

Analysing the Tweets about Distance Learning during COVID-19 Pandemic using Sentiment Analysis

Nimasha Arambepola
Software Engineering Teaching Unit
Faculty of Science
University of Kelaniya
Kelaniya, Sri Lanka
nimashandk94@gmail.com

Abstract — With the COVID-19 pandemic situation, distance learning has become a mandatory alternative for the education system in almost every country. Even though this approach seems beneficial for both students and teachers, its effectiveness depends on various factors. For example, the availability of internet facilities and the economy of the users because of the cost of the required tools and components. Under these circumstances, several recent research has investigated the effectiveness of distance learning using different approaches. In this study, we analyze the effectiveness of distance learning using sentiment analysis. In our experimental analysis, we used a twitter data-set which consists of 202,645 tweets posted between 23rd July and 14th August in 2020. This is the period where many countries were locked down due to the pandemic. Our analysis consists of three stages. First, preprocessed the data to remove irrelevant tweets. Second, classified the tweets into three main classes namely positive, negative and neutral based on their polarities. There we observed that 54% sentiments were classified as positive, 30% as negative and 16% as neutral. As the final stage, further, we performed statistical data analysis to find whether there is a relationship between user experience on distance learning in different countries and the economic stability of those countries.

Keywords — *Distance Learning, Sentiment Analysis*

I. INTRODUCTION

With the ever-changing world of technology, people experience a greater impact on every aspect on the day to day lives. Align with that, the education sector has been changed drastically with the advancement of technology. The way the students are taught and interact with teachers has been changed a lot with the internet, online classes, mobile applications and websites. The term, e-learning or electronic learning is a type of distance learning that deliver learning and training through digital resources. Even though e-learning is nothing new in this century, people have to adapt to this education system with the effect of the COVID-19 pandemic. In this pandemic situation, people have to follow the COVID-19 guidelines and precautions such as maintain social distance. Therefore, students cannot attend schools or universities physically. But at the same time, they need to learn to live with this situation and continue education. As a result of that, most of the educational institutes in the world move to online platforms for conducting lectures. There are several online platforms and online course providers such as Coursera, edX and Udacity. They play a major role in this modern education system by offering online courses from reputed universities in the world [1].

Distance learning provides a considerable number of benefits for both students and lecturers such as they can manage their time effectively for studies without wasting time for travelling [2]. At the same time, when considering the economic feasibility in different countries in the world, practically some of the distance learning approaches are not

fruitful. For example, continues technical issues because of not having proper infrastructure and internet bandwidth as required, the difficulty of evaluation and assessment, the limited facility for collaborative learning and delivery of course content may not productive as productive in physical lecture hall [3]. Lately, the natural language processing(NLP), Sentiment Analysis(SA) has used to data mining related to sentiments or opinions for a specific topic [4]. Therefore, in this study, we use sentiment analysis to analyze the effectiveness of distance learning during the COVID-19 pandemic by extracting users' opinions from recently posted tweets.

We have presented our research in detail in the following sections. In the next section, we discuss the objectives of this study. Then, we present our analysis of prediction models and the performance of each model. After that, we present the results and discuss the findings of the study. Finally, we conclude this paper with the future directions of our research.

II. OBJETIVES

In recently, various methods have been used to predict and analyze the effectiveness of various distance learning methods such as e-learning. However, we hardly find existing research which has used sentiment analysis to classify users' opinion of e-learning into three classifiers namely positive, negative and neutral during the user experience of COVID-19 pandemic. It is more accurate if the analysis is done using the opinions of people who are engaging with distance learning. In this situation, collecting social media data is the most suitable method to collect relevant information. For that, we chose twitter dataset for this study. Moreover, it is not possible to get accurate insights from more than two hundred thousand user statements manually. As a result of that, we used sentiment analysis in NLP for extracting insights from tweets. Therefore, in this research, we analyze the effectiveness of distance learning during the COVID pandemic using the opinions extracted from Tweets [5].

The main objectives of this research are as follows:

- Studying different distance learning methods use worldwide.
- Classifying the users' opinion about distance learning as positive, negative and neutral during the COVID-19 pandemic in the world.
- Classifying the users' opinion about distance learning as positive, negative and neutral during the COVID-19 pandemic in country wise.

III. METHODOLOGY

This section presents the details of experimental design and analysis of the study. Sentiment analysis of micro-



blogging posts can proceed in many ways. In our study, we used the way of identifying the most suitable classifiers out of a large pool of mined data [6]. In the first stage, we retrieved data from the data source. In the next stage, preprocessed the data for handling missing values and removing irrelevant data. Then, the sentiments were extracted and classified them based on the polarity into three classes namely positive, negative and neutral. Finally, performed a statistical data analysis as explained in below. The flow of methodology is shown in Fig. 1.

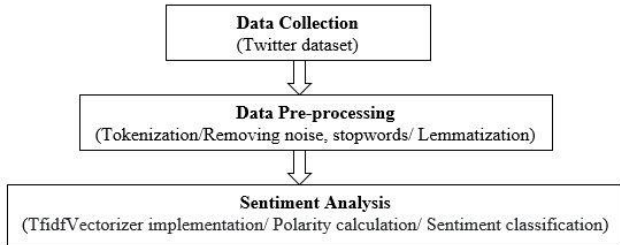


Fig. 1. Flow of the methodology

Twitter has become a popular social media that has a large and rapidly growing user base where users create status messages called tweets. Twitter users are useful for real-world up to date situations such as the current situation like COVID-19 pandemic because tweets express their opinion towards their experience of usage in products, services and new trends [6]. Therefore, we used a dataset which consists of 202,645 tweets posted between 23rd July and 14th August in 2020. This is the period when a considerable number of countries has been under the Lockdown situation. These tweets have been collected with Twitter API by using hashtags and keywords such as "#distancelearning", "#onlineschool", "#onlineteaching" and "#virtuallearning".

Due to there are irrelevant features included and the varying language structures included in tweets, it is important to preprocess the data to standardize certain tokens of tweets. Therefore, first, we removed three irrelevant features and keep the rest of the six (06) features for the analysis. Then missing values in the country column are replaced with 'unknown'. Secondly, performed the tokenization, removing stop-words, punctuation marks and hashtags. Then performed lemmatization to achieve the root words to improve the results.

- tokenization - tweets are split into individual words based on the space
- removing stopwords - remove common words such as a, is and has
- lemmatization - grouping together the inflected forms of a word so they can be analyzed as a single item

For the second stage, sentiments were extracted from tweets using TfidfVectorizer module from sklearn library. After assigning individual scores to all the words, polarity and subjectivity were calculated using TextBlob. Then classified the sentiments into three classes namely positive, negative and neutral based on their polarities as shown in Table 1. As the final stage, statistical data analysis was performed to find whether there is a relationship between user experience on distance learning in different countries and the economic stability of those countries.

Table 1. Polarity classification

Polarity value	Sentiment
> 0	Positive
0	Neutral
< 0	Negative

The results of above every stage are discussed in the following section.

IV. RESULTS AND DISCUSSION

This section presents the results of the study. In the first stage of the methodology, after removing irrelevant features, six features remain namely, "tweet content", "location(country)", "username", "retweetCount", "favourites" and "cretaedDateTime".

After classifying the sentiments based on their polarities, we obtained results as shown in Fig. 2. It shows that the majority of people have stated their opinion about distance learning positively. In percentage, 54\% of tweets are in positive, 16\% of tweets are in neutral and 30\% are negative. However, it is worth to mention that according to Fig. 3, the highest number of tweets from developed countries or from the countries where part of that country is developed. Moreover, when considering their polarities, it has been clearly shown that a considerable number of them are positive. It reveals that people with sufficient facilities get true benefit from the distance learning approach.

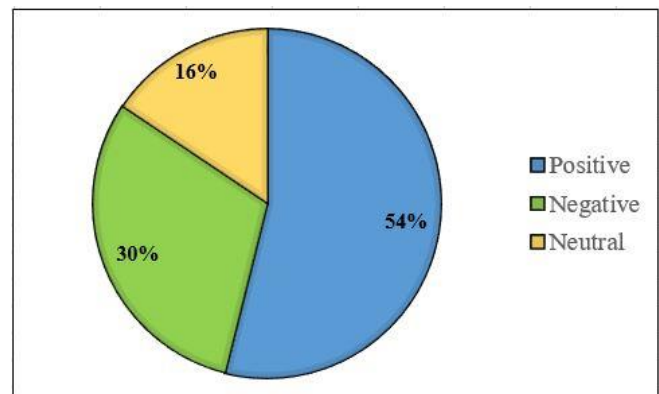


Fig. 2. Sentiment classification based on the polarity

According to the statistical data analysis performed, the highest number of tweets have been posted from the United States as shown in Fig. 3. Then we calculated the positive, negative and neutral polarity counts by country and we obtained the results as shown in Table 2. Moreover, sentiment classification and internet usage of few selected countries were comparatively summarized as shown in Table 2. Overall, users who are in developed countries where internet usage is comparatively high, have a high percentage of positive opinions about distance learning. For example, United States and Canada have above 92\% internet usage percentage. Nevertheless, the users who are in Yemen has a strongly negative opinion about distance learning and also they have less internet usage which is 26.72\% in percentage. According to these results, users from South Africa states the highest percentage of positive opinions about distance learning. Even though internet usage of South Africa is comparatively low, it has a highly developed economy and advanced infrastructure [7]. Moreover, other than the developed countries, Phillipine and India have a strong economy and their internet usage is also comparatively sufficient. Therefore, it reveals that there



is a relationship between user experience on distance learning in different countries and economic stability and internet usage of those countries.

effectiveness of distance learning for developing countries and also the Sri Lankan context separately. Further, this study can be performed using different datasets to obtain more accurate results.

V. CONCLUSION

In this research, we analyzed the effectiveness of distance learning during the COVID-19 pandemic using sentiment analysis of NLP. Here we used a recently released twitter dataset to extract the user opinion on distance learning. Then we classified them into three classes based on their polarity. According to the sentiment analysis and the statistical analysis performed, it can be concluded that if sufficient internet and other required facilities are available, distance learning is best suited as an approach for continuing the education system.

Table 2. Sentiment classification and internet usage comparison by country

	United States	Canada	Philippine	South Africa	India	Yemen
Positive	53%	64%	43%	84%	52%	2%
Negative	21%	17%	13%	2%	10%	84%
Neutral	26%	19%	44%	14%	38%	14%
Internet Usage	96.26%	92.7%	60.5%	56.2%	54.4%	26.7%

REFERENCES

- [1] J. Jaskolski, F. Sieberg, T. Tibroni, P. Cimiano, and R. Klinger, "Opinion mining in online reviews about distance education programs," 07 2016.
- [2] P. Fidalgo, J. Thormann, O. Kulyk, and J. Lencastre, "Students' perceptions on distance education: A multinational study," *International Journal of Educational Technology in Higher Education*, vol. 17, pp. 1–18, 2020.
- [3] A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami, "Elearning: Challenges and research opportunities using machine learning data analytics," *IEEE Access*, vol. 6, pp. 39 117–39 138, 2018.
- [4] S. Saad and B. Saberi, "Sentiment analysis or opinion mining: A review," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, p. 1660, 10 2017.
- [5] B. Hasdemir, "Tweets about distance learning," 2020 (Accessed on: October 09, 2020). [Online]. Available: <https://www.kaggle.com/barishasdemir/tweets-about-distance-learning>.
- [6] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera, "Opinion mining and sentiment analysis on a twitter data stream," in *International Conference on Advances in ICT for Emerging Regions (ICTer2012)*, 2012, pp. 182–188.
- [7] T. H. Foundation, *2020 Index of Economic Freedom*, 2020 (Accessed on: October 11, 2020). [Online]. Available: <https://www.heritage.org/index/country/southafrica?version=986>

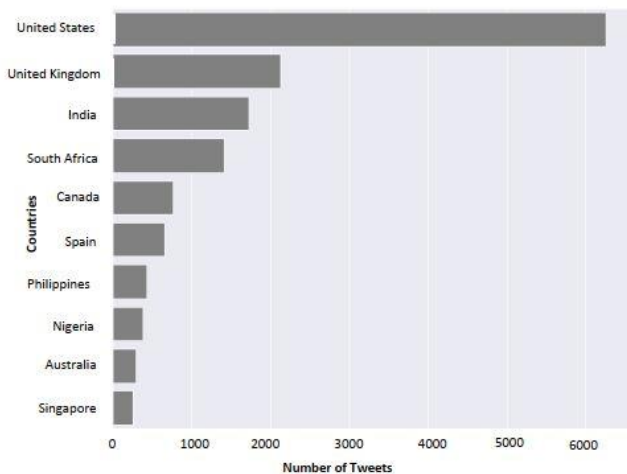


Fig. 3. Top Countries that posted tweets

However developed countries are dominant in this dataset and that has become a limitation in this study. For the future works, this research can be extended to analyze the

