EDITOR'S COMMENTS ▰▰▰▰▰▰

The Business of Business Data Science in IS Journals ▰▰▰▰▰▰

By: **Maytal Saar-Tsechansky**
        The McCombs School of Business
        The University of Texas at Austin
        2110 Speedway
        Austin, TX 78712
        maytal@mail.utexas.edu

Introduction

Data science is at the core of a host of ongoing business transformations and disruptive technologies. The application of data science methods to new and old business problems presents a wealth of research opportunities upon which the information systems (IS) data science community is uniquely positioned to focus. Strong demand for business data science education by both students and recruiters has also given rise to new business analytics programs, often lead by IS groups. This mission will be well served by an active business data science research within IS. While contributions by IS data science scholars are being published in the premier reference discipline journals, and although such a relationship with the reference disciplines is immensely important to maintain, it may be difficult to sustain IS data science research without a healthy presence in the leading IS journals. Furthermore, IS journals are arguably best positioned to publish research at the nexus of data science, business, and society. The objective of this editorial is to inspire a discussion on opportunities to facilitate IS data science reviewing and publication in top IS journals, and to capitalize on these opportunities effectively as a community. This discussion is based on the understanding that both writing and reviewing practices are fundamental to our ability to assess the quality of data science contributions and to draw upon and publish high-quality and impactful research. These practices also affect the potential to sustain and grow an impactful data science community. Toward that end, this editorial also hopes to initiate a discussion on sustaining and growing a data science community within IS.

The focus of data science, the extraction of informative patterns from data, differs from that of other streams of IS research, and thus calls for different guidelines to assess it. Because data science is a design science field of research, the guidelines outlined by Hevner et al. (2004) can be applied to produce and assess data science contributions. However, given data science is a broad, interdisciplinary field, and perhaps also because data science contributions are fairly recent in IS, there remains some uncertainty on how to apply the guidelines to data science in particular. It is beneficial, therefore, that we develop recommendations for authors and review teams on aspects of data science where uncertainty arises.

What Constitutes a Meaningful and Significant Data Science Contribution in IS?

The work by Hevner et al. (2004) offers a framework for determining what constitutes a significant design science contribution, and this framework applies to data science as well. Below we look at particular aspects of data science where additional discussion may be beneficial.

The development and evaluation of the IT artifact is the focus of design science research (Hevner et al. 2004). For data science in particular, it is useful we acknowledge the importance of contributions that focus exclusively on the evaluation and analysis of *existing* data science methods. Advances in data science research rely heavily on our understanding of the properties of existing methods. Such research includes studies of the performance of important methods under conditions of interest, or contributions

that advance our understanding of the causes for variation in performance using analytical or empirical frameworks. Advancing our knowledge of the settings under which existing methods are expected to yield better outcomes than alternatives, or of important pathologies and the conditions under which they arise, constitute significant contributions with implications to both practice and future research. Research that develops new methods includes evaluations and analyses of the key properties of the method; however, the complexity of settings to which methods can be applied and the different objectives they can be used to achieve are such that it is rarely reasonable for a single contribution to study all these aspects of a method. Similarly, analyses of existing methods are valuable in offering new insights that *explain* meaningful variation in performance. To date, such contributions have had significant impact because they provide important guidelines on applying the methods in practice, or because the insights can be built on in future research (e.g., Friedman 1997; Friedman et al. 2000).

In business, such contributions are valuable if they advance our understanding of the performance of important methods when they are applied to inform important business problems. For example, the appeal and popularity of tree-based predictive models in practice and the common challenge of using them to generate predictions for instances with missing values have motivated studying the performance of popular and less-known missing value treatments when applying tree-based predictive models (Saar-Tsechansky and Provost 2007). This study drew the attention of scholars and practitioners to a less-known and rarely used approach that performs remarkably well, and it advanced our understanding of the settings under which this approach is most advantageous. Similar publications in IS journals ought to also explore how the reported variations in performance bear potentially important business implications.

Indeed, IS data science is uniquely positioned to focus on challenges at the nexus of data science, business, and society. Insightful observations on the shortcomings of state-of-the art methods to address old and new business challenges can give rise to new data science problems and research that addresses them (Kong and Saar-Tsechansky 2013; Prawesh and Padmanabhan 2014; Sheng et al. 2008). However, some uncertainty exists in determining whether a research study that improves an existing problem by a novel use of an exiting solution, or by a novel combination of methods, constitutes a significant contribution. Data science contributions that make use of existing methods in novel ways are significant if they are derived from nontrivial insights and can potentially impact practice or research in meaningful ways. In particular, novel contributions are trivial when they are likely to be developed and deployed in practice, independently of publishing the contribution. It is also useful for the review team to establish that the contribution can be potentially impactful such that the ideas improve the solution to an important problem, or that future research can build on these ideas.

The IS data science community may also benefit from a focus on emerging or future environments, such as future education or energy systems (Peters et al. 2013; Bauman and Tuzhilin 2015). Forward-looking research offers problem-rich domains with the potential for significant impact. Using real business data for evaluating this research, however, may not always be meaningful or possible. In assessing such contributions, it is useful that a review team acknowledges that evaluating methods designed for future environments may entail artificial construction of key properties of these settings, such as via simulation.

Note that it may also not always be reasonable to embed a proposed method in its intended setting for evaluation purposes, even if such a setting already exists. These evaluations may involve significant risk, as in, for example, health care applications (Meyer et al. 2013), or require cooperation with commercial or government entities. Securing cooperation can be invaluable; however, it would be unfortunate if this poses a hurdle for publishing the research, because this might draw us away from advancing research based on its merit and potential impact. In such cases as well, appropriate evaluations can be offered in carefully simulated settings (Atahan and Sarkar 2011; Sheng et al. 2008).

Our community's business and organizational focus positions us uniquely to identify new opportunities to advance data science. It is useful we acknowledge the objectives and appreciate the challenges and potential impact of this research in order to advance and publish quality contributions.

Reviewing and Writing of Data Science Contributions

The work by Hevner et al. (2014) and Gregor and Hevner (2013) developed a thorough set of guidelines on the elements that make up design science contributions. These thoughtful guidelines apply to data science as well. Reviewers who follow these

guidelines, however, are also faced with the challenge of applying them to evaluate contributions, and of effectively communicating the critique to yield a productive review process that improves the research. Below, we discuss some data science research writing and review practices that may aid the assessment of data science contributions and can facilitate the publication of impactful business data science research in top IS journals.

Recommendations for Reviewers and Editors

Perhaps the most basic element of a critique is that it is well-founded and clearly explained. It is important, therefore, that we strive to offer critiques that are carefully established, detailed, and clearly communicated. Such attention is important for communicating the judgments to authors and editors. However, perhaps particularly because of the technical nature of data science, this also presents a useful check for the review team. For example, consider that reviewers find that (some of) the claimed contributions have already been established in prior work. In addition to listing relevant references, a critique is well-established if it details what aspects of the referenced method(s) achieve the claimed contributions, as well as how. It is similarly important that reviewers' key judgments, such as on the significance of contributions, are communicated by outlining the principles guiding the judgment, so that it is clear how the judgment is derived from them. Doing so consistently is invaluable to our review process and culture. This is because it offers authors unambiguous insights on which they can build, and it offers reviewers an opportunity to revisit prior work and the principles underlying the judgments to confirm the critique is correct and clearly communicated. Such careful attention to details is equally invaluable to editors' efforts to critically evaluate reviewers' comments, and to form their decisions.

An editor's proactive leadership in guiding the review team can prove important to sustain a reviewing culture for data science contributions. Communication between senior and associate editors, and between editors and reviewers, is particularly essential when a relatively new stream of research becomes part of the landscape of a discipline. It is useful that editors consider interacting with reviewers before the paper is assigned, and discuss the critique after the reviews are in, should the need arise to improve the clarity or content of the critique.

It is also worth considering what review team composition can best deliver a careful critique of data science contributions. Note that data science is a broad, interdisciplinary field, and includes contributions from such disciplines as databases, artificial intelligence, machine learning, and statistics. Given this breadth of disciplines, it is useful that judges of contributions have strong relevant expertise, and, even better, research interest in the subject area. Expertise is necessary to appreciate the data science challenge the paper addresses and to critically evaluate the contributions. Editors' expertise is important as well in order to critically evaluate reviewers' substantive comments, and for guiding the review process in general. When editorial board members with relevant data science expertise are unavailable to take on a submission, it is beneficial to consider *ad hoc* expert editors to oversee the review, or the inclusion of experts in the editorial board that reflect the breadth of data science research in IS.

Recommendations for Authors

Authors of data science contributions in IS also have opportunities to facilitate the assessment by IS journals of their contributions. It is essential that contributions are clearly communicated (Hevner et al. 2004). In addition, it is beneficial that authors carefully discuss two key aspects of their work that are often overlooked by authors, and yet are important to facilitate reviewers' assessment of the contribution. First, the review team's assessment of contribution is greatly facilitated if authors state explicitly (preferably in the introduction) what the claimed contributions are. A clear statement minimizes ambiguity regarding the claimed contributions, and this clarity allows the review team to focus its efforts on establishing whether the claimed contributions are significant and on how convincingly the research establishes these contributions.

It is also important that authors of data science contributions discuss the rationale for the choice of prior work against which empirical or analytical evaluations are presented. Such rationale should to be based on how the subsequent evaluations offer evidence toward all or some of the claimed contributions. As before, such clarity facilitates the review team's assessment of the adequacy of authors' choices. The review team may similarly suggest that an alternative rationale be adopted instead, and discusses how such evaluations are expected to improve the evaluation of the claimed contributions.

Closing Remarks

The possible use of data science methods to address new and old business problems presents a wealth of research opportunities that the IS data science community is well positioned to explore. Drawing more quality data science contributions and facilitating their publication in leading IS journals has the potential to enrich our field, well serve our students and stakeholders, and may extend our community's impact. We hope that this editorial inspires a discussion on opportunities for authors, reviewers, and editors to facilitate meaningful assessment of data science research contributions in important business contexts. More than the comments made here, a discussion is crucial to develop shared understandings and practices that our community will embrace. Whatever the conclusions of this discussion might be, having it will invariably better us as a community and benefit our mission.

References

Atahan, P., and Sarkar, S. 2011. "Accelerated Learning of User Profiles," *Management Science* (57:2), pp. 215-239.

Bauman, K., and Tuzhilin, A. 2015. "Recommending Learning Materials to Students by Identifying Their Knowledge Gaps," Working Paper, Stern School of Business, New York University.

Friedman, J. H. 1997. "On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality," *Data Mining and Knowledge Discovery* (1:1), pp. 55-77.

Friedman, J., Hastie, T., and Tibshirani, R. 2000. "Additive Logistic Regression: A Statistical View of Boosting," *The Annals of Statistics* (28:2), pp. 337-407.

Gregor, S., and Hevner, A. R. 2013. Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.

Kong, D., and Saar-Tsechansky, M. 2014. "Collaborative Information Acquisition for Data-Driven Decisions," *Machine Learning* (95:1), pp. 71-86.

Meyer, G. , Adomavicius, G., Johnson, P., Elidrisi, M., Rush, W., Sperl-Hillen, J., and O'Connor, P. 2014. "A Machine Learning Approach to Improving Dynamic Decision Making," *Information Systems Research* (25:2), pp. 239-263.

Peters, M., Ketter, W., Saar-Tsechansky, M., and Collins, J. 2013. "A Reinforcement Learning Approach to Autonomous Decision-Making in Smart Electricity Markets," *Machine Learning* (92, pp. 5-39.

Prawesh, S., and Padmanabhan, B. 2014, "The 'Most Popular News' Recommender: Count Amplification and Manipulation Resistance," *Information Systems Research* (25:3), pp. 569-589.

Saar-Tsechansky, M., and Provost, F. 2007. "Handling Missing Values When Applying Classification Models," *Journal of Machine Learning Research* (8), pp. 1623-1657.

Sheng, V., Provost, F., and Ipeirotis, P. G. 2008. "Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers," in *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, August 24-27.