Self-Attention-Based Bi-LSTM Model for Sentiment Analysis on Tweets about Distance Learning in Higher Education

https://doi.org/10.3991/ijet.v18i12.38071

Imane Lasri(✉), Anouar Riadsolh, Mourad Elbelkacemi
Mohammed V University in Rabat, Rabat, Morocco
imane_lasri@um5.ac.ma

Abstract—For limiting the COVID-19 spread, countries around the world have implemented prevention measures such as lockdowns, social distancing, and the closers of educational institutions. Therefore, most academic activities are shifted to distance learning. This study proposes a deep learning approach for analyzing people's sentiments (positive, negative, and neutral) from Twitter regarding distance learning in higher education. We collected and pre-processed 24642 English tweets about distance learning posted between July 20, 2022, and November 06, 2022. Then, a self-attention-based Bi-LSTM model with GloVe word embedding was used for sentiment classification. The proposed model performance was compared to LSTM (Long Short Term Memory), Bi-LSTM (Bidirectional-LSTM), and CNN-Bi-LSTM (Convolutional Neural Network-Bi-LSTM). Our proposed model obtains the best test accuracy of 95% on a stratified 90:10 split ratio. The results reveal generally neutral sentiments about distance learning for higher education, followed by positive sentiments, particularly in psychology and computer science, and negative sentiments in biology and chemistry. According to the obtained results, the proposed approach outperformed the state-of-art methods.

Keywords—COVID-19, distance learning, higher education, sentiment analysis, deep learning, Twitter

1 Introduction

The coronavirus (COVID-19) was detected in late-December 2019 in Wuhan city of China and caused respiratory illness, severe illness, and even death. It has been declared by the world health organization (WHO) as a global pandemic that presents a serious threat to public health and the economy worldwide. Most governments around the world have applied several prevention measures to fight the COVID-19 spread, such as limiting international travel, adopting lockdowns, and temporary school and university closures [1]. As a result, educational institutions were forced to shift to distance learning.

Distance learning also called online learning, e-learning, and distance education is a form of education in which students receive learning materials online via internet-enabled mobile devices like laptops, mobile phones, and tablets [2]. Its advantages include accessibility, flexibility [3], non-confinement of time and place, recording of lectures, and reducing financial costs [4]. However, various challenges have been reported, such as internet access problems, difficulty to adapt to distance learning technologies, and the lack of teacher-student interactions [5]. Consequently, the level of stress, fear, and anxiety has increased among students [6].

Since the beginning of the pandemic, some higher education institutions have continued to deliver only distance courses, other institutions have decided to use blended learning, and others have resumed in-person learning [7]. This transition to distance education has opened new research directions in the higher education field, where several survey-based studies [8,9,10] have aimed to analyze the impact of distance learning on students and teachers. However, survey-based studies may not provide a comprehensive view due to their limitations in sample size, time, and audience. To address this, sentiment analysis techniques can process large amounts of data, including text [11,12], images [13,14], and speech, to extract people's opinions and emotions towards distance learning. By leveraging artificial intelligence techniques, sentiment analysis provides a more nuanced and comprehensive understanding of the impact of distance learning on students and teachers. Social media platforms like Twitter, Facebook, and Instagram are primary channels where people express their opinions and sentiments on various topics, including online learning, services, instructors, courses, and teaching methods. Studies have utilized sentiment analysis to examine these opinions, such as So and Oh's study [15] which examined students' perceptions of online learning during the COVID-19 pandemic, and Kim et al.'s study [16] which analyzed Twitter posts related to massive open online courses (MOOCs). This approach allows for a larger and more diverse sample size than traditional survey-based studies [17]. Higher education institutions can use sentiment analysis techniques to analyze the vast amount of data generated on these platforms in real-time. Analyzing social media data can provide insights into areas for improvement in the online learning experience for students and teachers [17], thus enabling higher education institutions to make data-driven decisions.

Many existing research studies on sentiment analysis about distance learning in higher education rely on machine learning algorithms [18,19]. However, these algorithms have limitations in terms of accuracy and advanced processing of large datasets. Therefore, it is essential to explore new approaches for sentiment analysis that can provide a more advanced and accurate analysis. Deep learning-based algorithms are gaining popularity in various research fields due to their ability to process large datasets and provide accurate predictions. Hence, we propose for the first time, a study that uses a deep learning-based method to analyze people's sentiments (positive, negative, and neutral) from Twitter regarding distance learning in higher education. The proposed approach consists of seven main phases: data collection from Twitter, data pre-processing, data annotation, feature extraction using GloVe [20], stratified data splitting with a 90:10 ratio, sentiment classification using a self-attention-based Bi-LSTM model, and model performance evaluation and comparison to LSTM [21], Bi-LSTM [22], and CNN-Bi-LSTM models. The objective of this study is to understand the public sentiment expressed on Twitter about distance learning in higher education after the pandemic. And to identify the factors that explain variation in this public sentiment regarding six academic disciplines: mathematics, physics, biology, chemistry, computer science, and psychology.

The key contributions of our research study are as follows:

- Introducing a deep learning-based method for predicting the sentiments from Twitter about distance learning in higher education.
- Collecting data from Twitter in English concerning distance education in higher education, then pre-processing and labeling them.
- Predicting the sentiment of tweets using a self-attention-based Bi-LSTM model with GloVe word embedding.
- Comparing the proposed model performance to three others deep learning models.

2 Related work

Several studies have focused on analyzing public opinion from Twitter regarding distance learning during the COVID-19 pandemic using machine learning algorithms. Kharde and Sonawane [23] used machine learning algorithms to classify positive and negative tweets about distance learning. Senadhira et al. [24] performed a study to analyze tweets about online learning during the pandemic situation. The authors collected and pre-processed 8976 tweets (4486 positive and 4490 negative) to be passed into an artificial neural network and support vector machine [25] algorithms. According to the obtained results, ANN achieved the best performance. Das et al. [26] proposed a hybrid approach to predict the sentiment of tweets about online learning using the lexicon-analysis technique for sentiment analysis and four machine learning models for sentiment prediction. The multinomial naive Bayes model [27] outperformed the other models. Sahir et al. [28] carried out a study to analyze public tweets about online learning in Indonesia during October 2020 pandemic using the naive Bayes algorithm. Their results reveal that 'stress' and 'covid' were the most frequently occurring words in the tweets.

Three different studies were carried out using machine learning algorithms for analyzing Arabic tweets concerning distance learning in Saudi Arabia. First, Almalki [29] proposed a model that uses Twitter API for collecting 14000 tweets, a regex-based technique for data pre-processing, a logistic regression algorithm [30] for sentiment prediction, and Flask API for getting tweet sentiment. The logistic regression outperformed the others models. It was also used in the second study by Aljabri et al. [31] with the term frequency-inverse document frequency [32] and unigram. The proposed model gave the best results on a total of 20827 tweets. In the third study, Althagafi et al. [33] analyzed 8176 Arabic tweets about distance learning in Saudi Arabia using three machine learning algorithms. The random forest [34] with multi-class classification gave the best test accuracy.

On the other hand, deep learning algorithms have shown great success in sentiment classification of tweets about distance learning. Mujahid et al. [35] analyzed 17155 tweets about e-learning using lexicon-based approaches for labeling tweets, the bag of words and term frequency-inverse document frequency for features extraction, synthetic minority over-sampling technique (SMOTE) for data balancing, and nine machine learning models for sentiment classification. The term frequency-inverse document frequency combined with the support vector machine model shows the best result. Further, the authors used deep learning models like LSTM, CNN [36], CNN-LSTM, and Bi-LSTM for sentiment classification. According to their results, deep learning models had the highest performance compared to machine learning models. Waheeb et al. [37] used the extreme learning machine autoencoder (ELM-AE) with LSTM to analyze tweets concerning e-learning and to detect COVID-19 fake news. Their first dataset contained 60000 tweets labeled as (very positive, positive, neutral, negative, and very negative), and their second dataset contained 6000 articles from several websites about fake and true COVID-19 news. Their results show that ELM-AE with SMOTE gave the best performance.

In particular, several studies have been carried out to analyze public tweets concerning online learning for higher education using machine learning algorithms. Remali et al. [18] introduced a machine learning-based method for analyzing tweets about online learning in higher education. They used four different machine learning algorithms to classify the sentiment of 38602 tweets posted between 23 July and 14 August 2020. Their obtained results reveal that the support vector machine with an 80:20 split ratio and Vader had the best test accuracy. Baragash et al. [19] used a support vector machine model to classify 1201 tweets about online learning posted by Malaysian university students. Their proposed model achieved the best performance.

While machine learning algorithms have been widely used in various studies to analyze public opinion on distance learning during the COVID-19 pandemic from Twitter, some studies have demonstrated better performance using deep learning models compared to traditional machine learning models. These studies have been conducted in various countries, such as Indonesia and Saudi Arabia. Although several studies have focused on analyzing tweets about online learning in higher education using machine learning, no previous studies have utilized deep learning algorithms to analyze the sentiment of tweets about distance learning in higher education. Therefore, this study aims to fill this gap by investigating the sentiment of tweets about distance learning in higher education after the pandemic and evaluates the performance of deep learning algorithms in sentiment analysis. Specifically, it aims to identify the factors that contribute to variations in this sentiment across six academic disciplines: mathematics, physics, biology, chemistry, computer science, and psychology. The post-pandemic period was chosen to investigate how distance learning is perceived and utilized in an environment where the restrictions on in-person education have eased and distance learning has become more commonplace. This would provide valuable insights into the current perceptions and utilization of distance learning in a post-pandemic context.

3 Proposed methodology

The proposed architecture for sentiment analysis of tweets related to distance learning consists of seven stages, as depicted in Figure 1. The first stage is responsible for data collection from Twitter. The second stage is focused on data pre-processing, followed by data annotation using Text Blob [38] in the third stage. Feature extraction using GloVe is performed in the fourth stage. The fifth stage is responsible for data splitting into training and test sets in order to be processed by the self-attention-based Bi-LSTM model in the next stage. And the last stage is the model evaluation using performance metrics. A brief description of each of these stages is given in the following subsections.

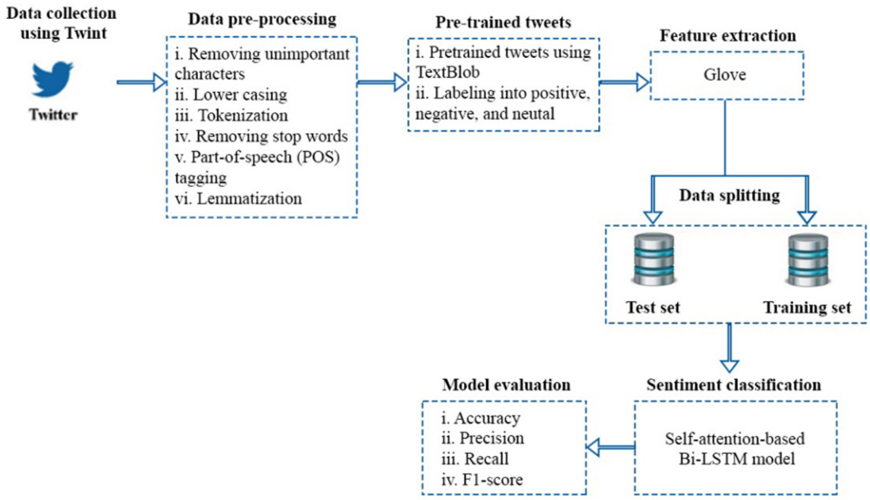

Fig. 1. Sentiment analysis architecture

Data collection. To collect tweets about distance learning, we used Twint [39], an open-source python library for Twitter scraping. Unlike the standard search application programming interface (API) that restricts access to tweets older than one week. Twint can extract almost all tweets without using Twitter's API and authentication. A total of 24642 historical tweets in English about distance learning written between July 20, 2022, and November 06, 2022, were collected based on keywords such as "distance learning", "remote education", "online learning", "virtual learning", "online education", "e-learning", and "distance education". This time span was selected because it falls after the COVID-19 pandemic, which had a significant impact on the education sector. By selecting this time frame, we anticipated that the pandemic-related restrictions on in-person education would have eased, and distance learning would have become more normalized. Therefore, the tweets collected during this period may provide valuable insights into how distance learning is being perceived and utilized in the post-pandemic context.

Data pre-processing. A big volume of unstructured data is generated daily on social media sites in forms of text, images, video, and audio. These data often are noisy and unstructured. Thus, it is necessary to clear the noise from the social media data with the following text pre-processing techniques in order to properly apply machine learning and deep learning algorithms.

1. *Removing emoticons, URLs, punctuation, numbers, usernames, special characters, and hashtags:* in this step we removed punctuation, numbers, usernames, special characters, and hashtags from the tweet because they don't have a role in our analysis.
2. *Lower casing:* uppercase letter is replaced by its corresponding lowercase letter.
3. *Tokenization:* sentences are divided into tokens with white space characters as delimiters.
4. *Removing the stop words:* we removed stop words from the tweet, which are words with no semantic value in natural language such as (by, with, is, the, etc.).

5. *Part-of-speech (POS) tagging:* gives the contextual information of a word (verbs, nouns, adverbs, conjunctions, etc.) that is essential in the lemmatization process.
6. *Lemmatization:* transforms a word to its root form using NLTK [40] lemmatization.

Table 1 illustrates an example of implementing the pre-processing steps on the following tweet: "We know that online learning is not only the solution for pandemic times, but also for the uncertainty like this. It is time to bring it back to the table. #onlinelearning #remotelearning #lms #learningplatform".

Pre-trained tweets using Text Blob. After the pre-processing step, every tweet is labeled as 'Positive', 'Negative', and 'Neutral' using the Text Blob [38] in order to be passed to the proposed deep learning model. The Text Blob is an open-source Python NLP tool that represents the emotional sentiment of a tweet based on the polarity score, which lies between −1 and 1. In our work, the sentiment is 'Positive' if the polarity score is higher than 0.2. Also, the sentiment is 'Neutral' if the polarity score is between 0 and 0.2. Then, the sentiment is 'Negative' if the polarity score is less than 0. The polarity score estimation is demonstrated in Eq. (1).

$$S_{t_i} = \left\{ \begin{array}{ll} \mathit{Positive} & P > 0.2 \\ \mathit{Neutral} & 0 \leq P \leq 0.2 \\ \mathit{Negative} & P < 0 \end{array} \right\} \tag{1}$$

The result of this step was: 6980 positive tweets, 14048 neutral tweets, and 3614 negative tweets. Figure 2 shows the number of tweets by sentiment in the period between July 20, 2022, and November 06, 2022.

Table 1. Example of implementing the pre-processing steps on a Tweet

Pre-Processing Step	Tweet
After removing emoticons, URLs, punctuation, numbers, usernames, special characters, and hashtags	We know that online learning is not only the solution for pandemic times, but also for the uncertainty like this. It is time to bring it back to the table
After lower casing	We know that online learning is not only the solution for pandemic times but also for the uncertainty like this it is time to bring it back to the table
After tokenization	['we', 'know', 'that', 'online', 'learning', 'is', 'not', 'only', 'the', 'solution', 'for', 'pandemic', 'times', 'but', 'also', 'for', 'the', 'uncertainty', 'like', 'this', 'it', 'is', 'time', 'to', 'bring', 'it', 'back', 'to', 'the', 'table']
After removing the stop words	['know', 'online', 'learning', 'solution', 'pandemic', 'times', 'also', 'uncertainty', 'like', 'time', 'bring', 'back', 'table']
After part-of-speech (POS) tagging	[('know', 'VERB'), ('online', 'NOUN'), ('learning', 'VERB'), ('solution', 'NOUN'), ('pandemic', 'ADJ'), ('times', 'NOUN'), ('also', 'ADV'), ('uncertainty', 'NOUN'), ('like', 'ADP'), ('time', 'NOUN'), ('bring', 'VERB'), ('back', 'ADV'), ('table', 'ADJ')]
After lemmatization	['know', 'online', 'learning', 'solution', 'pandemic', 'time', 'also', 'uncertainty', 'like', 'time', 'bring', 'back', 'table']

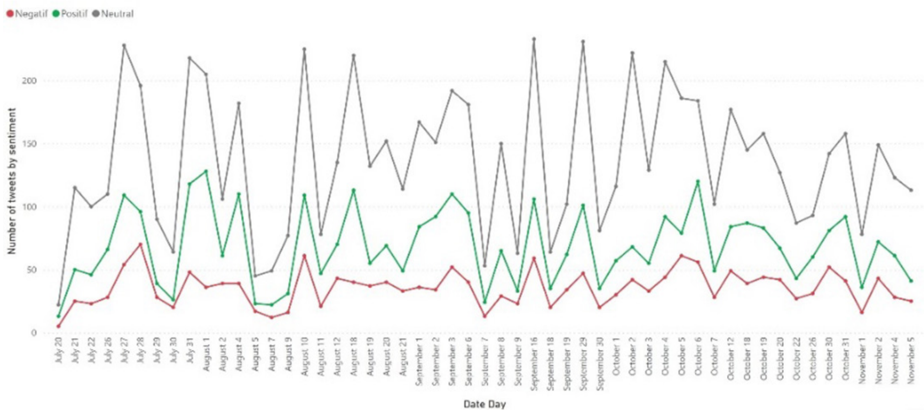

Fig. 2. Timeline showing sentiment of tweets about distance learning

Feature extraction. There are many feature extraction techniques like Word embedding, GloVe, and Term Frequency-Inverse Document Frequency (TF-IDF). In this study, the GloVe (Global Vectors) word embedding technique [20] is used in the proposed approach to obtain semantic features from the tweet. The GloVe is an unsupervised learning algorithm for distributed word representation of the text. It is successful in capturing semantic relations among words in the vector space. The GloVe objective is to minimize the objective function indicated in Eq. (2).

$$J = \sum_{i,j=1}^{|V|} f(X_{ij})(v_i^T v_j' + b_i + b_j' - \log(X_{ij}))^2 \tag{2}$$

Where v_i is the word representation of w_i, b_i is the bias of v_i, and $f(x)$ is the weighting function described in Eq. (3).

$$f(x) = \begin{cases} (x / x_{max})^{\alpha}, \; if \; x < x_{max}, \\ 1, \qquad otherwise \end{cases} \tag{3}$$

In this work, we used the Glove word vectors with 300 dimensions pre-trained on Common Crawl 42B (glove.42B.300d.zip) [41] to create a co-occurrence matrix X for training the GloVe model.

Data splitting. In our study, we applied stratified data splitting with a 90:10 ratio where 90% of data are used for training the model while 10% are used for testing. Then, we shuffled the data to ensure that the training set is more representative of the overall distribution of the data. The number of tweets in the training set is 22177 and in the test set is 2465.

Sentiment classification using self-attention-based Bi-LSTM model. The general architecture of the proposed model comprises the input layer, Bi-LSTM layer, self-attention mechanism [42], dropout layer, batch normalization, and output layer. The description of the model's components is outlined as follows.

The input layer receives data as a sequence of tokens. Each token is converted into a word vector. Tokens are padded using the zero-padding strategy or truncated based on the max length of the model. Then, they are fed to the Glove embedding layer in order to produce the word embedding vector.

In our work, we used the Bi-LSTM network that consist of forward and backward LSTMs to get features and learn their contexts. Forward and backward LSTMs are defined in Eq. (4) and Eq. (5), where m is the maximum feature-length.

$$\overrightarrow{h}_{f_{lstm}} = \overrightarrow{LSTM}(p_i),\, i \in [1, m] \tag{4}$$

$$\overleftarrow{h}_{b_{lstm}} = \overleftarrow{LSTM}(p_i),\, i \in [m, 1] \tag{5}$$

h_t, defined in Eq. (6), is the concatenating output of Eq. (4) and Eq. (5).

$$h_t = LSTM\,[\overrightarrow{h}_{f_{lstm}}, \overleftarrow{h}_{b_{lstm}}] \tag{6}$$

Since not all words in a tweet have the same meaning, we applied the self-attention mechanism to Bi-LSTM generated features by giving the most important words a higher weight to get their influence on the emotion of the tweet. The h_t word annotation produced by Bi-LSTM is fed to one layer perceptron MLP to obtain u_t as the hidden representation of h_t for *t*th tweet, as defined in Eq. (7), where W is a weight matrix of the MLP, and b is a bias vector of the MLP.

$$u_t = \tanh(w * h_t + b) \tag{7}$$

Then we measured the importance of words through the similarity between u_t and context vector v_s, which is randomly initialized. In addition, we got a normalized importance weight A_{t_i} through a Softmax function. A_{t_i}, as represented in Eq. (8), is the normalized weight of the *i*th word in the *t*th tweet, and exp(.) is the exponential function. The bigger A_{t_i} is, the more important the *i*th word is for emotional representation. Finally, the sentence vector s_t defined in Eq. (9) is the weighted sum of the word annotations.

$$A_{t_i} = \frac{\exp(u_t * v_s)}{\sum_{i=1}^{m} \exp(u_t * v_s)} \tag{8}$$

$$s_t = \sum_i \alpha_{t_i} h_{t_i} \tag{9}$$

The self-attention mechanism is followed by a dropout layer and batch normalization layer to avoid overfitting, then an output layer with three neurons, Softmax actvation function, and the sparse categorical cross-entropy as a loss function to perform the sentiment classification. Figure 3 shows the summary of our proposed self-attention-based Bi-LSTM model.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 20, 300)           2842500

bidirectional_1 (Bidirectio  (None, 20, 32)            40576
nal)

seq_self_attention_1 (SeqSe  (None, 20, 32)            2113
lfAttention)

dropout_1 (Dropout)          (None, 20, 32)            0

batch_normalization_1 (Batc  (None, 20, 32)            128
hNormalization)

flatten_1 (Flatten)          (None, 640)               0

dense_1 (Dense)              (None, 3)                 1923

=================================================================
Total params: 2,887,240
Trainable params: 44,676
Non-trainable params: 2,842,564
_________________________________________________________________
```

Fig. 3. Model summary of the proposed self-attention-based Bi-LSTM model

Model evaluation. We used the following performance metrics to evaluate the proposed model:

- Accuracy, defined in Eq. (10), indicates the percentage of correctly annotated tweets.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

- Precision, defined in Eq. (11), measures the percentage of correctly positive tweets relative to all positive tweets of each class.

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

- Recall, defined in Eq. (12), calculates the percentage of positive tweets in the dataset well predicted by our model.

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

- $F1$-score, defined in Eq. (13), measures the harmonic mean between recall and precision.

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{13}$$

Where TP means true positive, TN represents true negative, FP signifies false positive, and FN characterizes false negative.

4 Results

This section describes the setup of our experiments, following by the performance results of the deep learning classifiers, and the sentiment analysis result.

4.1 Experimental setup

We run the experiment on Windows 10 with 16 GB RAM, 64-bit, 3.20 GHz in CPU, Nvidia GTX 1070, and Intel Core i7. The proposed model is built using the Keras [43] and a TensorFlow backend [44]. In training, the dataset used in this study is divided into 90% for the training and 10% for the test. We used GloVe with 300 dimensions at the input layer for producing the word embedding vector. The size of the hidden layer in the Bi-LSTM neural network is 32 with ReLU activation function. The Bi-LSTM output is transmitted to the self-attention mechanism to get attention weight. Then, to avoid overfitting a dropout of 0.2 is applied, followed by batch normalization. The output layer has three neurons with the sparse categorical cross-entropy and the activation function Softmax. The batch size is 16, the model optimizer is Adam [45] with a learning rate of 0.01, and the number of epochs is 100. We used the RandomizedSearchCV method from the Sklearn library [46] to get the best hyperparameters values. Table 2 describes the hyperparameters value used in our approach.

Table 2. Hyperparameters of the self-attention-based Bi-LSTM model

Parameter	Value
Embedding size	300
Word embeddings	GloVe
Bi-LSTM hidden units	32
Bi-LSTM layers	1
Attention	Self-attention mechanism
Dropout	0.2
Bach normalization	Yes
Batch size	16
Optimizer	Adam
Learning rate	0.01
Number of epochs	100
Loss function	Sparse categorical cross-entropy
Classifer	Softmax
Activation function	ReLU

4.2 Experimental results

Performance results of deep learning models. We trained the self-attention-based Bi-LSTM model on 22177 tweets and tested on 2465 tweets. This resulted in an overall test accuracy of 95%. The proposed model performance is shown by plotting the confusion matrix validated with a stratified 90:10% split validation scheme, which is used to determine the relationship between true and predicted classes as illustrated in Figure 4.

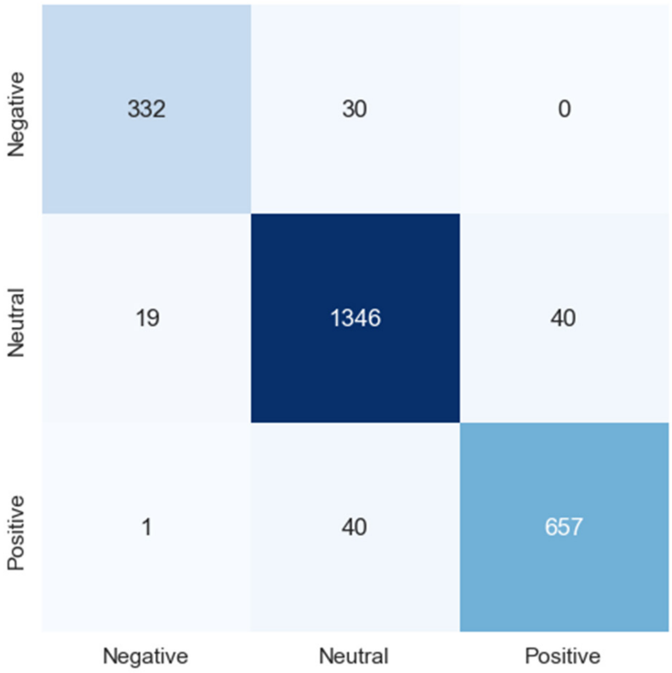

Fig. 4. Confusion matrix of the self-attention-based Bi-LSTM model

Accuracy, precision, F1-score, and recall are employed in the study to compare our model performance with LSTM, CNN-Bi-LSTM, and Bi-LSTM, models, as depicted in Figure 5 and Table 3.

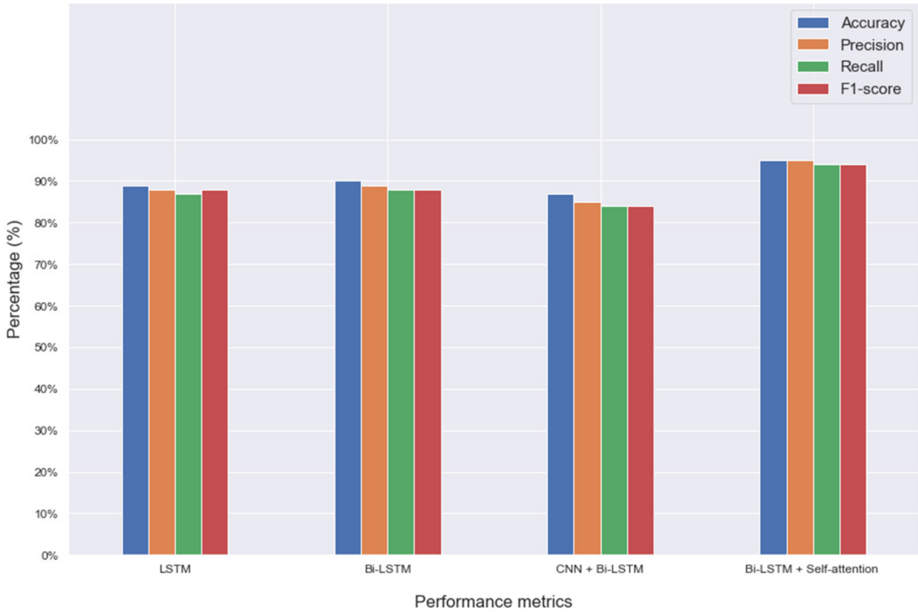

Fig. 5. Performance metrics of the proposed model with different deep learning models

Table 3. Performance comparison of the proposed model with different deep learning models

Model	Accuracy	Precision	Recall	F1-Score
LSTM	89.00%	88.00%	87.00%	88.00%
Bi-LSTM	90.00%	89.00%	88.00%	88.00%
CNN + Bi-LSTM	86.81%	85.00%	84.00%	84.00%
Bi-LSTM + Self-attention	95.00%	95.00%	94.00%	94.00%

The architectures of those three deep learning models are shown in Figures 6, 7 and 8. In comparison, our proposed self-attention-Bi-LSTM model outperformed the other models in terms of all performance metrics with an overall test accuracy of 95%. Meanwhile, the Bi-LSTM model acquired a test accuracy of 90%, followed by LSTM model with an accuracy of 89%. However, the CNN-Bi-LSTM model achieved the lowest test accuracy of 86.81%. It enables us to conclude that our proposed model attains the best performance among all the other models.

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_1 (Embedding)     (None, 20, 300)           2842500

 bidirectional (Bidirectiona  (None, 20, 32)           40576
 l)

 dropout_1 (Dropout)         (None, 20, 32)            0

 batch_normalization_1 (Batc  (None, 20, 32)           128
 hNormalization)

 flatten_1 (Flatten)         (None, 640)               0

 dense_1 (Dense)             (None, 3)                 1923

=================================================================
Total params: 2,885,127
Trainable params: 42,563
Non-trainable params: 2,842,564
_________________________________________________________________
```

Fig. 6. Summary of the LSTM model

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 20, 300)           2842500

 lstm (LSTM)                 (None, 20, 32)            42624

 dropout (Dropout)           (None, 20, 32)            0

 batch_normalization (BatchN  (None, 20, 32)           128
 ormalization)

 flatten (Flatten)           (None, 640)               0

 dense (Dense)               (None, 3)                 1923

=================================================================
Total params: 2,887,175
Trainable params: 44,611
Non-trainable params: 2,842,564
_________________________________________________________________
```

Fig. 7. Summary of the Bi-LSTM model

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_3 (Embedding)     (None, 20, 300)           2842500

 conv1d_1 (Conv1D)           (None, 20, 32)            19232

 max_pooling1d_1 (MaxPooling  (None, 10, 32)           0
 1D)

 bidirectional_2 (Bidirectio  (None, 10, 32)           6272
 nal)

 dropout_3 (Dropout)         (None, 10, 32)            0

 batch_normalization_3 (Batc  (None, 10, 32)           128
 hNormalization)

 flatten_3 (Flatten)         (None, 320)               0

 dense_3 (Dense)             (None, 3)                 963

=================================================================
Total params: 2,869,095
Trainable params: 26,531
Non-trainable params: 2,842,564
_________________________________________________________________
```

Fig. 8. Summary of the CNN-Bi-LSTM model

Sentiment analysis result. A word cloud of the most frequent words in the collected dataset is illustrated in Figure 9. The font sizes of the words represent their occurrence frequency. The most frequent words depicted in the figure are 'learning', 'education', 'remote', 'course', 'virtual', 'free', 'digital', and 'school'.

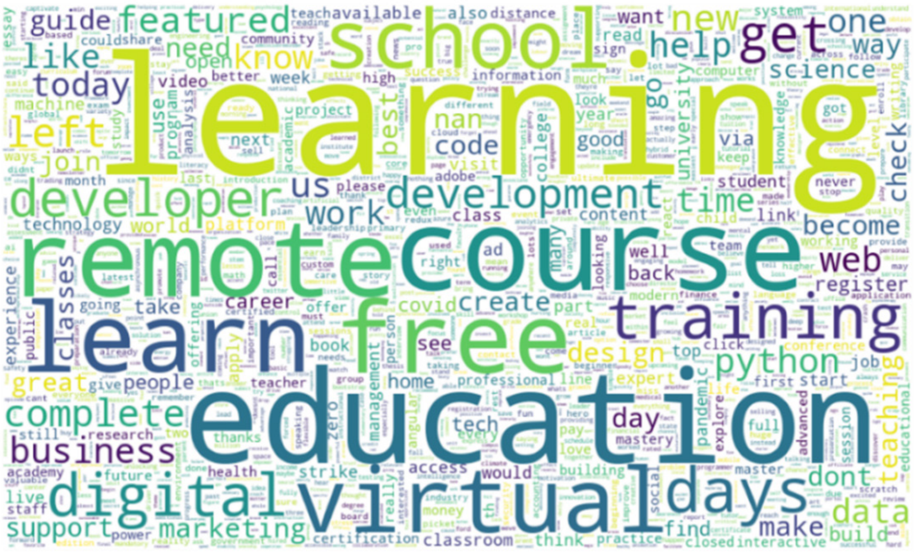

Fig. 9. Word cloud of frequent words in the dataset

The segmentation of tweets related to distance learning according to their predicted sentiment in the test data set is represented in Figure 10. It is seen that 57% of tweets were neutral, 28.3% were positive, and 14.7% were negative. Further, we classified the tweets in the test data set according to their corresponding educational stage (higher education, intermediate and high school, primary school, and kindergarten). And if the educational stage was not mentioned in the tweet, the tweet is classified as unspecified. Table 4 shows the number of tweets concerning each educational stage with the number of their predicted sentiment (positive, negative, and neutral) for the test dataset.

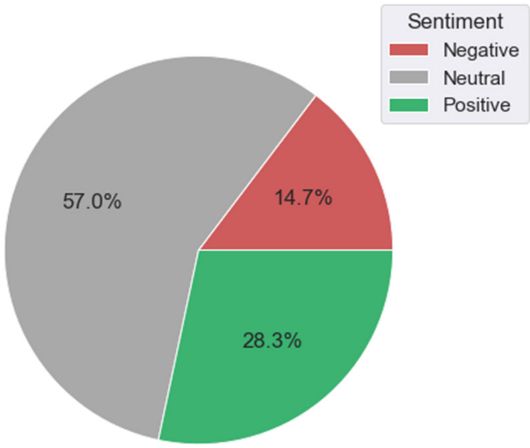

Fig. 10. Distribution of predicted sentiments of tweets related to distance learning

Table 4. Number of tweets in each educational stage by sentiment for the test dataset

Sentiment	Higher Education	Intermediate and High School	Primary School and Kindergarten	Unspecified
Positive	108	10	13	569
Neutral	192	13	21	1180
Negative	26	7	6	320
Total number of tweets	326	30	40	2069

Figure 11 shows a horizontal bar chart that represents the comparison between the educational stages by the number of tweets and the predicted sentiments. It is clearly visible that the higher education stage had the highest number of tweets and the highest percentage of positive tweets about distance learning, followed by primary school and kindergarten. However, the intermediate and high schools have the lowest percentage of positive tweets. In terms of neutral tweets, the percentage was higher in all the educational stages.

An example of tweets related to higher education with their predicted sentiment in the test dataset is shown in Table 5.

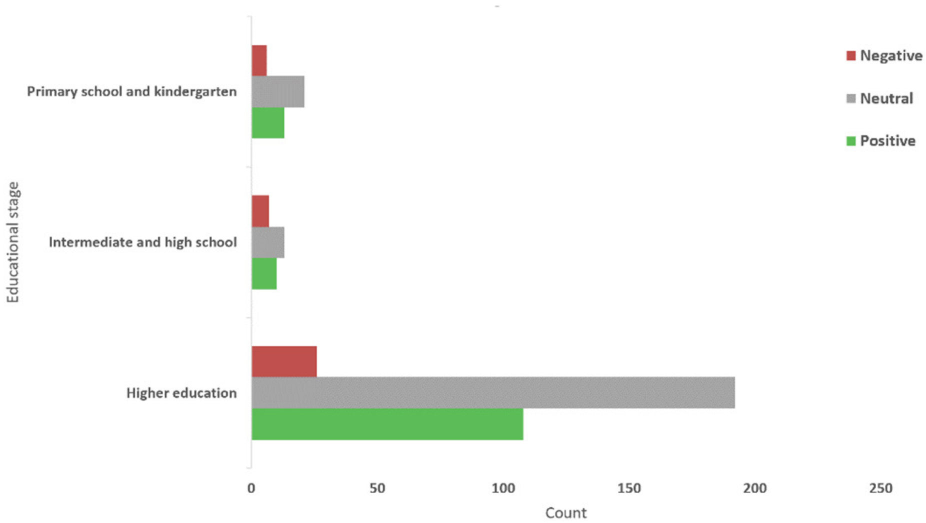

Fig. 11. Predicted sentiments regarding distance learning for different educational stages

Table 5. Example of tweets related to higher education with their predicted sentiment in the test dataset

Tweet	Predicted Sentiment
taught many people turn learning flexibility learning virtually anywhere team pursue without commute physical campus	Positive
remote student sometimes feel lonely however experience full community building support faculty teaching staff	Negative
faculty last chance apply virtual exchange training spring virtual exchange provide global collaboration apply	Neutral

We constructed a comparison between the disciplines of the higher education stage. Figure 12 shows a bar chart demonstrating the percentage of positive, negative, and neutral tweets about distance learning in six academic disciplines: mathematics, physics, biology, chemistry, computer science, and psychology. We can see that psychology had the highest percentage of positive tweets 41.17%, followed by 34.2% in computer science. Meanwhile, mathematics had 23.5% of positive tweets. However, physics, chemistry, and biology have the lowest percentage of positive tweets 20.8%, 20%, and 19%, respectively. On the other hand, the percentage of negative tweets was very high in biology, chemistry, and physics with 28%, 26%, and 24.6%, respectively. In addition, all six academic disciplines have a high percentage of neutral tweets, which varies between 53% and 67%.

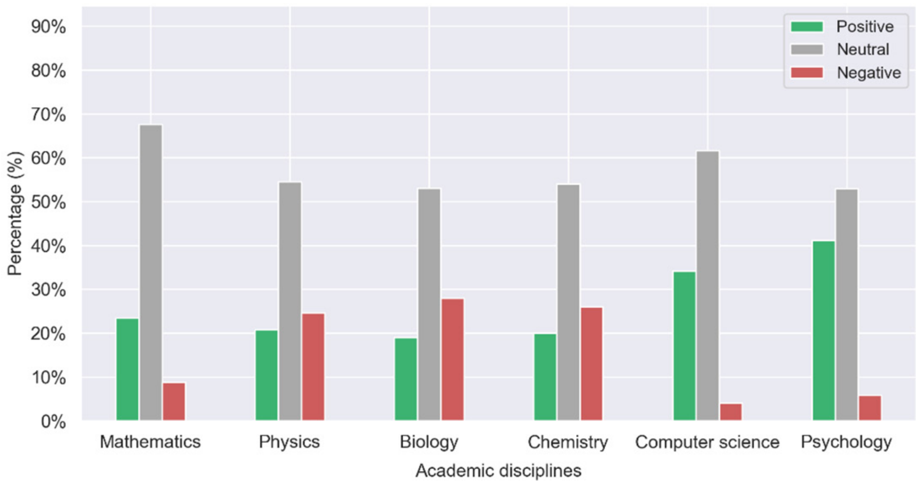

Fig. 12. Distribution of predicted sentiment from tweets about distance learning by six academic disciplines

As Figure 13 illustrates, the most frequently occurring bigrams related to distance learning in higher education are web development, understanding speak, remote assistance, effective communication, and easy access. We can see that people were talking positively about web development as an application of computer science and about effective communication as an essential skill a psychology student needs to develop during their undergraduate studies. These explain why the percentage of tweets in psychology and computer science disciplines is very high.

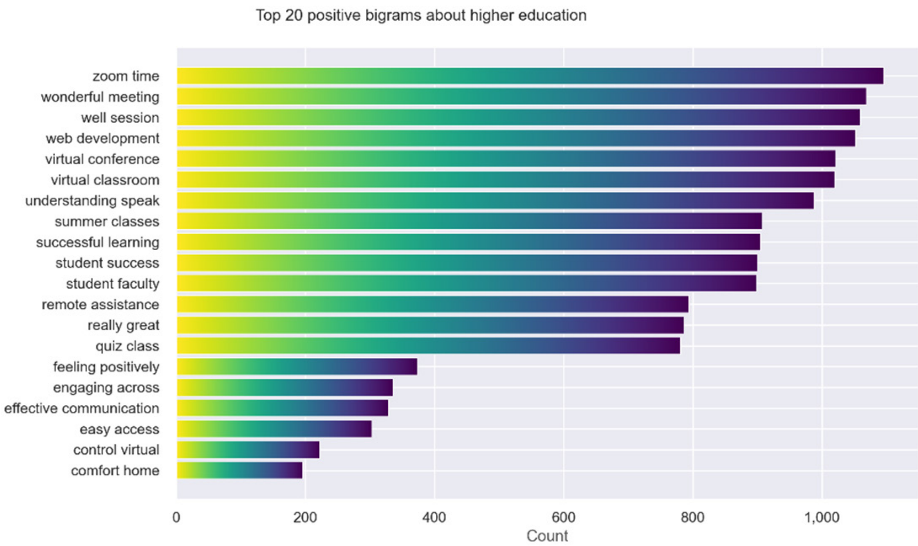

Fig. 13. Top 20 frequent bigrams from the positive tweets about distance learning in higher education

In contrast, the top 20 frequent bigrams from negative tweets about distance learning in higher education, as shown in Figure 14, are money funds, method study, physics essay, anatomy accounting, following biology, due biology, and everyone go. Therefore, it is revealed that the main reasons for the negative sentiment in these tweets were related to funds, lack of engagement, and the need for in-person laboratories for biology, chemistry, anatomy, and physics courses where students can see concepts in action. These underline why the percentage of negative tweets is very high in biology, chemistry, and physics.

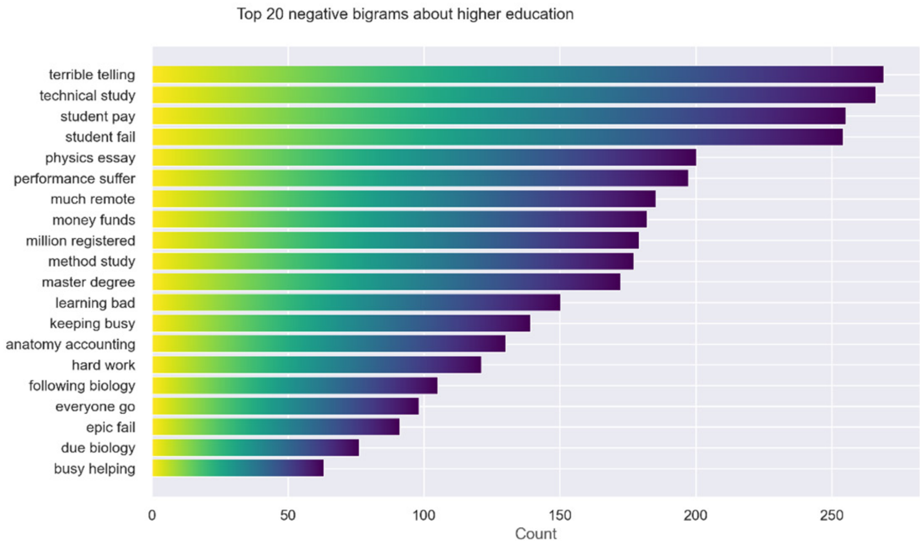

Fig. 14. Top 20 frequent bigrams from the negative tweets about distance learning in higher education

5 Discussion

Our study on public opinions towards distance learning on Twitter after the pandemic found that 59% of people expressed neutral sentiments, 33.12% of people felt positively about it, and 7.88% of people expressed negative sentiments. The highest positive sentiment was observed in psychology (41.17%) and computer science (34.2%). In contrast, the highest negative sentiment was observed in biology (28%) and chemistry (24.6%). Our findings are different from previous studies that mostly investigated negative attitudes towards distance learning during the pandemic [28,47,48, 49,50].

The top 20 frequent bigrams from negative tweets about distance learning in higher education were related to funds, lack of engagement, and the need for in-person laboratories for biology, chemistry, anatomy, and physics courses where students can see concepts in action. This is in line with previous studies that have shown concerns about the quality of distance learning, especially in practical fields such as science and engineering [51,52]. Additionally, the need for engagement and interaction in distance learning has been emphasized in the literature [53,54]. On the other hand, the most frequently occurring bigrams related to distance learning in higher education that were positively perceived by Twitter users were web development, understanding speak, remote assistance, effective communication, and easy access. These bigrams show that people were talking positively about the application of computer science in web development and the essential skill of effective communication in psychology. This is consistent with the literature that highlights the potential benefits of distance learning, such as flexibility, accessibility, and the use of technology [55,56].

Our study's results offer a more balanced view of public opinions towards distance learning post-pandemic, highlighting the effectiveness of distance learning in courses that do not require practical work. Furthermore, our study fills a gap in the literature by examining public opinions towards distance learning after the pandemic, which has not been studied before. These findings have significant implications for stake-holders in higher education and provide valuable insights for improving distance learning experiences for students in different disciplines.

6 Conclusion and future work

Our research proposes a deep learning-based approach for analyzing people's sentiments from Twitter concerning distance learning for higher education. We used Twint to collect 24642 tweets about distance learning posted between July 20, 2022, and November 06, 2022. The collected tweets were pre-processed and labeled positive, negative, and neutral to be passed to our proposed self-attention-based Bi-LSTM model with Glove word embedding for feature extraction. Then, the proposed model showed the best test accuracy of 95% with Adam optimizer compared to LSTM, Bi-LSTM, and CNN-Bi-LSTM models. The results showed generally neutral sentiments about distance learning for higher education, followed by positive sentiments, particularly in psychology and computer science, and negative sentiments in biology and chemistry. Further, an analysis was applied to identify the main reasons behind positive and negative tweets about distance learning for six academics disciplines. We can see that the blended learning of some face-to-face higher education courses, such as practical work, and others at a distance will make it possible to reduce the problem of massification and the availability of rooms and amphitheaters (mainly those coinciding with exam periods). This new pedagogical form offers more flexibility to educators and students. The obtained results confirm that our proposed approach can be used by decision-makers for producing reliable insights and making decisions. The current study has two limitations. First, we only studied posts from Twitter. It may be worthwhile to analyze sentiment through other social media platforms such as Facebook. Second, we have extracted only English tweets during the data collection stage, but users use more languages to express their sentiments, especially in a multi-cultural environment. Hence, studying sentiments in different languages may help analyze the demographic composition. In the future, we intend to plan to analyze Arabic tweets and posts from other social media platforms.

7 References

[1] Lasri, I., RiadSolh, A., El Belkacemi, M. (2021). Toward an Effective Analysis of COVID-19 Moroccan Business Survey Data using Machine Learning Techniques. In 13th International Conference on Machine Learning and Computing (ICMLC 2021), Shenzhen, China, 50–58. https://doi.org/10.1145/3457682.3457690

[2] Oliveira, M., Penedo, A., Pereira, V. (2018). Distance education: Advantages and disadvantages of the point of view of education and society. Dialogia, pp. 139–152. https://www.britannica.com/topic/distance-learning [Accessed: March 02, 2022]. https://doi.org/10.5585/dialogia.N29.7661

[3] Pakdaman, M., Moghadam, M., Dehghan, H. R, Dehghani, A., Namayandeh, S. M. (2019). Evaluation of the cost-effectiveness of virtual and traditional education models in higher education: A systematic review. Health Technology Assessment in Action, 3(1).

[4] Barrot, J. S., Llenares, I. I., Del Rosario, L. S. (2021). Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines. Education and Information Technologies, 26(6), pp. 7321–7338. https://doi.org/10.1007/s10639-021-10589-x

[5] Vagos, P., Carvalhais, L. (2022). Online versus classroom teaching: Impact on teacher and student relationship quality and quality of life. Front. Psychol., 13, p. 828774. https://doi.org/10.3389/fpsyg.2022.828774

[6] Pelikan, E. R. et al. (2021). Distance learning in higher education during COVID-19: The role of basic psychological needs and intrinsic motivation for persistence and procrastination–a multi-country study. PLoS One, 16(10), p. e0257346. https://doi.org/10.1371/journal.pone.0257346

[7] https://higheredpartners.co.uk/why-online-learning-in-higher-education-is-here-to-stay-a-trends-assessment/ [Accessed: March 15, 2022].

[8] Jebbour, M. (2022). The unexpected transition to distance learning at Moroccan universities amid COVID-19: A qualitative study on faculty experience. Social Sciences & Humanities Open, 5(1), p. 100253. https://doi.org/10.1016/j.ssaho.2022.100253

[9] Niu, M. (2022). Classification of learning sentiments of college students based on topic discussion texts of online learning platforms. International Journal of Emerging Technologies in Learning (iJET), 17(24), pp. 42–56. https://doi.org/10.3991/ijet.v17i24.35951

[10] Sun, J., Zhang, X. (2021). Exploring Chinese college students' emotions as they engage in online learning during a pandemic. Asia Pacific Journal of Education, pp. 1–12. https://doi.org/10.1080/02188791.2021.1965541

[11] Adwan, O. Y., Al-Tawil, M., Huneiti, A., Shahin, R., Abu Zayed, A., Al-Dibsi, R. (2020). Twitter sentiment analysis approaches: A survey. International Journal of Emerging Technologies in Learning (iJET), 15(15), pp. 79–93. https://doi.org/10.3991/ijet.v15i15.14467

[12] Lasri, I., Riadsolh, A., Elbelkacemi, M. (2023). Real-time Twitter sentiment analysis for Moroccan Universities using machine learning and Big Data technologies. International Journal of Emerging Technologies in Learning (iJET), 18(05), pp. 42–61. https://doi.org/10.3991/ijet.v18i05.35959

[13] Lasri, I., Riadsolh, A., Elbelkacemi, M. (2022). Facial emotion recognition of deaf and hard-of-hearing students for engagement detection using deep learning. Education and Information Technologies. https://doi.org/10.1007/s10639-022-11370-4

[14] Liang, Y. (2019). Intelligent emotion evaluation method of classroom teaching based on expression recognition. International Journal of Emerging Technologies in Learning (iJET), 14(04), pp. 127–141. https://doi.org/10.3991/ijet.v14i04.10130

[15] So, H. J., Oh, E. (2020). The effects of COVID-19 on online learning based on sentiment analysis. The Journal of Educational Technology Systems, 49(1), pp. 5–22. https://doi.org/10.1177/0047239520934018
[16] Kim, J., Kim, Y. J., Lee, H. W., Cho, D. (2015). Analyzing learners' tweets and perceptions in a massive open online course. Computers in Human Behavior, 51, pp. 1293–1302.
[17] Kessler, G. C., DeCarlo, J. (2017). Sentiment analysis in education: Opportunities, challenges, and future directions. Educational Researcher, 46(9), pp. 492–499.
[18] Remali, N. A. S., Shamsuddin, M. R., Abdul-Rahman, S. (2022). Sentiment Analysis on Online Learning for Higher Education During Covid-19. In the 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), pp. 142–147. https://doi.org/10.1109/AiDAS56890.2022.9918788
[19] Baragash, R., Aldowah, H. (2021). Sentiment analysis in higher education: a systematic mapping review. J. Phys. Conf. Ser., 1860(1), p. 012002. https://doi.org/10.1088/1742-6596/1860/1/012002
[20] Pennington, J., Socher, R., Manning, C. (2014). Glove: global vectors for word representation. In the Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, pp. 1532–1543. https://doi.org/10.3115/v1/D14-1162
[21] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Comput., 9(8), pp. 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
[22] Schuster, M., Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45(11), pp. 2673–2681. https://doi.org/10.1109/78.650093
[23] Kharde, V. A., Sonawane, S. S. (2016). Sentiment analysis of Twitter data: A survey of techniques. International Journal of Computer Applications, 139(11), pp. 5–15. https://doi.org/10.5120/ijca2016908625
[24] Senadhira, K. I., Rupasingha, R. A. H. M., Kumara, B. T. G. S. (2022). Sentiment Analysis on Twitter Data Related to Online Learning During the Covid-19 Pandemic. In International Research Conference on Smart Computing and Systems Engineering (SCSE), pp. 131–136. https://doi.org/10.1109/SCSE56529.2022.9905190
[25] Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), pp. 273–297. https://doi.org/10.1007/BF00994018
[26] Das, S. et al. (2022). Sentiment dynamics detection of online learning impact using hybrid approach. Special Education, 1(43), pp. 1225–1236.
[27] McCallum, A., Nigam, K. (1998). A comparison of event models for naive bayes text classification. In the AAAI-98 Workshop on Learning for Text Categorization.
[28] Sahir, S. H. et al. (2021). Online learning sentiment analysis during the covid-19 Indonesia pandemic using twitter data. IOP Conference Series: Materials Science and Engineering, 1156(1), p. 012011. https://doi.org/10.1088/1757-899X/1156/1/012011
[29] Almalki, J. (2022). A machine learning-based approach for sentiment analysis on distance learning from Arabic Tweets. PeerJ Comput. Sci., 8, p. e1047. https://doi.org/10.7717/peerj-cs.1047
[30] Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B (Methodological), 120(2), pp. 215–32. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x
[31] Aljabri, M. et al. (2021). Sentiment analysis of Arabic tweets regarding distance learning in Saudi Arabia during the COVID-19 Pandemic. Sensors, 21, pp. 5431. https://doi.org/10.3390/s21165431
[32] Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for idf. https://doi.org/10.1108/00220410410560582

[33] Althagafi, A., Althobaiti, G., Alhakami, H., Alsubait, T. (2021). Arabic Tweets sentiment analysis about online learning during COVID-19 in Saudi Arabia. International Journal of Advanced Computer Science and Applications (IJACSA), 12(3), pp. 620–625. https://doi.org/10.14569/IJACSA.2021.0120373
[34] Ho, T. K. (1995). Random decision forests. In 3rd international conference on document analysis and recognition, pp. 278–82.
[35] Mujahid, M. et al. (2021). Sentiment analysis and topic modelling on tweets about online education during COVID-19. Appl. Sci., 11(18), p. 8438. https://doi.org/10.3390/app11188438
[36] LeCun, Y. et al. (1989). Backpropagation applied to handwritten zip code recognition. Neural Computation, 1(4), pp.541–551. https://doi.org/10.1162/neco.1989.1.4.541
[37] Waheeb, S. A., Khan, N. A., Shang, X. (2022). Topic modelling and sentiment analysis of online education in the COVID-19 era using social networks based datasets. Electronics, 11(5), p. 715. https://doi.org/10.3390/electronics11050715
[38] Loria, S. (2018). Textblob Documentation. Release 0.15, 2.
[39] https://medium.com/analyticsvidhya/how-to-scrape-tweets-from-twitter-with-python-twint-83b4c70c5536 [Accessed: March 12, 2022].
[40] Bird, S., Klein, E., Loper, E. (2009). Natural language processing with Python: Analyzing text with the natural language toolkit. OReilly Media Inc.
[41] https://nlp.stanford.edu/projects/glove/ [Accessed: March 24, 2022].
[42] Vaswani, A. et al. (2017). Attention is all you need. arXiv, p. 11.
[43] Chollet, F. et al. (2015). Keras. https://github.com/fchollet/keras.
[44] Abadi, M. et al. (2016). Tensorflow: a system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283.
[45] Kingma, D. P., Ba, J. (2014). Adam: a method for stochastic optimization. arXiv.
[46] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, pp. 2825–2830.
[47] Sadigov, R., et al. (2023). Deep learning-based user experience evaluation in distance learning. Cluster Comput. https://doi.org/10.1007/s10586-022-03918-3
[48] Stracqualursi, L., Agati, P. (2022). Tweet topics and sentiments relating to distance learning among Italian Twitter users. Sci Rep, 12, p. 9163. https://doi.org/10.1038/s41598-022-12915-w
[49] Ali, M. M. (2021). Arabic sentiment analysis about online learning to mitigate Covid-19. Journal of Intelligent Systems, 30(1), pp. 524–540. https://doi.org/10.1515/jisys-2020-0115
[50] Qaqish, E., Aranki, A., Etaiwi, W. (2023). Sentiment analysis and emotion detection of post-COVID educational Tweets: Jordan case. Social Network Analysis and Mining, 13(1), pp. 1–11. https://doi.org/10.1007/s13278-023-01041-8
[51] Cavanagh, A. J., Chen, X., Bathgate, M., Frederick, J., Hanauer, D. I. (2016). Teaching scientific writing in an introductory lab course. Journal of Microbiology & Biology Education, 17(3), pp. 417–419.
[52] Graham, C. R., Woodfield, W., Harrison, J. B. (2018). A framework for institutional adoption and implementation of blended learning in higher education. Internet and Higher Education, 18, pp. 4–14. https://doi.org/10.1016/j.iheduc.2012.09.003
[53] Dixson, M. D. (2010). Creating effective student engagement in online courses: What do students find engaging? Journal of the Scholarship of Teaching and Learning, 10(2), pp. 1–13.

[54] Zawacki-Richter, O., Marín, V. I., Bond, M., Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? International Journal of Educational Technology in Higher Education, 16(1), pp. 1–24. https://doi.org/10.1186/s41239-019-0171-0

[55] Means, B., Toyama, Y., Murphy, R., Bakia, M., Jones, K. (2013). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. US Department of Education.

[56] Simonson, M., Smaldino, S., Albright, M., Zvacek, S. (2015). Teaching and learning at a distance: Foundations of distance education (6th ed.). Information Age Publishing.

8 Authors

Imane Lasri is a PhD student at the Laboratory of Conception and Systems (Electronics, Signals, and Informatics), Faculty of Sciences Rabat, University Mohammed V in Rabat, Morocco. She received a Master's degree in Big Data Engineering from the Faculty of Sciences Rabat. She received the awards of excellence of the major winners from the Mohammed V University in Rabat in 2019. Her current field of research is pattern recognition applied to higher education using machine learning and deep learning algorithms (email: imane_lasri@um5.ac.ma).

Anouar Riadsolh is a Professor at the Faculty of Sciences Rabat, University Mohammed V in Rabat, Morocco. He received his PhD in Computer Science from the Faculty of Sciences Rabat. He is a member of the Laboratory of Conception and Systems (Electronics, Signals, and Informatics). His current research interests are focused on data mining, big data, and machine learning (email: a.riadsolh@um5r.ac.ma).

Mourad Elbelkacemi is a Professor at the Faculty of Sciences Rabat, University Mohammed V in Rabat, Morocco. He received his PhD in Computer Science. He is a member of the Laboratory of Conception and Systems (Electronics, Signals, and Informatics). His main research interests are focused on electronics, education, and data mining (email: mourad_prof@yahoo.fr).

Article submitted 2023-01-14. Resubmitted 2023-03-27. Final acceptance 2023-04-11. Final version published as submitted by the authors.