Analyzing Tweets Using Topic Modeling and ChatGPT: What We Can Learn About Teachers and Topics During COVID-19 Pandemic-Related School Closures

Anna C. Weigand[1,2][a], Maj F. Jacob[2], Maria Rauschenberger[2][b] and Maria José Escalona Cuaresma[1][c]

[1]*Department of Computer Languages and Systems, University of Seville, Seville, Spain*

[2]*Faculty of Technology, University of Applied Sciences Emden/Leer, Emden, Germany*

Keywords: Machine Learning, Data Set, Topic Modeling, Twitter, X, Twitterlehrerzimmer, Twlz, Covid, ChatGPT.

Abstract: This study examines the shifting discussions of teachers within the *#twlz community* on Twitter across three phases of the COVID-19 pandemic – before school closures and during the first and second school closures. We analyzed tweets from January 2020 to May 2021 to identify topics related to education, digital transformation, and the challenges of remote teaching. Using machine learning and ChatGPT, we categorized discussions that transitioned from general educational content to focused dialogues on online education tools during school closures. Before the pandemic, discussions were generally focused on education and digital transformation. During the first school closures, conversations shifted to more specific topics, such as online education and tools to adapt to distance learning. Discussions during the second school closures reflected more precise needs related to fluctuating pandemic conditions and schooling requirements. Our findings reveal a consistent increase in the specificity and urgency of the topics over time, particularly regarding digital education.

1 INTRODUCTION

Among teachers in Germany, the *#twitterlehrerzimmer* or *#twlz community* on *Twitter* is an established forum for digital exchange (Fütterer et al., 2021). In the following, we refer to it as the #twlz community. The social media platform Twitter (https://twitter.com) is a microblogging service that, since 2006, has allowed users to write short posts (called tweets) of up to 280 characters. Although it was renamed *X* in July 2023 (Britannica, The Editors of Encyclopaedia, 2024), we refer to it as Twitter in this study because our data set was collected before this change.

During the COVID-19 pandemic, participation in the #twlz community grew, especially during the school closures (Fütterer et al., 2021). The pandemic changed the situation in the schools immediately, forcing schools to take measures such as completely closing down for several weeks (Huber, 2021) or organizing rotating classes (Grill, M., Mascolo, G., Munzinger, P., Zick, T., 2022). Rotating classes meant that some of the students were homeschooled (*i.e.,* distance education by their school teachers) while the others were physically in the classroom to ensure small groups of students and reduce the risk of COVID-19 infection. The *School Barometer*, which monitors the situation at schools in Germany, Austria, and Switzerland from different perspectives, depicts high stress, struggles, and challenges for teachers during this time (Huber, 2021). Consequently, there was an increased demand among teachers for knowledge exchange. We hypothesize that the importance of specific topics varied according to the timing of their tweet publication – either before or during the COVID-19 pandemic.

In this paper, we explore two primary areas with a mixed-methods approach: (1) the evolution of discussion topics in the #twlz community before and during the first nationwide and second school closures in Germany due to the COVID-19 pandemic and (2) the different methodologies (*i.e.,* topic modeling and ChatGPT) employed to analyze the Twitter data sets.

[a] https://orcid.org/0000-0003-2674-0640
[b] https://orcid.org/0000-0001-5722-576X
[c] https://orcid.org/0000-0002-6435-1497

Weigand, A., Jacob, M., Rauschenberger, M. and Escalona Cuaresma, M.
Analyzing Tweets Using Topic Modeling and ChatGPT: What We Can Learn About Teachers and Topics During COVID-19 Pandemic-Related School Closures.
DOI: 10.5220/0013036900003825

In *Proceedings of the 20th International Conference on Web Information Systems and Technologies (WEBIST 2024)*, pages 350-357
ISBN: 978-989-758-718-4; ISSN: 2184-3252

2 RELATED WORK

Tweets related to the German school closures during the COVID-19 pandemic and before have already been analyzed using a mixed-methods approach (Fütterer et al., 2021). The study combined quantitative and qualitative research approaches to examine discussions within the #twlz community during the first nationwide school closures. They explored the differences between before and during the first school closures and the opportunities and challenges discussed during the first school closures.

The authors applied the *tf-idf-analysis* (*term frequency-inverse document frequency*) to statistically identify the importance of strings to a document (Arya, 2022) on two subsets of data: before the nationwide school closures (January 6 to February 17, 2020) and during the first nationwide school closures (March 16 to April 27, 2020). This analysis identified keywords by their frequency of appearance and bigrams that were especially relevant within the tweets. The three most distinctive words were selected to calculate correlations with all other words used in the tweets in order to evaluate the significance of the content based on these bigram networks. In addition, to address the research question of opportunities and challenges, a manual, resource-intensive content analysis was conducted on just 270 tweets. These were selected from an initial data set of 128,422 tweets with high interaction metrics such as retweets, likes, and comments, following Mayring's methodology (Mayring, 2015). The bigram networks revealed that the topic *digital education* had already been discussed before the Germany-wide school closures but that the exchange increased during the first school closures. They found that, before the pandemic, users discussed more general topics, such as *education and learning*, *classes and school life*, and *educational revolution and crisis*. While the schools were closed, topics such as *mutual help* and *specific software and tools for teaching and learning* became popular (Fütterer et al., 2021). In addition, *distance learning*, *live streaming*, *flipped learning*, and *homeschooling* were often discussed within the #twlz community during the first school closures in Germany. According to their manual content analysis, the biggest challenges during the nationwide school closures in Germany were *good digital classes*, *missing software*, and the *lack of digital know-how for digital teaching*. Opportunities included the *opportunity for networking and exchange* within Twitter's teacher community as well as the *offering of digital material* and *explanations and tricks*.

Other studies (Xue et al., 2020a; Xue et al., 2020b) used a machine learning (ML) approach to analyze COVID-19-related tweets but not specifically #twitterlehrerzimmer or #twlz. *Latent Dirichlet Allocation* (LDA) was applied to find the discussed topics. These topics were the basis for the authors' manual content analysis (Braun and Clarke, 2006) to identify *themes*, such as *public health measures to slow the spread of COVID-19*, *social stigma associated with COVID-19*, *COVID-19 new cases and deaths*, *COVID-19 in the United States*, and *COVID-19 cases in the rest of the world*. They also conducted a sentiment analysis, which is a natural language processing method, by applying the *NRC Emotion Lexicon* (Mohammad and Turney, 2013).

In summary, existing research has already examined Twitter data through ML approaches such as LDA models. The Twitter data from the #twlz teacher community concerning COVID-19-related topics was analyzed using the statistical approach of tf-idf-analysis together with manual content analysis. However, there has been no analysis using ML to highlight the main topic changes within the #twlz teacher community over the course of the pandemic. Specifically, there has been no comparison of the content from before the pandemic to the content during the first and second school closures. A comparison of results from different studies using various methods could also provide new methodological insights.

3 METHODOLOGY

In this section, we first describe our mixed-methods research approach. We then explain our process of data collection, data preparation, and modeling.

3.1 Research Approach

As an overall research approach, we apply the *Design Science Research Methodology* (Peffers et al., 2007). Therefore, we go through the following steps: We *identify the problem* through a literature review, *define objectives for problem-solving* by our study design, *design and develop solutions for the problem* by training an LDA model with our Twitter data set, *demonstrate the solution for the problem* by inferring themes personally and with the help of ChatGPT-3.5 (OpenAI, 2022), *evaluate the solution for the problem* by comparing our results with the findings of Fütterer et al. (2021), and *communicate the problem and its solution* with this work.

Therefore, from our literature review and the related work (Fütterer et al., 2021; Xue et al., 2020a; Xue et al., 2020b), we derive the following research

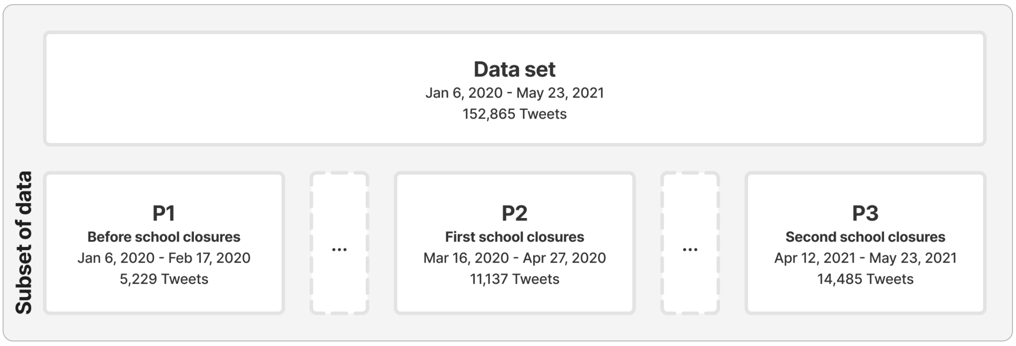

Figure 1: We cut out three subsets of data (P1, P2, and P3) from the raw data set.

questions (RQ), which include both content (RQ1 - RQ3) and technical (RQ4) perspectives:

RQ1. What topics were frequently being discussed using the hashtag #twitterlehrerzimmer and #twlz in January and February 2020, before the pandemic?

RQ2. What topics were frequently being discussed during the first nationwide school closures in March and April 2020, and are there obvious changes from the period of January and February 2020?

RQ3. Are there any differences between the topics that were being discussed during the first nationwide school closures (March and April 2020) and the second school closures (April and May 2021)?

RQ4. Comparing our methods with those of the existing study (Fütterer et al., 2021), are there any substantive differences in terms of the results?

Since there is no publicly available data set to answer our research questions, we collected and curated our own data set from Twitter.

3.2 Data Collection

Using the *Twitter API* (Twitter, 2023) and the *twarc2* Python library (twarc, 2024), we generate a raw data set of 152,865 tweets with the hashtags #twitterlehrerzimmer and/or #twlz. These tweets will later be made anonymous for ethical reasons (Webb et al., 2017). We did not include retweets. To answer our research questions, we chose the time period from January 6, 2020 (two months before the first school closures) to May 23, 2021 (second school closures). This resulted in three subsets of data (P1, P2, P3), each covering a period of 42 days (see also Figure 1), as in the study by Fütterer et al. (2021):

P1. from January 6 to February 17, 2020, when no COVID-19 measures were in effect in Germany, as the first measures were taken in schools by the government on March 16, 2020 (5,229 tweets)

P2. from March 16 to April 27, 2020, when all schools in Germany closed (11,137 tweets)

P3. from April 12 to May 23, 2021, when the schools closed again or organized rotating classes (14,485 tweets)

3.3 Data Preparation

In line with previous studies (Fütterer et al., 2021; Xue et al., 2020a; Xue et al., 2020b), we clean the subsets of data P1, P2, and P3 to prepare them for further processing by removing:

- all *space characters* exceeding the character length of one, all single characters, and all URLs (to remove irrelevant information)
- all *numbers*, *punctuation*, and *special characters* (to remove irrelevant information such as emojis)
- all *@-mentions* of persons (to remove irrelevant information and make tweets anonymous)
- *#-characters* (to consider the word of the used hashtag as a topic as well)
- all * and *:* (to convert inclusive German language forms into standard feminine forms without special characters)
- *stop words* according to the *NLTK* Python library (Aarsen et al., 2023), plus the following additional words: *twitterlehrerzimmer*, *twlz*, *gehen*, *ja*, *nein*, *ab*, *für*, *hallo*, and *liebes* (to remove words with less informative value, such as articles, pronouns, and prepositions)

Then, to split the tweets (strings) into single words (sub-strings), we apply *tokenize* from the *NLTK* Python library (Aarsen et al., 2023) to each subset of data, or *corpus* (P1, P2 and P3). From these

Table 1: An example of our data preparation process for a tweet, from the original to the bigrammed to the lemmatized tweet.

Original tweet	https://t.co/VcdT4Za2WD Auch für den #Impfstoff von #Moderna soll die Zulassung für #Kinder ab 12 Jahre bei der #EMA beantragt werden. Eine gute Nachricht, je mehr zugelassen ist, umso schneller sollten die #Impfungen klappen. #SichereBildung #twlz #Schulen #ImpfenRettetLeben
Tokenized tweet	['Impfstoff', 'Moderna', 'Zulassung', 'Kinder', 'Jahre', 'EMA', 'beantragt', 'gute', 'Nachricht', 'je', 'mehr', 'zugelassen', 'umso', 'schneller', 'sollten', 'Impfungen', 'klappen', 'SichereBildung', 'Schulen', 'ImpfenRettetLeben']
Bigrammed tweet	['Impfstoff', 'Moderna', 'Zulassung', 'Kinder', 'Jahre', 'EMA', 'beantragt', 'gute_Nachricht', 'je', 'mehr', 'zugelassen', 'umso', 'schneller', 'sollten', 'Impfungen', 'klappen', 'SichereBildung', 'Schulen', 'ImpfenRettetLeben']
Lemmatized tweet	['Impfstoff', 'Moderna', 'Zulassung', 'Kind', 'Jahr', 'ema', 'beantragen', 'Gute_nachricht', 'je', 'mehr', 'zulassen', 'umso', 'schnell', 'sollen', 'Impfung', 'Klappen', 'Sicherebildung', 'Schule', 'Impfenrettetlebe']

sub-strings, we build bigrammed tweets using *models.phrases* of the *gensim* Python library (Řehůřek, 2022). This means finding sequences of two contiguous words to analyze their relationships and probabilities. These word pairs are combined into one word divided by an underscore. Furthermore, we apply the *HanoverTagger* (Wartena, 2019) to lemmatize the tweets. Thereby, the words are changed into their basic forms. An example of the data preparation process is shown in Table 1.

3.4 Modeling

For topic modeling, we use the *LDA* (Blei et al., 2003) model, as it has already been applied in the context of tweet topic analysis (Xue et al., 2020a; Xue et al., 2020b). This algorithm considers each document (*i.e.,* tweet) as a collection of latent topics and calculates the weights of the topics within the document as well as their probability of appearance over the whole corpus (*i.e.,* subset of data).

First, a vector called *bag of words* is generated for each subset of data. It stores the words and their frequency of each corpus.

Second, for each corpus, an optimal number of topics is defined by calculating the *perplexity* and the *coherence score* (see Table 2). The goal is to find the number of topics with the lowest perplexity and the highest coherence score at the same time. Hence, we defined the following optimal numbers of topics: *P1* = 4 topics, *P2* = 3 topics, and *P3* = 2 topics. For each topic, the top 15 related keywords are listed in descending order according to rank. To extract the most dominant topic from each time period, we calculate the topic weightage per tweet, and put it in relation to the number of all the tweets within the specific time period.

3.5 Analysis of the Topics

First, two of our project members manually analyze the listed keywords to define an overall theme for each topic, and we merge their results into a common theme. We then use ChatGPT-3.5 to examine the extracted keywords. All prompts are documented in our research protocol (Weigand et al., 2024). This additional approach is an audit for our manually extracted themes.

4 RESULTS

To answer our research questions, we use the subsets of data P1, P2, and P3, along with their individual numbers of optimal topics. For each subset of data, the LDA model returns the 15 highest weighted lemmatized keywords for each topic. These keywords provide insights into the content of each topic (see Table 3).

Concerning RQ1 and the topics discussed before the first measures were taken, we investigate the subset of data P1. For example, in P1, *topic 1* (see Table 3) appears as the dominant topic in $n = 4,168$ tweets. However, *topic 2* is dominant in $n = 283$ tweets, *topic 3* in $n = 516$ tweets, and topic 4 in $n = 262$ tweets. To make the keywords easier to understand, we abstract them into themes based on our understanding and with the help of ChatGPT (see Table 4). Overall, the keywords of *topic 1* indicate that *general idea exchange within the community* was in high demand before the COVID-19 pandemic. ChatGPT summarizes this topic as *education and learning*. We summarize the keywords of *topic 2* as dealing with *new and up-to-date concepts for digital education*, while ChatGPT abstracts it as *digital transformation in education*. *Topic 3* we cluster in a group related to *school projects and additional*

Table 2: Results of the *perplexity* and the *coherence score* for P1, P2, and P3 to define the optimal number (see the bold values) of topics for each subset of data.

	P1		P2		P3	
topics	*perplexity*	*coherence score*	*perplexity*	*coherence score*	*perplexity*	*coherence score*
2	-8.973	0.276	-8.852	0.315	**-9.049**	**0.365**
3	-9.151	0.257	**-8.995**	**0.323**	-9.244	0.311
4	**-9.294**	**0.332**	-9.118	0.292	-9.385	0.297
5	-9.437	0.306	-9.242	0.29	-9.512	0.294
6	-9.571	0.33	-9.358	0.292	-9.634	0.293
7	-9.723	0.318	-9.486	0.303	-9.792	0.296
8	-9.913	0.405	-9.659	0.329	-9.969	0.35
9	-10.219	0.393	-9.884	0.335	-10.236	0.33
10	-10.598	0.426	-10.224	0.294	-10.599	0.338
11	-11.174	0.384	-10.718	0.358	-11.144	0.279
12	-11.998	0.441	-11.473	0.314	-11.981	0.308

Table 3: LDA results for all subsets of data (P1, P2, and P3) and their topics. Topic 1 appears as the dominant topic in P1 that occurs most often. In P2, topic 3 is the dominant topic. In P3, the two topics are almost evenly distributed among the tweets.

Subset of data	Topic	Lemmatized keywords within topic
P1	**1**	**Schule, geben, Mal, schon, jemand, Thema, danke, Frage, heute, amp, Sus, Neue, Idee, mehr, viel**
	2	Unterricht, Digitalebildung, neu, digital, Arbeit, Lehrkraft, Bildung, erstellen, Tolle, Edupnx, warum, statt, Zeit, müssen, Medium
	3	Jahr, Bayernedu, erst, finden, gut, haben, gerne, Gute, wer, Klasse, Projekt, dabei, einfach, Wunsch, vielleicht
	4	Lehrerleben, Schüler, Lehrer, Schülerin, immer, Lehrerin, freuen, sein, tipps, kommen, gleich, Mensch, können, Podcast, letzt
P2	1	digital, Idee, viel, amp, Gute, Schüler, schon, Aufgabe, Frage, Unterricht, gerne, jemand, Schulschließung, Lehrer, tipps
	2	Coronaviru, Neue, müssen, immer, kommen, Server, Plattform, kostenlos, Spiel, Kurs, letzt, schnell, Via, Geben, Twitter
	3	**Schule, Corona, Mal, Sus, Zeit, Online, heute, Schulschliessung, Lernen, geben, Homeschooling, gerade, covid, gut, Schülerin**
P3	1	amp, Bildung, Thema, Unterricht, Uhr, Schule, geben, Bayernedu, digital, freuen, Lernen, Online, Tool, Idee, Moodle
	2	Schule, Mal, Test, Kind, Inzidenz, Sus, heute, mehr, schon, gut, Klasse, Corona, Woche, sein, haben

wishes, especially in Bavaria, while ChatGPT labels it as *education initiatives and collaboration*. From our perspective, *topic 4* is about *teachers' lives*, while ChatGPT summarizes it as *teaching and learning dynamics*. ChatGPT describes P1 as *education and learning environments*.

To evaluate RQ2 and the topics discussed during the first nationwide school closures, we examine the subset of data P2. Three topics are identified for P2 (see Table 3). *Topic 3* occurs as the dominant topic in $n = 11,136$ tweets. *Topic 1* only occurs in $n = 1$, and *topic 2* is never the dominant topic. From our perspective, the keywords of the dominant *topic 3* during the first nationwide school closures due to the COVID-19 pandemic in March and April 2020 are related to *good online education during homeschooling and school closures*. ChatGPT summarizes this as *Education amidst the pandemic*. *Topic 1* relates to the *search for advice-related tasks in digital education during school closures*. ChatGPT abstracts this as *education in the digital age*. Regarding *topic 2*, the keywords imply discussion about *(free) platforms and tools during the COVID-19 pandemic*. ChatGPT calls this *adapting to change in the pandemic era*. ChatGPT describes P2 as *adapting education in the face of crisis*.

Regarding RQ3, we analyze the subset of data P3, which contains the topics discussed in April and May 2021. The two topics of P3 (see Table 3) are relatively even in terms of their distribution: *Topic 1* is dominant in $n = 6,319$ tweets, and *topic 2* is dominant in $n = 8,166$ tweets. We summarize the keywords of *topic 1* as *digital education and tools, especially in Bavaria*, which ChatGPT labels as *digital education in Bavaria*. Regarding the keywords of *topic 2*, we find the overall theme to be *school life during the COVID-19 pandemic influenced by the current incidence levels*, while ChatGPT characterizes it as *schooling during the pandemic: challenges and adaptations*. In summary, ChatGPT describes P3 as *digital education and pandemic adaptations*.

Table 4: Themes of the LDA results defined by the authors and ChatGPT per topic of the subsets of data P1, P2, and P3.
* CG = ChatGPT.

Overall CG*	Subset of data	Topic	Theme defined by authors	Theme defined by CG*
Education and learning environments	P1	1	General idea exchange within the community	Education and learning
		2	New and up to date concepts for digital education	Digital transformation in education
		3	School projects and additional wishes, especially in Bavaria	Education initiatives and collaboration
		4	Teachers' life	Teaching and learning dynamics
Adapting education in the face of crisis	P2	1	Search for advice related tasks in digital education during school closures	Education in the digital age
		2	(Free) platforms and tools during the COVID-19 pandemic	Adapting to change in the pandemic era
		3	Good online education during homeschooling and school closures	Education amidst the pandemic
Digital Education and Pandemic Adaptations	P3	1	Digital education and tools, especially in Bavaria	Digital education in Bavaria
		2	School life during the COVID-19 pandemic influenced by the current incidence levels	Schooling during the pandemic: challenges and adaptations

5 DISCUSSION

To discover the dominant topics within the #twlz teacher community at different times during the COVID-19 pandemic (RQ1 – RQ3), we examined three Twitter subsets of data (P1, P2, P3). In addition, we reflected on different methodologies used to examine these research questions (RQ4).

For RQ1, we examined P1 to find topics that were discussed using the hashtags #twitterlehrerzimmer and #twlz before the COVID-19 pandemic in January and February 2020. From our understanding, complemented by ChatGPT, the general exchange of ideas about education and learning was prevalent in the #twlz community at that time. There was also a focus on digital transformation of the educational environment. Education initiatives and collaboration, especially in Bavaria, as well as teaching and learning dynamics, were also topics of discussion within the community. The analysis of Fütterer et al. (2021) also shows that digital education was discussed before the school closures in Germany and that education and learning, classes and school life, and the educational revolution and crisis were also common themes. The results regarding the topics *education and learning* and *digital transformation of education* are nearly identical in the two analyses. In addition, *education initiatives and collaboration (in relation to Bavaria)* and *teaching and learning dynamics* may be similar to *classes and school life* in the results of Fütterer et al. (2021). Although our topic analysis did not reveal any discussion related to educational revolution or crisis, we only examined the four highest weighted topics within P1, so it may have been that this topic was just not considered.

The insights for RQ2 regarding the topics discussed using the hashtags #twitterlehrerzimmer and #twlz during the first nationwide school closures due to the COVID-19 pandemic in March and April 2020 were extracted from our subset of data P2. In Germany, different approaches were applied in different regions due to the federal governance of education (Huber, 2021). According to our findings, complemented by ChatGPT, the discussion within the #twlz community was still about digital education, but it was focused more specifically on online education in times of homeschooling and school closures due to the COVID-19 pandemic. The discussion also included an exchange on (free) platforms and tools to adapt to the situation. Fütterer et al. (2021) also found that digital education was an ongoing topic, but their data set of topics also included *specific software and tools for teaching and learning* and *distance learning* or *homeschooling*. Furthermore, they found that mutual help was essential in these times, though this is more of an implicit topic. They extracted the main challenges and opportunities through a manual content analysis. They found the following challenges: good digital classes, missing software, and the lack of digital know-how for digital teaching. Opportunities included networking and sharing possibilities, offering digital material, and explanations or tricks. Concerning the second part of our RQ2 (whether there are obvious changes from the period of January and February 2020), we determine that the exchange within the #twlz community became more specific during the COVID-19 pandemic. Online education became relevant overnight, and teachers were expected to change how they taught. This caused them to have more concrete questions about online education and homeschooling and therefore look for help, exchange, and tips within the #twlz community.

Regarding the second school closures in April and May 2021 (RQ 3), we found two more or less evenly distributed topics. Schools were affected by the so-called “federal emergency brake” (German: “Bundesnotbremse”), which limited in-classroom teaching to schools in counties with COVID-19 incidence levels below 200 and then 165 in the relative county (Grill, M., Mascolo, G., Munzinger, P., Zick, T., 2022). According to our findings, the main topics were digital education and tools, especially in Bavaria, and school life during the pandemic. Both were influenced by the fluctuations in incidence levels. Since Fütterer et al. (2021) published their work in 2021, their work does not include this period. In contrast to the period of the first nationwide school closures in Germany (March and April 2020, P2), the topics discussed in relation to school life became even more precise in terms of the requirements for schooling in times of short-term adjustments based on incidence levels. Again, we have a reference to a specific region (Bavaria), which may indicate that COVID-19 measures were especially strong in Bavaria during this time, increasing the local exchange on the topic. However, we can already see the Bavarian influence before the pandemic in topic 3 of P1, so we assume there is a strong #twlz community in Bavaria.

In terms of content, our findings show that the topics did not change completely over time (from P1 to P2 to P3). Before the COVID-19 pandemic, the exchange in the #twlz community was about education and learning in general. Digital education played a role, but it was not the dominant topic. Over time, the exchange shifted toward digital education, the tools needed, and how to adapt schooling to settings such as homeschooling or short-term changes in educational conditions. This is also underlined by the overall themes of ChatGPT, which range from education and learning environments (P1) to adapting education in the face of crisis (P2) to digital education and pandemic adaptations (P3).

Regarding RQ4, we have the following insights: It is not always possible to understand a complete data set in a reasonable amount of time. Hence, techniques such as ML are helpful for reducing the time, funding, and personnel needed. Although the data come from different sources, a comparison of our findings with ChatGPT and the topics extracted by Fütterer et al. (2021) for the period before and during the first nationwide school closures in Germany reveals no major differences.

Given the time-sensitive nature of research and the urgent need for results [*e.g.,* health-related analysis (Rauschenberger and Baeza-Yates, 2020)], using methods such as those suggested by Mayring (Mayring, 2015) or the examination of detailed bigram networks may not be ideal because these methods are resource-intensive even for smaller data sets. Since the outcomes are comparable in our case, less time-critical methods (such as using the ChatGPT interface) may be preferable, especially when resources are limited. In addition, since the results are similar, ChatGPT can enhance the personnel’s point of view with its objectivity. ChatGPT also summarizes the topics in a shorter and more precise way, so the combination of both perspectives can enrich the result.

Our findings are limited by the fact that we do not have the same raw Twitter data set as Fütterer et al. (2021), which may occur due to deletion of user accounts. Since ChatGPT can only handle a limited amount of data input, we only used lemmatized keywords within topics as input. Therefore, ChatGPT only had a very small view of the data. In addition, school closures in period P3 were not the same for each region. This may have affected the urgency or time users spent on Twitter in general. We did not find any major effect on the topics themselves but rather on the number of tweets (P1: 5,229 tweets; P2: 11,137 tweets; P3: 14,485 tweets). Finally, there are various biases in the data sets, and it is important to consider that Twitter is not representative of the entire population (Graells-Garrido et al., 2019). These biases must be acknowledged and addressed when utilizing these insights for decision-making.

6 CONCLUSION

We conducted an analysis of our Twitter data set from three distinct time periods (before school closures, during the first nationwide COVID-19 school closures in Germany, and during the second school closures) within the #twitterlehrerzimmer or #twlz community. We used ChatGPT to extract themes and compared the outcomes with those of a previous study. The results from various research methodologies yielded similar insights regarding the exchanges of teachers on Twitter. However, we observed that ChatGPT provides comparable results with greater ease of use and less effort.

The next step is to conduct a systematic analysis comparing ML techniques to traditional manual methods to explore their respective limitations in content analysis, whether for small or large data sets. Furthermore, the limitations of using ChatGPT in terms of reliability and accuracy should be the subject of further investigation.

ACKNOWLEDGMENTS

In our work, we used ChatGPT prompts as a tool to abstract themes from specific keywords. This research was supported by the EQUAVEL project PID2022-137646OB-C31, funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE.

REFERENCES

Aarsen et al. (2023). Documentation Natural Language Toolkit. Accessed July 3, 2023, from https://www.nltk.org/.

Arya, N. (2022). TF-IDF Defined. Accessed April 11, 2024, from https://www.kdnuggets.com/2022/10/tfidf-defined.html.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101. DOI: 10.1191/1478088706qp063oa.

Britannica, The Editors of Encyclopaedia (February 9, 2024). X. Accessed April 10, 2024, from https://www.britannica.com/money/Twitter.

Fütterer, T., Hoch, E., Stürmer, K., Lachner, A., Fischer, C., and Scheiter, K. (2021). Was bewegt Lehrpersonen während der Schulschließungen? – Eine Analyse der Kommunikation im Twitter-Lehrerzimmer über Chancen und Herausforderungen digitalen Unterrichts. *Zeitschrift fur Erziehungswissenschaft : ZfE*, 24(2):443–477. DOI: 10.1007/s11618-021-01013-8.

Graells-Garrido, E., Baeza-Yates, R., and Lalmas, M. (2019). How representative is an abortion debate on twitter? In Boldi, P., editor, *Proceedings of the 10th ACM Conference on Web Science*, ACM Digital Library, pages 133–134, New York,NY,United States. Association for Computing Machinery.

Grill, M., Mascolo, G., Munzinger, P., Zick, T. (2022). Deutschlands Problemzone: Corona und Schulen. Accessed April 14, 2024, from https://www.sueddeutsche.de/projekte/artikel/politik/corona-und-die-schulen-deutschlands-problemzone-e671108/.

Huber, S. G. (2021). Schooling and Education in Times of the COVID-19 Pandemic: Food for Thought and Reflection Derived From Results of the School Barometer in Germany, Austria and Switzerland. *International Studies in Educational Administration (Commonwealth Council for Educational Administration & Management (CCEAM))*, 49(1).

Mayring, P. (2015). Qualitative Content Analysis: Theoretical Background and Procedures. *Approaches to Qualitative Research in Mathematics Education*, pages 365–380. DOI: 10.1007/978-94-017-9181-6_13.

Mohammad, S. M. and Turney, P. D. (2013). NRC emotion lexicon. DOI: 10.4224/21270984.

OpenAI (2022). Introducing ChatGPT. Accessed April 15, 2024, from https://openai.com/blog/chatgpt.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45–77. DOI: 10.2753/MIS0742-1222240302.

Rauschenberger, M. and Baeza-Yates, R. (2020). How to Handle Health-Related Small Imbalanced Data in Machine Learning? *i-com*, 19(3):215–226. DOI: 10.1515/icom-2020-0018.

Řehůřek, R. (December 21, 2022). GENSIM: Topic modelling for humans. Accessed July 3, 2023, from https://radimrehurek.com/gensim/index.html.

twarc (2024). twarc2. Accessed March 25, 2024, from https://twarc-project.readthedocs.io/en/latest/twarc2_en_us/.

Twitter (2023). Twitter API. Accessed July 8, 2023, from https://developer.twitter.com/en/docs/twitter-api.

Wartena, C. (2019). A Probabilistic Morphology Model for German Lemmatization. DOI: 10.25968/OPUS-1527.

Webb, H., Jirotka, M., Stahl, B. C., Housley, W., Edwards, A., Williams, M., Procter, R., Rana, O., and Burnap, P. (2017). The ethical challenges of publishing twitter data for research dissemination. In Boldi, P., editor, *Proceedings of the 2017 ACM on Web Science Conference*, ACM Digital Library, pages 339–348, New York, NY. ACM. DOI: 10.1145/3091478.3091489.

Weigand, A. C., Jacob, M. F., Rauschenberger, M., and Escalona Cuaresma, M. J. (2024). Research Protocol for Analyzing Tweets Using Topic Modeling and ChatGPT: What We Can Learn About Teachers and Topics During COVID-19 Pandemic-Related School Closures. DOI: 10.13140/RG.2.2.12205.91367.

Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., and Zhu, T. (2020a). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PloS one*, 15(9):e0239441. DOI: 10.1371/journal.pone.0239441.

Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., and Zhu, T. (2020b). Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. *Journal of medical Internet research*, 22(11):e20550. DOI: 10.2196/20550.