Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.778007

SENTIMENT ANALYSIS OF ONLINE EDUCATIONAL PLATFORMS AMONG USERS IN SAUDI ARABIA

Hawazin Z. Albalawi[1]*, Aisha Y. Alsobhi[2]
[1]*Information Systems Department King Abdulaziz University Jeddah, Saudi Arabia,
E-mail:- halbalawe@stu.kau.edu.sa
[2]Information Systems Department King Abdulaziz University Jeddah, Saudi Arabia,
E-mail:- ayalsobhi@kau.edu.sa
***Corresponding Author: -** Hawazin Z. Albalawi
*Information Systems Department King Abdulaziz University Jeddah, Saudi Arabia,
E-mail:- halbalawe@stu.kau.edu.sa

Abstract
In this era, the great development in technology and communication media transformed the whole world into a small village. The education sector was one of the most important sectors that were highly affected by the digital revolution, especially in E-learning. Online educational platforms are considered one of the most recent fields that are helping generate a great amount of data. E-learning means the use of electronic tools and techniques for learning new information and skills. The demand increase in using digital media has had a significant impact since the COVID-19 pandemic. Online educational platforms were an effective solution as a sole alternative for students' inability to continue their educational achievement. Although it is an ideal solution in the pandemic crisis to adopt distance learning technology, many students have faced problems and challenges in the E-learning system. These problems are related to the technical difficulty students have in dealing with this type of education, especially in pre-university education, and even the teachers themselves did not possess the qualifications to deal with this sudden change in the method of learning.
This study aims to provide a sentiment analysis based on user preferences. This research presents an exploration of a researched area within a specific geographical and cultural context. In this paper, we focus on the platforms for educational purposes. This study provides context-specific insights that can guide the development of online learning platforms, and shape educational outlines. The findings can potentially enhance user engagement and learning outcomes, promoting better quality of education in the digital century.
Keywords: - Sentiment Analysis, Online educational platforms, Pre-university, culture-context.

1. INTRODUCTION

The continuing growth of digital transformation has impacted our lives, including education. The advent of online learning platforms has become a necessity since the waves of the COVID-19 pandemic. Digital educational platforms have emerged as critical tools in ensuring the continuity of education during times of social distancing and quarantine. However, understanding user sentiment towards these platforms remains a relatively under-researched area, especially within specific geographical and cultural contexts such as Saudi Arabia.

This research aims to fill this gap, examining user sentiment towards online educational platforms in Saudi Arabia. The research seeks to evaluate the effectiveness of various digital educational platforms. This study aims to contribute to this body of knowledge by providing a comparative analysis of the performance of a variety of algorithms in the context of online educational platforms.
This study is based on survey data gathered from users who have already dealt with a variety of popular online platforms for pre-university education in Saudi Arabia. The comparative analysis of the performance of these algorithms can help further research regarding the feasibility of online business, in addition to sentiment analysis.
The study's main contribution is the further potential to influence the development of online educational platforms and adapt educational policies according to user's preferences. The findings could potentially increase user engagement and learning outcomes for educational purposes, promoting the quality of education in the digital age, especially for Pre-university education.

2. LITERATURE REVIEW

The presence of online educational platforms in the pandemic period has changed the field of education, offering educational opportunities beyond the traditional on-campus classroom setting. The literature on online educational platforms has largely focused on aspects such as their effectiveness, convenience, and flexibility [1]. However, understanding user sentiment towards these platforms remains a relatively under-researched area, especially within specific geographical and cultural contexts such as Saudi Arabia.

A. User Sentiment Towards Online Educational Platforms

Sentiment analysis is a method used to identify and categorize opinions expressed in a piece of text, especially to determine whether the writer's attitude towards a particular topic is positive, negative, or neutral [2]. In the context of online educational platforms, user sentiment can provide insights into the perceived advantages and disadvantages of these platforms from a user perspective. A few studies have explored user sentiment towards online educational platforms [3], [4]. Still, these have largely been conducted in Western contexts and may not be fully applicable to the Saudi Arabian context.

B. Machine Learning for Sentiment Analysis

Machine learning, a subset of artificial intelligence, involves the use of algorithms that improve automatically through experience [5]. It has been used in a wide range of applications, including sentiment analysis. Several machine learning algorithms, such as Naive Bayes, Support Vector Machines, and Random Forests, have been employed for sentiment analysis [6], [7]. However, there is a lack of studies comparing the performance of different machine learning algorithms in this context.

C. Deep Learning for Sentiment Analysis

In [8], the study compared pre-trained word embeddings with custom embeddings, revealing that ARABERT-based custom embeddings consistently outperformed other options across all datasets. Custom-generated Fast Text embeddings are closely followed as the second-best choice, emphasizing the critical role of context-aware embeddings in the realm of sentiment classification. the classical word embeddings, such as FastText, Word2Vec, and GloVe, were pitted against contextualized BERT embeddings. The findings demonstrated that BERT, whether pre-trained or trained on dataset-specific information, consistently surpassed traditional word embeddings. This highlights the superior performance of BERT's context-aware embeddings in the context of Arabic sentiment analysis. Lastly, the choice between deep learning models, BiLSTM and CNN, was found to substantially influence performance. BiLSTM exhibited superior performance in accuracy, precision, recall, and F1-measure for the HARD, Khooli, and ArSAS datasets. The HARD dataset, akin to our collected data, comprises balanced positive and negative labels in Arabic. As our dataset expands, we expect improved performance, mirroring HARD's notable results. Conversely, CNN excelled in the AJGT and ASTD datasets, underscoring the importance of dataset-specific considerations when selecting the appropriate deep-learning model. In sum, this research not only provides valuable insights into the complex world of word embeddings and deep learning models for Arabic sentiment analysis but also emphasizes the critical importance of context-aware embeddings, such as BERT, and the thoughtful selection of deep learning models, adding depth to our understanding of sentiment classification.

D. Contextual Factors

The literature suggests that user sentiment toward online educational platforms can be influenced by various factors, including usability, accessibility, and the quality of content and interaction [9]. Cultural and geographical factors can also play a role [10]. However, these factors have not been extensively explored within the Saudi Arabian context as our study focuses on.

3. METHODOLOGY

This section demonstrates the methods, tools, and techniques applied in assessing the distance learning experience. The study employed various embedding techniques using machine learning and deep learning models. The machine learning classifiers include Decision Tree, Random Forest, Extra Trees, K-Neighbors, SVC, and XGB Classifier. Deep learning models were constructed using Arabert and Fasttext, and a classification layer was appended at the end for prediction purposes.

A. Dataset Collection:

The dataset gathering was considering the inclusion of a different distance learning platform, whether national in Saudi Arabia, such as Madrasty and Ain, or international, such as Microsoft Teams and Zoom.

The study utilized two data sources: users' feedback on Twitter and a custom survey collecting opinions of E-learning platforms.

1. Twitter Data:

It's clear that a meticulous approach was taken in curating the dataset, focusing on tweets relevant to distance learning in Saudi Arabia and categorizing them based on their relevance levels and sentiments. Specific accounts, particularly semi-official ones providing updates on distance learning in Saudi Arabia, were intentionally targeted for data collection. Tweets were manually labeled based on their relevance to distance learning research. Three levels of relevance (0, 1, 2) were assigned, indicating irrelevant tweets, strongly relevant tweets with justification, and tweets with partial relevance without justification, respectively. Tweets were annotated into two classes: Class 0 for dissatisfaction and Class 2 for satisfaction. This binary classification provides a clear distinction between positive and negative sentiments regarding distance learning.

A preprocessing phase was conducted to refine the dataset. Irrelevant or spam tweets (12,620) were excluded. Partially relevant tweets (8,213) were identified, indicating concise responses without substantial elaboration. The final dataset used for analysis comprised 467 highly relevant tweets with a focus on substantial elaboration and opinions. The dataset reflects a careful selection process, emphasizing quality and relevance over quantity. By targeting specific accounts and focusing on relevance, the dataset aims to provide a balanced representation of opinions and sentiments related to distance learning. The emphasis on substantial elaboration and justification in relevant tweets indicates a desire for detailed insights into the opinions expressed. The final dataset, consisting of 467 highly relevant tweets, suggests a focus on quality and precision in the analysis.

This detailed approach to dataset curation and annotation lays a solid foundation for meaningful sentiment analysis and insights into the attitudes and opinions regarding distance learning in Saudi Arabia. The inclusion of justification and the focus on relevance levels add depth to the analysis, allowing for a nuanced understanding of public sentiments on the topic.

2. Survey Data:

A survey was conducted among secondary school students, their parents, and teachers in Saudi Arabia. The objective was to evaluate the E-learning platform through a series of questions. Secondary school students were chosen for their ability to assess whether distance learning suited them or not. Parents' responses were deemed crucial, as they could observe changes in their children's behavior during this process. Furthermore, teachers were enlisted to provide an objective perspective after evaluating student performance and comparing E-learning to traditional methods.

The survey was designed to collect textual reviews and user information. It included multiple-choice questions to assess user sentiment towards distance learning platforms. The users were then asked to provide an explanation for their selection. Notably, instances where users expressed positive sentiments but wrote negative comments, and vice versa were excluded to avoid confusion.

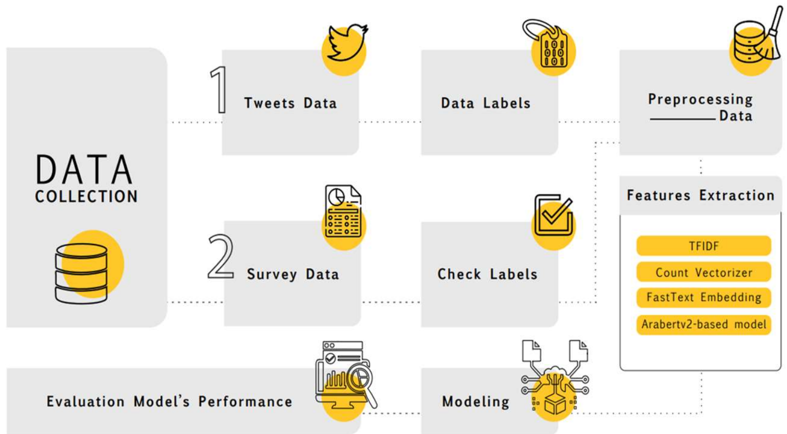

Fig. 1. Methodology pipeline

1) Preprocessing

The preprocessing of text data originating from social media sources, involves a series of steps. We eliminated dates presented in the "2K19" format, purged URLs, and stripped away redundant information such as usernames, "RT", and "CC". In addition, we removed all hashtags, emojis, punctuation, and excessive whitespace. Linguistic elements like stop words and Tashkeel were also discarded.

2) Features Extraction

As shown in Fig. 1, upon the completion of text preprocessing, the next step involved the extraction of features. This is an essential phase as it transforms the text data into a numerical format, which is understandable for processing by machine learning models. We used four distinct encoding techniques for this transformation: Term Frequency Inverse Document Frequency (TF-IDF), Count Vectorizer, FastText, and Arabert. Each of these techniques served to convert the textual information into numerical data, thereby facilitating the subsequent stages of analysis.

2.1) Term Frequency Inverse Document Frequency (TF-IDF)

TF-IDF is a statistical method frequently used in natural language processing and information retrieval. It measures the significance of a term within a document relative to a corpus of documents. TF-IDF scores a word by multiplying the term frequency (TF) by the inverse document frequency (IDF). The term's significance increases when it appears frequently in one text and infrequently in others.[17].

2.2) Count Vectorizer

Count Vectorizer encodes text or documents into a matrix. Each row represents a document, and each column represents a unique word in the corpus. The values in the matrix indicate the frequency of the word or term in the document.

2.3) FastText

FastText is a library developed by Facebook's AI Research team for text classification and language processing tasks. Fasttext provides two models for computing word representation: Skip-gram and CBOW. The Skip-gram model uses a nearby word to predict a target word, the CBOW model uses the bag of words found in a fixed-size window around the target word to predict it. In this study, we used the CBOW model.

B. Modeling:

Data is split for training and testing, 80% of the survey data and the Twitter data were taken as the training set. The 20% left from the dataset is used to test the performance of the machine learning models on this dataset. As for the training dataset, we split it again to take a validation and test set from it for training the Arabertv2-base model. It was trained on only 70% of the dataset, validated on 15%, and tested on 15%, .This split was done to be able to find the best hyperparameters. While the other machine learning models were trained on 80% of the dataset and evaluated on the remaining 20%.

The study utilizes a variety of both machine learning and deep learning models for the classification tasks. These models are briefly described below, with their respective hyperparameter settings tabulated.

1) Decision Tree

The Decision Tree algorithm constructs a tree-like model of decisions using the most significant attribute to split the dataset recursively. It starts from the root node, which represents the entire dataset, and splits data based on features such as Gini impurity or entropy. The splitting continues until a stopping criterion is met, and during inference, data traverses from the root node to a leaf, where the predicted class is represented [11].

2) Random Forest

Random Forest is an ensemble learning method that builds multiple decision trees using bootstrap sampling. To counteract overfitting, it only considers a random subset of features for splitting at each node. It combines individual tree predictions to make the final class prediction [12].

3) Extra Trees

Extra Trees or Extremely Randomized Trees is another ensemble learning method that creates multiple decision trees. Unlike the Random Forest algorithm, Extra Trees randomly selects feature thresholds for splits, speeding up the training process. It also uses bootstrapped samples for training and combines predictions from individual trees through majority voting [13].

4) K-Nearest Neighbors

The K-Nearest Neighbors (K-NN) model assigns a label to a new data point based on its K-nearest neighbors in the training data. It uses distance metrics such as Euclidean distance to calculate the distance between the new data point and all data points in the training set. The class label assigned is the most frequently observed one among the K neighbors [14].

5) Support Vector Classification

Support Vector Classification (SVC) finds a hyperplane that best separates the data points into distinct classes in a high-dimensional space. The SVC can transform non-linear data into a higher-dimensional space using a kernel function where the optimal hyperplane can be identified. The regularization parameter (C) controls the trade-off between maximizing the margin and minimizing classification errors [15].

6) XGBoost Classifier

The XGBoost (Extreme Gradient Boosting) classifier, an ensemble learning algorithm, builds a sequence of decision trees, each designed to correct the errors of the preceding one. The model minimizes the loss function through gradient descent and incorporates a regularization term to control the model's complexity. XGBoost is also capable of providing a feature importance score [16].

7) AraBERT

AraBERT shares the same architecture as BERT but was pre-trained on approximately 8.6 billion Arabic words. The model starts by encoding and tokenizing the input text into subword tokens and includes special [CLS] and [SEP] tokens at the beginning and end of the text. It uses multi-layer bidirectional transformers and a feedforward neural network to refine token representations. The [CLS] token is passed to a fully connected layer that serves as a classifier [17].

In Table.1, it’s shown the hyperparameters for the utilized fine tuned models:

TABLE .1 MODELS HYPER-PARAMETERS

Model	Hyperparameters	
Decision Tree	criterion	gini
	splitter	best
	max_depth	None
	min_samples_split	2
Random Forest	n_estimators	100

	criterion	gini
	max_depth	2
	min_samples_split	2
Extra Trees	n_estimators	100
	criterion	gini
	max_depth	2
	min_samples_split	2
K-Nearest Neighbors	n_neighbors	3
	weights	uniform
	algorithm	auto
	leaf_size	30
SVC	C	1.0
	kernel	rbf
	degree	3
	gamma	auto
XGBoost Classifier	booster	gbtree
	n_estimators	100
	learning_rate	0.3
	max_depth	6
	eval_metric	mlogloss
Arabert V2	Learning Rate	0.00005

	Batch Size	32
	Epochs	3

4. RESULTS AND DISCUSSION

In this section, we are going to discuss the experimental results of models used with respect to a variety of different embedding techniques. The models were evaluated based on their accuracy and F1 score for further model performance exploration according to each class.

A. Machine Learning Approaches

Using FastText embeddings [18], the Decision Tree model resulted in an accuracy of 0.7172 and an F1 score of 0.7146. For the same embeddings, the Random Forest model with an accuracy of 0.8073 and an F1 score of 0.8059. The Extra Trees model had an accuracy of 0.8114 and an F1 score of 0.896. The KNN and SVC models followed with accuracies of 0.7213, and the SVC model outperformed with 0.8512 , and F1 scores of 0.7212 and 0.8512 respectively. Lastly, the XGB model achieved an accuracy of 0.8360 and an F1 score of 0.8353.

When we used the CountVectorizer embeddings [19], the Decision Tree presented an accuracy of 0.8032 and an F1 score of 0.8019. The Random Forest model showed an accuracy of 0.8237 and an F1 score of 0.8229. The Extra Trees and KNN models recorded values of accuracies of 0.8196 and 0.6885, and F1 scores of 0.8188 and 0.6799 respectively. SVC and XGB models followed with accuracies of 0.8155 and 0.8278, and F1 scores of 0.8125 and 0.8264, respectively.

In addition to the TF-IDF embeddings [20], the Decision Tree model resulted in an accuracy of 0.7950 and an F1 score of 0.7948. The Random Forest model achieved an accuracy of 0.8114 and an F1 score of 0.8102, while the Extra Trees model had an accuracy of 0.8196 and an F1 score of 0.8192. The KNN model showed an accuracy of 0.7786 and an F1 score of 0.7777. The SVC model outperformed in this category with the highest accuracy of 0.8524 and an F1 score of 0.8507. Finally, the XGB model achieved an accuracy of 0.8278 and an F1 score of 0.8271. These results are summarized in Table 2, providing insights into the effectiveness of different machine learning models and embedding techniques for sentiment analysis text binary classification. The XGB model shows the strongest performance when FastText embeddings are used. The Extra Trees model demonstrates consistent performance across all three types of embeddings.

To conclude, the SVC model outperforms the other models' performance recording the best metrics with the use of FastText and TF-IDF embeddings.

TABLE 2. MODELS ACCURACY VS. F1

Model	Embeddings					
	Fasttext Embeddings		Count-Vectorizer		TF-IDF	
	Accuracy	**F1**	**Accuracy**	**F1**	**Accuracy**	**F1**
Decision Tree	**0.717**	**0.714**	**0.803**	**0.801**	**0.795**	**0.794**
Random Forest	**0.807**	**0.805**	**0.823**	**0.822**	**0.811**	**0.810**
Extra Trees	**0.811**	**0.809**	**0.819**	**0.818**	**0.819**	**0.819**
KNN	**0.721**	**0.721**	**0.688**	**0.679**	**0.778**	**0.777**
SVC	**0.852**	**0.851**	**0.815**	**0.812**	**0.852**	**0.850**
XGB	**0.836**	**0.835**	**0.827**	**0.826**	**0.827**	**0.827**

B. Deep Learning Approach - AraBERTv2:

We set the model parameters including a batch size of 32, a learning rate of 5.00E-05, and a number of epochs of 3 to avoid overfitting. In Fig. 2, and Fig. 3, the Arabert model loss and accuracy, respectively, across data epochs are plotted, The model accuracy on average was 90.16%, in addition to an F1 score of 90.14%.

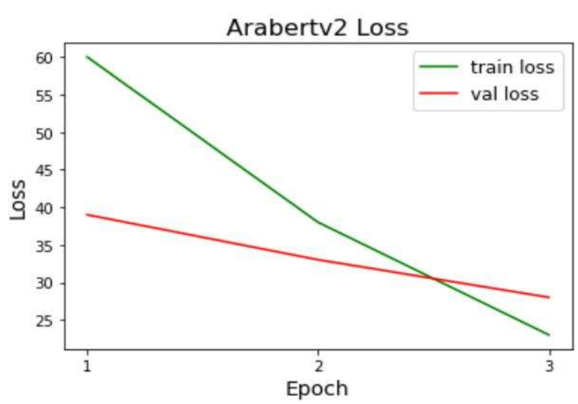

Fig. 2. Arabert Model Loss

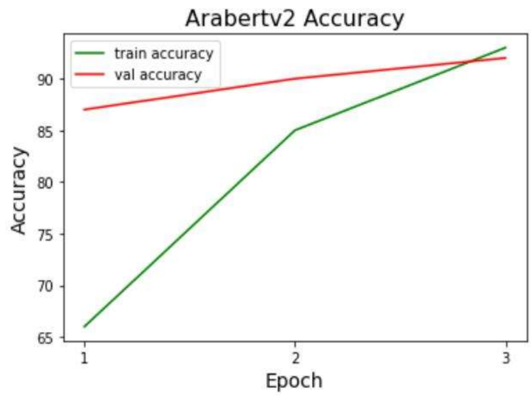

Fig. 3. Arabert Model Accuracy

V. CONCLUSION

In this paper we have presented a sentiment analysis using a combination of machine learning and deep learning techniques to help both businesses of online pre-university educational platforms and users demand coverage within a specific geographical and cultural context in Saudi Arabia. Exploring the effectiveness of our analysis based on text embedding techniques for text classification tasks. The experimental results demonstrated that the specific choice of machine learning model and embedding technique can significantly impact the performance of the task at hand.

In our machine learning experiments, the Random Forest model demonstrated the best performance when paired with Fast Text embeddings, while the Support Vector Classifier (SVC) model showed superior performance when TF-IDF embeddings were utilized. The Extra Trees model, on the other hand, exhibited consistent performance across all three types of embeddings, namely Fast Text, Count Vectorizer, and TF-IDF. The optimal configuration may vary depending on the dataset's characteristics, and it is crucial to choose a combination that best aligns with the unique characteristics of the dataset and the objectives of the sentiment analysis problem. In addition, the results we have obtained showed the importance of continued research in the realm of text classification. Further exploration into novel techniques and models can also enhance accuracy and efficiency in text classification tasks is of paramount importance.

VI. FUTURE WORK

In future work, we aim to explore other embedding techniques and machine learning models. Furthermore, we plan to explore the impact of integrating multiple models or techniques (an ensemble approach) on text classification tasks. In addition to using data augmentation techniques to increase our data set size, we can teach the machine a variety of sentiment patterns. We can also widen our data collection era by gathering people's opinions on social media platforms and doing a survey with a wide area coverage in Saudi Arabia.

VII. REFERENCES

1. A Smith, "Online Learning in Higher Education: A review," Education Journal, vol. 7, no. 2, pp. 59-64, 2018.
2. B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge: Cambridge University Press, 2015.
3. J. Brown and D. Jones, "User sentiment towards e-learning: A UK-based study," Technology in Education Journal, vol. 12, no. 1, pp. 22-30, 2019.
4. S. Wilson, "Sentiment Analysis for E-Learning Systems," in Proceedings of the International Conference on E-Learning, Paris, 2020, pp. 125-132.
5. E. Alpaydin, Introduction to Machine Learning. Cambridge, MA: MIT Press, 2020.
6. P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero- one loss," Machine Learning, vol. 29, no. 2-3, pp. 103-130, 1997.
7. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.
8. S. F. Sabbeh and H. A. Fasihuddin, "A Comparative Analysis of Word Embedding and Deep Learning for Arabic Sentiment Classification," in Electronics 2023, 12, 1425 10 of 16.
9. T. Anderson, The Theory and Practice of Online Learning. Edmonton: Athabasca University Press, 2008.
10. H. Alshammari, "Cultural aspects in designing e-learning platforms: A Saudi Arabian study," International Journal of E-Learning, vol. 14, no. 1, pp. 71-80, 2016.
11. Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81–106.
12. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
13. Geurts, P., Ernst, D., Wehenkel, L. (2006). Extremely randomized trees. Machine Learning, 63(1), 3–42.
14. P. Cunningham and S. J. Delany, ‘k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples)’, ACM Comput. Surv., vol. 54, no. 6, pp. 1–25, Jul. 2022, doi: 10.1145/3459665.
15. Cortes, C., Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297.
16. Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
17. Antoun, W., Baly, F., Hajj, H. (2020). AraBERT: Transformer-based Model for Arabic Language Understanding. arXiv preprint arXiv:2003.00104.
18. Nurul Khasanah, I., "Sentiment Classification Using fastText Embedding and Deep Learning Model," in Procedia Computer Science, vol. 189, pp. 343-350, 2021. DOI: 10.1016/j.procs.2021.05.103. doi: 10.1109/ICoDT252288.2021.9441508.
19. G. M. Raza, Z. S. Butt, S. Latif and A. Wahid, "Sentiment Analysis on COVID Tweets: An Experimental Analysis on the Impact of Count Vectorizer and TF-IDF on Sentiment Predictions using Deep Learning Models," 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), Islamabad, Pakistan, 2021, pp. 1-6, doi: 10.1109/ICoDT252288.2021.9441508.

20. P. Bafna, D. Pramod and A. Vaidya, "Document clustering: TF-IDF approach," 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 2016, pp. 61-66, doi: 10.1109/ICEEOT.2016.7754750