inwood, Daniel Lizotte, and J. Andrew Ross

Linking records from different census databases to identify individuals over time allows for greater insights into evolving trends than analyzing each database in isolation. This paper describes a record linkage system that combines Canadian data from the 1870 and 1880 censuses, then reports on how this system has enabled new insights for historians.

These papers tackle a wide diversity of application domains, but in many cases the same obstacles or lessons emerged. Here we highlight some common themes for the benefit of future researchers:

*Communication* Several papers (McGovern et al. 2013; Menon et al. 2013; Williams 2013; Kong and Saar-Tsechansky 2013; Li et al. 2013) suggest that for a true interdisciplinary collaboration, both sides need to understand each other's specialized terminology and together develop the definition of success for the project. We ourselves must be willing to acquire at least apprentice-level expertise in the domain at hand to develop the data and knowledge discovery process necessary for achieving success. However, McGovern et al. (2013) and Menon et al. (2013) note that surprises are still possible, in that intuitions about the domain can sometimes be refuted by data. On the other hand, domain knowledge can also help us better phrase the problems to be solved; Li et al. (2013) give an example where dividing the data into two separate problems corresponding to two geographic regions gives better results than combining them.

*Raw classification accuracy is uninformative for imbalanced problems... and most problems of interest are imbalanced* For any task where the goal is to predict rare occurrences, raw classification accuracy is not a good measurement of overall prediction quality. Most of the papers in this special issue aim to predict something rare, including severe weather events, turbulence, privacy breaches, and equipment failures (Williams 2013; McGovern et al. 2013; Menon et al. 2013; Li et al. 2013). Li et al. (2013) point out that for maintenance applications, only a small portion of the ROC curve (the leftmost part) is useful for practitioners, because maintenance actions can only be taken on a small percentage of the total equipment. This observation holds for many, many real-world problems. One important related point is that, for such problems, area under the curve (AUC) is also an inappropriate metric to use, since it may be dominated by performance regimes that do not matter (Japkowicz and Shah 2011). This is the basis of the "learning to rank" subfield of machine learning, and it has also motivated "partial AUC" evaluations that limit reporting to the operating regime of interest (Dodd and Pepe 2003).

*Going beyond prediction accuracy* One interesting point that Menon et al. (2013) and Kong and Saar-Tsechansky (2013) illustrate is that predictive performance is not always the highest priority for the end-user. Kong and Saar-Tsechansky (2013) discuss the importance of how the model will be used: for tax audits, prediction accuracy is not as important as the amount of money gained from performing audits on non-complying firms. The collaborators of Menon et al. (2013) in a hospital system view the importance of their model as a decision support guiding tool, to combat the most extreme potential violations of privacy.

*Interpretability—no black boxes allowed* For many domains, a predictive model cannot truly be useful unless a human understands it, regardless of how accurate it is. The system of McGovern et al. (2013) exemplifies this, as human forecasters need to deeply understand their model. Tornado warnings are issued by humans, not by a computer, and if a forecaster does not understand the model, they are very unlikely to use it. This sentiment was echoed by statistician Greg Ridgeway, the head of the National Institute of Justice, as he discussed an interpretable scoring system that is used to determine which Los Angeles Police Department recruits had the highest change of becoming officers (Ridgeway 2013). He stated that: "This simplicity gets at the important issue: A decent transparent model that is actually used will outperform a sophisticated system that predicts better but sits on a shelf. If the researchers had created a model that predicted well but was more complicated, the LAPD likely would have ignored it, thus defeating the whole purpose."

*Choosing data carefully* In the paper “Tracking People over Time” (Antonie et al. 2013) the authors’ use of machine learning was to create a data set (a linkage system) that could be used for studying demographic trends. To do this, the authors cautioned against using data in a way that might bias their results for subsequent studies that use the dataset created by the ML system. Attaining low bias data, in this case, induced the authors to make design decisions that are atypical for many machine learning settings. In particular, they needed to explicitly *not* use certain information, which could inadvertently lead to a bias in subsequent studies involving their linkage system. Another perspective comes from Perlich et al. (2013), who discuss that data from outside the exact target domain can be useful. Specifically, they state that “acquiring a large amount of data that is not from the optimal data generating distribution can be better than acquiring only a small amount of data from the optimal data generating distribution.”

#### 4 The bigger context: what is Machine Learning good for?

While it has been a positive experience for us to put together this special issue, and to observe so much energy and enthusiasm from authors and reviewers involved, it is somewhat surprising at this point in time that such a special issue is truly necessary. As things currently stand, it is clear that *our research efforts are not distributed according to the needs of society*. In the communities of machine learning, data mining, and statistics, we spend most of our effort on novel algorithms, novel models, and novel theory, and relatively little effort on the other aspects of the knowledge discovery process, such as data understanding, data processing, feature development, development of new machine learning problems and formulations, practical evaluation and deployment, and how all of these pieces work together to uncover new knowledge in new domains. This continues despite the admonitions of many authors about the dangers of minimizing these essential aspects of the knowledge discovery process (Hand 1994; Fayyad et al. 1996; Brodley and Smyth 1997; Saitta and Neri 1998; Provost and Kohavi 1998; Wagstaff 2012).

Yet the debate continues about whether we should accept decidedly applied papers in the top machine learning publishing venues and whether there is a need for applied work in statistics journals (Peng 2011). Provost et al. (1996) discussed the perceived bias against application papers in their essay regarding decisions made by the program committee at the 1996 International Machine Learning Conference, writing “What we take from the program committee’s decision, given the reviews, is that the important lessons from the real world are not as important as the ability to bundle up the complexity of a real-world problem into a nice, neat conference paper; applied work is interesting only if it looks like good academic work.” This observation is, for the most part, just as apt now, seventeen years later. It means that we as a community do not prioritize impact to science and society.

We are now in the era of “big data,” which is more than a funding buzzword. There is a growing recognition outside the field of machine learning that scalable ML methods can solve problems that are otherwise insoluble or impractical to attack. We stand to benefit as a community by meeting this demand with pioneering machine learning work in modern applications, which can help people beyond our community find solutions to the big data problems they face.

We strongly encourage our top ML publishing venues to find ways to publish more high-quality application papers. If applied research is not considered publishable in top ML venues, our field faces the following disadvantages:

1. We lose the flow of applied problems necessary for stimulating relevant theoretical work (Saitta and Neri 1998; Provost and Kohavi 1998), leaving us in a theoretical and algorithmic echo chamber. An existing example of this is our field's myopic focus on classification accuracy, which is frequently not the metric of importance in a real application (Provost and Kohavi 1998; Wagstaff 2012) but is the main focus of a vast number of academic papers. Provost and Kohavi (1998) pointed out that applied work is part of an important cycle in which the applied world provides important problems, which feeds into principled algorithmic approaches, which in turn helps applications, and that this cycle may already be broken in ML as a community.
2. We further exacerbate the gap between theoretical work and practice. As noted above, some ML researchers may feel that applied work should be published in a corresponding application-specific journal rather than an ML venue. Pushing applied work outside of ML journals or conferences virtually ensures that the advances will not be seen by the ML community, and it can have the deleterious side effect of *forcing* such papers to exclude relevant theory or algorithms, since these aspects may not be of interest to the non-ML audience. It also breaks down mechanisms for communications between fields and, even worse, may make the advances difficult for other ML researchers to understand if they only appear in a highly domain-specific format. Terminology and communication issues between fields are discussed by McGovern et al. (2013). Brodley and Smyth (1997) noted that these "human factors" are absolutely vital for successful application of ML in practice.

We do not suggest that ML venues should be the only repository of applied ML papers. As noted by Peng (2011), sometimes it is better to address a natural applied audience. However, this should not preclude publishing in ML venues. Peng (2011) supposes such a natural audience always exists, which is unfortunately not the case as our next point discusses.

3. We may prevent truly new applications of ML to be published in top venues *at all* (ML or not). For instance, one of the editors of this special issue uses ML to predict manhole fires and explosions on the NYC power grid (Rudin et al. 2010, 2012). There is no existing applied journal that would be a natural fit for this work, since it is a novel application. If we as a community want to encourage the expansion of ML into new application areas, we need to encourage top venues to publish new application papers. Otherwise we artificially limit the reach of our techniques in the world, and at the same time, limit our theoretical advances to accompany only those applied problems that are already established.
4. We strongly discourage applied research by machine learning professionals. For a researcher to work on applied problems with no top avenue for publication leads to problems with career advancement. This could (and already does) present researchers with an unfortunate choice: work on problems that are either important to society or beneficial to their own career, but not both. Further, the dearth of top ML researchers working on applied problems could prevent work done by engineers and companies from being performed by those with a deep understanding of ML; there may simply not be enough ML researchers to go around.

Our goal with this special issue is to raise the profile of excellent ML work currently being done that takes the extra steps needed to develop and implement solutions that make a difference for the world outside of ML. The editors of this special issue have worked on both theoretical and applied topics, where the applied topics between us include criminology (Wang et al. 2013), crop yield prediction (Wagstaff et al. 2008), the energy grid (Rudin et al. 2010, 2012), healthcare (Letham et al. 2013b; McCormick et al. 2012), information retrieval (Letham et al. 2013a), interpretable models (Letham et al. 2013b; McCormick et al.

2012; Ustun et al. 2013), robotic space exploration (Castano et al. 2007; Wagstaff and Bornstein 2009; Wagstaff et al. 2013b), and scientific discovery (Wagstaff et al. 2013a). In our experience, working in applied areas strongly motivates the development of algorithms and theory that can go beyond the single application domain for which they were designed. The academic/applied cycle discussed by Provost et al. (1996) and Provost and Kohavi (1998) is real. We hope that this special issue will help re-invigorate the part of this cycle that links machine learning to real applications, and to encourage researchers who aim to perform work with direct import to science or society.

**Acknowledgements** We would like to thank the reviewers for our special issue who did an excellent job providing high quality, thorough, and relevant reviews. We also thank our colleagues and the editor for helpful suggestions on this editorial. Partial funding for this work was provided by the National Science Foundation under grant IIS-1053407. This work was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Copyright 2013. Government sponsorship acknowledged.

## References

- Antonie, L., Inwood, K., Lizotte, D., & Ross, J. A. (2013). Tracking people over time in 19th century Canada. *Machine Learning, Special Issue on ML for Science and Society*.
- Brodley, C., & Smyth, P. (1997). Applying classification algorithms in practice. *Statistics and Computing*, 7, 45–56.
- Castano, R., Wagstaff, K. L., Chien, S., Stough, T. M., & Tang, B. (2007). On-board analysis of uncalibrated data for a spacecraft at Mars. In *Proceedings of the thirteenth international conference on knowledge discovery and data mining* (pp. 922–930).
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. Tech. rep., SPSS.
- Dodd, L. E., & Pepe, M. S. (2003). Partial AUC estimation and regression. *Biometrics*, 59(3), 614–623.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17, 37–54.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: an overview. *AI Magazine*, 13(3), 57–70.
- Hand, D. J. (1994). Deconstructing statistical questions. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 157(3), 317–356.
- Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: a classification perspective*. Cambridge: Cambridge University Press.
- Kong, D., & Saar-Tsechansky, M. (2013). Collaborative information acquisition for data-driven decisions. *Machine Learning, Special Issue on ML for Science and Society*.
- Letham, B., Rudin, C., & Heller, K. A. (2013a). Growing a list. *Data Mining and Knowledge Discovery*, 27(3), 372–395.
- Letham, B., Rudin, C., McCormick, T. H., & Madigan, D. (2013b). An interpretable stroke prediction model using rules and Bayesian analysis. In *Proceedings of AAAI late breaking track*.
- Li, Z., Zhang, M. B., Chen F, W. Y., Whiffin, V., Taib R, Vicky W, & Wang, Y. (2013). Water pipe condition assessment: a hierarchical beta process approach for sparse incident data. *Machine Learning, Special Issue on ML for Science and Society*.
- McCormick, T. H., Rudin, C., & Madigan, D. (2012). Bayesian hierarchical modeling for predicting medical conditions. *The Annals of Applied Statistics*, 6(2), 652–668.
- McGovern, A., DavidJ Williams J, G. I., Brown, R., & Basara, J. (2013). Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning. *Machine Learning, Special Issue on ML for Science and Society*.
- Menon, A. K., Jiang, X., Kim, J., Vaidya, J., & Ohno-Machado, L. (2013). Detecting inappropriate access to electronic health records using collaborative filtering. *Machine Learning, Special Issue on ML for Science and Society*.
- Peng, R. (2011). Do we really need applied statistics journals? Simply statistics blog. <http://simplystatistics.tumblr.com/post/11655593971/do-we-really-need-applied-statistics-journals>.
- Perlich, C., Dalessandro, B., Raeder, T., Stitelman, O., & Provost, F. (2013). Machine learning for targeted display advertising: transfer learning in action. *Machine Learning, Special Issue on ML for Science and Society*.

- Provost, F., & Kohavi, R. (1998). Guest editor's introduction: on applied research in machine learning. *Machine Learning*, 30, 127–132.
- Provost, F., Fawcett, T., Danyluk, A., & Riddle, P. (1996). On the value of applied research in machine learning. [http://home.comcast.net/~tom.fawcett/public\\_html/papers/essay.html](http://home.comcast.net/~tom.fawcett/public_html/papers/essay.html).
- Ridgeway, G. (2013). The pitfalls of prediction. *NIJ Journal*, 271, 34–40.
- Rudin, C., Passonneau, R., Radeva, A., Dutta, H., Ierome, S., & Isaac, D. (2010). A process for predicting manhole events in Manhattan. *Machine Learning*, 80, 1–31.
- Rudin, C., Waltz, D., Anderson, R. N., Boulanger, A., Salleb-Aouissi, A., Chow, M., Dutta, H., Gross, P., Huang, B., Ierome, S., Isaac, D., Kressner, A., Passonneau, R. J., Radeva, A., & Wu, L. (2012). Machine learning for the New York City power grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2), 328–345.
- Saitta, L., & Neri, F. (1998). Learning in the “real world”. *Machine Learning*, 30, 133–163.
- Ustun, B., Tracà, S., & Rudin, C. (2013). Supersparse linear integer models for predictive scoring systems. In *Proceedings of AAAI late breaking track*.
- Wagstaff, K. L. (2012). Machine learning that matters. In *Proceedings of the twenty-ninth international conference on machine learning* (pp. 529–536).
- Wagstaff, K. L., & Bornstein, B. (2009). K-means in space: a radiation sensitivity evaluation. In *Proceedings of the twenty-sixth international conference on machine learning* (pp. 1097–1104).
- Wagstaff, K. L., Lane, T., & Roper, A. (2008). Multiple-instance regression with structured data. In *Proceedings of the 4th international workshop on mining complex data*.
- Wagstaff, K. L., Lanza, N. L., Thompson, D. R., Dieterich, T. G., & Gilmore, M. S. (2013a). Guiding scientific discovery with explanations using DEMUD. In *Proceedings of the twenty-seventh conference on artificial intelligence*.
- Wagstaff, K. L., Thompson, D. R., Abbey, W., Allwood, A., Bekker, D. L., Cabrol, N. A., Fuchs, T., & Ortega, K. (2013b). Smart, texture-sensitive instrument classification for in situ rock and layer analysis. *Geophysical Research Letters*, 40.
- Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2013). Detecting patterns of crime with Series finder. In *Proceedings of the European conference on machine learning and principles and practice of knowledge discovery in databases (ECML-PKDD 2013)*.
- Williams, J. K. (2013). Using random forests to diagnose aviation turbulence. *Machine Learning, Special Issue on ML for Science and Society*.