Social Aspects of Machine Learning Model Evaluation: Model Interpretation and Justification from ML-practitioners' Perspective

Victoria Zakharova[a] and Alena Suvorova[a]

[a] *HSE University, 3A Kantemirovskaya Street, St Petersburg, 194100, Russia*

Abstract

Machine Learning (ML) is now widely applied in various life spheres. Experts from different domains become involved in the decision-making on the basis of complex machine learning models that causes in-creased interest in the research in model explainability. However, little is known about the ways that ML-practitioners use to describe and justify their models to others. This work aims to fill the research gap in understanding how data specialists evaluate machine learning models and how they communicate results to third parties. To explore that, the qualitative research design is suggested and semi-structured interviews with ML-practitioners are conducted. The decision-making process will be explored from a sociological perspective according to which data specialists are considered as actors who tend to construct knowledge rather than passively take it. The potential result of this work is to reveal the role of data specialists in model explanation and justification and describe methods they could use to explain complex models to domain experts with non-technical backgrounds.

Keywords

Machine Learning, Algorithm Evaluation, Knowledge Sharing

1. Introduction

Digitalization promotes innovations and facilitates a process of globalization. With that, ongoing digital transformation causes the emergence of new tasks together with new methods for their solutions which are rarely clear for a wide audience but accepted since they provide solutions for urgent issues [1]. This tendency is noticeable in the applied domains when medium-size companies, large corporations, and small start-ups appeal to non-traditional digital solutions to present unique values of their works to strengthen competitiveness and take an outstanding position among the other market players. Data-driven approaches have achieved their recognition in customer-oriented settings that are thought to have an impact on society and its characteristics causing far-reaching effects [2]. For example, the banking sphere has changed with the help of the implementation of chat-bots based on machine learning algorithms, that give answers to clients quicker or send personalized notifications that are also already used in such industries as retail [3], healthcare [4], and insurance [5].

As one of the consequences, being motivated by the up-growing demand for analytical expertise at the labour market some people adhere to follow trends and take roles of problem-solvers to deal with latter-day challenges [6]. Expanding knowledge to boost expertise and diving into the data science sphere, such roles become diverse and barely clearly defined due to uncertainty. Moreover, specialists have to collaborate with each other to reach the commonly established goals such as releasing new digital products or upgrading existing infrastructure with advanced algorithms. Simplifying the concepts, model builders, model breakers, and consumers can be distinguished [7]. Considering a ground stage of the technological development and knowledge formation about that, the first two mentioned roles are taken by actors who are interested in facilitating innovations initially and make

IMS 2021 - International Conference "Internet and Modern Society", June 24-26, 2021, St. Petersburg, Russia
EMAIL: vv.vict26@gmail.com (A. 1); suvalv@gmail.com (A. 2)
ORCID: 0000-0003-3641-6326 (A. 1); 0000-0002-5392-4683 (A. 2)

decisions based on data introducing state-of-the-art project results to the mass. Specifically, data specialists have not only to develop and evaluate statistical models such as machine learning ones but also come to an agreement with stakeholders who are far from direct work with mathematical algorithms, but they are who can ensure the promotion and assistance of the further project realization.

Motivation for conducting this research is based on the recent increase in scientific papers that emphasize the importance of machine learning practitioners’ expertise dealing with algorithms promoting transparency analysis as an integral part of the work with algorithms and, by this, paying precise attention to a need for clarification of the data- and machine-learning-based solutions to providing understanding for all involved parties [7, 8, 9, 10]. As for a potential work contribution, this works attempts to present theoretical justification of model evaluation process with consideration of practices of model interpretation and sharing the knowledge to the other involved actors supported with qualitative data provided by machine learning practitioners seeing a case from their perspective.

2. Related Research and Problem Statement

Presently, there is plenty of research papers proving scientific interests towards data-driven innovations from the managerial, economical, and social perspectives. One of the research themes is related to studying a working process performed by data practitioners. In particular, these research are more focused on data-oriented skill-set [11], data science role division [12, 13, 14], team collaborations [15, 16], tools and practices within the workflow with the notion of practical settings [17, 18, 19]. In addition, other studies focus on the role of explanation in decision-making revealing that data specialists tend to trust algorithms too much and make decisions in a biased manner [20, 21]. However, little is known how data specialists, who implement complex models (i.e., machine learning ones), evaluate models in non-academic settings, and how they translate the obtained information to the others involved directly or indirectly in their work. Actually, several studies related to that issue have been aimed at a direction of practitioner’s work investigation, but they are much more empirical rather than theoretically justified [7], and experts’ needs and opinions about interpretability is rarely provided.

Thus, this research is aimed at studying practices (i.e., practical actions based on the real-life working experience) of data science specialists with the focus on the model evaluation stage and communicating their knowledge about model quality and other characteristics with the third parties. The following research questions are proposed: How do specialists perform model evaluation and what they pay attention at? How do data practitioners explain complex statistical models to other people without a deep understanding of data science principles and ideas? The relevance of studying this issue stems from the idea that data-experts are the first who interact with algorithms, who have specific knowledge to understand them, and their decisions are initial for promoting the use of algorithms in production, which might have a significant impact on society over time [22, 23].

3. Research Design

In the framework of this research, the description of the practical work of specialists is planning to be supported with empirical data collected via semi-structured interviews with practitioners working in different business spheres with data-intense applications. An interview-based approach is used to understand experience, positions, attitudes, and to know opinions of industry practitioners who are direct guides to the world of technology [24]. Variability sample or, in other words, interviewing practitioners from various domains is thought to be applicable for reviewing common (domain-independent) patterns and discrepancies to provide explanations of the performed actions and formed viewpoints with the help of shared real-life. As for sampling technique, convenient and snowball samplings were performed, and, as a result, 16 interviews with 11 men and 5 women have been conducted. The main criteria for recruiting participants were that they had to have at least one year of practical experience in the industry, as well as they had to practice machine learning algorithms for problem-solving at their work.

The obtained results will be analyzed with the help of thematic qualitative analysis in order to explore the general case from the perspective of the applied theoretical framework. In detail, this work is planning to be based on theory in order to justify its results by grounded interpretation of empirics.

As for the theory, a concept of "worlds" introduced by Boltanski and Thévenot in 2006 [25] is chosen for the elaboration of data practitioners' work. According to that, there are a few "worlds" or ways of thinking related to how people and objects dwell together being guided by their own interests, intentions, and perception of particular issues. These "worlds", that are prone to experience conflicts, reach compromises, and collaborate on justification, are the following: inspired, fame, domestic, civic, market, and industrial. Taking into account a fact that data scientists generally work in a business sphere, an idea that these practitioners have to work together not only with each other but also with managers and stakeholders that are more likely to be related to the other "worlds", especially market one, seems to be straightforward.

4. Plans and Preliminary Results

Further plans of this research are mainly focused on data analysis to obtain justified answers to the research questions. In beforehand, findings emphasize the difficulty of contacting a few "worlds". Precisely, data practitioners actually evaluate models with the help of mathematical metrics that are understandable for them, and further, they have to consider interests of the others such as managers who are more likely to concern about financial payoffs and stakeholders who decide whether they should invest to an ML-based project or not. The situation becomes more complicated when there is a necessity to review the models in a social context (e.g., whether obscene content that was unblocked by mistake is causing moral injury to users). In addition, data practitioners support the idea that one of the managerial purposes is to sell projects reeling in superiors. Moreover, sometimes managers can attempt to take part in market tenders offering technical solutions that hardly can be realized by data specialists. In general, these insights strengthen the idea that there is a high need in building effective communication between the "worlds" to inform about the capabilities of each of the parties, in particular converting mathematical metrics to business ones to demonstrate the efficiency and potential benefits justifiably.

As for interpretable machine learning methods (which appear to be one of the highly debatable topics in data science communities), several practitioners mentioned the usefulness of such tools for revealing model transparency with a certain degree of confidence since there were cases when they helped to define which model would be better in terms of its algorithm or even elaborate on a project case considering it step-by-step making representation of the work easier for experts from the other "worlds". The others pointed that they did not use interpretable machine learning methods in their project workflow since they are not worth it: strict explanations are required by stakeholders but computationally and timely expensive.

5. Acknowledgements

The work is supported by the Russian Science Foundation grant (project No. 19-71-00064).

6. References

[1] J.J. Kassema, Products and Services Improvement through Innovation and Creativity: Case of IT Business Sector. Social Science Research Network, Rochester, NY, 2019. https://doi.org/10.2139/ssrn.348581111.

[2] A. Mugrauer, J. Pers, Marketing managers in the age of AI: A multiple-case study of B2C firms, 2019.

[3] T. Calle-Jimenez, B. Orellana-Alvear, R. Prado-Imbacuan, GIS and User Experience in Decision Support for Retail Type Organizations. In: 2019 International Conference on Information Systems and Software Technologies (ICI2ST), 2019, pp. 156–161. https://doi.org/10.1109/ICI2ST.2019.00029.

[4] L. Syed, S. Jabeen, S. Manimala, H.A. Elsayed, Data Science Algorithms and Techniques for Smart Healthcare Using IoT and Big Data Analytics. In: Mishra, M.K., Mishra, B.S.P., Patel, Y.S., and Misra, R. (eds.) Smart Techniques for a Smarter Planet: Towards Smarter Algorithms,

Springer International Publishing, Cham, 2019, pp. 211–241. https://doi.org/10.1007/978-3-030-03131-21124.
[5] A. Singh, K. Ramasubramanian, S. Shivam, Building an Enterprise Chatbot: Work with Protected Enterprise Data Using Open Source Frameworks. Apress, 2019.
[6] S. Kampakis, Problem Solving. In: S. Kampakis (ed.) The Decision Maker’s Handbook to Data Science: A Guide for Non-Technical Executives, Managers, and Founders. Apress, Berkeley, CA, 2020, pp. 89–95.
[7] S.R. Hong, J. Hullman, E. Bertini, Human Factors in Model Interpretability: Industry Practices, Challenges, and Needs. Proc. ACM Hum.-Comput. Interact. 4,1–26, 2020. https://doi.org/10.1145/33928789.
[8] C. Molnar, G. Casalicchio, B. Bischl, Interpretable Machine Learning A Brief History, State-of-the-Art and Challenges. arXiv:2010.09337 [cs, stat], 2020.
[9] W.J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, B. Yu, Interpretable machine learning: definitions, methods, and applications. Proc Natl Acad Sci USA.116, 2019, pp. 22071–22080. https://doi.org/10.1073/pnas.190065411616.
[10] H. Suresh, S.R. Gomez, K.K. Nam, A. Satyanarayan, Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and their Needs. arXiv:2101.09824 [cs], 2021. https://doi.org/10.1145/3411764.344508823.
[11] T. Stadelmann, K. Stockinger, G. Heinatz B¨urki, M. Braschler, Data Scientists. In: Braschler, M., Stadelmann, T., and Stockinger, K. (eds.) Applied Data Science: Lessons Learned for the Data-Driven Business, Springer International Publishing, Cham, 2019, pp. 31–45. https://doi.org/10.1007/978-3-030-11821-1322.
[12] S. Ba˘skarada, A. Koronios, Unicorn data scientist: the rarest of breeds. Program, 51, 2017, pp. 65–74. https://doi.org/10.1108/PROG0720160053.
[13] M. Kim, T. Zimmermann, R. DeLine, A. Begel, Data Scientists in Software Teams: State of the Art and Challenges. IEEE Transactions on Software Engineering. 44, 2018, 1024–1038. https://doi.org/10.1109/TSE.2017.275437413.
[14] J.S. Saltz, N.W. Grady, The ambiguity of data science team roles and the need for a data science workforce framework. In: 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 2355–2361. https://doi.org/10.1109/BigData.2017.825819019.
[15] A.Y. Wang, A. Mittal, C. Brooks, S. Oney, How Data Scientists Use Computational Notebooks for Real-Time Collaboration. Proc. ACM Hum.-Comput. Interact., 3, 2019, 39:1-39:30. https://doi.org/10.1145/335914125.
[16] A.X. Zhang, M. Muller, D. Wang, How do Data Science Workers Collaborate? Roles, Workflows, and Tools. Proc. ACM Hum.-Comput. Interact., 4, 2020, 022:1-022:23. https://doi.org/10.1145/3392826.
[17] N. Boukhelifa, M.-E. Perrin, S. Huron, J. Eagan, How Data Workers Cope with Uncertainty: A Task Characterisation Study. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2017, pp.3645–3656. https://doi.org/10.1145/3025453.3025738.
[18] A. Crisan, B. Fiore-Gartland, M. Tory, Passing the Data Baton: A Retrospective Analysis on Data Science Work and Workers. IEEE Transactions on Visualization and Computer Graphics, 27, 2021, 1860–1870. https://doi.org/10.1109/TVCG.2020.3030340.
[19] P. Pereira, J. Cunha, J.P. Fernandes, On Understanding Data Scientists. In: 2020IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2020, pp. 1–5 https://doi.org/10.1109/VL/HCC50065.2020.912726918.
[20] D.-A. Ho, O. Beyan, Biases in Data Science Lifecycle. arXiv:2009.09795 [cs], 2020.
[21] H. Kaur, H. Nori, S. Jenkins, R. Caruana, H. Wallach, J. Wortman Vaughan, Interpreting Interpretability: Understanding Data Scientists’ Use of Interpretability Tools for Machine Learning. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–14. https://doi.org/10.1145/3313831.337621912.
[22] U. Garzcarek, D. Steuer, Approaching Ethical Guidelines for Data Scientists. In: Bauer, N., Ickstadt, K., L¨ubke, K., Szepannek, G., Trautmann, H., and Vichi, M.(eds.) Applications in Statistical Computing: From Music Data Analysis to Indus-trial Quality Improvement, Springer

International Publishing, Cham, 2019, pp. 151–169. https://doi.org/10.1007/978-3-030-25147-510.
[23] S. Passi, S.J. Jackson, Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. Proc. ACM Hum.-Comput. Interact. 2, 2018, pp. 1–28 https://doi.org/10.1145/327440517.
[24] N. Seaver, Algorithms as culture: Some tactics for the ethnography of algorithmic systems. Big Data & Society, 4, 2053951717738104, 2017. https://doi.org/10.1177/205395171773810420.
[25] L. Boltanski, L. Th´evenot, On Justification: Economies of Worth. Princeton University Press, Princeton, 2006.