

Bachelorarbeit

Aufbau eines Social-Media-Datenkorpus zur Analyse der Wahrnehmung des Lehrberufs

Ausgangslage und Problemstellung

An (berufsbildenden) Schulen besteht ein anhaltender Lehrkräftemangel, der mittelfristig zunehmend auch über Quer- und Seiteneinstiege gedeckt werden muss. Ob Personen unterschiedlicher Lebensphasen den Lehrberuf als berufliche Option in Betracht ziehen, hängt wesentlich davon ab, wie der Beruf wahrgenommen wird, sein „Berufsimage“. Diese Wahrnehmung wird in zunehmendem Maße im digitalen Raum geprägt und ausgehandelt: auf Plattformen wie TikTok, YouTube und Reddit wird über Arbeitszeit, Belastung, Verbeamtung, Quereinstieg oder die Sinnhaftigkeit der Tätigkeit diskutiert. Eine systematische, wiederverwendbare Datengrundlage, um diese Darstellung und Kommunikation empirisch zu untersuchen, fehlt bislang.

Ziel der Bachelorarbeit ist es, einen strukturierten Datenkorpus zur Wahrnehmung des Lehrberufs in sozialen Medien (z. B. Reddit, TikTok, YouTube) aufzubauen und über eine reproduzierbare Pipeline zu erheben, zu bereinigen und aufzubereiten. Auf Basis dieses Korpus sollen erste explorative Analysen durchgeführt werden, um wiederkehrende Themen und Zuschreibungen sichtbar zu machen, etwa durch Themenidentifikation und (LLM-gestützte) Clusterbildung. Der Korpus bildet damit die Grundlage für weiterführende Untersuchungen (z. B. Schulformvergleiche).

- Einarbeitung in den Forschungskontext (Berufsimage, Lehrkräftegewinnung) sowie in die relevanten Plattformen und deren Kommunikationsformen
- Recherche und Auswahl geeigneter Datenquellen und Zugangswege (offizielle Schnittstellen wie TikTok Research API, YouTube Data API, Reddit API o. Ä.) unter Berücksichtigung rechtlicher und datenschutzrechtlicher Rahmenbedingungen
- Konzeption und Implementierung einer Pipeline zur Datenerhebung, -bereinigung und -strukturierung (Aufbau eines reproduzierbaren Korpus inkl. Metadaten)
- Aufbereitung und ggf. Annotation der Daten (z. B. Plattform, Sprache, Beitragstyp)
- Durchführung erster explorativer Analysen: Themenidentifikation, (LLM-gestützte) Clusterbildung sowie ggf. Sentiment- bzw. Emotionsanalyse
- Aufbereitung, Visualisierung und Einordnung der Ergebnisse sowie Diskussion von Grenzen und Potenzialen des Vorgehens

Anforderungen

- Grundkenntnisse in Python
- Interesse an API-/web-basierter Datenerhebung und Natural Language Processing
- Grundverständnis von LLMs und Text-Mining
- Sorgfältiger Umgang mit rechtlichen und ethischen Aspekten
- Fähigkeit zur eigenständigen Recherche und strukturierten Arbeitsweise

Ansprechpartner

Andreas Burgdorf | E-Mail: burgdorf@uni-wuppertal.de

Bachelor's thesis

Building a social media corpus to analyze perceptions of the teaching profession

Initial Situation and Problem Definition

(Vocational) schools face a persistent teacher shortage that will increasingly have to be addressed through lateral and career-change entrants. Whether people at different stages of life consider teaching as a career option depends substantially on how the profession is perceived, its "occupational image". This perception is increasingly shaped and negotiated in the digital space: on platforms such as TikTok, YouTube and Reddit, users discuss working hours, workload, civil-servant status, lateral entry or the sense of purpose of the work. To date, a systematic, reusable data basis for empirically studying how the profession is portrayed there is lacking.

The aim of this bachelor's thesis is to build a structured data corpus on the perception of the teaching profession in social media (e.g. Reddit, TikTok, YouTube) and to collect, clean and prepare it via a reproducible pipeline. Based on this corpus, initial exploratory analyses are to be carried out in order to surface recurring themes and attributions, for example through topic identification and (LLM-supported) clustering. The corpus thus forms the basis for further investigations (e.g. comparisons across school types).

- Familiarization with the research context (occupational image, teacher recruitment) as well as with the relevant platforms and their forms of communication
- Research and selection of suitable data sources and access methods (official interfaces such as the TikTok Research API, YouTube Data API, Reddit API, etc.), taking legal and data-protection requirements into account
- Design and implementation of a pipeline for data collection, cleaning and structuring (building a reproducible corpus including metadata)
- Preparation and, where applicable, annotation of the data (e.g. platform, language, post type)
- Conducting initial exploratory analyses: topic identification, (LLM-supported) clustering and, where applicable, sentiment or emotion analysis
- Preparation, visualization and interpretation of the results as well as a discussion of the limitations and potential of the approach

Requirements

- Basic knowledge of Python
- Interest in API-/web-based data collection and Natural Language Processing
- Basic understanding of LLMs and text-mining methods
- Careful handling of legal and ethical aspects
- Ability to work independently and in a structured manner with a research focus

Contact Person

Andreas Burgdorf | E-Mail: burgdorf@uni-wuppertal.de