

WORKSHOP

Qualitätsaspekte Maschinellen Lernens –

Amtliche Statistik zwischen methodischer Aktualität und besonderer Qualitätsverpflichtung

06. bis 08. September 2022 in München

Mit dem European Code of Practice unterliegt die amtliche Statistik in Europa besonderen Qualitätsanforderungen. Eine dieser Anforderungen ist, dass die statistischen Prozesse zur Erhebung, Aufbereitung und Verbreitung von Statistiken internationalen Standards und Leitlinien in vollem Umfang genügen und zugleich dem aktuellen Stand der wissenschaftlichen Forschung entsprechen. Andererseits spielen Ansätze des maschinellen Lernens eine zunehmend größere Rolle in der amtlichen Statistikproduktion. Um das Erfahrungswissen der amtlichen Statistik in diesem Feld mit den Erkenntnissen der Wissenschaft unter dem Dach der Qualität zusammenzubringen, veranstaltet das Statistik Netzwerk Bayern einen wissenschaftlichen Workshop vom **6. bis 8. September 2022 in München**. Der Workshop wird für das Statistik Netzwerk Bayern vom Bayerischen Landesamt für Statistik, dem Institut für Statistik der Ludwig-Maximilians-Universität München und dem ifo Institut München unter Mitwirkung des Statistischen Bundesamtes organisiert.

Veranstaltungsdetails:

Datum	06. bis 08. September 2022
Räumlichkeiten	Der Workshop findet im Gebäude des Departments Psychologie und Pädagogik der Ludwig-Maximilians-Universität München statt (Raum 2U01, 1.UG, Leopoldstr. 13, 80802 München).
Anfahrt	Das Gebäude des Departments Psychologie und Pädagogik der Ludwig-Maximilians-Universität München liegt direkt an der U-Bahn-Haltestelle Giselastraße (Linien U3 und U6). Die Anreise mit dem Auto zum Veranstaltungsort ist nicht zu empfehlen. Die LMU liegt in der Innenstadt, weshalb außer den üblichen zeitlich begrenzten und kostenpflichtigen Parkplätzen leider keine Parkmöglichkeiten gegeben sind.
Anmeldung	Um Anmeldung wird per E-Mail bis spätestens 30. August 2022 gebeten unter: anmeldung@statistik.bayern.de
Kosten	Die Teilnahme an der Veranstaltung ist kostenlos.
Abstracts	Abstracts zu den Vorträgen stehen hier zur Verfügung: https://www.statistiknetzwerk.bayern.de/service/aktuelles/maschinellen-lernens/
Conference Dinner	Das Conference Dinner findet am 06. September um 19 Uhr statt. Die Teilnahme erfolgt auf eigene Kosten. Aufgrund begrenzter Kapazität wird um Anmeldung bis spätestens 30. August 2022 gebeten unter: anmeldung@statistik.bayern.de
Kontakt	Bei Fragen können Sie sich gerne an anmeldung@statistik.bayern.de wenden.

Die Veranstalter sind mit den Editoren der Zeitschrift **AStA Wirtschafts- und Sozialstatistisches Archiv** bzgl. eines Sonderhefts zum Thema im Gespräch. Paper zu Vorträgen oder Diskussionsbeiträgen bei diesem Workshop bieten sich – vorbehaltlich des erfolgreichen Durchlaufens des Reviewprozesses – für eine Veröffentlichung in diesem Sonderheft an. Einsendeschluss wird voraussichtlich Ende Januar 2023 sein. Sprechen Sie uns bei Interesse gerne an.

Programm:**Tag 1 Dienstag, 06. September 2022**

12:30	Registrierung und Begrüßungskaffee
13:30	Grußworte
13:50	Qualitätsdimensionen für den Einsatz von ML in der amtlichen Statistik Johannes Rohde (IT.NRW) und Christian Salwiczek (Statistik Nord)
14:50	Qualität, ML, Destatis und internationale Projekte Florian Dumpert (Statistisches Bundesamt)
15:20	<i>Diskussion, Transfer und Weiterentwicklung</i>
16:00	Equity, inclusion, and fairness in data-driven decision making in the public sector Christoph Kern und Frauke Kreuter (Uni Mannheim und LMU München)
17:00	Fashions of Artificial Intelligence Rudolf Seising (Forschungsinstitut für Technik- und Wissenschaftsgeschichte; Deutsches Museum München)
18:10	Ende Tag 1

Tag 2 Mittwoch, 07. September 2022

09:00	Begrüßung
09:10	Ten propositions on machine learning in official statistics Arnout van Delden, Joep Burger und Marco Puts (CBS – Statistics Netherlands)
10:10	Challenges and solutions when adopting ML (Machine Learning) and AI (Artificial Intelligence) in large organisations Joni Karanka und Eleanor Law (ONS – Office for National Statistics)
10:25	<i>Diskussion, Transfer und Weiterentwicklung</i>
11:10	What is Fairness? Implications for FairML Ludwig Bothmann, Kristina Peters, Bernd Bischl (LMU München)
12:10	Mittagspause
13:40	Three ML-assisted strategies for coding diverse data sources Malte Schierholz (LMU München)
14:10	Record Linkage of Company Data Sets Valentin Reich (ifo Institut – Leibniz-Institut für Wirtschaftsforschung)
15:10	<i>Diskussion, Transfer und Weiterentwicklung</i>
15:50	The SearchEngine: a Holistic Approach to Matching Thorsten Doherr (ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung)
16:20	Climate data for official statistics – 3 machine learning applications Hendrik Doll (Deutsche Bundesbank)
17:30	Ende Tag 2

Tag 3 Donnerstag 08. September 2022

09:00	Begrüßung
09:10	Befristungen in der Statistik der gemeldeten Arbeitsstellen: Analysen mit Text Mining Arsen Çelikel, Joachim Seitz, Jörg Szameitat (Bundesagentur für Arbeit)
09:40	Data Science Informed by Survey Science: Collecting More Accurate Labels Jacob Beck, Stephanie Eckman, Frauke Kreuter (LMU München)
10:10	<i>Diskussion, Transfer und Weiterentwicklung</i>
10:50	Combining Bayesian estimation and machine learning methods for handling missing values and model selection in complex sampling designs Christian Aßmann, Christoph Gaasch, Doris Stingl (Otto-Friedrich-Universität Bamberg)
11:20	Zusammenfassung, Feedback und Verabschiedung
Ca. 12:00	Veranstaltungsende

Hinweise:

Vorträge werden in deutscher oder englischer Sprache gehalten, Diskussionen in deutscher oder englischer Sprache (oder gemischt) geführt. Die Vortragsfolien werden auf Englisch sein.

Vorträge, die mit einer Stunde im Zeitplan angekündigt sind, bestehen aus ca. 40 Minuten Präsentationszeit und sich anschließenden 20 Minuten für direkte Fragen und erste Diskussionen.

Vorträge, die mit einer halben Stunde im Zeitplan angekündigt sind, bestehen aus ca. 20 Minuten Präsentationszeit und sich anschließenden 10 Minuten für direkte Fragen und erste Diskussionen.

Darüber hinaus finden immer wieder längere Sessions mit dem Titel „Diskussion, Transfer und Weiterentwicklung“ statt. Die Idee ist hier, dass die Teilnehmerinnen und Teilnehmer untereinander und mit den Vortragenden in den Austausch treten, in tiefergehende Diskussionen einsteigen und Ergebnisse idealerweise auch sichern. Hierfür werden entsprechende Materialien bereitgestellt. Der Workshop lebt davon, dass gemeinsam an den Themen gearbeitet wird.

Ein Sonderformat bietet die Session des britischen statistischen Amtes (ONS). Diese am Ende der Reihe der Vorträge aus der amtlichen Statistik angesiedelte – und mithin den Charakter einer Gesamtschau aufweisende – Session beginnt mit einem einführenden Input, an den sich unmittelbar eine umfangreiche Diskussions-, Transfer- und Weiterentwicklungsgelegenheit anschließt. Mit dieser Ankündigung möchten wir allen Teilnehmerinnen und Teilnehmern eine entsprechende Vorbereitung zu folgenden Aspekten ermöglichen:

- a) Organisation: Wie lässt sich ML in der Organisation am besten organisieren? Und wo werden welche Rollen (aus Methodik, Softwareentwicklung, Trainingsdatenerstellung etc.) wahrgenommen? Eher zentral oder eher dezentral? Inhouse oder durch externe Vergaben? Und wie wirkt sich das auf die Qualität aus?
- b) Modellqualität: Wann ist ein Modell gut? Wann sollte es überarbeitet, aktualisiert oder ersetzt werden? Wie kann mit Drift umgegangen werden? Was ist zu tun, damit die Qualität dauerhaft hoch bleibt?
- c) Wartung, Pflege und Kompetenzaufbau: Wie werden ML-Lösungen über ihre Einsatzdauer hinweg effizient gewartet und gepflegt? Wie eine kontinuierliche Qualitätssicherung effizient möglich? Wie wird das dafür erforderliche Fachwissen aufgebaut und über längere Zeiträume erhalten?
- d) Werkzeuge, Standards, Automatisierung: Welche Tools können bei der Nutzung von ML unterstützen? Wie lässt sich der Deployment-Prozess automatisieren?

Ansprechpartner Workshop:

- Thomas Augustin (Institut für Statistik, LMU München)
- Sebastian Wichert (ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München e.V.)
- Nina Storfinger (Bayerisches Landesamt für Statistik, Statistik Netzwerk Bayern)
- Florian Dumpert (Statistisches Bundesamt, Referat „Künstliche Intelligenz, Big Data“)