

## WORKSHOP

### Quality Aspects of Machine Learning - Official Statistics between Specific Quality Requirements and Methodological Innovation

06. bis 08. September 2022 in Munich

With the European Code of Practice, official statistics in Europe are subject to special quality requirements. 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. On the one hand, machine learning methods offer new possibilities for analysis and insight – not only but also against the background of new types of data (sources), some of which are immense in scope – and thus enable higher-quality statistics and can contribute to efficiency improvements in the processes of official statistics. On the other hand, 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. In order to bring together the already built-up empirical knowledge from the practice of official statistics with the findings of science the Statistics Network Bavaria intends to organize a scientific workshop **from 6th to 8th September 2022 in Munich**. The workshop will be organized for Statistics Network Bavaria by the Bavarian State Statistical Office, the Institute of Statistics of the Ludwig-Maximilians-University Munich and the ifo Institute Munich with the cooperation of the Federal Statistical Office of Germany.

#### Details:

<b>Date</b>	September, 06 to 08 2022
<b>Venue</b>	The workshop will take place in the building of the Department of Psychology and Education of the Ludwig-Maximilians-University Munich (Room 2U01, 1.UG, Leopoldstr. 13, 80802 Munich, Germany).
<b>Directions</b>	The building of the Department of Psychology and Education of the Ludwig-Maximilians-University Munich is located directly at the subway stop Giselastraße (lines U3 and U6). Travel by car to the venue is not recommended. The LMU is located in the city centre, which is why there are unfortunately no parking possibilities apart from the usual time-limited and paid parking spaces.
<b>Registration</b>	Registration is requested by email no later than August 30, 2022 at: <a href="mailto:anmeldung@statistik.bayern.de">anmeldung@statistik.bayern.de</a>
<b>Fee</b>	Participation in the event is free of charge.
<b>Abstracts</b>	Abstracts of the presentations are available here: <a href="https://www.statistiknetzwerk.bayern.de/service/aktuelles/maschinellen-lernens/">https://www.statistiknetzwerk.bayern.de/service/aktuelles/maschinellen-lernens/</a>
<b>Conference Dinner</b>	The Conference Dinner will be held on September 06 at 7 pm. Participation is at your own expense. Due to limited capacity, registration is requested no later than August 30, 2022 at: <a href="mailto:anmeldung@statistik.bayern.de">anmeldung@statistik.bayern.de</a>
<b>Contact</b>	If you have any questions, please feel free to contact <a href="mailto:anmeldung@statistik.bayern.de">anmeldung@statistik.bayern.de</a> .

The organizers are in discussion with the editors of the journal AStA Wirtschafts- und Sozialstatistisches Archiv regarding a special issue on the topic. Papers on presentations or discussion contributions at this workshop are eligible for publication in this special issue, subject to successful completion of the review process. The deadline for submissions is expected to be the end of January 2023. Please contact us if you are interested.

**Program:****Day 1 Tuesday, September 06, 2022**

12:30	Registration and Welcome Coffee
13:30	Welcoming
13:50	<b>Quality dimensions for the use of ML in official statistics</b> Johannes Rohde (IT.NRW), Christian Salwiczek (Statistik Nord)
14:50	<b>Quality, ML, Destatis und International Projects</b> Florian Dumpert (Federal Statistical Office of Germany)
15:20	<i>Discussion, transfer and further development</i>
16:00	<b>Equity, inclusion, and fairness in data-driven decision making in the public sector</b> Christoph Kern and Frauke Kreuter (University of Mannheim and LMU Munich)
17:00	<b>Fashions of Artificial Intelligence</b> Rudolf Seising (Research Institute for the History of Science and Technology; Deutsches Museum Munich)
18:10	End of Day 1

**Day 2 Wednesday, September 07, 2022**

09:00	Welcoming
09:10	<b>Ten propositions on machine learning in official statistics</b> Arnout van Delden, Joep Burger und Marco Puts (CBS – Statistics Netherlands)
10:10	<b>Challenges and solutions when adopting ML (Machine Learning) and AI (Artificial Intelligence) in large organisations</b> Joni Karanka und Eleanor Law (ONS – Office for National Statistics)
10:25	<i>Discussion, transfer and further development</i>
11:10	<b>What is Fairness? Implications for FairML</b> Ludwig Bothmann, Kristina Peters, Bernd Bischl (LMU Munich)
12:10	Lunch break
13:40	<b>Three ML-assisted strategies for coding diverse data sources</b> Malte Schierholz (LMU Munich)
14:10	<b>Record Linkage of Company Data Sets</b> Valentin Reich (ifo Institute – Leibniz Institute for Economic Research at the University of Munich)
15:10	<i>Discussion, transfer and further development</i>
15:50	<b>The SearchEngine: a Holistic Approach to Matching</b> Thorsten Doherr (ZEW – Leibniz Centre for European Economic Research)
16:20	<b>Climate data for official statistics – 3 machine learning applications</b> Hendrik Doll (Deutsche Bundesbank)
17:30	End of Day 2

**Day 3 Thursday, September 08, 2022**

09:00	Welcoming
09:10	<b>Temporary employment in the statistics of registered jobs: Analyses with Text Mining</b> Arsen Çelikel, Joachim Seitz, Jörg Szameitat (Federal Employment Agency)
09:40	<b>Data Science Informed by Survey Science: Collecting More Accurate Labels</b> Jacob Beck, Stephanie Eckman, Frauke Kreuter (LMU Munich)
10:10	<i>Discussion, transfer and further development</i>
10:50	<b>Combining Bayesian estimation and machine learning methods for handling missing values and model selection in complex sampling designs</b> Christian Aßmann, Christoph Gaasch, Doris Stingl (University of Bamberg)
11:20	Summary, feedback and farewell
12:00	End of workshop

**Notes:**

Presentations will be in German or English, discussions will be in German or English (or mixed). The presentation slides will be in English.

Presentations scheduled for one hour consist of approximately 40 minutes of presentation time followed by 20 minutes for direct questions and initial discussion.

Presentations announced with a half hour in the schedule consist of approximately 20 minutes of presentation time followed by 10 minutes for direct questions and initial discussion.

In addition, longer sessions entitled "Discussion, Transfer and Further Development" are on the schedule. The idea here is for participants to engage in exchange with each other and with the presenters, to enter into more in-depth discussions and ideally also to save results. Appropriate materials are provided for this purpose. The workshop thrives on the fact that people work together on the topics.

A special format is offered by the session of the British statistical office (ONS). This session, which is located at the end of the series of presentations from official statistics – and thus has the character of an overall view – begins with an introductory input, which is immediately followed by an extensive opportunity for discussion, transfer and further development. With this announcement we would like to enable all participants to prepare accordingly on the following aspects:

- a) Organization: What is the best way to organize ML in the organization? And where will which roles be located (from methodology, software development, preparation of training data etc.)? More centralized or more decentralized? In-house or through external outsourcing? And how does this affect quality?
- b) Model Quality: When is a model good? When should it be revised, updated or replaced? How can drift be dealt with? What needs to be done to keep the quality permanently high?
- c) Maintenance and competence building: How are ML solutions efficiently maintained and serviced over their deployment duration? How is continuous quality assurance efficiently possible? How is the expertise required for this built up and maintained over longer periods?
- d) Tools, standards, automation: Which tools can support the use of ML? How can the deployment process be automated?

**Contact:**

- Thomas Augustin (Institute of Statistics, LMU Munich)
- Sebastian Wichert (ifo Institute – Leibniz Institute for Economic Research at the University of Munich)
- Nina Storfinger (Bavarian State Statistical Office, Statistics Network Bavaria)
- Florian Dumpert (Federal Statistical Office of Germany, AI and Big Data Unit)



Bayerisches Landesamt für Statistik



STATIS  
Statistisches Bundesamt