International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Classification of Spanish Fake News about Covid-19 using Text Augmentation and Transformers

Anibal Flores*[1,2], Hugo Tito-Chura[1,3], Victor Yana-Mamani[1,4], Alejandro Ecos-Espino[1,5]

Submitted: 25/08/2023 **Revised**: 16/10/2023 **Accepted**: 29/10/2023

Abstract: This paper presents the results of five models based on transformers such as DistilBERT, ALBERT, BETO, DistilBETO, and ALBETO for the classification of fake news about covid-19 in the Spanish language. Two text augmentation processes based on GPT-3 are compared, the first TA1 consists of the most common way of increasing the records of the training data, that is, increase all training data; and the second TA2, is more selective in the sense that it only increases the records that could not be learned by the models in the training phase, thus optimizing the training time of the models with respect TA1. The results show that both text augmentation techniques allow improvement, however, TA2 has a better performance in the models based on the Spanish language such as BETO, DistilBETO, and ALBETO, improving on average 1.15%, 11.12%, 2.44%, and 7.50% in terms of accuracy, recall, precision and f1-score respectively.

Keywords: *Covid-19 fake news, transformers, spanish language models, text augmentation*

1. Introduction

Fake news in social networks is a critical problem since it can harm the population [1]. Fake news can be created with the intent to deceive, manipulate, defame, or create fear in the audience. The World Health Organization has stated that fake news or infodemic is more deadly than the Covid-19 virus itself, it has affected people physically as well mentally [2].

Fake news detection can be approached from different perspectives, as a text classification task, graph classification or as a hybrid task [3], considering that extra information can also be extracted from images and videos, in this work it is approached as a text classification task.

The COVID-19 pandemic [4] has generated a large volume of fake news, misinformation, and conspiracy theories around the world. Some of the most common fake news related to the pandemic include conspiracy theories[5] about the origin of the virus, such as claiming that the virus was man-made; false claims about remedies or cures for disease[6]; misinformation about preventive measures, such as stating that the use of masks is harmful to health or that vaccination is not safe [7]; numbers of cases and deaths, such as claiming that the numbers are exaggerated or manipulated for political reasons.

[1] *Universidad Nacional de Moquegua, Urb. Ciudad Jardin-Pacocha-Ilo, Moquegua 18611, PERU*
[2] *Departamento de Ingeniería de Sistemas e Informática, UNAM, PERU ORCID ID : 0000-0003-4638-8242*
[3] *Departamento de Ingeniería de Sistemas e Informática, UNAM, PERU ORCID ID : 0000-0001-8269-0655*
[4] *Departamento de Ingeniería de Sistemas e Informática, UNAM, PERU ORCID ID : 0000-0003-0982-2353*
[5] *Departamento de Ciencias Básicas, UNAM, PERU ORCID ID : 0000-0002-3452-5388*
* *Corresponding Author Email: afloresg@unam.edu.pe*

Fake news about Covid-19 and other topics in English have been studied regularly and there are plenty of frameworks and models that help to detect them, however, there are not many works about fake news in the Spanish language.

In this work, fake news about Covid-19 in Peru is used as a study. For this, a 1015 record dataset was built including true and fake news from social networks such as Facebook and YouTube, they were labelled using platforms such as chequeado.com and factcheck.org.

For experimentation five transformers-based models have been implemented, DistilBERT [8], Albert [9], BETO [10], DistilBETO [11], and ALBETO [11] which were fine-tuned with the collected dataset. DistilBERT and Albert are multilingual models and they have been used in several works related to text classification such as [12], [13], [14] and, others. BETO, DistilBETO, and ALBETO are monolingual, they were trained just in Spanish language and, they are less popular than DistilBERT and ALBERT because as it was mentioned before, there are fewer works in the Spanish language than in the English one.

The main difficulty encountered in this work was the small size of the collected dataset compared to free datasets in English used in related work. To overcome this limitation, in this work, text augmentation is used, it is commonly used in Natural Language Processing (NLP) works. For this task, there are many options, but GPT-3 model was chosen, due to the quality of the synthetic text it generates. In works that use text augmentation, the procedure is generally applied to all training data without exception, this allows all texts to obtain similar synthetic texts, significantly increasing the size of the training data. What is observed is that not all texts require text augmentation, that is, only those texts that are difficult to learn should be augmented. Therefore, in this work, two text augmentation processes TA1 and TA2 are experimented with. The first TA1 augments text in the most common way, that is, synthetic texts are generated for all the training data without exception. The Second TA2 selectively increases the text, that is, synthetic texts are only generated for the registers that could not be learned in the training phase of models. The main idea of comparing TA1 and TA2 is to demonstrate that the results of the models to be implemented can be improved to the same or greater extent, without the need to augment all the training texts.

In summary, the main contributions of this work are the proposal of a Spanish language corpus for fake news about Covid-19, and the proposal of a new strategy for text augmentation TA2 using results of training phase of models.

Regarding the structure of this work. In section II, the related work that constitutes the state of the art of the study is described. Section III describes the methodology addressed for the implementation of transformers-based models. Section IV describes the results achieved by the transformers-based models. Finally, in section V, the conclusions reached and the future work that can be carried out based on the results of this work are shown.

2. Related Work

In this section, the related works for fake news about Covid-19 are briefly described. All these works were performed for datasets in English language, no related work could be found for Spanish language.

In [15] the authors used a dataset of 6424 social media posts, and implement classical machine learning models to predict fake COVID-19 news, these models are compared with the model based on Transformers known as DistilBERT.

In [16] the authors use a public dataset of 10202 records, 8202 are used for training and 2,000 for testing. They implement eight machine learning models such as Naive Bayes, Adaboost, KNN, Random Forest, Logistic Regression, Decision Tree, Neural Networks, and Support Vector Machine. And, four Deep learning models CNN, LSTM, RNN, and GRU to predict fake news. The best machine learning model in terms of accuracy, precision, recall, and f1-score is Random Forest and presents 97.00%, 99.00%, 98.00%, and 98.00% respectively. The best deep learning models are BiLSTM and CNN both with 97.00%, 97.00%, 97.00%, and 97.00% respectively.

In [17] proposed the use of n-grams of POS tags for Covid-19 fake news classification. the best f1-score is 0.761.

In [18] an ensemble model known as SCLAVOEM is proposed which is compared with machine learning models such as Naive Bayes, Function-Sequential Minimal Optimization, Voted Perceptron, and KNN. The best accuracy achieved is 87.26%.

In [19] the authors experiment with the Koirala dataset. They propose three wrapper feature selection techniques (Particle Swarm Optimization PSO, Genetic Algorithm GA, and Salp Swarm Algorithm SSA) to reduce symmetrical features. The best accuracy is obtained by KNN with 75.4% including a precision of 66.22%.

In [20] the authors proposed a combined deep learning model based on the ideal distance weighting method for fake news detection. To validate the model, they used two datasets: the ISOT and the COVID-19 fake news dataset. In data Covid-19 dataset the best accuracy is obtained by a combination of 2 and 3 models with 73.72%.

In [8] the authors as in [18] use a dataset of 984 claims and propose DistilBERT and Shapley Additive exPlanations (SHAP) to predict Covid-19 fake news, they also use text augmentation based on back-translation. The results reach an accuracy of 97.2% and an AUC of 99.3%.

In [21] the authors propose a framework based on RNN and CNN models. They experiment in 2 Covid-19 fake news datasets, reaching an accuracy of 100.00% in dataset 1 and 93.55% in dataset 2.

Table 1. Results of related works for Covid-19 fake news detection

Work	Dataset	Model	Metric	Result (%)
Bangyal et al., 2021 [16]	10202	Random Forest	Accuracy	97.00
			F1-Score	98.00
Ayoub et al, 2021 [8]	984	DistilBERT	Accuracy	97.20
Al-Ahmad et al,2021 [19]	3002	KNN	Accuracy	75.40
Kapusta et al, 2021 [17]	1100	Decision Trees	F1-Score	76.10
Olaleye et al, 2022 [18]		Ensemble	Accuracy	87.26
Gonwirat et al, 2022 [20]		Ensemble	Accuracy	73.72
Raj et al, 2022 [21]		RNN+CNN	Accuracy	93.55
Agarwal et al, 2022 [2]				

According to Table 1, it can be seen that in terms of Accuracy, the best results have been achieved for a Transformers-based model (DistilBERT [8]), it was the main motivation to work with Transformers-based models in this work.

Regarding text augmentation, none of the related work about Covid-19 fake news uses it; however, its use is common in other works related to NLP. Below some of them are briefly described.

Several text augmentation related-works focus on the synonym replacement method, works such as [22], [23] and [24] used WordNet as a synonym database where the synonym selection is randomly based or based on geometric distribution and others. Another group of works focuses on embedding replacement, so [25], [26], [27], and [28] used different techniques for embedding selection, for example random, cosine similarity, and others; neural networks such as DNN, CNN, and LSTM/GRU are used as base models. Other works use replacement by language models, works such as [29],[30],[28],[31], and [32] used BERT for this purpose. Round-trip translation is another strategy for text augmentation, works such as [33], [34] and [30] use google translate API for this task. Generative methods are also used in different works such as [35], [36], [26], and [37], they used BiLSTM, LSTM, BERT, CNN, and others as base models. Other works use interpolation in the feature space; so [30], [35], and [39] interpolate embedding matrices and padded word embeddings according to the used model.

According to the text augmentation related works described above, it can be seen that in none of the cases, the model results in the training phase are used to augment incorrectly learned sentences. Thus, the proposal text augmentation TA2 constitutes one of the contributions of the work described in this paper. Also, in recent work, the GPT-3 model is being used for text augmentation generating realistic synthetic data, works such as [40] and [41] used this model for text augmentation but not in the way it is proposed with TA2 in this work.

Another important aspect to highlight about the text augmentation techniques described above is that they do not always improve the accuracy or f1-score of the implemented models, in some cases, the text augmentation techniques worsen the results.

3. Methodology

Graphically Fig. 1 summarizes the methodology followed for the implementation of the five transformers-based models in this work.

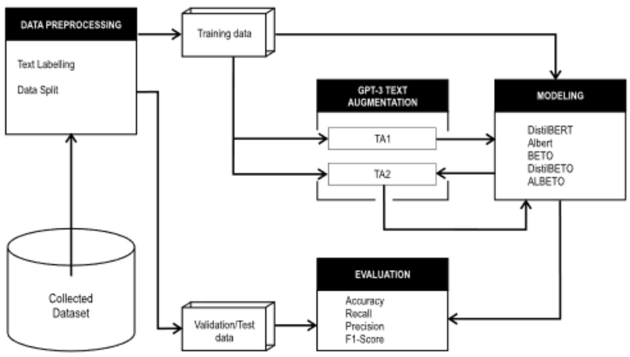

Fig. 1. Methodology

3.1. Data preprocessing

The dataset for this study has 1015 records including true and fake news in the Spanish language, they were collected from social networks such as Facebook and YouTube from March 2020 to May 2023. It includes 914 records (90.05%) for true news and 101 records (9.95%) for fake news.

3.1.1 Text labeling

The first data preprocessing activity was to label every collected record, this task was performed by the research team because information about the truth or falsehood of each piece of news was extracted from sites such as chequeado.com and factcheck.org. Fake news was labeled with 1 and true news was labeled with 0. Table 2 shows some examples of the dataset records.

Table 2. Examples of collected news

N°	Text	Label
1	en vivo se agotan pasajes para distintos puntos del país. varias personas tratan de viajar antes de que inicie la cuarentena en lima	0
2	ya me contagié de covid-19 y es improbable que me pueda volver a infectar	1
3	variante delta plus: la nueva mutación del coronavirus que podría ser más mortal y contagiosa, según alerta la oms	0
4	vacunas contra la covid-19 en perú: enfermeras de essalud se preparan para jornada de inoculación histórica a miles de peruanos y peruanas	0
5	victor zamora pide postergar reapertura de bares y discotecas: tenemos el ejemplo de europa	0
6	El nuevo coronavirus muere en el agua de mar	1
…	…	

Once the texts of the dataset were labeled, the training and test data were generated. For this, 80% (812 records) are randomly considered for training and, 20% for test (203 records).

3.2 Text augmentation

Text augmentation is a strategy that has been used quite frequently to enrich the available data, especially when there is scarcity. Some of the most common techniques are described below:

Synonym change [42]: This technique consists of replacing some words in the text with synonyms to create variations in the wording of the text.

Syntactic rules [43]: New text is generated based on certain syntactic rules.

Translation [44]: The text is translated into another language and then translated back into the original language to create variations in the text.

Text Generation: Text generation models are used to create additional text that can be added to the training data set. Thus the work [28] uses the GPT-2 generative model, and [40] uses the GPT-3 model for this task.

In this work GPT-3 model [45] is used for text augmentation, the openai library was installed and model engine "text-davinci-003" was used, Fig. 2 shows the Python function for this purpose. Two different processes are implemented for text augmentation, TA1, and TA2.

TA1 consists of the most common way to augment the text, that is, a certain number of synthetic rows are generated for each training data record.

TA2 is a selective text augmentation process, that is, a certain number of synthetic data is generated as TA1, but not all the training data is considered as in TA1, only the records that could not be learned in the training phase are considered during No TA model implementation. For this, once the No TA model is trained, training data is used to classify news and from this FPs and FNs are selected to be augmented.

Ten (10) synthetic texts were generated for every record in TA1 and TA2.

```
def getTextA(row,oakey):
    openai.api_key = oakey
    model_engine = "text-davinci-003"
    prompt="10 formas de decir '"+row+"'"
    completions = openai.Completion.create(
        engine=model_engine,
        prompt=prompt,
        max_tokens=1024,
        n=1,
        stop=None,
        temperature=0.5,
        )
    message = completions.choices[0].text
    return message
```

Fig. 2. Python function to get 10 synthetic texts with GPT-3.

Being 812 original records and 10 synthetic records per original record, the number of synthetic records for TA1 is 8120, having in total 8932 records. And, for TA2 there are 812 original records, of which 44 records were not adequately learned by the implemented models, these were augmented getting 440 synthetic records, finally for TA2 there are 1252 records. Table 2 shows 10 synthetic records for one record using GPT-3 and function code of Fig.2.

Table 2. Examples of GPT-3 synthetic texts

N°	Text	Label	Type
1	Mujer muere tras trasplante de pulmones infectados con covid-19	0	Original
2	Mujer fallece por complicaciones de trasplante de pulmones contaminados con el coronavirus.	0	Synthetic
3	Una mujer pierde la vida luego de recibir pulmones infectados con el covid-19.	0	Synthetic
4	Una mujer fallece tras recibir pulmones infectados con el virus.	0	Synthetic
5	Una mujer muere a causa de un trasplante de pulmones infectados con el covid-19.	0	Synthetic
6	Una mujer pierde la vida por complicaciones de un trasplante de pulmones con el coronavirus.	0	Synthetic
7	Una mujer se ve afectada por un trasplante de pulmones infectados con el covid-19.	0	Synthetic
8	Una mujer pierde la vida por un trasplante de pulmones infectados con el covid-19.	0	Synthetic
9	Una mujer muere como consecuencia de un trasplante de pulmones infectados con el virus.	0	Synthetic
10	Una mujer se ve afectada trágicamente por un trasplante de pulmones infectados con el covid-19.	0	Synthetic
11	Una mujer fallece luego de recibir un trasplante de pulmones infectados con el coronavirus.	0	Synthetic

3.3 Modeling

For the implementation of models, Google Colab is used including libraries such as Transformers-4.27.3

The selected models for experimentation are from huggingface.co are: DistilBERT, Albert, BETO, DistilBETO and ALBETO.

DistilBERT [8]. It is a light version of BERT that uses 60% of BERT model parameters (66 million).

Albert [46]. Similar to DistilBERT, it is a light version of BERT with approximately 12 million parameters in its base version.

BETO [10]. It is a BERT model but trained with Spanish data with 110 million parameters.

DistilBETO [11]. Similar to DistilBERT, it is a light version of the BETO model, using the distillation technique for knowledge transfer. It includes 6 layers instead of 12.

ALBETO [11]. It is a Spanish Albert model because it is trained with Spanish data, it includes different versions (tiny, base, large, xlarge, and xxlarge), in this work the base version is used.

For model building, the dataset is transformed into the shape required to fine-tune the transformers-based model. So, first, the pretrained model tokenizer is loaded, then the dataset is transformed using the loaded tokenizer. Next, padding is applied, for this DataCollatorWithPadding is used. Finally, the pre-trained model for sequence classification is loaded, and evaluation metrics, and training args are passed to compile. Table 3 shows the training arguments used for the implemented models in this work.

Table 3. Training arguments for transformers-based models

Model	**Training arguments**	**Num_train_epochs**	**Huggingface Model**
DistilBERT	Learning_rate:2e-5, weight_decay:0.01, per_device_train_batch_size=16, per_device_eval_batch_size=16, *	10	distilbert-base-uncased
AlBERT		2	albert-base-v2
BETO		2	bert-base-spanish-wwm-uncased
DistilBETO		10	distilbert-base-spanish-uncased
ALBETO		10	albert-base-10-spanish

* The same training arguments were used for all models

3.4 Evaluation

For the evaluation of models, metrics such as Accuracy, Recall, Precision, and F1-Score are considered, which are estimated through equations (1), (2), (3), and (4) respectively.

$$accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{1}$$

$$recall = \frac{(TP)}{(TP+FN)} \tag{2}$$

$$precision = \frac{(TP)}{(TP+FP)} \tag{3}$$

$$f1 - score = \frac{(2*recall*precision)}{(recall+precision)}) \tag{4}$$

Where:
TP : True Positives
TN : True Negatives
FP : False Positives
FN : False Negatives

Table 4. Results of implemented models with No TA (No text augmentation), TA1 and TA2

Model	Accuracy			Recall			Precision			F1-Score		
	No TA	TA1	TA2	No TA	TA1	TA2	No TA	TA1	TA2	No TA	TA1	TA2
DistilBERT	0.9212	0.9507	0.9064	0.6667	0.6667	0.5238	0.6087	0.8235	0.5500	0.6364	0.7368	0.5365
ALBERT	0.9064	0.9212	0.8916	0.2381	0.5714	0.2857	0.6250	0.6316	0.4615	0.3448	0.6000	0.3529
BETO	0.9360	0.9360	0.9507	0.4762	0.6667	0.6667	0.7623	0.7000	**0.8235**	0.5882	0.6829	0.7368
DistilBETO	**0.9557**	0.9458	**0.9557**	0.7619	0.6190	**0.8095**	0.8000	0.8125	0.7727	**0.7804**	0.7027	**0.7907**
ALBETO	0.9360	0.9409	0.9507	0.6190	0.5714	0.7143	0.7222	0.8000	0.7895	0.6667	0.6667	0.7500

4. Results

In this section, the results achieved by implemented models are described.

Once the models are fine-tuned with our collected data, they are evaluated with test data. Table 4 shows the results of model evaluations in terms of Accuracy, Recall, Precision and F1-Score.

According to Table 4, it can be seen that in terms of Accuracy, the proposal text augmentation technique (TA2) allows improving two of the five implemented models, the models that improve with text augmentation are BETO (1.47%) and ALBETO (1.47%). The model that does not improve its accuracy with TA2 is DistilBETO. And, the models that worsen their accuracy are Albert and DistilBERT. While using common text augmentation (TA1) improves three of the five implemented models (DistilBERT 2.95%, Albert 1.48%, and ALBETO 1.47%), but in the case of Spanish language models such as BETO, DistilBETO, and ALBETO, TA2 gets better improvements than TA1, on average 0.98%.

In terms of Recall or Sensitivity, according to Table 4, it can be seen that the proposal text augmentation TA2 allows improving four of five implemented models ALBERT (4.76%), BETO (19.05%), DistilBETO (4.76%) and ALBETO (9.53%). Just DistilBERT doesn't improve with TA2. With TA1 just ALBERT and BETO improve.

Also, according to Table 4, it can be seen that in terms of Precision, with the proposal text augmentation TA2, two of five implemented models manage to improve: ALBETO (6.73%) and BETO (6.12%), being DistilBETO, ALBERT, and DistilBERT the models that don't improve. TA1 gets to improve four of the five implemented models.

Regarding the F1-Score, according to Table 4, TA2 got to improve four of five implemented models: ALBETO (8.33%), BETO (14.86%), DistilBETO (1.03%), and ALBERT (0.81%). DistilBERT does not improve. With TA1 also three models are improved: BETO (9.47%), Albert (25.52%), and DistilBERT (10.04%).

Also, according to the obtained results, several aspects can be discussed.

Regarding accuracy, three of the five implemented models improve with TA1; however, the highest accuracy is obtained with TA2 (95.57%). Something similar occurs with the Precision metric, where the highest precision is achieved with TA2 (87.50%).

Regarding recall, the highest result is achieved with No TA for the DistilBETO model (76.19%), however, the next best results are achieved for TA2 with DistilBETO and ALBETO with 71.43%.

The f1-score is the main weakness of all models. For all cases, this is less than 80%. Neither TA1 nor TA2 achieves results above 80%. In the best of the TA2 cases, it reaches 76.92% for the DistilBETO model.

Table 5. Results of implemented models in terms of TP, FP, TN, FN using TA2 in test data.

Model	TP	FP	TN	FN
DistilBERT	11	9	173	10
ALBERT	6	7	175	15
BETO	14	2	180	7
DistilBETO	17	5	177	4
ALBETO	16	4	178	5

According to Table 5, in columns FP and FN the prediction errors of the implemented models can be seen. Here, the superiority of the Spanish-based models (BETO, DistilBETO, and ALBETO) concerning the multilingual models (DistilBERT and ALBERT) in terms of the prediction of TPs and TNs is appreciated.

On the other hand, in terms of runtime, for training, TA1 (8932 rows) is much more expensive than TA2 (2064 rows) due to the number of synthetic texts generated by each one. According to Table 6, TA1, increases training runtime in more than 10-times (>1000%) while TA2 just in around 50%. Table 6 shows the required times to train models including no text augmentation (No TA) and text augmentation (TA1 and TA2) in terms of hours, minutes and seconds.

Table 6. Required time to train models

Model	**DistilBERT**	**ALBERT**	**BETO**	**DistilBETO**	**ALBETO**
No TA	00:01:02	00:25:07	00:15:26	00:00:50	00:01:04
TA1	00:11:08	03:54:38	05:18:31	00:08:14	00:11:34
TA2	00:01:36	00:42:25	02:21:00	00:01:15	00:01:42

5. Conclusions

According to the results of this work, although TA2 does not improve all the implemented models, it is very important for models based on Spanish language such as BETO, DistilBETO, and ALBETO, where in most metrics TA2 outperforms TA1. Averaging the results of each metric in the Spanish language models, it is appreciated that on average Spanish language models improve 1.15%, 11.12%, 2.44%, and 7.50% in terms of Accuracy, Recall, Precision, and F1-Score respectively. Fig. 3 shows the superiority of TA2 over TA1, being Recall and F1-Score the most outstanding metrics.

Comparing computational cost TA1 increases training runtime by more than 1000%, while TA2 increases runtime by just around 50% regarding not using text augmentation (No TA). It represents another of the main advantages of TA2 with respect TA1.

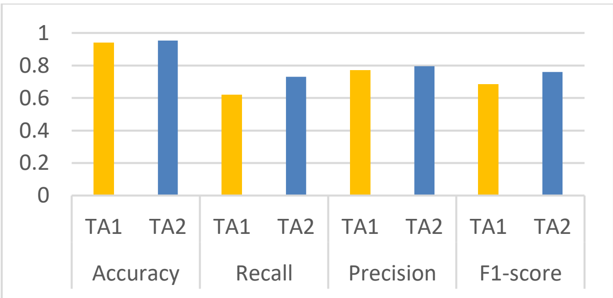

Fig. 3. Average of metrics for Spanish language models

References

[1] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manag.*, vol. 57, no. 2, 2020, doi: 10.1016/j.ipm.2019.03.004.

[2] I. Y. Agarwal, D. P. Rana, M. Shaikh, and S. Poudel, "Spatio-temporal approach for classification of COVID-19 pandemic fake news," *Soc. Netw. Anal. Min.*, vol. 12, no. 1, 2022, doi: 10.1007/s13278-022-00887-8.

[3] J. Li and M. Lei, "A Brief Survey for Fake News Detection via Deep Learning Models," in *Procedia Computer Science*, 2022, vol. 214, no. C, doi: 10.1016/j.procs.2022.11.314.

[4] WHO, "Coronavirus." https://www.who.int/es/health-topics/coronavirus (accessed Aug. 10, 2022).

[5] K. M. Douglas, "COVID-19 conspiracy theories," *Gr. Process. Intergr. Relations*, vol. 24, no. 2, 2021, doi: 10.1177/1368430220982068.

[6] M. Mohammed *et al.*, "Assessment of COVID-19 Information Overload Among the General Public," *J. Racial Ethn. Heal. Disparities*, vol. 9, no. 1, 2022, doi: 10.1007/s40615-020-00942-0.

[7] L. E. Smith *et al.*, "Side-effect expectations from COVID-19 vaccination: Findings from a nationally representative cross-sectional survey (CoVAccS – wave 2)," *J. Psychosom. Res.*, vol. 152, 2022, doi: 10.1016/j.jpsychores.2021.110679.

[8] J. Ayoub, X. J. Yang, and F. Zhou, "Combat COVID-19 infodemic using explainable natural language processing models," *Inf. Process. Manag.*, vol. 58, no. 4, 2021, doi: 10.1016/j.ipm.2021.102569.

[9] H. Luo, "Yeti at fakedes 2021: Fake news detection in spanish with albert," in *CEUR Workshop Proceedings*, 2021, vol. 2943.

[10] J. Cañete *et al.*, "Spanish Pre-Trained BERT Model and Evaluation Data," 2020.

[11] J. Cañete, S. Donoso, F. Bravo-Marquez, A. Carvallo, and V. Araujo, "ALBETO and DistilBETO: Lightweight Spanish Language Models," in *13th Conference on Language Resources and Evaluation (LREC 2022)*, 2022, pp. 4291–4298.

[12] Z. H. Kilimci, "Prediction of user loyalty in mobile applications using deep contextualized word representations," *J. Inf. Telecommun.*, vol. 6, no. 1, 2022, doi: 10.1080/24751839.2021.1981684.

[13] J. F. Low, B. C. M. Fung, F. Iqbal, and S. C. Huang, "Distinguishing between fake news and satire with transformers," *Expert Syst. Appl.*, vol. 187, 2022, doi: 10.1016/j.eswa.2021.115824.

[14] M. Abadeer, "Assessment of DistilBERT performance on Named Entity Recognition task for the detection of Protected Health Information and medical concepts," 2020, doi: 10.18653/v1/2020.clinicalnlp-1.18.

[15] A. Ul Hussna, I. I. Trisha, M. S. Karim, and M. G. R. Alam, "COVID-19 Fake News Prediction on Social Media Data," 2021, doi: 10.1109/TENSYMP52854.2021.9550957.

[16] W. H. Bangyal *et al.*, "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," *Comput. Math. Methods Med.*, vol. 2021, 2021, doi: 10.1155/2021/5514220.

[17] J. Kapusta, M. Drlik, and M. Munk, "Using of n-grams from morphological tags for fake news classification," *PeerJ Comput. Sci.*, vol. 7, 2021, doi: 10.7717/PEERJ-CS.624.

[18] T. Olaleye, A. Abayomi-Alli, K. Adesemowo, O. T. Arogundade, S. Misra, and U. Kose, "SCLAVOEM: hyper parameter optimization approach to predictive modelling of COVID-19 infodemic tweets using smote and classifier vote ensemble," *Soft Comput.*, 2022, doi: 10.1007/s00500-022-06940-0.

[19] B. Al-Ahmad, A. M. Al-Zoubi, R. A. Khurma, and I. Aljarah, "An evolutionary fake news detection method for covid-19 pandemic information," *Symmetry (Basel).*, vol. 13, no. 6, 2021, doi: 10.3390/sym13061091.

[20] S. Gonwirat, A. Choompol, and N. Wichapa, "A combined deep learning model based on the ideal distance weighting method for fake news detection," *Int. J. Data Netw. Sci.*, vol. 6, no. 2, 2022, doi: 10.5267/j.ijdns.2022.1.003.

[21] C. Raj and P. Meel, "ARCNN framework for multimodal infodemic detection," *Neural Networks*, vol. 146, 2022, doi: 10.1016/j.neunet.2021.11.006.

[22] D. Guo, Y. Kim, and A. M. Rush, "Sequence-level mixed sample data augmentation," 2020, doi: 10.18653/v1/2020.emnlp-main.447.

[23] A. V. Mosolova, V. V. Fomin, and I. Y. Bondarenko, "Text augmentation for neural networks," in *CEUR Workshop Proceedings*, 2018, vol. 2268.

[24] X. Zhang, J. Zhao, and Y. Lecun, "Character-level convolutional networks for text classification," in *Advances in Neural Information Processing Systems*, 2015, vol. 2015-January.

[25] X. Wang, Y. Sheng, H. Deng, and Z. Zhao, "Charcnn-svm for chinese text datasets sentiment classification with data augmentation," *Int. J. Innov. Comput. Inf. Control*, vol. 15, no. 1, 2019, doi: 10.24507/ijicic.15.01.227.

[26] O. Kashefi and R. Hwa, "Quantifying the Evaluation of Heuristic Methods for Textual Data Augmentation," 2020, doi: 10.18653/v1/2020.wnut-1.26.

[27] M. Wei, N. S. Harzevili, Y. Huang, J. Wang, and S. Wang, "CLEAR: Contrastive Learning for API Recommendation," in *Proceedings - International Conference on Software Engineering*, 2022, vol. 2022-May, doi: 10.1145/3510003.3510159.

[28] S. Y. Feng, V. Gangal, D. Kang, T. Mitamura, and E.

Hovy, “GenAug: Data Augmentation for Finetuning Text Generators,” 2020, doi: 10.18653/v1/2020.deelio-1.4.
[29] N. Mrkšić *et al.*, “Counter-fitting word vectors to linguistic constraints,” 2016, doi: 10.18653/v1/n16-1018.
[30] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, vol. 1.
[31] C. Olah and S. Carter, “Attention and Augmented Recurrent Neural Networks,” *Distill*, vol. 1, no. 9, 2017, doi: 10.23915/distill.00001.
[32] S. T. Aroyehun and A. Gelbukh, “Aggression Detection in Social Media: Using Deep Neural Networks, Data Augmentation, and Pseudo Labeling,” 2018.
[33] A. W. Yu *et al.*, “QaNet: Combining local convolution with global self-attention for reading comprehension,” 2018.
[34] G. Rizos, K. Hemker, and B. Schuller, “Augment to prevent: Short-text data augmentation in deep learning for hate-speech classification,” 2019, doi: 10.1145/3357384.3358040.
[35] T. Miyato, A. M. Dai, and I. Goodfellow, “Adversarial training methods for semi-supervised text classification,” 2017.
[36] E. Chang, X. Shen, D. Zhu, V. Demberg, and H. Su, “Neural data-to-text generation with LM-based text augmentation,” 2021, doi: 10.18653/v1/2021.eacl-main.64.
[37] M. Jungiewicz and A. Smywiński-Pohl, “Towards textual data augmentation for neural networks: Synonyms and maximum loss,” *Comput. Sci.*, vol. 20, no. 1, 2019, doi: 10.7494/csci.2019.20.1.3023.
[38] L. Sun, C. Xia, W. Yin, T. Liang, P. S. Yu, and L. He, “Mixup-Transformer: Dynamic Data Augmentation for NLP Tasks,” 2020, doi: 10.18653/v1/2020.coling-main.305.
[39] H. Guo, “Nonlinear mixup: Out-of-manifold data augmentation for text classification,” 2020, doi: 10.1609/aaai.v34i04.5822.
[40] K. M. Yoo, D. Park, J. Kang, S. W. Lee, and W. Park, “GPT3Mix: Leveraging Large-scale Language Models for Text Augmentation,” 2021, doi: 10.18653/v1/2021.findings-emnlp.192.
[41] C. Zhou, M. Li, and S. Yu, “Intelligent Grouping Method of Science and Technology Projects Based on Data Augmentation and SMOTE,” *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2145637.
[42] Abdurrahman and A. Purwarianti, “Effective use of augmentation degree and language model for synonym-based text augmentation on Indonesian text classification,” 2019, doi: 10.1109/ICACSIS47736.2019.8979733.
[43] J. Min, R. T. McCoy, D. Das, E. Pitler, and T. Linzen, “Syntactic Data Augmentation Increases Robustness to Inference Heuristics,” 2020, doi: 10.18653/v1/2020.acl-main.212.
[44] C. Shorten, T. M. Khoshgoftaar, and B. Furht, “Text Data Augmentation for Deep Learning,” *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00492-0.
[45] R. Dale, “GPT-3: What’s it good for?,” *Natural Language Engineering*, vol. 27, no. 1. 2021, doi: 10.1017/S1351324920000601.
[46] H. Choi, J. Kim, S. Joe, and Y. Gwon, “Evaluation of BERT and Albert sentence embedding performance on downstream NLP tasks,” 2020, doi: 10.1109/ICPR48806.2021.9412102.