

## Call for Abstracts

Gemeinsamer Kongress der DGS & ÖGS, 23.-25. August 2021, Wien

Sektionsveranstaltung

### **Big Data und maschinelles Lernen in der Soziologie**

Erkenntnisse und Herausforderungen in der Kombination von Umfrage- und Twitter-Daten: Eine Untersuchung der gesellschaftlichen Polarisierung in der COVID-19 Debatte im deutschsprachigen Raum

Beate Klösch<sup>1\*</sup>, Markus Reiter-Haas<sup>2</sup>, Markus Hadler<sup>1</sup>, Elisabeth Lex<sup>2</sup>

Die Nutzung von Big Data bietet der Soziologie eine Erweiterung zu klassischen sozialwissenschaftlichen Methoden. Die Kombination dieser unterschiedlichen Datentypen wurde aufgrund der Neuheit dieses Forschungsansatzes, fehlender gemeinsamer Datenquellen und ethischer Hürden in der aktuellen Literatur bisher wenig behandelt (vgl. Al Baghal et al. 2020). Diesen Herausforderungen widmet sich unser interdisziplinäres Forschungsprojekt mithilfe einer Kombination von soziologischen Umfragedaten und Twitter-Daten. Dazu wurde im Sommer 2020 eine für Internetnutzer\*innen repräsentative Online-Umfrage zu diversen gesellschaftspolitischen Themen in der DACH-Region durchgeführt, in deren Rahmen zugleich private Twitter-Benutzernamen sowie das Einverständnis zur wissenschaftlichen Verarbeitung der öffentlich geteilten Tweets erhoben wurden. Es zeigt sich, dass die Nutzung von Twitter im deutschsprachigen Raum gering ist und der Großteil der befragten Personen der Analyse ihrer Twitter-Daten nicht zustimmt. So nahmen insgesamt 2560 Personen an der Umfrage teil, wovon lediglich 79 Twitter-Accounts für die Analyse verblieben. Liegt der Fokus darüber hinaus auf spezifischen Themen, so wird die Anzahl der zu analysierenden Twitter-Accounts noch geringer, da nicht alle Nutzer\*innen zum jeweiligen Thema tweeten. Weiters können die verschiedenen Datentypen nicht direkt miteinander verglichen werden, da Meinungen auf Twitter in Form von unstrukturierten Texten anstatt von strukturierten Antworten wiedergegeben werden. Dieser Herausforderung treten wir mit Approximationsverfahren (z.B. Sentiment Analyse), qualitativen Inhaltsanalysen und statistischen Kennwerten entgegen, da andere gängige Methoden wie Netzwerkanalysen hier nicht direkt anwendbar sind. Schlussendlich drängt sich die Frage nach diversen Verzerrungen und der fehlenden Generalisierbarkeit der Ergebnisse auf. Am Beispiel der Polarisierung der öffentlichen Meinung hinsichtlich COVID-19-Maßnahmen werden Vorteile und Schwierigkeiten dieser interdisziplinären Daten- und Methodenkombination präsentiert und diskutiert.

Quelle:

Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., and Burnap, P. (2020): Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates Across Three UK Studies. In: *Social Science Computer Review*, 38(5), S.517-532.

<sup>1</sup> Institut für Soziologie, Universität Graz, Graz

<sup>2</sup> Institute of Interactive Systems and Data Science, Graz University of Technology, Graz

\* Korrespondierende Autorin: beate.kloesch@uni-graz.at