
Solidarität im Migrationsdiskurs: Klassifizierung von Leserinnen- und Leserbriefen mit DistilBERT

Lisanne Rüh

Universität Hildesheim
ruehli@uni-hildesheim.de

Soll eine sehr große Menge an textlichen Daten klassifiziert werden, muss dies automatisiert erfolgen. Können die Daten zudem nicht nach strukturellen Merkmalen gefiltert werden, weil ein so abstraktes und schwer fassbares Konzept wie *Solidarität* untersucht werden soll (enthält ein Text *Solidarität* oder nicht?), kann ein Machine-Learning-Algorithmus wie DistilBERT (Sanh et al. 2020) für die Textklassifikation verwendet werden. In meiner Projektarbeit habe ich daher untersucht, ob DistilBERT in der Lage ist, solidarische Äußerungen im Migrationskontext in Leserinnen- und Leserbriefen zu erkennen.

Basierend auf der Begriffskonzeption von *Solidarität* nach Kneuer et al. (2021) sollte dies mithilfe zweier Klassifizierer erreicht werden, die jeweils auf einen der Solidaritätsbestandteile *Adversity* (eine Widrigkeit, die einer Person widerfährt) und *Contribution* (ein Beitrag, der zur Erreichung eines gemeinsamen Ziels geleistet wird) trainiert wurden. In Kombination sollten diese beiden Klassifizierer dann zur Identifikation der Gesamtkategorie *Solidarität* genutzt werden. Grundlage für die Trainings- und Testdaten bildete ein bereits zum Themenfeld Migration vorgefiltertes Korpus mit Leserinnen- und Leserbriefen aus verschiedenen deutschen Zeitungen.

Die Untersuchung hat gezeigt, dass die Klassifikation von kleineren Textausschnitten bereits gute bis sehr gute Ergebnisse erzielen konnte (F-Maß = 0,786 u. 0,908), aber die Gesamtkategorie *Solidarität* in vollständigen Briefen noch nicht in einem zufriedenstellenden Maße erkannt wurde (F-Maß = 0,364). Die Erfassung der Teilkategorien *Adversity* und *Contribution* funktionierte hier dagegen besser (F-Maß = 0,630 bzw. 0,625). Insgesamt wurden aber auch viele positive Beispiele, die Äußerungen der jeweiligen (Teil-)Kategorie enthalten, nicht von den Klassifizierern identifiziert. Bei der Erkennung von *Solidarität* gilt außerdem, dass Briefe, die als *Solidarität* enthaltend eingeordnet wurden, oftmals das falsche Label erhielten. Es ist hier somit noch weitere Arbeit notwendig, um die Genauigkeit der Klassifizierungsergebnisse, insbesondere bei der Bewertung gesamter Briefe, zu verbessern.

References: • Kneuer, M.; Corsten, M.; Schammann, H.; Kahle, P.; Wallaschek, S. & Ziegler, F. (2021). Claiming Solidarity. A multilevel discursive reconstruction of solidarity. *European Journal of Social Theory*. • Sanh, V.; Debut, L.; Chaumond, J. & Wolf, T. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In *EMC2, the 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing*.