



Editorial issue 3 + 4, 2023

**Florian Dumpert · Sebastian Wichert · Thomas Augustin ·
Nina Storfinger**

Accepted: 21 November 2023 / Published online: 7 December 2023
© The Author(s) 2023

This special issue of ASStA Wirtschafts- und Sozialstatistisches Archiv with two volumes aims to offer an overview of some relevant quality aspects of machine learning methods, especially but not only when applied to official statistics.

Triggered by advances in new possibilities for analysis and insight, and thus enabling higher-quality statistics, machine learning methods have gained strong momentum in recent years. Machine learning techniques are convincing with regard to more accurate classification, regression, and clustering, especially against the background of new types of data (sources), some of which are immense in scope. However, when these algorithms are applied to official statistics it has not yet been generally clarified whether and how the typically prediction-oriented, non-model-based approach of machine learning methods can be reconciled with the special quality requirements and framework conditions of official statistics as formulated, for instance, in the European Code of Practice. One of these requirements is that the statistical processes for collecting, processing and disseminating statistics fully comply with international standards and guidelines and at the same time reflect the current state of scientific research.

✉ Florian Dumpert
Statistisches Bundesamt, Gustav-Stresemann-Ring 11, 65189 Wiesbaden, Germany
E-Mail: florian.dumpert@destatis.de

Sebastian Wichert
ifo Institut, Poschinger Str. 5, 81679 München, Germany

Thomas Augustin
Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 München, Germany

Nina Storfinger
Bayerisches Landesamt für Statistik, Nürnberger Straße 95, 90762 Fürth, Germany

In order to shed more light on different aspects of this problem and to find first answers, we organized a scientific workshop held in September 2022 in Munich. The workshop enabled the exchange between official statisticians working on applications of modern machine learning methods and researchers from the field of machine learning.¹ Due to the wide range of participants, it was possible to look at the quality aspects of machine learning from different perspectives: organizational and technical as well as conceptual and methodological. In addition, the user's perspective was also taken into account. Furthermore, the reflections of the official statisticians on the concrete quality dimensions of the use of machine learning were discussed with international guests from the Office for National Statistics (ONS) as well as Statistics Netherlands (CBS).

Already during the organization of the workshop, we developed the idea of bringing together approaches and document current work on the topic of quality aspects of machine learning methods in a special issue. We are pleased that we can now publish Volume I of this special issue continuing and deepening the discussion at the workshop and enriching it with further contributions. Volume II of this special issue will be published as paper version in summer 2024. Already accepted contributions will be successively made available at <https://link.springer.com/journal/11943/online-first>.

The Volume I of the special issue contains the following six articles showing from different perspectives how the preparation, implementation and application of machine learning techniques raise critical questions, in official statistics and far beyond it.

While the use of machine learning in official statistics is no longer new territory, there are still open points that need to be clarified for quality-assured use. The statistical offices have also made observations in recent years that point to pitfalls or need for further development. Van Delden et al. (2023) wrote down their experiences from practical work at Statistics Netherlands in the form of ten propositions, thereby contributing to a reflective approach to machine learning in official statistics. Their particular focus is on explainability, deployment, inference, and quality aspects in general.

The paper by Molladavoudi and Yung (2023) also addresses explainability and looks in addition in greater depth at the field of uncertainty measurement. Here, too, a special focus is placed on official statistics. The authors examine in detail approaches to fulfilling these interrelated quality dimensions and, in this context, also design conceivable quality indicators for official statistics. They also discuss the integration of subject matter experts into automated production processes in statistical offices. Molladavoudi and Yung thus make a targeted contribution to implementing the abstract quality dimensions in the production processes of official statistics.

Indeed, official statistical offices in Europe and Germany follow a strict codified and internationally harmonized quality framework when producing official statistics, which is reviewed regularly. However, the more frequent use of machine learning methods in official statistics requires a concretization of these quality dimensions.

¹ The program of the workshop can be viewed online at the following link: <https://www.statistiknetzwerk.bayern.de/service/aktuelles/maschinellen-lernens/>.

Saidani et al. (2023) contribute to this task by applying traditional quality principles to machine learning-based statistics taking into account additional aspects to be considered such as robustness, MLOps and fairness.

Accuracy and robustness are two possible quality dimensions that should be taken into account when using machine learning, whereby it is intuitively clear that they are antagonists in some way. Trinkaus and Kauermann (2023) show this in their article investigating the question whether or not machine learning should be used in the creation of rental guides. On the one hand, they examine the accuracy of the estimates of various methods on the data material provided, and on the other hand, the accuracy when the methods were trained on disturbed data. The chosen perturbation represents a conceivable scenario with respect to which the procedures should be robust—and can thus serve as a model for a concrete, easy-to-implement robustness check.

Reviewing the available literature on aspects that influence different human annotation behaviour, Beck (2023) takes up a very important but often neglected topic when it comes to the quality of machine learning. Supervised learning indispensably requires labels (frequently, but not only, in the target variable). While these labels are often technically assumed to be objectively and neutrally “correct” or “true”, the fulfilment of this assumption can be rather questionable. An at least almost perfect machine learning algorithm, if trained on incorrect data, would reproduce this bias originating in improper annotations, i.e. improper labels, which emphasizes the actual need for good annotations.

Central banks now increasingly use novel data sources to provide policymakers and the public with information about financial risks and, more recently, climate risks in the economy. In this context, often text-based data sources, such as ESG reports from companies, are used in combination with machine learning methods to generate (new) official statistics. However, this comes again with challenges and questions regarding the reliability and quality of these new tools and data sources. Against this background, Doll and Alves Werb (2023) describe how the Deutsche Bundesbank and its newly founded Sustainable Finance Data Hub help to tackle these questions and improve quality assessment.

We hope you enjoy reading this issue.

Florian Dumpert, Sebastian Wichert, Thomas Augustin and Nina Storfinger

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Beck J (2023) Quality aspects of annotated data: A research synthesis. *AStA Wirtsch Sozialstat Arch.* <https://doi.org/10.1007/s11943-023-00332-y>
- van Delden A, Burger J, Puts M (2023) Ten propositions on machine learning in official statistics. *AStA Wirtsch Sozialstat Arch.* <https://doi.org/10.1007/s11943-023-00330-0>
- Doll HC, Alves Werb G (2023) Innovation for improving climate-related data-Lessons learned from setting up a data hub. *AStA Wirtsch Sozialstat Arch.* <https://doi.org/10.1007/s11943-023-00326-w>
- Molladavoudi S, Yung W (2023) Exploring quality dimensions in trustworthy Machine Learning in the context of official statistics: model explainability and uncertainty quantification. *AStA Wirtsch Sozialstat Arch.* <https://doi.org/10.1007/s11943-023-00331-z>
- Saidani Y, Dumpert F, Borgs C et al (2023) Qualitätsdimensionen maschinellen Lernens in der amtlichen Statistik. *AStA Wirtsch Sozialstat Arch.* <https://doi.org/10.1007/s11943-023-00329-7>
- Trinkaus O, Kauermann G (2023) Can Machine Learning Algorithms deliver superior Models for Rental Guides? *AStA Wirtsch Sozialstat Arch.* <https://doi.org/10.1007/s11943-023-00333-x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.