ORIGINAL ARTICLE

Does technology assist to continue learning during pandemic?
A sentiment analysis and topic modeling on online learning in south asian region

Abdul Raheem Fathima Shafana1 · Sahabdeen Mohamed Safnas2

Received: 22 January 2022 / Revised: 4 June 2022 / Accepted: 6 June 2022 / Published online: 26 June 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract
Online mode of education has been identified as the subtle solution to continue learning during the pandemic. However, the accessibility to online platforms, suitable devices, and connections are not equal across the globe thus raising the question of whether the opinion of the public in the South Asian region where the technology is not comparatively higher as in the western world would be the same as that to the global perspective. This study involves the sentiment analysis of natural language processing on recently tweeted data and concludes that the sentiment of the South Asian public remains positive as online education is the most suitable approach to overcome the learning difficulties during a pandemic. The study performs a ternary classification based on the polarity scores obtained from two robust lexicon-based sentiment analyzer tools namely VADER and TextBlob and observes that 63.2% of the tweets were positive, 30.5% of the tweets were neutral and around 6.3% of them were negative. Finally, topic modeling was also performed using the Latent Dirichlet Allocation method to gain insight into each of the classes.

Keywords Technology-blended LEARNING · Online education · South Asian education · COVID19 sentiment analysis · Natural language processing

Abbreviations
COVID-19 Coronavirus disease 2019
API Application programming interfaces
BNC Bayesian network classifier
VADER Valence aware dictionary and sentiment reasoner
URL Universal resource locator
NLTK Natural language toolkit
EDA Exploratory data analysis
LDA Latent dirichlet allocation
ERL Emergency remote learning

1 Introduction
Technology-assisted teaching–learning methodology has been consistently affected by its continuous development ever since the late 1990’s when the term was first coined (Pishva et al. 2010). Online education was first initiated to improve the quality of learning while reducing the cost and extending access to students across the globe (Twigg 2003). It is one such digital transformation that has been gaining popularity as an efficient method of technology-blended learning. Currently, the outbreak of the COVID19 global pandemic has redefined the scope of online education on a massive scale. Distance learning a.k.a Online Education has become the most promising solution to live a new normal amidst the pandemic to continue learning without a disrupt (Arambepola 2020). However, accessibility to online platforms is not equal across the globe and the lack of online devices and remote connections in the South Asian region has made the situation even worse (Reuters 2021). This arises a necessity to explore the opinion of South Asian people in adopting an online mode of learning.

Online Education has been witnessing a variety of feedbacks for its benefits since the time of its inception. The
principal benefit of online education is its flexibility of time and space (Mupinga 2005). Furthermore, the higher level of motivation, expansion of access, provision of choices for learning are certain other benefits pointed out in literature (Cavanaugh et al. 2009). On the other hand, technical issues faced by students, lesser importance towards collaborative learning, and lesser interactions with the faculty are pointed as disadvantages of online education (Dumford & Miller 2018). Thus, the review on the compendium of the literature suggests that there exists an equal weightage of advantages and disadvantages. Therefore, it is beneficial to mine the opinion based on its context of use rather than exploring the merits and demerits worldwide as there can be many other reasons as to why people do not like to adopt online learning in addition to the obvious causes as stated above.

To be concise, the review of literature suggests that the scientific investigations in this direction are quite minimal. Thus, it seems hard to derive some strong insights on the use of online learning in place of face-to-learning and to assume how public would accept this paradigm shift. In addition, we could note that there are substantial controversies about the outcome of distance learning which calls forth further scientific investigations to assess if the online learning serves as an appropriate substitute for the face-to-face learning, especially in the event of a pandemic.

In order to extract the public opinion towards the online learning system, various attempts have been undertaken. From the literature, Natural Language Processing and Sentiment Analysis are proved to be successful domains in obtaining the opinions for a specific topic (Saberi & Saad 2017). In this context, social media can be considered as the most suitable channel to obtain opinions since netizens use social media greatly not only to consume the content but also to express their views and opinions readily (Jasti & Mahalakshmi 2019). The tweets from Twitter users are such a wealth of information that enables the data scientists to explore and get insights into a specific topic (Jasti & Mahalakshmi 2019). Opinion Mining of Twitter data enables us to classify the public opinion as either being positive or negative or even neutral depending on the polarity of the tweet.

The objective of the study is to identify the public opinion in South Asia towards transforming to online education during the COVID19 pandemic and to perform topic modeling to obtain the context on Twitter contents. For this, the study utilizes the lexicon-based sentiment analysis in conjunction with Natural Language Processing on Twitter data. Tweets relevant to online learning were scraped using Twitter Application Programming Interfaces (API) from 10th of December 2021 to 31st of December 2021 during which OMICRON, the new variant of COVID19 started to spike. Then, the tweets from the South Asian region were specifically filtered out. Exploratory Data Analysis (EDA) was first performed to gain insight into the data. Later, using two sentiment analyzing tools, the tweets were classified as either positive, negative, or neutral. The tweets that were classified similarly using both the analyzers were only used for further analysis and interpretation. Topic modeling based on the Twitter content was done to retrieve the common topics in relevant to the online learning. The detailed analysis is discussed in later sections of this paper.

To the best of our knowledge, this is the first work performing sentiment analysis and topic modeling for tweets about online learning and COVID-19 specific to the South Asian region. Although the usage of Twitter among the South Asian public was low a few years back, an increase in usage could be witnessed recently as public realized the potential of social media for digital diplomacy (Ittefaq 2019). Thus, the applied methodology is justified, and we believe that this study would perhaps be a pioneer in analyzing the Twitter data among the South Asian public for mining opinions, that could also help other practitioners to apply this methodology in various other contexts. The study was able to derive two important conclusions regarding the public opinion on online learning during pandemic. Firstly, the opinion among the South Asian public towards online learning is still positive, though they lack access to online devices and connections. Secondly, the cost and the lack of devices are the potential causes of creating a negative viewpoint towards online learning, the outcome of this research is reliable, and the findings could be utilized to devise a proper methodology to introduce advanced technology in the South Asian region to avail online education to this society as well.

2 Related works

2.1 Sentiment analysis on Twitter data

Twitter is an excellent platform to be used for text analysis mainly due to its limited character count and the Twitter users are from different ages, races, gender, culture, etc. (Alharbi & de Doncker 2019). Sailunaz and Alhajj (2019) analyzed the sentiment and emotions from Twitter posts to produce generalized and personalized recommendations for users depending on their activity on Twitter. The authors reveal that text is the most common method of revealing emotions and sentiments in social network posts. However, Alharbi and de Doncker (2019) used an enhanced approach to the above that includes the behavioral information of the user in the specific tweet in addition to the tweet text alone. The deep learning model used for the study outperformed other baseline methods in analyzing sentiments. Bayesian Network Classifier (BNC) is yet another approach to analyze the sentiments in Twitter data (Ruz et al. 2020). BNC has been used to analyze the sentiment of people during
natural disasters or critical events. Sentiments Analysis on such twitter data have led to the findings on event dynamics as well. Twitter data have been used to perform sentiment analysis on people’s reactions to a lawsuit (Sukma et al. 2020), to detect student depression (Mowery et al. 2017), and on many other contemporary issues. Thus, it can be concluded that Sentiment Analysis on social media data, especially on Twitter data has provided sufficient insights on the target context since the past (Agarwal et al. 2011; Budiharto & Meiliana 2018; Paltoglou et al. 2010; Paltoglou & Giachanou 2014; Zhou et al. 2013).

### 2.2 Sentiment analysis in response to pandemic

The impact of COVID-19 on education has created an urgent need to evaluate the opinion of distance learning ever than before especially due to the wide mixed perception towards remote learning. Public opinion on the use of Emergency Remote Learning (ERL) was analyzed using the data collected from Twitter (Asare et al. 2021). The study was able to detect positive and negative sentiments and revealed topics for each of the clusters thus serving as a recommendation to improve the online educational systems amidst the pandemic. Makalesi et al. (2021) also examined the public opinion regarding online learning during the pandemic. The study used the polarity, subjectivity, and emotion of the tweets and concluded that the tweets were more negatively relevant to online education. Batubara et al. (2021) also made use of the Twitter data to perform sentiment analysis to detect the feasibility of face-to-face education during the COVID19 pandemic in Indonesia using tweets from many countries.

However, the above studies are not confined to a specific region or an institute. Yet, Garcia and Berton (2021) undertook a similar study concerning two different geographical locations namely USA and Brazil. The study aimed at identifying sentiments and topic modeling in two different languages and found that the negative sentiments were superseding the positive sentiments. The study emphasizes that the investigation in this selective approach can provide many insights for further studies.

### 3 Materials and methods

The methodology employed for the study is presented in Fig. 1. The methodology for the study has four phases. Phase 1 is the acquisition of Twitter data. The study employed the public streaming Twitter API to retrieve tweets concerned with Distance Learning and COVID19. The query consists of the keywords such as ‘covid’, ‘covid19’, ‘pandemic’, ‘distance learning’, ‘online learning’, ‘virtual learning’ and ‘remote education’. The extracted tweets belong to the period between the 10th of December 2021 and the 31st of December 2021 during which OMICRON, the new variant of COVID19 started to spike. The query resulted in 211,624 tweets out of which 2103 tweets from the South Asian region (Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka) were filtered based on the location attribute of the tweet.

Phase 2 is the preprocessing and exploratory data analysis (EDA) on the filtered tweets. The tweets were cleaned and preprocessed by removing the null and duplicate values. Irrelevant information to the study was also removed since tweets consist of embedded URLs, images, usernames, and emoticons. The preprocessing step is crucial since the cleaner data yields accurate results. Natural Language Toolkit (NLTK) was used to obtain the processed data that contains the main message and the respective location. The regular expression in Python was used to remove special characters, Universal Resource Locators (URLs), retweets, user mentions, and unnecessary punctuation marks. The consecutive steps utilize many functionalities such as converting the tweets to lower case, removing the stop words, tokenizing the words, and eventually stemming them. The resultant tweets have now been processed and can be fed for sentiment classification.

Phase 3 is an important phase where the sentiment analysis was performed. The sentiment analysis was done using VADER and TextBlob tools. VADER is a parsimonious rule-based tool for sentiment analysis and a VADER lexicon that is used to perform well in the social media domain (Hutto & Gilbert 2014). TextBlob is another sentiment analyzing tool that uses the NLTK corpora (Laksono et al. 2019). TextBlob...
has been used for sentiment analysis in many of the previous studies as well. This study used both VADER and TextBlob Sentiment Analyzer tools to obtain a ternary classification of emotions such as positive, negative, and neutral. The threshold values for the VADER and TextBlob are based on previous studies and presented in Table 1. Later, the results from both tools were compared and the tweets where both tools return the same sentiment were kept for further analysis.

The next phase is topic modeling. This was mainly done to identify the prominent topics in each of the sentiment classes to obtain insights into the specific class. The list of tweets was classified based on the sentiments obtained from the tools. Topic modeling was done using the Latent Dirichlet Allocation (LDA) method and the words from each topic were also visualized using the matplotlib library. The documents for the topic modeling were extracted based on the sentiment classification from previous step. Since fine-tuning is an important step for topic modeling, hyperparameters were optimized. Number of topics was set at three while the number of words in each topic was set to 20. The other parameters that were fine-tuned includes the number of iterations (i.e., set at 10) and the learning method (i.e., online). Topic Modeling was done for each of the sentiments in the South Asian region. Further modeling was done for Sri Lanka specifically.

The results from the study are presented in the next section.

4 Results and discussion

4.1 Exploratory data analysis

The filtered tweets were first analyzed to observe the frequency of tweets from each of the countries in the South Asian region. The results show that (Fig. 2) India has the maximum number of tweets and Bhutan has the least number of tweets. The number of tweets has a significant correlation with the number of Twitter users in the country as the results were compared with the recent statistics (Degenhard 2021).

4.2 Sentiment analysis

The sentiment analysis on the Twitter data revealed that nearly 63.2% of the people have positive sentiment towards online learning during a pandemic. As the situation of the pandemic is not getting over more than two years, it seems that people are getting adapted to the online mode of learning. Despite the lack of connectivity and devices, the public tends to provide a negative sentiment to only about 6.3%. The sentiment was further analyzed based on the country for a detailed overlook of the opinion. The results are in Fig. 3a, b, respectively. The analysis by country revealed that the sentiment breakdown has a similar tendency where positive sentiment superseded the negative and neutral sentiments. The percentage mapping of sentiments showed that more than 50% of the opinions are positive in each of the countries under study as shown in Table 2.

| Table 1  | Threshold values for the sentiment tools |
|----------|-----------------------------------------|
| Vader (compound score from analyzer) | TextBlob (polarity score from analyzer) |
| Positive | >= 0.05 | > 0 |
| Neutral  | < 0.05 AND > -0.05 | = = 0 |
| Negative | <= -0.05 | < 0 |

Fig. 2 Country wise ranking based on the number of tweets
4.3 Topic modeling

The topic modeling suggested that people were much concerned with educational technology (EdTech) and used this pandemic to learn courses online especially the programming courses like Python. Specific topics much prevalently discussed in the context of Sri Lanka is the perspective of students doing Ordinary Level Examination on online learning which is an important examination in Sri Lanka. Furthermore, the lack of emotional connection in online classes and the use of bilingual and monolingual in distance learning have been discussed.

The sentiment-wise topic modeling in the South Asian region suggested that the reduction in tuition fees gains a positive sentiment of people for online learning, whereas the lack of complete attention in online classes and the issues in internet access could be inferred as negative topics. The topics such as the government job and applying for courses online could be perceived as a neutral topic among the public. Similar topics appeared when the topic modeling was done for each of the countries for each of the sentiments suggesting that the nature of public sentiment of all the South Asian countries are much correlated. The respective word clouds are as shown in Fig. 4 where Fig. 4a represents the word clouds of the topic from South Asia and Figs. 4b–d are the word clouds of positive, neutral, and negative sentiments, respectively.

5 Conclusion

This study aims to gain insight into the opinion of the public especially in the South Asian region towards adopting distance education during the COVID19 pandemic. Twitter has been identified as an appropriate platform where people express their opinions exclusively. Thus, Sentiment Analysis and Topic Modeling have been undertaken on acquired twitter data on distance learning from South Asian users. The results suggest that the public opinion towards distance learning remains positive as it is the optimal solution to continue learning during the pandemic. Despite the reduction of cost for online learning people still consider that the lack of internet access causes a negative view on online learning. Therefore, it can be concluded that the provision of
a sophisticated infrastructure on devices and connections could yield the fullest benefit of technology blended online education even after the pandemic. Moreover, the studies utilizing the Twitter data seems to be sparse for South Asian context. Thus, this study serves as a model for future investigations to utilize tweets for opinion mining in various other contexts.

Acknowledgements Not Applicable

Funding Not Applicable.

Availability of data and materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R (2011) Sentiment analysis of twitter data. In: Proceedings of the workshop on language in social media (LSM 2011), pp. 30–38

Alharbi ASM, de Doncker E (2019) Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. Cogn Syst Res 54:50–61. https://doi.org/10.1016/j.cogsys.2018.10.001

Arambepola N (2020) Analysing the tweets about distance learning during COVID-19 pandemic using sentiment analysis. In: Proc. of the International Conference on Advances in Computing and Technology (ICACT), pp. 169–171

Asare AO, Yap R, Truong N, Sarpong EO (2021) The pandemic semesters: examining public opinion regarding online learning amidst COVID-19. J Comput Assist Learn 37(6):1591–1605. https://doi.org/10.1111/JCAL.12574

Batubarra HS, Ambiyar A, Syahril S, Fadhilah F, Watrianthos R (2021) Sentiment analysis of face-to-face learning based on social media. JURNAL PENDIDIKAN TEKNOLOGI KEJURUAN 4(3):102–106. https://doi.org/10.24036/PTK.V4I3.22623

Budiharto W, Meiliana M (2018) Prediction and analysis of Indonesia presidential election from twitter using sentiment analysis. Journal of Big Data 5(1):1–10. https://doi.org/10.1186/S40537-018-0164-1/FIGURES/6

Cavanaugh CS, Barbour MK, Clark T (2009) Research and practice in K-12 online learning: a review of open access literature. Int Rev Res Open Distance Learn. https://doi.org/10.19173/IRRODL.V10I1.607

Degenhard J (2021, July 20) Twitter users in Asia 2020, by country Statista. https://www.statista.com/forecasts/1169294/twitter-users-in-asia-by-country

Dumford AD, Miller AL (2018) Online learning in higher education: exploring advantages and disadvantages for engagement. J Comput High Educ 30:452–465. https://doi.org/10.1007/s12528-018-9179-z

Garcia K, Berton L (2021) Topic detection and sentiment analysis in twitter content related to COVID-19 from Brazil and the USA. Appl Soft Comput 101:107057. https://doi.org/10.1016/J.ASOC.2020.107057

Hutto C, Gilbert E (2014) VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the International AAAI Conference on Web and Social Media, 8(1):216–225, https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Ittefaq M (2019) Digital diplomacy via social networks: A cross-national analysis of governmental usage of Facebook and twitter

Springer
for digital engagement. J Contemp Eastern Asia 18(1):49–69. https://doi.org/10.17477/jcea.2019.18.1.049

Jasti S, Mahalakshmi TS (2019) A Review on sentiment analysis of opinion mining. Adv Intell Syst Comput 768:603–612. https://doi.org/10.1007/978-981-13-0617-4_58

Laksono RA, Sungkono KR, Sarno R, Wahyuni CS (2019) Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes. In: Proceedings of 2019 international conference on information and communication technology and systems, ICTS 2019, 49–54. https://doi.org/10.1109/ICTS.2019.8850982

Makalesi A, Çiğdem Ekin C, Üniversitesi A, Mühendisliği B (2021) Examining public opinion regarding online learning during covid19 outbreak: sentiment analysis. European J Sci Technol 29(29):425–431. https://doi.org/10.31590/EJOSAT.1035267

Mowery D, Bryan C, Conway M (2017) Feature studies to inform the classification of depressive symptoms from twitter data for population health. https://arxiv.org/abs/1701.08229v1

Mupinga DM (2005) The clearing house: a journal of educational strategies Issues and Ideas 78(3):105–109. https://doi.org/10.3200/TCHS.78.3.105-109

Paltoglou G, Gobron S, Skowron M, TheWall M, Thalmann D (2010). Sentiment analysis of informal textual communication in cyberspace. In: Proceedings of ENGAGE 2010, 13–25. https://infoscience.epfl.ch/record/164428

Paltoglou G, Giachanou A (2014) Opinion retrieval: searching for opinions in social media. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8830:193–214. https://doi.org/10.1007/978-3-319-12511-4_10

Pishva, D., Nishantha, G. G. D., & Dang, H. A. (2010). A Survey on How Blackboard is Assisting Educational Institutions around the World and the Future Trends. 2010 The 12th International Conference on Advanced Communication Technology (ICACT), 2.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.