Online Learning Communities in Covid-19 Days: Mining Twitter Data

Mahmut BOZKURT, mahmutbozkurt@sinop.edu.tr, Sinop University, Turkey, https://orcid.org/0000-0002-3169-6819

SUMMARY

As a teacher educator who provides online education during the pandemic process, I wanted to examine what is needed and what is shared in online learning communities. The aim of this research is to reveal the most shared concepts and sentiments in online learning networks using the data obtained from Twitter. In this context, the #edtech hashtag, one of the most shared hashtags, was chosen and a total of 134,101 unique tweets were analyzed in the specified time period with some data mining and sentiment analysis techniques. As a result of these analyzes, the most shared websites, words and bigrams were extracted. In addition, sentiment analysis based on NRC and Bing lexicons was also performed. According to the findings about the most shared websites, it has been seen that, in addition to current learning approaches such as digital learning designs, STEAM, coding, robotics, artificial intelligence and augmented reality, various mathematical calculation applications and websites based on fun math activities are shared. The words learning, student, school, education, online, free, classroom, teacher, join and check are the most used words. The most used bigrams are activity pack, remote learning, pupil activity, online learning and google classroom. Considering the sentiment analysis, it was seen that the most prominent sentiment based on the NRC lexicon was positive, followed by the sentiments of trust, anticipation and joy. According to Bing, positive sentiments were dominant.

Keywords: Educational technology, edtech, text mining, sentiment analysis, COVID-19

INTRODUCTION

Data mining, according to the Baker (2010), is the field of discovering useful information from large amounts of data. Social media data mining allows obtaining information about people's opinions or feelings. Also, social media data can be used to make predictions about a wide variety of topics and also help to understand the sentiments and thoughts of others (Bovet, Morone & Makse, 2018; Tumasjan, Sprenger, Sandner & Welpe, 2010; Asur & Huberman, 2010). In social movements such as the Arab Spring showed that people do not use social media only to express their feelings and thoughts, but also for call-to-action (Choudhary, Hendrix, Lee, Palsetia & Liao, 2012). Today, people can communicate, learn, and share their knowledge through divergent digital environments. Such social media platforms are increasingly becoming image-based environments (pics and videos) rather than texting. On these microblogging platforms, people share selective and limited posts (e.g., 140 characters) about their lives or interests.

Twitter is used for creating classroom community, developing collaborative writing, initiating online question-and-answer forums, establishing cooperation between schools and even countries, project management, evaluating a student by interactions on a given topic, thinking or reflecting thoughts on a particular topic, posting about a conference, presentation, or workshop and thus allow interacting with the attendees, and particularly for increasing interaction in higher education. Also, Twitter can be used to facilitate online class discussions, to provide learning experiences, and to help creating personalized learning networks (Grosseck & Holotescu, 2008).

So, it can be argued that social media is used for a variety of purposes from marketing, medicine, politics to education. Regarding education, researchers are trying to examine Twitter for learning communities established by stakeholders in education, professional development, higher education (Ricory & Feliz, 2016). In this regard, many papers are available on Twitter's contributions to the teachers' professional development, for online and blended learning, and how to design (Macià & García, 2016; (Lantz-Andersson, Lundin & Selwyn, 2018; Macià & García, 2016).

It is stated that Twitter has positive impacts on informal learning, class dynamics, motivation, and college students' academic and psychological development (Dhir, Buragga & Boreqqah, 2013). Davis (2015) examined the use of Twitter by teachers regarding professional development and communities of practice and found that teachers use Twitter for sharing their practices, knowledge and experience exchange, and benefit of the professional expertise of experienced colleagues. Furthermore, it was found that although teachers were surprised with the massive size and speed of Twitter data, they get used to it and they understood that those discussions are learning-centric and supportive. Accordingly, teachers commented that learning communities on Twitter offer them a sense of belonging more than their real-life and their advantages far outweigh their disadvantages.

Recently, social media and internet usage has increased with covid-19 (Yıldırım & İpek, 2020; Yıldızgörür, 2020). And some reports state that pre-service teachers use social media and the internet for professional learning

networks (Davis, 2015). Online learning communities can be classified as formal and informal (Lantz-Andersson, Lundin & Selwyn, 2018; Macià & García, 2016). Regarding educational purposes, some reports indicate that Twitter is used for communication, professional development, and class activities (Krutka & Milton, 2013). Twitter attracts the attention of stakeholders in education as it allows thousands of people with common interests to make real-time knowledge sharing (Chen, 2011), to maintain interaction, to organize discussions on a given problem immediately, and fast information flow (Britt & Paulus, 2016) especially during the Covid-19 pandemic.

One of the most important connection and collaboration facilitators on Twitter is hashtags. For example, it was reported that over 320,000 unique users shared education-related 25 hashtags in March 2017 (Carpenter, Tani, Morrison & Keane, 2020). Some of these hashtags are #Edchat, #Edtech, #collaborativePD, #globaledchat. In the current paper, the #edtech hashtag is selected since the researcher (as an educator) aims at identifying educators' wishes and needs during the pandemic. In this context, it is thought that a topic related to the increased use of technology in education during the covid-19 pandemic and educational technologies can be discussed.

Although many studies examining Twitter as online learning networks were published (Krutka & Milton, 2013; Davis, 2015; Trust, Krutka & Carpenter, 2016), the present study focuses on a different perspective using a large amount of data and data mining techniques. And more importantly, since the dataset was collected during the COVID-19 process, the aim is to reveal which concepts, materials, or activities are most frequently shared and which sentiments stand out in the era of digitalized learning during the pandemic. The primary purpose of this study is to reveal the most frequently commented concepts and sentiments on online learning environments using Twitter data. The answers were sought to the following questions about Twitter data collected during the covid-19 pandemic;

a.   What are the descriptive features of Tweets?

b.   What are the most frequent words, bigrams, and shared websites?

c.   What are the sentiments of shared Tweets?

METHOD

Since data mining applications are used in this paper, the method section is organized to introduce these applications. In this regard, at first, some general information and top ten considerations regarding the tweets are presented. Second, word counting and bigrams of tweets were calculated and analyzed. In the third and last step, sentiments in the Twitter posts were analyzed. Thus, what is posted about educational technologies during the covid-19, what the learners seek, and sentiments regarding the big picture were revealed.

Data preprocessing

Tweets were collected at random time intervals between May and October 2020. No language filter was implemented on the data collected using Twitter API. Furthermore, there were both duplicates (since the tweets were collected at random time intervals) and retweets. Therefore, these duplicates were excluded. Further, since it is aimed to examine the sentiments and thoughts of the unique users and the concepts used by these users, retweets were also excluded. The number of tweets retrieved and the amount of data after exclusion are shown in Graph 1.

Graph 1. Total and Clean Data

Graph 1. Total and Clean Data
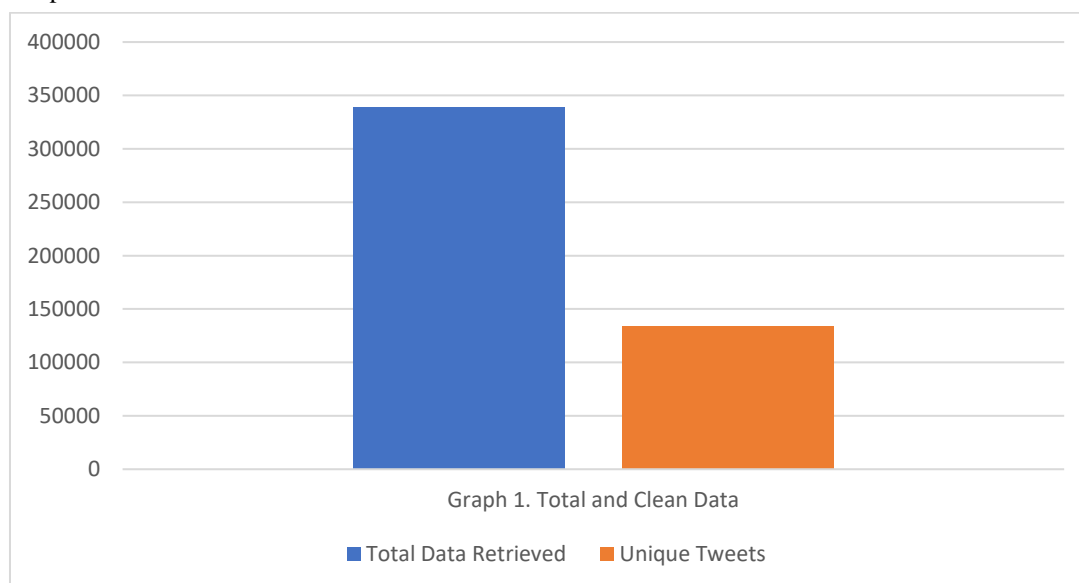
In the data gathering process, all tweets within the specified time period were searched and extracted from Twitter using the twitteR package, written in the R programming language. The data were collected between 13 May 2020-17 October 2020. During the data extraction process, 338,926 tweets were collected and archived in CSV format. Then, among this dataset, a total of 134,101 unique tweets posted by 23,833 unique users were further analyzed. The texts in the dataset were examined and the findings were extracted using various text mining techniques.

Following the exclusion of retweets and duplicates, the data were organized for further analyzes using tidyverse software written in R programming language. All uppercase characters were converted into lowercase characters, punctuations, numbers, and links were removed. So, only the words were kept for the analyses. After cleaning the data, the words and bigrams were calculated separately for Word counting and sentiment analysis.

Data Analysis (1), Word counting and bigrams

Data mining provides a viewpoint based on statistics and numbers in both word counting and sentiment analysis. In general, this procedure relies on segmenting the texts into words and analyzing the frequency of these words, analyzing the words used together (bigrams), or revealing sentiments through most common lexicons or classifications.

Data Analysis (2), sentiment analysis

This technique is based on the idea "words are associated with emotions" (Mohammad & Turney, 2013). In the text mining processes, the texts are treated as a bag of words (Misuraca, Forciniti, Scepi & Spano, 2020). Accordingly, while the words in a given dataset can be analyzed using known lexicons such as Bing, AFINN, NRC, they can also be classified as supervised or unsupervised using a Bayesian approach. Also, a researcher or user can develop their own lexicon word database and perform customized analyzes. In this context, lexicon-based sentiment analyzes were conducted in the current paper using syuzhet. The most well-known sentiment analysis lexicons, Bing and NRC were used.

FINDINGS

Descriptive Features of Tweets

The dataset consists of 338,926 tweets archived in the CSV format. Then, a total of 134,101 unique tweets were selected for further analysis. Accordingly, the findings regarding the most frequently shared websites, words, bigrams, as well as sentiment analysis results are presented in the following sections.

The Most Shared Websites, Frequent Words, and Bigrams

Top Ten Shared Websites

The dataset was analyzed to determine the most frequently shared websites and the results obtained are presented in Table 1.

Table 1. *Top Ten Web URLs Shared Most Frequently*

Top Shared Web URLs	About	Number of Shares
1. Rachelle Dene Poth's books on Amazon	She has been described herself as an ISTE certified educator, consultant, and author.	274
2. Personal website of Rachelle Dene Poth	Poth's personal web page	257
3. Oodlu.org	A website that provides support material to teachers for play-based learning	162
4. Daniel Down's YouTube channel	This URL is about digital learning environments. He studies STEAM and robotics and he focuses on the design of digital learning	122
5. THRIVEinEDU podcasts by Rachelle Dene Poth	This URL introduces new trends and topics about teaching and learning such as AI, AR/VR, global collaborations, preparing students for the future, PBL, SEL, coding, PLN.	116
6. An app about quadratic roots calculator	An app related to calculating quadratic roots	98

7.	An app about Pythagorean calculation	An app related to finding any missing side of a right triangle	97
8.	An app about geometry area calculation	An app related to finding the area of geometric shapes	97
9.	An app called subtraction fun house	An app about fun activities for kindergarten	94
10.	An app called addition fun house	An app about addition activities in math especially for kindergarten	94

As seen in Table 1, the top Web URL shared most frequently is Dene Poth's books and it is followed by her personal website. The third most frequently shared website provides support material to teachers for play-based learning. The fourth website is Daniel Down's YouTube channel about innovative topics in education and design of digital learning. The fifth website is Poth's podcasts on trending topics such as AI, AR/VR, global collaborations, professional learning networks (PLNs), coding, STEAM. Put simply, the top five websites are about trending learning approaches such as design of digital learning and STEAM, coding, robotics, artificial intelligence and augmented learning. Surprisingly, the remaining five URLs are about some mathematical calculation apps and fun math activities.

Top Ten Most Common Words

The most common words in the posts are listed in Table 2.

Table 2. *Top Ten Most Common Words*

Words	n
Learning	23409
Student	13741
School	11194
Education	8760
Online	7181
Free	6502
Classroom	5841
Teacher	7224
Join	5083
Check	4979

Following the pre-processing, the most common words and bigrams were counted. According to the results obtained, the most frequently used word is *learning* (n=23409) and followed by *student* (n=13741) and *school* (n=11194). As seen in Table 2, besides educational concepts such as *education*, *classroom*, and *teacher*, technology-related concepts such as *online*, *free*, *join*, and *check* were also used.

Top Ten Most Common Bigrams

The most frequently used bigrams given in Table 3 provide some insight into the recent period. The common bigrams can be used to make interpretations about the related period.

Table 3. *Top Ten Most Common Bigrams*

Bigrams	n
Activity pack	1794
Remote learning	903
Pupil activity	746
Online learning	722
Google classroom	716
Video tutorials	680
Distance learning	566
Primary computing	509

| Media4math resources	488
Home-schooling	444

In this regard, the concepts, *activity pack*, *pupil activity*, *primary computing*, *video tutorials*, and *media4math resources* can be associated with the idea that education stakeholders are seeking answers to the question "what should we teach" during the pandemic. Also, *online*, *remote*, *distance learning*, and *home-schooling* concepts draw attention to the environment. Finally, the bigram, *Google classroom* indicates the topic of tools.

Sentiment Analysis of Shared Tweets

In this paper, sentiment analysis was also performed based on NRC and Bing lexicons. The NRC sentiment analysis is presented in Graphs 2 and 3.

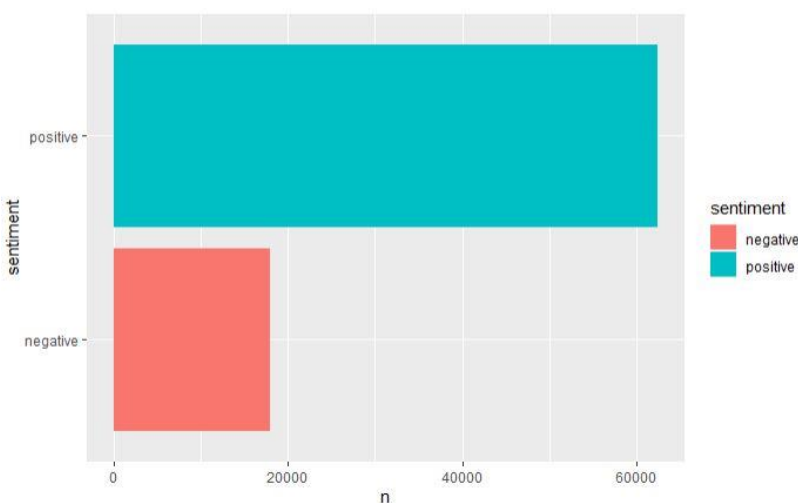Graph 2. *NRC Sentiment Analysis Results*
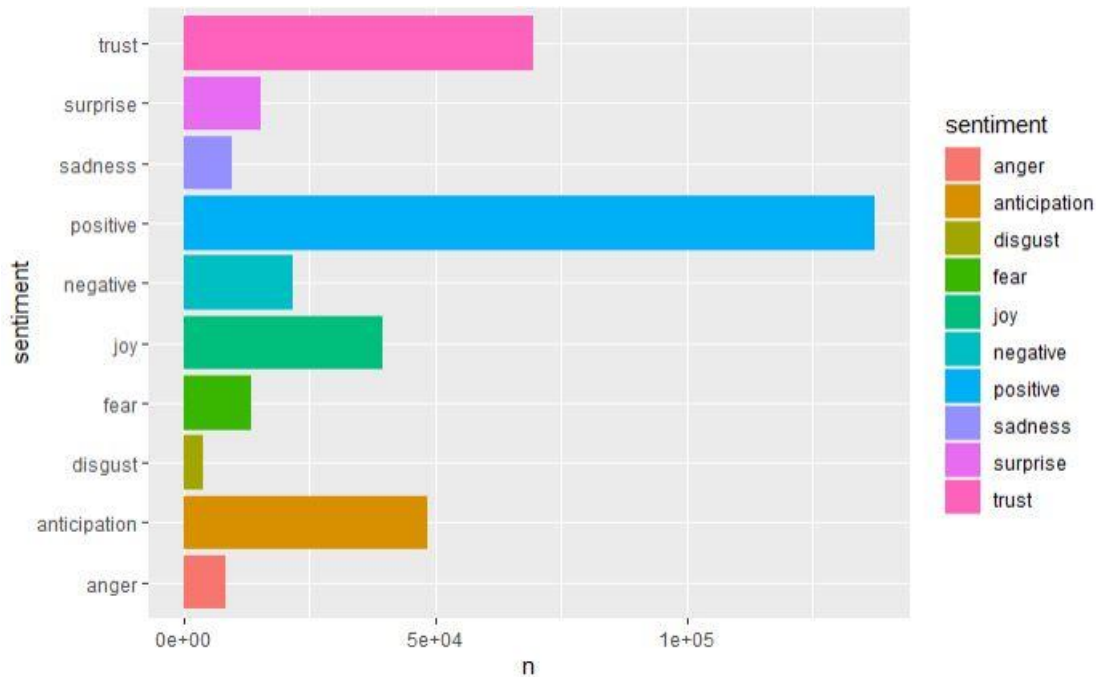
NRC sentiment lexicon (Mohammad & Turney, 2013) analyzes a given data by comparing it with 10 different sentiments. As seen in Graphs 2, the most common sentiment is *positive* and followed by *trust*, *anticipation*, and *joy*.
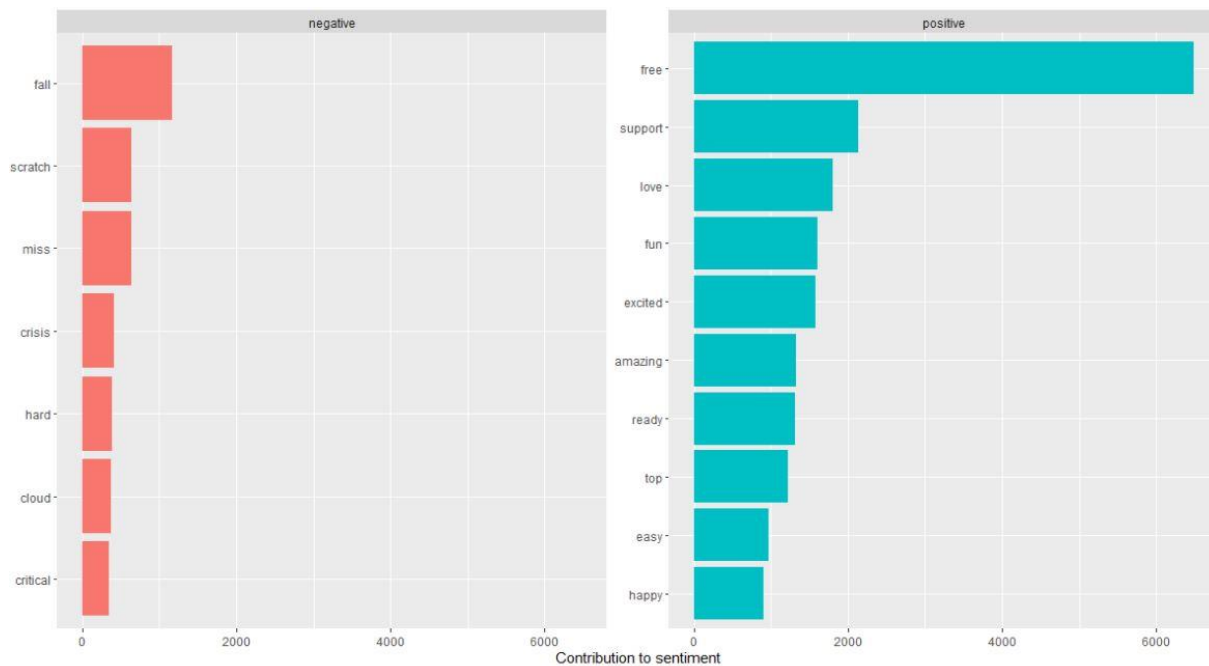
Bing Sentiment Analysis Results

Bing lexicon sentiment analysis was performed on two different sentimental states, positive and negative. Accordingly, the analysis results and most positive and negative terms are presented in Graphs 3 and 4.

Graph 3. *Bing Sentiment Analysis Results*

As shown in Graph 3, the Bing analysis results indicate that a majority of the postings about the Edtech hashtag are positive. The most frequently used words for positive and negative sentiments are shown in Graph 4.

Graph 4. *The Most Frequently Used Sentiment Words*

The most contributing words to positive and negative sentiments were examined by Bing lexicon sentiment analysis. As seen in Graph 4, the most frequently used positive word is *free* and followed by *support*, *love*, *fun*, *excited*, *amazing*, *ready*, *top*, *easy*, and *happy*. On the other hand, it was determined that *fall*, *scratch*, *miss*, *crisis*, *hard*, *cloud*, and *critical* are the most used negative words. Regarding the positive terms, the use of terms *free* and *support* can be explained that free resources and support in specific issues are needed for educational technologies. It is seen that most of the remaining positive terms express positive sentiments. Considering negative terms, it can be argued that the term *fall* was used in postings of the users to express what they will do in the forthcoming education semester in which the data was collected. It is seen that the users make frequent postings in the first academic semester following the Covid-19 pandemic about that period. Here, the word *scratch* draws attention as the second most frequently shared term.

CONCLUSION AND DISCUSSION

In the current study, a series of text mining and sentiment analysis were performed using Twitter data in the context of teachers' online learning communities. Sentiment analysis based on NRC and Bing lexicons were performed and it was determined that according to NCR, feelings of positive and trust are the most frequently used terms and according to Bing, positive postings are more frequent. In a study conducted by Trust et al. (2020), the authors found that a majority of the postings consisted of neutral and positive sentiments. Therefore, it can be argued that the findings of the current study are consistent with the previous results.

Furthermore, Trust et al. (2020) stated that educators used the concepts of encouragement or gratitude in their postings about the solutions to the problems they encountered, and there are also postings about the difficulties they overcome or postings supporting the distance education method. In the current study, the most frequently used words were extracted and sentiment analyzes were performed and accordingly, it was found that the positive sentiments far outweighed the negative ones according to both lexicons.

Regarding the most frequently shared websites, it was found that most of the postings were about stem, coding, robotic, and mathematics. Moreover, the most frequently shared concepts were learning, education, and school. Similarly, Kimmons et al. (2018) found that the concept of STEM is one of the most frequently shared hashtags. Also, in the same study of Kimmons et al. (2018), concepts were grouped under particular themes such as affinity spaces for edtech, edchat; education-related topics for education, learning; groups of people for teacher, student. Based on these results, it can be claimed that the obtained concepts in the current study are similar to those obtained in the study of Kimmons et al. (2018).

Carpenter et al. (2021) determined that educators shared various tools, apps, and resources through a common hashtag with regard to professional knowledge sharing and they shared various webinars or communities about information broadcasting, as well as different resources, event information and links. In the present study, the analyzes of the shared links revealed that various apps, tools, and trending topics were shared by users.

In the current paper, the most frequently shared websites were also extracted and it was found that the most frequently shared URLs and apps were related to STEM, AI, AR/VR, robotics, and mathematics. Similar to these, in a review paper published by Lantz-Anderson et al. (2018), the authors found that Twitter is used by educators for steering teachers to new and trending ideas and resources. So, it can be argued that the results obtained in this paper regarding both the most frequently shared websites and bigrams are consistent with the previous reports.

Sentiment analyzes were also conducted in the current study and it was determined that a majority of the analyzed data consist of terms associated with positive sentiments. Many reports in the literature state that Twitter supports educators socially and emotionally (Davis, 2015; Chen, 2011; Carpenter & Krutka, 2014; Staudt Willet, 2019). Similarly, the analysis of the Twitter data collected using data mining techniques indicates mostly positive sentiments. As a result of the research, it can be suggested that researches about different courses (such as social studies, math, etc), educational stages (primary, secondary, higher education, etc.) and stakeholders (teachers, students, administrators, etc.) can be done in the context of online learning communities.

REFERENCES

Asur, S., & Huberman, B. A. (2010, August). *Predicting the future with social media*. In 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (Vol. 1, pp. 492-499). IEEE.

Baker, R. S. J. D. (2010). Data mining for education. *International Encyclopedia Of Education*, 7(3), 112-118.

Britt, V.G. & Paulus, T. (2016). "Beyond the Four Walls of My Building": A Case Study of #Edchat as a Community of Practice. *American Journal of Distance Education*, 30(1), 48-59.

Bovet, A., Morone, F. & Makse, H. A. (2018). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific Reports*, 8, 8673.

Carpenter, J. P., Trust, T., Kimmons, R., & Krutka, D. G. (2021). Sharing and self-promoting: An analysis of educator tweeting at the onset of the COVID-19 pandemic. *Computers and Education*, 2, 100038.

Carpenter, J. P., Krutka, D. G., & Kimmons, R. (2020). # RemoteTeaching & # RemoteLearning: Educator tweeting during the COVID-19 pandemic. *Journal of Technology and Teacher Education*, 28(2), 151-159.

Carpenter, J., Tani, T., Morrison, S., & Keane, J. (2020). Exploring the landscape of educator professional activity on Twitter: An analysis of 16 education-related Twitter hashtags. *Professional Development in Education*, 1-22.

Carpenter, J. P., & Krutka, D. G. (2014). How and why educators use Twitter: A survey of the field. *Journal of Research on Technology in Education*, 46(4), 414–434.

Chen, G.M. (2011). Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behavior*, 27(2), 755-762.

Choudhary, A., Hendrix, W., Lee, K., Palsetia, D., & Liao, W. K. (2012). Social media evolution of the Egyptian revolution. *Communications of the ACM*, 55(5), 74-80.

Davis, K. (2015). Teachers' perceptions of Twitter for Professional development. *Disability and Rehabilitation*, 37(17), 1551-1558.

Dhir, A., Buragga, K. & Boreqqah, A.A. (2013). Tweeters on campus: Twitter a learning tool in classroom? *Journal of Universal Computer Science*, 19(5), 672-691.

Grosseck, G. & Holotescu, C. (2008, 17-18 April). *Can we use Twitter for educational purposes*. The 4th International Scientific Conference eLearning and software for education, Bucharest, Romania.

Royce Kimmons, Jeffrey P. Carpenter, George Veletsianos & Daniel G. Krutka (2018) Mining social media divides: an analysis of K-12 U.S. School uses of Twitter, *Learning, Media and Technology*, 43:3, 307-325

Krutka, D., & Milton, M. K. (2013). The enlightenment meets twitter: using social media in the social studies classroom. *Ohio Social Studies Review*, 50(2), 22-29.

Lantz-Andersson, A., Lundin, M., & Selwyn, N. (2018). Twenty years of online teacher communities: A systematic review of formally-organized and informally-developed professional learning groups. *Teaching and Teacher Education*, 75, 302-315.

Macià, M., & García, I. (2016). Informal online communities and networks as a source of teacher professional development: A review. *Teaching and teacher education*, 55, 291-307.

Mohammad, S. M., & Turney, P. D. (2013). Nrc emotion lexicon. National Research Council, Canada, 2.

Misuraca, M., Forciniti, A., Scepi, G. & Spano, M. (2020). Sentiment analysis for education with R: packages, methods, and practical applications. arXiv preprint arXiv:2005.12840.

Ricory, M.C. & Feliz, T. (2016). Twitter as a learning community in higher education. *Journal of Educational Technology & Society*, 19(1), 237-248.

Staudt Willet, K. B. (2019). Revisiting how and why educators use Twitter: Tweet types and purposes in #Edchat. *Journal of Research on Technology in Education*, 51(3), 273–289.

Trust, T., Krutka, D. G., & Carpenter, J. P. (2016). "Together we are better": Professional learning networks for teachers. *Computers & Education*, 102, 15-34.

Tumasjan, A.T., Sprenger, T.O., Sandner, P.G. & Welpe, I.M. (2010). *Predicting elections with Twitter: What 140 characters reveal about political sentiment*. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, D.C.

Yıldırım, O. & İpek, İ. (2020). Yeni koronavirüs salgını dolayısıyla gündeme gelen sosyal izolasyon ve gönüllü karantina döneminde internet ve sosyal medya kullanımı. *İletişim Kuram ve Araştırma Dergisi*, 52, 69-94.

Yıldızgörür, M.R. (2020). Küresel salgın sürecinde sosyal medyanın habercilik rolü: yararlı mı? Zararlı mı? Kullanıcı deneyimleri üzerine bir araştırma. *Kastamonu İletişim Araştırmaları Dergisi*, 5, 44-58.