Journal of Engineering and Computer Sciences
Vol. 16, No. 2, pp. 38-54 (Jul 2024)

Sentiment Analysis of people's opinions in E-Learning based on Support Vector Machine

Athar Albogami
Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia,
E: 421200373@qu.edu.sa

Athbah Alrugaibah
Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia,
E: 421200031@qu.edu.sa

Gareebah Alrasheedi
Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia,
E: 421200394@qu.edu.sa

Afef Selmi*
Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia,
E: a.selmi@qu.edu.sa*, ORCID: 0000-0001-9363-7702

[Received: 04 March 2024, Revised: 14 June 2025, Accepted: 16 Jun 2025, Published Sep 2025]

Abstract: The rapid development and popularity of social media have allowed people to express their ideas and views, and feelings through social networks such as Twitter and Meta. This research focused on sentiment analysis of social media content about electronic learning (E-learning) since public opinions can help organizations and individuals. In this area, the social media content is abundant and unregulated. Therefore, sentiment analysis has become mainly an area of research interest. This study explores the machine learning approach for sentiment analysis of Twitter content to analyze people's opinions about E-learning. Using the Twitter API, the number of tweets collected was 42368. Tweets that involved E-learning were selected by using a programmable code written in Python. Then, a Support Vector Machine classifier was trained on the pre-processed data for analysing the tweets. The result of this study showed that the classifier's accuracy applied to the dataset was 92%. Generally, the people's opinions were positive toward E-learning.

Keywords: Social Network, Sentiment Analysis, E-learning, Twitter, Machine Learning

1. INTRODUCTION

Nowadays, the Education process has shifted from physical (face-to-face) to online learning. E-learning is "the learning supported by digital electronic tools and media" [1]. The learners no longer need to gather physically. They just need a desktop, laptop, or smartphone and the internet. With E-learning, instructors and students can communicate efficiently and continue education outside the educational institutions. Therefore, the distance issue led people to communicate through virtual social media. Social media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content [2].

In general, social media analytics refers to gathering data from social media platforms and analysing it to help decision-makers address specific problems [3]. People worldwide are using different social media platforms to communicate, express their opinions, and share new information related to trending topics. Each social media platform has different functionality and uses [4]. The social media platforms include collaborative projects (e.g., Wikipedia), blogs and microblogs (e.g., Twitter), content communities (e.g., YouTube), social networking (e.g., Facebook and LinkedIn), media sharing (YouTube, Flickr, and Instagram), blogging (Tumblr and WordPress), virtual worlds (Second Life), commerce communities (SnapDeal and PepperFry), and social bookmarking (StumbleUpon) [5] [6]. Among all social media networks, Facebook and Twitter are the most widely used by businesses [7]. Social media aims to utilize big data to gather important information regarding public opinion. Social Media Analytics (SMA) is about doing some activities such as searching, collecting, filtering, summarizing, identifying, visualizing, analyzing, and generating insights from social media data [8, 9].

Social Media Analytics encompasses various techniques, including Sentiment Analysis. Sentiment Analysis deals with analysing users' opinions and sentiments expressed in a written text [10]. Sentiment orientation/polarity is the popular approach that assigns scores (+1, 0, −1) to all words: positive opinion (+1), neutral opinion (0), and negative opinion (−1) [11]. According to [3], sentiment analysis, which is also called opinion mining, refers to applying computational technologies such as natural language processing and computational linguistics to identify and extract information from vast amounts of user-generated content. Sentiment analysis uses two methods: a machine-learning method and a lexical-based method. Sentiment analysis mainly relies on machine learning techniques such as Support Vector Machine (SVM), Naive Bayes, Maximum Entropy, and Matrix Factorization to classify texts into positive or negative categories [6].

Many research studies have utilized questionnaires to collect data and assess their targeted students' opinions. To the best of our knowledge, no other studies have used sentiment analysis on social media data in E-learning, such as Twitter, that were found in the literature. People have used social platforms to express their opinions freely and with no unnecessary compliments. Therefore, this research study focuses on utilizing the Twitter platform dataset to obtain the public's insights into E-learning. The purpose of this research is to propose a sentiment analysis model for people's opinions about E-learning through social media.

1.1. Research contributions of this work

The contributions of this research study are as follows:

- We reviewed several proposed Machine Learning based approaches for Sentiment Analysis in the Learning process.
- We collected and generated a clean version of the dataset of users' opinions on Twitter using the API tool and extracted the dataset features.
- We proposed a sentiment Analysis model for people's opinions about E-learning based on an SVM classifier.
- We evaluated the accuracy of the proposed model.

The remaining structure of the paper is as follows: Section 2 discusses the major related studies dealing with the research problem. Section 3 illustrates the followed methodology. Section 4 demonstrates the experimental results. Section 5 discusses the results. Finally, section 6 concludes this study.

2. RELATED WORK

Numerous studies related to sentiment Analysis on Twitter data in E-Learning during COVID-19 were collected. All those studies have used the questionnaire and survey as a data collection process for analysing and evaluating the tweets. In [11], the authors have used the Word2vec technique and three Machine Learning classifiers for sentiment analysis in the learning process during COVID-19 of Egyptian Students at the Arab Academy for Science and Technology and Maritime Transport University. The three used classifiers are Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT). In this study, the Authors have classified the E-learning process's problems; the largest number of sentiments is communication between students and instructors. The other problems were online lecture voice latency, online examination limited time, and Internet connection.

Lau and Sim [12] have evaluated the students' anxiety in one of the Top 20 universities in China. A total of 3611 students participated in the questionnaire. The self-Rating Anxiety Scale was used to analyse the students' anxiety. In this study, the authors built a Chinese sentiment analysis model based on Bi-directional Long Short-Term Memory. They have used K-Means Clustering to categorize all samples into three types and analysed each type's scoring characteristics. The results have shown that 62.54% of the students expressed a positive attitude.

On the other hand, 37.46% of the students had some concerns about online learning. One of the concerns that some students expressed was that the original study plan could not be completed on time. The study [13] investigated students' online learning experiences and influencing factors during COVID-19. The authors have surveyed quantitative and qualitative questions to over 400 students in a private university in Malaysia. For sentiment analysis, the authors used QDAP2 (Quantitative Discourse Analysis Package) in R to analyse the study students' feedback. The authors got positive feedback for the rapid changes in the education process due to the pandemic. Although studying from home is flexible and convenient, most students prefer face-to-face interaction and communication among themselves.

Many medical schools also quickly shifted the entire pre-clinical curriculum to online formats [14]. On March 17, 2020, the Association of American Medical Colleges (AAMC) released guidelines to suspend any clinical activities involving patient contact for 2 weeks [15]. Later on, the suspension was extended to April 14, 2020, with new guidelines. Some schools prevented any patient interaction, while others recruited the students for hospitals to serve as frontline clinicians [16]. In [16], the authors obtained 330 participants after distributing a questionnaire to 60 medical students. The method used for item-by-item analysis is Likert-type items. The participating students had positive perceptions of the online learning environment with moderate disruptions to concentration and sleep.

In addition, there was another study [17] about identifying the barriers to online learning during COVID-19. The pandemic forced medical schools in the Philippines to stop face-to-face learning activities and unexpectedly shift to online classes. In this study, a National Survey was conducted for Medical Students in the Philippines. A total of 3670 participants were collected, and descriptive statistics were calculated. The findings showed that 41% of the students considered themselves physically and mentally capable of engaging in online learning. The barriers were divided into technological, individual, domestic, institutional, and community barriers. Most of the students encountered some difficulty adjusting learning styles, performing responsibilities at home, and communicating effectively with their educators. For analysis, authors have used Kruskal-Wallis and Dunn's multiple nonparametric pairwise tests.

The study [18] aimed to suggest an analytical framework for retail pharmacy organizations to use social media, and show the most discussed topics by consumers and indicate the main areas for improvement based on the negative comments received. The researchers have conducted a deep analysis of the retail pharmacy organizations' Twitter platforms in the United Kingdom: Boots, Lloyds, and Superdrug. The findings showed that marketing, customer service, and product issues are the main improvement areas for retail pharmacies. Boots, in particular, received a better sentiment performance than Lloyds and Superdrug.

In [19], the study aimed to identify customer knowledge on social media through data analytics, encompassing customer knowledge, customer knowledge about customers, and customer knowledge from customers, and to encourage social media-based customer knowledge management. It also explored how to turn large-scale social media data into valuable consumer information. This study has conducted a case study to analyse people's online conversations on Twitter regarding laptop brands and manufacturers. The researchers applied statistical analysis, text mining, and sentiment analysis techniques to the relevant tweets collected using Twitter search APIs, analyzing the social media dataset and visualizing relevant patterns to determine customer knowledge. The study results have shown that RAM and pricing are the most debated subjects for Internet users; Dell is the most commonly discussed brand on social media; HP is leading the market with the most favourable perceptions; Lenovo's battery received a higher percentage of encouraging feedback, and Asus's laptop received a higher rate of positive pricing comments. Also, HP's battery received a higher percentage of negative feedback, and Toshiba received a higher percentage of negative price comments than other brands.

In [20], the researchers proposed a big-data analytics-based approach. This proposed approach considered social media (Twitter) data to identify supply chain management issues in the food industry. The researchers used a method that includes text analysis using a support vector machine (SVM) sentiment analysis, hierarchical cluster analysis with multi-scale bootstrap sampling techniques to categorize each tweet as either positive or negative. As a case study, the researchers used this proposed approach to analyse the beef supply chain using three weeks of Twitter data. The outcome set of words in this research enabled the supply chain decision-makers to improve food products based on customer opinions.

To align this study with the most up-to-date developments in sentiment analysis and E-learning, we have incorporated recent research (published after 2020) that emphasizes Twitter-based sentiment detection, social media data analysis, and the application of machine learning methods in educational contexts.

Alsuraihi et al. (2023) conducted a focused sentiment analysis on students' attitudes toward E-learning platforms in Saudi Arabia using Twitter posts. Their study highlights key challenges, like usability and support, expressed by students in Arabic and English tweets, further validating Twitter as a rich data source for understanding public opinion in the education sector [21].

Zhang et al. (2022) proposed a deep learning approach for evaluating student sentiment on E-learning during the COVID-19 pandemic using Twitter data. Their model effectively detected nuanced emotions and identified trends in user satisfaction and concern with online education. This work confirms the increasing role of real-time social media analysis in assessing digital learning environments [22].

Kumar et al. (2021) provided a machine learning-based review of social media sentiment and emotion classification techniques. Their comparative analysis of classifiers such as SVM, Random Forest, and neural networks supports the choice of SVM in our study. It highlights the effectiveness of ML in extracting meaningful patterns from unstructured Twitter data [23].

Patel and Mehta (2021) compared the performance of various machine learning algorithms, including SVM, Naïve Bayes, and Logistic Regression, in classifying sentiments of educational tweets. Their findings reinforce the effectiveness of SVM for accuracy and scalability in sentiment classification, aligning with the methodology of our proposed model [24]. Table 1 shows a summary of the related studies that applied sentiment analysis in the context of E-Learning, highlighting the main classifier used, data collection, and type of analysis.

Table 1: Summary of related research applied models

Study	**Field**	**Location**	**Dataset**		**Techniques used**	**Type of Analysis**		**Measures and techniques of SA**
			Data collection	**size**		**SA used?**	**Other analysis**	
[11]		Egypt	Questionnaire	1000	- NB - SVM - DT	Yes	_	Word2vec technique
[12]		China	Questionnaire	3611	- K-Means	Yes	_	Self-Rating Anxiety Scale
[13]	E-Learning	Malaysia	Quantitative and qualitative questions	400	_	Yes	_	QDAP2
[16]		United States	Questionnaire	330	_	Yes	_	Likert-type items
[17]		Philippine	National Survey	3670	- Descriptive statistics	Yes	_	Kruskal-Wallis and Dunn's multiple nonparametric pairwise tests
[18]	Retail pharmacy organizations	United Kingdom	Twitter			Yes	_	_
[19]	Customer knowledge on social media		Twitter			Yes	- Statistical analysis - Text mining	_
[20]	Supply chain management issues in the food industry.		Twitter		- SVM - Hierarchical cluster analysis - Multi-scale bootstrap sampling	Yes	- Text analysis	_

Based on our review, we can derive the following observations:

1. All studies ([11, 12, 13, 16, 17], related to Sentiment Analysis on Twitter data in E-Learning during COVID-19 have used the questionnaire and survey as a data collection process for analysing and evaluating the tweets in specific countries: Egypt, China, Malaysia, the United States, and the Philippines.
2. The size of samples collected in those studies is limited, it was 1000, 3611, 400, 330, and 3670, respectively for each country.

3. All research studies have focused on Sentiment Analysis to analyse the content of Twitter. Only limited studies [11] and [12] have been conducted to analyse Twitter data in E-Learning using Machine Learning techniques.
4. Alsuraihi et al. [21] and Zhang et al. [22] demonstrated that Twitter is a valuable platform for analysing student sentiment toward E-learning, using both deep learning and traditional sentiment analysis methods.
5. Kumar et al. [23] and Patel and Mehta [24] confirmed that Support Vector Machine (SVM) is a reliable and high-performing algorithm for classifying educational sentiment in social media data.

3. METHODOLOGY

The methodology of the current study consists of five main phases. The first phase is data collection. In data pre-processing, a Lexicon-based approach was applied to classify the extracted tweets. The third phase consists of feature extraction using the TF-IDF feature selection method. After feature extraction, the sentiment analysis model was built by using the SVM classifier. Finally, the model is evaluated.

Figure 1 illustrates the overall methodology of this study to build a Sentiment Analysis model. The following subsections describe the methodology phases in detail.

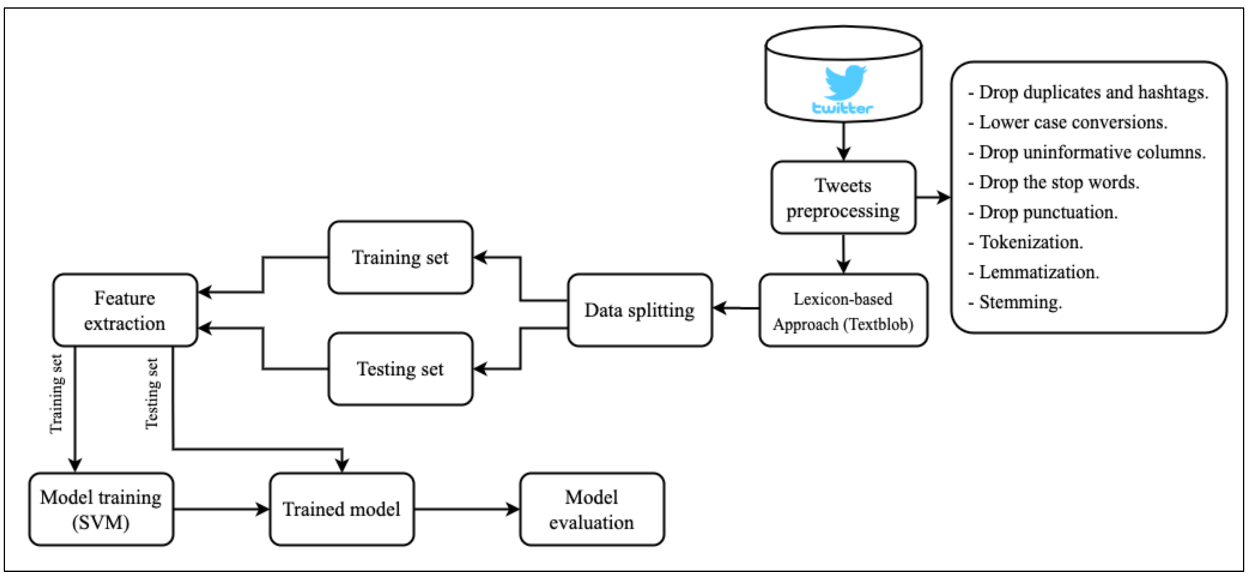

Figure 1: The overall methodology of this study

3.1. Data Collection

The Twitter platform is the source of data for this study because it generates many users' opinions. The dataset was built from scratch by using Tweepy, a Python library to access the Twitter API. Tweets that involved E-learning were selected by using a programmable code written in Python. The total number of extracted data was 42368 tweets in the English language. The data collection was conducted from 25 September to 04 October 2021. Figure 2 presents a sample of the extracted data.

	Unnamed: 0	Content	Location	Username	Retweet-Count	Favorites	Created at
0	0	Make homeschooling easier for you and the kids with these life changing tips :)\n#homeschooling #virtuallearning\n\nhttps://t.co/jhJeD4NoMy	California, USA	allaboutbaby00	0	0	2021-10-04 16:58:28
1	1	If you're wondering how you can support early literacy, donate to Reach Out and Read! We're still handing out books through our amazing network of pediatric clinics! 📚💪 \n\nDONATE: https://t.co/gITko73Srd \n\n#ReachOutAndReadGNY #EarlyLiteracy #Donate #VirtualLearning https://t.co/M9WBhAskXz	New York, NY	ReachOutReadGNY	0	0	2021-10-04 16:50:03
2	2	What professional learning are educators most in need of right now? How can we work this into both our short-term and long-term plans? \n\n📘 Learn more here: https://t.co/2YYKaQJli6 \n\n#professionallearning #personalizedPL #PD #professionaldevelopment #virtuallearning https://t.co/CbyjeaT6zW	New Jersey & ✈️	thelearningloop	0	0	2021-10-04 16:39:26
3	3	It's not too late for Early Action or Early Decision! FREE Bootcamps! Sign up here: https://t.co/iMtzSBQpef \n\n#story2 #collegeadmissions #storytelling #backtoschool #personalstatement #highschool #college #virtuallearning #free #class #collegeessays #earlyaction #earlydecision https://t.co/25Cmz9O3Ft	New York, NY	Story2	1	1	2021-10-04 16:07:36
4	4	Big thanks to some of my awesome cohort members for introducing me to "Fat Bear Week"! The kids are loving this and it is giving us a fun twist during our morning meeting! #VirtualLearning #FirstGrade https://t.co/s0YX1UaXBs	Raleigh, NC	MarydithB456	0	0	2021-10-04 14:53:58

Figure 2: Sample of the extracted dataset from Twitter

3.2. Ethical Considerations

This study uses publicly available data collected from Twitter through its official API, in accordance with Twitter's Developer Policy and Terms of Service. Only public tweets were extracted, and no private or sensitive user information was collected or stored. The dataset was anonymized during pre-processing by excluding usernames, user IDs, and any identifiable metadata to preserve user privacy. As the data is publicly accessible and used for academic research, individual user consent was not required. However, all efforts were made to maintain ethical integrity in data handling, ensuring transparency, confidentiality, and responsible use of social media content.

3.3. Data Pre-processing

The data pre-processing is applied to exclude irrelevant data from the dataset. This step is necessary and significant because it reduces the computational time and increases the classifier's performance. The noisy data can slow the performance and reduce the efficiency of the system. Figure 3 presents the new column named "Processed" that contains the tweets after implementing pre-processing. To clean our dataset, our study applied the following steps:

- Drop duplicates and hashtags.
- Lower case conversions.
- Drop uninformative columns,
- Drop the stop words.
- Drop punctuation.
- Tokenization.
- Lemmatization.
- Stemming.

	Processed
0	[make, homeschooling, easy, kids, life, changing, tips, homeschooling, virtuallearning, jhjed, nomy]
1	[wondering, support, early, literacy, donate, reach, read, still, handing, books, amazing, network, pediatric, clinics, donate, gitko, srd, reachoutandreadgny, earlyliteracy, donate, virtuallearning, wbhaskxz]
2	[professional, learning, educators, need, right, work, short, term, long, term, plans, learn, yykaqjli, professionallearning, personalizedpl, professionaldevelopment, virtuallearning, cbyjeat]
3	[late, early, action, early, decision, free, bootcamps, sign, imtzsbqpef, story, collegeadmissions, storytelling, backtoschool, personalstatement, highschool, college, virtuallearning, free, class, collegeessays, earlyaction, earlydecision, cmz]
4	[big, thanks, awesome, cohort, members, introducing, fat, bear, week, kids, loving, giving, fun, twist, morning, meeting, virtuallearning, firstgrade, uaxbs]

Figure 3: The processing tweets

After the data pre-processing, the final number of datasets is 39131 tweets.

3.4. Feature Extraction

The feature extraction is used for training the classifiers. For classification, features are information taken from the text and given to the algorithm. For sentiment analysis, features can be words, terms, or phrases that express the opinion as negative or positive. Term Frequency - Inverse Term Frequency (TF-IDF) feature selection is the method that we used to select the most important features in a text. TF-IDF is defined as a statistical concept to be used to get the frequency of words. The scikit-learn TF-IDF Vectorizer was used to calculate each word's weight and return a TF-IDF matrix.

3.5. Model Building

Based on our comparative study presented above in Table 1, we have chosen the Support Vector Machine classifier (SVM) to build the model for analyzing Twitter data in E-Learning. SMV has shown comparatively high performance in related tasks in the state-of-the-art [11], [20].

To build the machine learning model, the data was split into training and testing sets. The training data has been used to train the SVM classifier, and the testing data has been used to test the model accuracy. The dataset was split according to the 80/20 rule, i.e., 80% of the dataset for the training set, and 20% for the testing set. The train_test_split method of the Sk-learn library was used in Python.

The TF-IDF vectors were used as the features and the classes as the target. We have encoded the classes as follows: "Positive" = 2, "Neutral" = 1, and "Negative" = 0. Finally, a new TF-IDF vectorizer was created and initialized because the earlier TF-IDF vectorizer fit the entire dataset by following these steps:

- Initialize the TF-IDF vectorizer.
- Fit the vectorizer with the training set.
- Fit the test data with the same vectorizer.
- Initialize the SVM classifier.
- Fit the model.

3.6. Model Evaluation

Model evaluation metrics are used to show how predictions are expected to be accurate. For this study, we focused on the accuracy and confusion matrix to evaluate the model's performance. A confusion matrix is a tabular layout that allows visualization of classifier performance. Columns in the array are a prediction class, while rows are an actual class.

In the following section, we discuss the evaluation results of the classifier based on the Confusion matrix and the accuracy metric.

4. RESULTS AND DISCUSSION

This section presents the expressive and visual results obtained using the Support Vector Machine (SVM) classifier for sentiment analysis of Twitter data. Various visualization techniques have been applied to improve interpretability, highlight sentiment trends, and offer a more intuitive understanding of the data distribution. The results are discussed through multiple analytical lenses, including temporal distribution, geographical origin, tweet popularity, keyword frequencies, and polarity distribution.

4.1. Data Analysis

Data analysis was conducted to explore the characteristics of the collected dataset. The creation time of tweets was examined to understand user engagement trends throughout the day, as illustrated in Figure 4, which shows the distribution of tweet hours in a histogram. Additionally, Figure 5 presents the geographical distribution of tweets by country, indicating global participation in discussions related to E-learning.

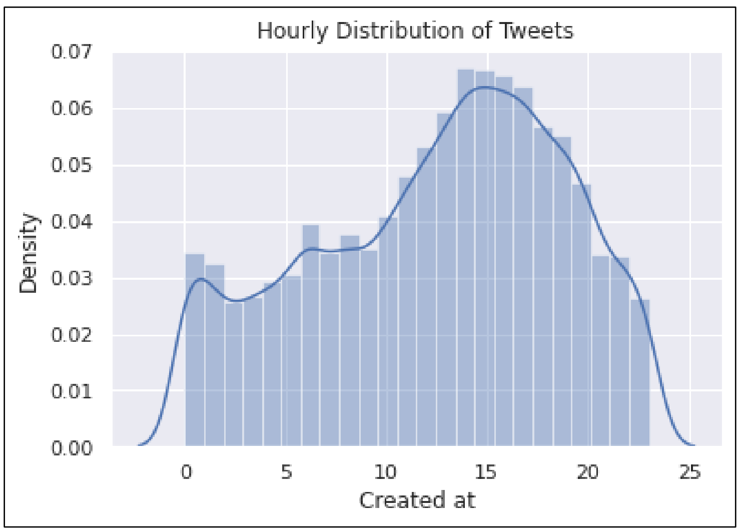

Figure 4: The hours' histogram

To identify the most engaging content, Figure 6 displays tweets with the highest numbers of retweets and likes. This provides insight into which types of E-learning-related content attract the most attention and interaction on Twitter.

Furthermore, Figure 7 includes a word cloud showing the most frequent terms found in the dataset. This helps highlight dominant themes and common expressions associated with E-learning.

To further refine sentiment-specific insights, Figure 8 presents separate word clouds for positive and negative tweets, showcasing what users appreciated and criticized.

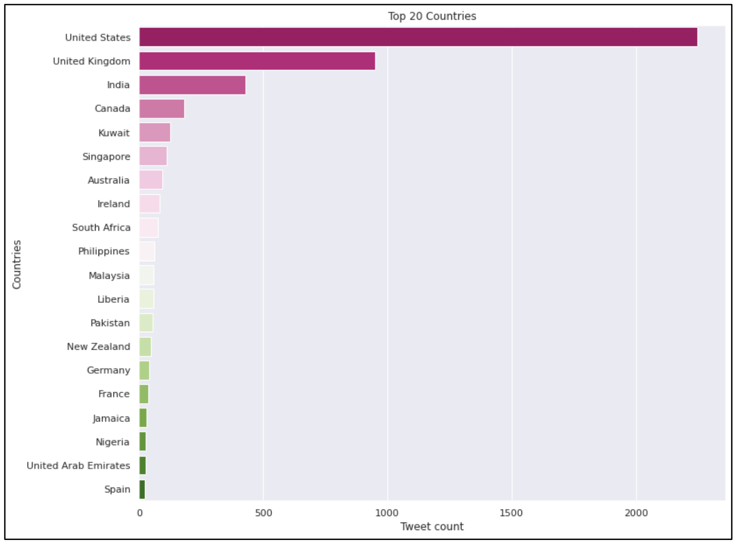

Figure 5: The visualization of the countries

	Content	Retweet-Count	Favorites
38801	Over 6500 people in Uttar Pradesh learned Sanskrit in 3 months.\n\nA total of 17,480 people were registered for the first level Sanskrit language teaching in the last 3 months. Out of these, 6,434 people were educated through 132 online classes.	3335	18612
21835	me during online class https://t.co/07I5dueSA1	5709	13731
32900	more JRPGs should adopt the xenoblade way of naming boss enemies because nothing beats seeing a big fucking frog in the distance followed by learning that it's called "MASSIVE GEOFFREY" or some shit	1751	9363
20769	sad hours bc i haven't learned anything since online class started and im worried that i'll be an incompetent adult after i graduate https://t.co/TMyGR4Cj7o	2853	7676
10368	Varsity students after realizing that level 1 means semester tests and exams will no longer be online. #FamilyMeeting https://t.co/ysQ3BlQVVo	600	4432
37151	If you are interested in learning about investing, I'm teaching a 4 week course online. 3-4 hours/week of content in audio/video/text: covering investing basics, drivers of value creation, the art of valuation, & developing the right mindset. Sign up here: https://t.co/t8Tw10zojh	146	3317
18860	This whole "waking up for online class" shit ain't for me	1060	3218

Figure 6: The most popular tweets

Figure 7: The word cloud of tweets

Figure 8: The word cloud for positive and negative tweets

After applying Text Blob, a lexicon-based sentiment analysis tool, the collected tweets were categorized into three sentiment classes: positive, neutral, and negative. The analysis revealed that 20,462 tweets (approximately 52%) expressed positive sentiment toward E-learning, indicating a general approval or appreciation of online education platforms, their flexibility, accessibility, or the innovation they offer. In contrast, 11,308 tweets (around 29%) were neutral, showing either objective observations, informational content, or unclear emotional tone. Lastly, 7,361 tweets (about 19%) were identified as negative, reflecting users' dissatisfaction, criticism, or concerns related to aspects such as internet connectivity, lack of interaction, or assessment challenges in E-learning.

To visualize these sentiments more intuitively, separate word clouds were generated for the positive and negative tweets, as illustrated in Figure 8. These visualizations display the most frequently occurring words in each sentiment category, with larger font sizes representing higher word frequencies. In the positive word cloud, terms such as "flexibility," "learning," "access," "innovative," and "support" were prominent, highlighting key aspects users value in E-learning systems. Conversely, the negative word cloud prominently featured terms such as "problem," "slow," "connection," "difficult," and "frustrating," pointing to common pain points and challenges faced by learners.

These word clouds serve as a powerful tool for identifying patterns in user opinions and understanding the qualitative dimensions behind the quantitative sentiment scores.

The visualization of the count of positive, neutral, and negative tweets is shown in Figure 9. It displays the polarity scores using a bar chart to illustrate the sentiment distribution across the dataset. The chart shows that positive tweets form the largest portion, reflecting a generally favourable perception of E-learning. Neutral tweets represent balanced or informational content without a strong emotional tone. Negative tweets, though fewer, highlight concerns such as technical issues, lack of engagement, or online learning fatigue. This visual summary clearly emphasizes the predominance of positive sentiment among users during the data collection period.

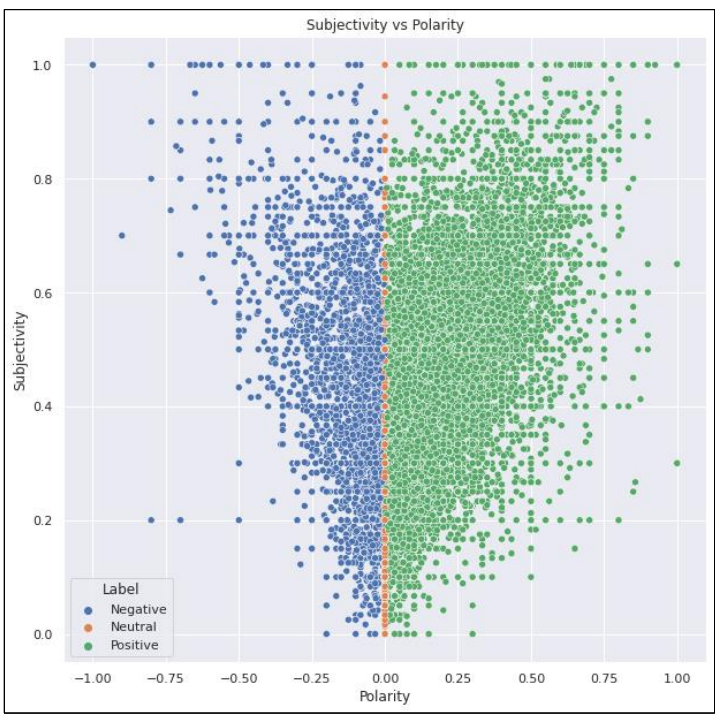

Figure 9: The polarity scores

Figure 10 includes sample tweets labelled as positive or negative, providing concrete examples of the language and tone associated with each category. These samples reinforce the sentiment classification process and demonstrate how the

SVM model interprets tweet content.

Negative Tweets

	Content	Retweet-Count	Favorites	Polarity
16082	@iCuddlyLouu Online class is boring sometimes yk 😭	0	3	-1.0
10584	Road to hell na ba kayo sa mga modules and activities niyo online? What are you waiting for? DM to get my attention! I can help you get through this dreadful semester. Di ka makasagot kasi maraming backlogs? pagawa mo na yan!\nTwt @ScarStudy \nshestudy023@gmail.com\nMOP: Gcash, load https://t.co/dgzFmUPTc9	1	2	-1.0
2435	University eliminating education choice if you don't get vaccinated. This is beyond insane. 😠 \nhttps://t.co/EXJr3A75i1	0	2	-1.0

Positive Tweets

	Content	Retweet-Count	Favorites	Polarity
28169	#irishrugby \n\nThe IRFU's 'GAINLINE' e-learning platform has won Silver in the Brandon Hall Excellence in Technology Awards in the category of "Best Advance in Creating a Learning Strategy" @D2L \n\nhttps://t.co/jwbc15pXD0	5	23	1.0
31058	A3 - You all are sharing some awesome strategies to make learning stick that are connected to research https://t.co/a6nbQVj9AS #CultureED https://t.co/kn9z6aoBAY	10	19	1.0
37539	1/ The best forward is to engage w/ their works & HK society. Hui Po-keung 許寶強 & Law Wing-sang 羅永生 retired from the uni in 2017 (https://t.co/x0w8qifnnL) but they've not stopped teaching or writing. Their online college Mobile Co-Learning 流動共學 https://t.co/871MczHRkL	11	17	1.0

Figure 10: The positive and negative tweets samples

4.2. Measurement Model

To evaluate the effectiveness of the proposed sentiment analysis model, we used two key metrics: accuracy and the confusion matrix. The SVM classifier was trained and tested using an 80/20 data split. Figure 11 displays the confusion matrix and the resulting accuracy score.

```
Confusion Matrix

[[1128  143  201]
 [   7 2166   89]
 [  42  179 3872]]


Classification Report

              precision    recall  f1-score   support

           0       0.96      0.77      0.85      1472
           1       0.87      0.96      0.91      2262
           2       0.93      0.95      0.94      4093

    accuracy                           0.92      7827
   macro avg       0.92      0.89      0.90      7827
weighted avg       0.92      0.92      0.91      7827
```

Figure 11: The evaluation results

The SVM model achieved a high accuracy of 92%, demonstrating its robustness and effectiveness in correctly classifying tweets into positive, neutral, or negative sentiment categories. This high accuracy reflects the model's strong ability to handle real-world, noisy, and diverse social media data, which often includes informal language, abbreviations, and varying contexts. The performance also highlights the careful pre-processing steps taken during data preparation. These results validate the selection of the SVM classifier as an appropriate machine learning approach for this task. They also confirm that the overall methodology, ranging from data collection to feature engineering and model training, is well-suited for large-scale sentiment analysis in the context of E-learning discussions on Twitter.

4.3. Limitations

Despite the promising results of the proposed model, this study has some limitations that should be acknowledged. First, the dataset is derived exclusively from Twitter, which may introduce sampling bias, as Twitter users do not represent the full demographic diversity of E-learning participants. Consequently, certain populations, such as younger students or those in regions with limited internet access, may be underrepresented. Second, the use of the Support Vector Machine (SVM) classifier, while effective, may not capture deeper contextual or emotional nuances in text compared to more advanced models like LSTM or BERT. Finally, the analysis is limited to English-language tweets, excluding opinions expressed in other languages, which could have provided a more comprehensive global perspective.

5. CONCLUSION

With a dataset comprising 42,368 tweets collected over 10 days, this research presents a robust and expressive foundation for building and validating hypotheses in the domain of sentiment analysis. Compared to previous studies that relied on smaller samples obtained via surveys or questionnaires (e.g., 400 - 3,600 participants), our approach leverages real-time, large-scale, user-generated data from a widely used social media platform. This significantly enhances the representativeness, diversity, and generalizability of our findings on public sentiment towards E-learning.

This study focused on social media platforms to monitor and gather people's opinions about E-Learning. It focused on the Twitter platform as a source for the data that was studied and analysed. The results indicated a great trend towards social media platforms to communicate with each other and express themselves. After the pre-processing of the collected data from Twitter using different E-learning keywords, 39,131 tweets related to E-learning were retained for analysis. A lexicon-based approach was applied to classify the sentences extracted from Twitter into positive, neutral, and negative categories. As a result, people's opinions were generally positive about E-learning during the tweet extraction period, as the number of positive tweets reached 20,462, while negative tweets numbered 7,361, and neutral tweets were 11,308.

Subsequently, a sentiment analysis model based on a Support Vector Machine (SVM) classifier was built using the cleaned Twitter dataset. The performance of the SVM classifier was high, achieving an accuracy rate of 92%. Generally, the results indicated that the percentage of people accepting E-learning was very high. These findings may help educational institutions to consider integrating E-learning with traditional physical attendance models. The main outcomes highlight the importance and necessity of sentiment analysis of social media opinions, which can be a powerful tool for organizations to enhance their services, align with user needs, and reshape their strategic directions.

For future work, we plan to enhance the model by applying advanced machine learning and deep learning techniques, such as BERT and LSTM, to better capture the contextual meaning of user opinions. We also intend to expand the dataset by including content from additional social media platforms like Reddit, YouTube, and Facebook, which can offer broader and more diverse viewpoints. These improvements aim to increase the model's accuracy, adaptability, and relevance in analysing sentiments within digital learning environments.

Acknowledgement

We thank the anonymous reviewers for their valuable comments, which helped us improve the quality, organization, content, and presentation of this paper.

Reference

[1] Hoppe, H.U.; Joiner, R.; Milrad, M.; Sharples, M. Guest Editorial: Wireless and Mobile Technologies in Education. Journal of Computer Assisted Learning 2003, 19, 255–259, doi:10.1046/j.0266-4909.2003.00027.x.

[2] Kaplan, A.M.; Haenlein, M. Users of the World, Unite! The Challenges and Opportunities of Social Media. Business Horizons 2010, 53, 59–68, doi:10.1016/j.bushor.2009.09.003.

[3] Lee, I. Social Media Analytics for Enterprises: Typology, Methods, and Processes. *Business Horizons* **2018**, *61*, 199–210, doi:10.1016/j.bushor.2017.11.002.

[4] Rathore, A.K.; Kar, A.K.; Ilavarasan, P.V. Social Media Analytics: Literature Review and Directions for Future Research. *Decision Analysis* 2017, *14*, 229–249.

[5] Mangold, W.G.; Faulds, D.J. Social Media: The New Hybrid Element of the Promotion Mix. *Business Horizons* **2009**, *52*, 357–365, doi:10.1016/j.bushor.2009.03.002.

[6] He, W.; Wang, F.K.; Akula, V. Managing Extracted Knowledge from Big Social Media Data for Business Decision Making. *Journal of Knowledge Management* **2017**, *21*, 275–294, doi:10.1108/JKM-07-2015-0296.

[7] Vázquez, S.; Muñoz-García, Ó.; Campanella, I.; Poch, M.; Fisas, B.; Bel, N.; Andreu, G. A Classification of User-Generated Content into Consumer Decision Journey Stages. *Neural Networks* **2014**, *58*, 68–81, doi:10.1016/j.neunet.2014.05.026.

[8] Ganganwar, V.; Rajalakshmi, R. Implicit Aspect Extraction for Sentiment Analysis: A Survey of Recent Approaches. In Proceedings of the Procedia Computer Science; Elsevier B.V., January 1 2019; Vol. 165, pp. 485–491.

[9] Batrinca, B.; Treleaven, P.C. Social Media Analytics: A Survey of Techniques, Tools and Platforms. *AI and Society* **2015**, *30*, 89–116, doi:10.1007/s00146-014-0549-4.

[10] Mostafa, L. Egyptian Student Sentiment Analysis Using Word2vec During the Coronavirus (Covid-19) Pandemic. In Proceedings of the Advances in Intelligent Systems and Computing; Springer Science and Business Media Deutschland GmbH, October 19 2021; Vol. 1261 AISC, pp. 195–203.

[11] Wu, M.; Zhao, H.; Guo, Y. Analysis of College Students' Psychological Anxiety and Its Causes under COVID-19. In Proceedings of the 15th International Conference on Computer Science and Education, ICCSE 2020; Institute of Electrical and Electronics Engineers Inc., August 1 2020; pp. 107–111.

[12] Lau, S.L.; Sim, T.Y. Feedback of University Students on Online Delivery Learning during the COVID-19 Pandemic Period. In Proceedings of the 2020 IEEE Conference on e-Learning, e-Management and e-Services, IC3e 2020; Institute of Electrical and Electronics Engineers Inc., November 17 2020; pp. 13–18.

[13] Rose, S. Medical Student Education in the Time of COVID-19. *JAMA - Journal of the American Medical Association* **2020**, *323*, 2131–2132, doi:10.1001/jama.2020.5227.

[14] Catanese, V.M. *Guidance on Medical Students' Clinical Participation: Effective Immediately*; 2020;

[15] Villanueva, E.W.; Meissner, H.; Walters, R.W. Medical Student Perceptions of the Learning Environment, Quality of Life, and the School of Medicine's Response to the COVID-19 Pandemic: A Single Institution Perspective. *Medical Science Educator* **2021**, *1*, 3, doi:10.1007/s40670-021-01223-z.

[16] Baticulon, R.E.; Sy, J.J.; Alberto, N.R.I.; Baron, M.B.C.; Mabulay, R.E.C.; Rizada, L.G.T.; Tiu, C.J.S.; Clarion, C.A.; Reyes, J.C.B. Barriers to Online Learning in the Time of COVID-19: A National Survey of Medical Students in the Philippines. *Medical Science Educator* **2021**, *1*, 3, doi:10.1007/s40670-021-01231-z.

[17] Rothan, H.A.; Byrareddy, S.N. The Epidemiology and Pathogenesis of Coronavirus Disease (COVID-19) Outbreak. *Journal of Autoimmunity* 2020, *109*, 102433.

[18] Zhan, Y.; Han, R.; Tse, M.; Ali, M.H.; Hu, J. A Social Media Analytic Framework for Improving Operations and Service Management: A Study of the Retail Pharmacy Industry. *Technological Forecasting and Social Change* **2021**, *163*, 120504, doi:10.1016/j.techfore.2020.120504.

[19] He, W.; Zhang, W.; Tian, X.; Tao, R.; Akula, V. Identifying Customer Knowledge on Social Media through Data Analytics. *Journal of Enterprise Information Management* **2019**, *32*, 152–169, doi:10.1108/JEIM-02-2018-0031.

[20] Singh, A.; Shukla, N.; Mishra, N. Social Media Data Analytics to Improve Supply Chain Management in Food Industries. *Transportation Research Part E: Logistics and Transportation Review* **2018**, *114*, 398–415, doi:10.1016/j.tre.2017.05.008.

[21] Alsuraihi, D., Alshareef, N., & Alharthi, A. Mining sentiments of students toward E-learning platforms using Twitter data. Education and Information Technologies, 2023, 28, 5137–5154. https://doi.org/10.1007/s10639-022-11389-2

[22] Zhang, Y., Lin, Z., & Hu, Q. Deep learning-based sentiment analysis for online learning experience evaluation during the COVID-19 pandemic. Computers in Human Behavior Reports, 2022, 6, 100164. https://doi.org/10.1016/j.chbr.2022.100164

[23] Kumar, A., Singh, J. P., & Dwivedi, Y. K. Social media analytics for sentiment and emotion classification: A machine learning perspective. Information Systems Frontiers, 2021, 23(1), 1–18. https://doi.org/10.1007/s10796-020-10087-2

[24] Patel, H., & Mehta, Y. A comparative study of machine learning algorithms for sentiment analysis of tweets in the educational domain. Procedia Computer Science, 2021, 185, 267–274. https://doi.org/10.1016/j.procs.2021.05.029

تحليل مشاعر آراء الأفراد في التعلم الإلكتروني باستخدام خوارزمية المتجهات الداعمة

ملخص: أتاح التطور السريع لوسائل التواصل الاجتماعي وانتشارها الواسع للناس التعبير عن أفكارهم وآرائهم ومشاعرهم عبر منصات التواصلالاجتماعي، مثل تويتر وميتا. ركز هذا البحث على تحليل مشاعر محتوى وسائل التواصل الاجتماعي المتعلق بالتعلم الإلكتروني، إذ يمكن للآراء العامة أن تساعد المؤسسات والأفراد. في هذا المجال، يتميز محتوى وسائل التواصل الاجتماعي بوفرة وعدم تنظيم، مما جعل تحليل المشاعر مجالًا بحثيًا رئيسيًا. تستكشف هذه الدراسة نهج التعلم الآلي لتحليل مشاعر محتوى تويتر لتحليل آراء الأفراد حول التعلم الإلكتروني. باستخدام واجهة برمجة تطبيقات تويتر، بلغ عدد التغريدات التي جُمعت 42368 تغريدة. تم اختيار التغريدات التي تناولت موضوع التعلم الإلكتروني باستخدام شيفرة قابلة للبرمجة مكتوبة بلغة بايثون. بعد ذلك، تم تدريب مُصنِّف خوارزمية المتجهات الداعمة على البيانات المُعالجة مسبقًا لتحليل التغريدات. أظهرت نتائج هذه الدراسة أن دقة المُصنِّف المُطبَّقة على مجموعة البيانات بلغت 92%. بشكل عام، كانت آراء الناس إيجابية تجاه التعلم الإلكتروني.

الكلمات المفتاحية: الشبكات الاجتماعية، تحليل المشاعر، التعلم الإلكتروني، تويتر، التعلم الآلي