



Deep Learning Model for COVID-19 Sentiment Analysis on Twitter

Salvador Contreras Hernández¹ · María Patricia Tzili Cruz¹ ·
José Martín Espínola Sánchez¹ · Angélica Pérez Tzili²

Received: 4 April 2022 / Accepted: 23 February 2023 / Published online: 13 March 2023

© The Author(s), under exclusive licence to The Japanese Society for Artificial Intelligence and Springer Nature Japan KK, part of Springer Nature 2023

Abstract

The COVID-19 pandemic impacted the mood of the people, and this was evident on social networks. These common user publications are a source of information to measure the population's opinion on social phenomena. In particular, the Twitter network represents a resource of great value due to the amount of information, the geographical distribution of the publications and the openness to dispose of them. This work presents a study on the feelings of the population in Mexico during one of the waves that produced the most contagion and deaths in this country. A mixed, semi-supervised approach was used, with a lexical-based data labeling technique to later bring these data to a pre-trained model of Transformers completely in Spanish. Two Spanish-language models were trained by adding to the Transformers neural network the adjustment for the sentiment analysis task specifically on COVID-19. In addition, ten other multilanguage Transformer models including the Spanish language were trained with the same data set and parameters to compare their performance. In addition, other classifiers with the same data set were used for training and testing, such as Support Vector Machines, Naive Bayes, Logistic Regression, and Decision Trees. These performances were compared with the exclusive model in Spanish based on Transformers, which had higher precision. Finally, this model was used, developed exclusively based on the Spanish language, with new data, to measure the sentiment about COVID-19 of the Twitter community in Mexico.

Keywords Sentiment analysis · COVID-19 Twitter · BERT

María Patricia Tzili Cruz, José Martín Espínola Sánchez and Angélica Pérez Tzili have contributed equally to this work.

✉ Salvador Contreras Hernández
scontreras@upvm.edu.mx

Extended author information available on the last page of the article

1 Introduction

Sentiment analysis is an important topic to know what a group of people thinks about a special topic, a product, service, or person, among others, to obtain data that allow decision-making by those involved, companies, people or governments, or for the interpretation of a social or commercial phenomenon, among others.

The issue of COVID-19 became the subject of conversations on social networks, from the beginning of 2020, of ordinary people to experts in various disciplines related to this topic of public health. The point of interest in this research covers all the people who issue comments or opinions on the social network Twitter, to know the feeling of a community. We carried out the study in Mexico.

Knowing what is being said in Mexico on this issue is of great importance in determining what people perceive of the management of the pandemic by local authorities, the federal government, and particularly the health sector, as well as the perception of the severity of the pandemic and the measures taken to prevent the disease, among other things. Information on the increase in the number of cases of COVID-19, hospitalizations, the number of deaths and the distribution of vaccines, among others, is considered of fundamental importance for a well-informed population. According to the posts, users react and write their own comments and interests. It is in these comments that they show their moods, which are very valuable for research about the feeling of a community, on specific topics.

The sentiment analysis in Spanish with networks of Transformers can be carried out through pre-trained multilanguage models, or else exclusively in Spanish. This paper explores the performance of multilanguage and Spanish-only BERT (Bidirectional Encoder Representations from Transformers) models for sentiment analysis on Twitter posts that address the topic of COVID-19.

The research questions for this work are the following:

How to improve the accuracy of the evaluations on expressions or publications of short texts, in Spanish, in sentiment analysis, through machine learning?

Can the classification of texts in Spanish be improved with a corpus exclusively in Spanish?

How accurate is the proposal of this work, with a network of Transformers trained in Spanish, compared to other multilanguage models and other machine learning algorithms?

In this work, the study of the Transformer models based on Spanish and multilanguage is proposed as a novelty to determine, according to the accuracy data, which is the best approach for sentiment analysis on a specific topic, in our case COVID-19. For this, the proposal uses a semi-supervised approach in the labeling of the data for training that allows the model the best interpretation of the semantics of the input texts. As the texts to be analyzed are in Spanish, it was considered important to train a model exclusively in Spanish, preprocessing the input data, an automatic lexical-based labeling, and a subsequent manual adjustment.

This process is intended to provide a better construction of sentences in the Spanish language that serves as input for the learning of the model. The use of a pre-trained model exclusively in Spanish has the objective of training it to recognize the structure of the language, then it is complemented by adding a specialized layer in the analysis of sentiments with the characteristics assigned in the hybrid labeling.

Two pre-trained BERT models exclusively in Spanish and ten multi-language pre-trained models were used, which were retrained to provide them with the capacity for sentiment analysis on the topic of COVID-19. A similar configuration is used in all of them, that is, the parameters are the same for all cases. In all models, a layer was added to the neural network, which is fully connected to the last layer of BERT, linked to a *Softmax* activation function. The added linear layer has 768 nodes and an output of two nodes, one for the classification and one for the text processing in the specific training for sentiment analysis.

The contributions of this work are the following:

Comparison of the performance of multilanguage BERT pre-trained models that include the Spanish language and models exclusively in Spanish, with the same parameters and data set that address the topic of COVID-19.

Comparison of the performance of the BERT models in Spanish with those of the SVM, Decision Trees, Logistic Regression, and Naive Bayes models, with the same data set and parameters, on the subject of COVID-19.

Application of a BERT model trained in Spanish for sentiment analysis on the subject of COVID-19 in Mexico.

The document is organized as follows: Sect. 2 includes the related works, where studies developed with Transformer networks and others with Machine Learning are addressed, Sect. 3 explains the methodology used, Sect. 4 shows the results, Sect. 5 includes the discussion on the development of the study and the results and Sect. 6 presents the conclusions.

2 Related Works

The use of social networks, in particular Twitter, to know the opinions or feelings on a specific topic, is continuously presented in the scientific literature under various approaches, all of them valuable. These approaches are, for example, the detection of the feeling of users in relation to a topic, valuing the phrases as positive, negative and neutral, or studies that focus on the algorithms to make this valuation process more efficient, comparing the most usual ones and measuring the effectiveness of the proposed model. Other works focus on the metrics of texts published by users, as in [26], where the identification of topics, recognition of entities, and grouping of sentences are obtained. This classification of texts according to certain topics or categories, is very useful and is applied in systems that manage large amounts of documents, for example, publishers who receive many proposals for articles to publish, or universities that manage institutional repositories, to name only two cases.

In the exploration of the topic of sentiment analysis, we can find works that group some research or lines of research that have been presented over the years, as shown in [25]. These authors do a study they call tertiary research, in which they show the orientation of sentiment analysis. Here are the main alternatives that are currently available to carry out this process, including machine learning. In the study, the authors report that they found that Long–Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) are the most widely used algorithms for sentiment analysis.

In the case of the social network Twitter, it decided to open its database for research, with some restrictions. Texts or tweets are used to measure the opinion or feelings of users on various topics, from commercial products or services to the mood of a community at some point in time. The information obtained on Twitter is valuable, even though not everyone in a particular country uses this social network and many do not publish constantly. The distribution of Twitter users in a country-like Mexico is distributed throughout its territory, which allows for considering valid samples under certain parameters that consider the purpose of each research.

The publication of works on the exploration of texts on Twitter and other social networks to discover the opinion of users on diverse topics are many; however, the problems to be solved from the computational point of view are still great, despite the advances in Natural Language Processing (NLP). Examples of improvements that are required of algorithms are classifiers, topic or aspect detection [29], hate language detection, classification with multiple categories, and finer aspects, such as recognition of irony or sarcasm. Papers have been published on *hate speech* as in [27, 28], where the performance of machine learning models to identify this type of language in social networks is explored, including Bidirectional Encoder Representations from Transformers. Within these works is [38], where an analysis of Twitter texts on the discussion of the Citizenship Amendment Bill in India is shown, in which the Support Vector Machine algorithm (SVM) is used to perform the positive, negative and neutral classification.

Some of the most commonly used classification methods for sentiment analysis, with machine learning, are Support Vector Machines, Naive Bayes, N-Gram, Maximum Entropy, K-Nearest Neighbors K-NN and weighted KNN, Feature-Based and Rule-Based Multilingual Analysis [8], Decision Trees, Logistic Regression and the word embedding-based ones. In addition, machine-learning methods include supervised, semi-supervised and unsupervised learning modalities. By the number of layers of the neural network used, we also have deep learning, which is distinguished from simple machine learning using a greater number of them, which allows more precision in calculations. As an example of this [11] uses deep learning with the SVM method to determine the polarity of opinions in the field of drugs within the medical field. This work focuses on predicting the satisfaction of patients with respect to the drugs they consume, through an opinion-mining model based on neural networks.

The work in [29] shows a survey, which analyzes the main methods for sentiment analysis based on aspects, focusing on the mapping between these, their interactions, dependencies, and the context and semantic relationships between the elements

analyzed. These methods are among the most used for creating machine-learning models in opinion analysis.

Another interesting work on affective coding in Twitter texts is the one that appears in [12], which focuses on content that reflects the health status of the people who broadcast the tweets. According to the authors of this study, important information can be obtained on Twitter about the health of users, such as descriptions of medical examinations, surgical interventions, or treatments. The importance of deep learning models highlights over those based on LSTM, both developed for their research. In [16] a similar work is carried out, but on two specific corpora for medical information and doctors, the first for a medication review and the latter for patient opinions about health professionals, both in the Spanish language.

One of the problems of supervised learning tasks of sentiment analysis is the labeling of the texts with which you will train the model, as it requires human intervention in the review or reading of each text and classifying it positive and negative, or, by adding the type neutral, depending on the objective to be achieved. In semi-supervised learning, automatic annotators are used that can do this work instead of humans, or with very little intervention by humans [5, 22, 45]. The hybrid approach combines lexical-based analysis with machine learning. In this type of solution, the lexicon-based technique is used for the scoring of the texts, and after these are used for the training of the machine-learning model [2].

A work related to the analysis of feelings in publications about COVID-19 in social networks is shown in [15], where Recurrent Neural Networks (RNN) is used for the classification of texts in positive, negative and neutral. The authors mention that the trained model showed an accuracy of 81.15%. Researchers use the Bayesian Network method for classifying tweets in the sentiment analysis task. They compare this method with Support Vector Machines and Random Forests, and the result is that Bayesian Networks achieve superior accuracy. These have been widely used for the classification of opinions of Twitter texts, in [32] two cases are reported, where they applied it after the earthquake in Chile in 2010 and the independence referendum of Catalonia in 2017. In the first case, it can be inferred that the mood of the Chilean population was not very high, due to the devastation caused by this natural phenomenon. Even so, it is important to measure the feelings of the inhabitants of the community in which these disasters occur.

In [46] there is further research on the impact of the COVID-19 pandemic in the city of New South Wales, in Australia. This work collects tweets from this area and identifies those that address the COVID-19 issue and subtopics that contain, for example, the spread of the virus or confinement. The article highlights the importance of focusing the study on a specific small area, because you can better know the mood of its inhabitants. However, Twitter communities in any country are limited communities, where the validity of the information is accepted just for those groups of users.

Two methods that have been widely used for text classification are Support Vector Machines and Decision Trees in sentiment analysis as shown in [1, 3, 13, 18, 30, 36, 41]. One of the reasons why they are frequently used is because of the speed of training and because they provide acceptable accuracy. In addition, classifiers based on Bayesian Networks are frequently used in the classification task for sentiment

analysis, as we can see in [4, 13, 14, 19, 42]. On sentiment analysis, there are also additional contributions as the extraction of features, which are mentioned in [9], where, in addition to classifying the opinions gathered for the research, as positive, negative, or neutral, the semantic orientation that gave rise to the sentiments expressed in the texts analyzed is obtained.

Sentiment analysis is traditionally performed with text classifiers of types, such as those described above, and some others. The scientific community for the results accepts these methods in their accuracy. The algorithm of this work is compared with some of these methods using the same data set for training and testing Tweet text classification models. Works on machine learning for creating text classification models use the types described above, but more recently, complex neural networks are applied. This type of network allows deep learning and has had much higher results than its predecessors have. Some works of this class are described in [37], where a deep learning model is used for the classification of documents.

A recent technology that came to revolutionize natural language processing is BERT. It uses sentence processing in contexts from left to right and vice versa, to improve text comprehension. It has the characteristic of working with pre-trained models, because it configures the output of these, in such a way that it can be reconfigured to implement additional layers for a specific task, this process is also known as fine-tuning.

Works on Bidirectional Encoder Representations from Transformers have been published since the emergence of this technology recently, with efficient results for written language management. Models can be generated from other pre-trained for tasks, such as finding phrases in large texts, also called semantic similarity, or, in question–answer, or sentiment analysis, word prediction, and translation, to mention a few [24, 34, 37, 39, 40, 43, 44]. The transformers technique has been applied in the detection of false news [17, 20, 31], where it is compared with traditional methods of classification, obtaining results with greater accuracy. Fake news detection is a feature that can be performed with supervised machine learning models, which are trained with a large set of fake news and authentic news, making clear the difference between both of them.

The text [35] addresses the development of a deep learning model as a social media sentiment classifier for COVID-19, using BERT. The authors describe the importance of obtaining social information about the current pandemic and the interpretation of results. It has an accuracy of 93.89%. On the other hand, [23] proposes a mechanism to obtain the variation of the intensity of emotions over time. It uses the technique of self-attention to extract the dependence between the emotions considered in the text.

For text classification, BERT has also been used in combination with CNN. The learning process occurs with the interaction of the embedding of words and graphics through the process of self-attention. In [33] BERT is used in combination with a word graph for keyword extraction in hate language detection. This approach uses BERT's multi-head attention process and combines it with the word graph. This makes the model able to detect the relationships through the mentioned graph.

The aspect-based sentiment analysis can also be aided by the power of BERT models, for the identification of categories or aspects and their opinion about them.

As shown in [21], a work was developed that uses a hybrid approach for the annotation of texts, in which it employs a small group of seed words, for each class or aspect and sentiment. These tagged texts are taken for BERT deep network training specialized for aspect detection and sentiment analysis. The pre-trained Hugging Face machine learning models for BERT were trained with millions of data and allowed adding the layers in the neural network to do what is of interest to us in the resulting model. This represents the fine adjustment to the structure and operation of the pre-trained models. For the case of our research, we used a pre-trained model in Spanish, so that it knows the form of writing in this language. The fine-tuning consisted of training it to distinguish the polarity of Twitter texts, in Spanish, according to the sentiments detected.

3 Methodology

Each language has its particularities, in the case of English or Spanish, for example, text normalization involves things, such as the conversion of uppercase to lowercase, removal of unimportant words, removal of special characters, *emoticons*, *hashtags* and others. In the normalization of text in Spanish, there are accents, which are sometimes eliminated, but these have importance in the semantics of the sentence. Training with a corpus in Spanish implies providing capacity to the model, on how to write in this language. The bidirectional processing that is carried out with Transformer networks allows good results in Spanish, which is why this work is based on a corpus in Spanish with its natural syntax.

In some cases, when machine learning models or Natural Language Processing applications are developed for Spanish or other languages, first a translation into English is made, then resources such as SentiWordNet are used to carry out the sentiment analysis and finally it is translated to the original language.

This causes the accuracy to decrease, because there are some important parts of the text that can be omitted in the translation, or some well-formed sentences could be reduced to fragments [6]. In some cases, the problem of the structure of the language also depends on the objective, for example, in the determination of specific metric patterns on the text, as in poetry, where the rules of the language are implicit, as well as the poetic inventions of the author [7]. Taking these characteristics into account, in this work a data set was prepared for training with the texts published on Twitter, in Spanish, specifically for the topic of COVID-19 in Mexico. Two pre-trained BERT models exclusively in Spanish and ten multilanguage models containing Spanish were used.

3.1 Data Collection

We collected Tweets through an application created in Python, with the Twint library. The programming environment was Google Colaboratory with a GPU. We recovered 20,000 tweets, in Spanish, that mention some of the aspects related to COVID-19. The concepts for the search were COVID-19, COVID, coronavirus,

SARS-CoV-2 and *pandemia* (pandemic). The final file for training and testing was made up of 10,619 records after a balancing process that is explained later. The downloaded tweet logs were stored in a CSV file.

We used the tweets obtained to train the machine-learning model. The tweets collected are from the months of August, September and October of 2021, when the pandemic affected strongly with the third wave or rebound in Mexico. At this time, the Delta variant of the SARS-CoV-2 virus was predominating in practically all the Mexican territory, displacing the other variants, such as the Alpha, Beta and Gamma that circulated in previous months.

3.2 Pre-data Processing

In this phase, the data are cleaned of characters or words that are inconsequential for the task of sentiment analysis. We eliminated *emojicons*, *hashtags*, URLs, special characters, and words with @. We also deleted records with blank fields. We obtained the data on different dates, by blocks, then passed it to Pandas Dataframes (Python) and joined to leave a single file with all the collected tweets. A case conversion was performed for the pre-trained models that required it. All pre-processing, including cleaning, automatic labeling, and balancing of the data set, was done with Python routines, in the Google Colaboratory programming environment.

Labeling process. The data for training and testing were labeled with a lexicon-based software for sentiment analysis, named *Spanish Sentiment Analysis*, which is a library for Python, and the result was subsequently manually adjusted. The texts are labeled with a value between 0 and 1 that measures the intensity of the feeling according to the words that are included in the library's dictionary. These labeling and conversion procedures to binary values were done with Python routines, a column was added to the tweets Dataframe to store the intensity of the sentiment with a lambda function that calls the routine to measure sentiments, with a score of 0–1, and then, all data were exported to a CSV file.

Tweets are labeled with values between 0 and 1, negative for those with values close to 0, and positive for values close to 1. Then, the tweets are automatically re-labeled with 0 or 1 for negative or positive sentiment, respectively. For values from 0.5 to 1, they are labeled with a 1, and for values from 0 to 0.49, they are labeled as negative.

After carrying out the automatic labeling process, it was observed that the number of texts classified as negative was greater, so the data file was balanced to leave 50% positive tweets and 50% negative ones. In addition, a manual adjustment was made, where 210 tweets were modified, because their meaning was contrary or confusing to the classification obtained automatically and finally, the file was balanced again, obtaining a data set of 10,619 records.

Figure 1 shows the dispersion in the 0–1 range of the tweet valuation for the training and test data set. In the five groups of 1000 tweets taken as a sample, a trend

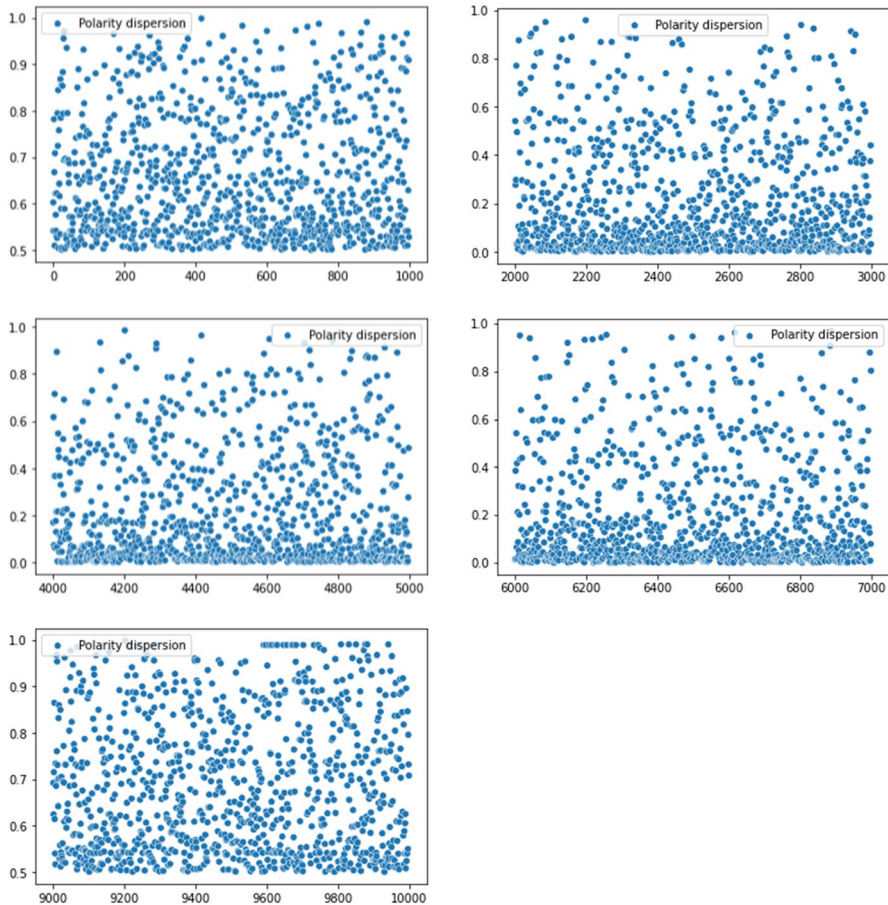


Fig. 1 Dispersion of texts valued as positive and negative

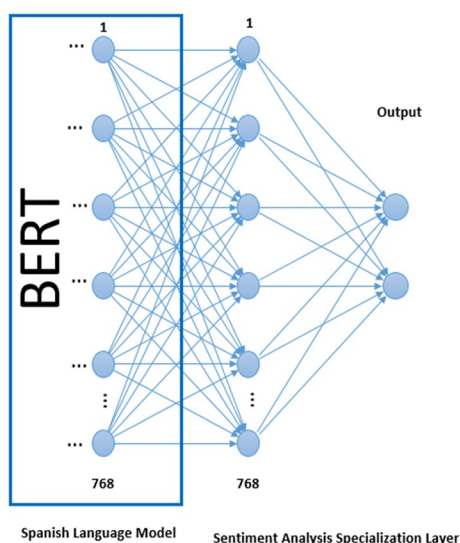
toward negative comments is evident. It is worth mentioning that the data set corresponds to the months of August, September, and October 2021.

The same data pre-processing was applied to the tweets retrieved in December 2021 that were used in the sentiment measurement of the Twitter community in Mexico. These data are the input of the Spanish-language model with more precision, trained in this work.

3.3 Model Training

We used the pre-trained models *dccuchile/bert-base-spanish-wwm-cased* and *dccuchile/bert-base-spanish-wwm-uncased* Hugging Face to take as a basis the Spanish language. We added a layer to the network of the pre-trained model for specific training in sentiment analysis about COVID-19, review Fig. 2. The added layer allowed the model to handle the texts referring to the topic of interest for

Fig. 2 Adjustment of the BERT network for sentiment analysis in Spanish on COVID-19



this project. The tweets that talk about the pandemic are very varied and address it from different angles, such as the management of the crisis by the government, prevention measures, number of deaths, cases of contagion, vaccines, and others. The model treats the information correctly, with high accuracy, 97% and 81% (train and test), which means that it can recognize the sentiment associated with a tweet, about COVID-19. There are also cases, where the text is written in a confusing manner, which the model considers a negative sentiment. We designed the project to rate tweets with two types of sentiment; positive or negative, so when a confusing text is analyzed in the tweet labeling process, it is rated with a value less than 0.5.

The BERT data set was configured with entries of 70 words for each tweet, which is more than enough for the 280 characters of each tweet. For the embedding of words in BERT, in case the tweet has fewer words than the configured maximum, the word vector is completed with a special token called [PAD].

Ten multilingual BERT models and two exclusively based on Spanish were trained. Models based on the Spanish language in general have slightly higher precision than others that use pre-trained Multilanguage models. BERT performance in languages other than English decreases somewhat with the use of pre-trained Multilanguage models as shown in [10]. The optimal accuracy in a Spanish-language BERT model is obtained when pre-trained models are used exclusively in Spanish and adding one or more layers to the neural network process the input data set also in Spanish. For this project, we experimented by adding two and three layers to the neural network in the setting for specialization in sentiment analysis on COVID-19. A *Softmax* activation function was also added, to try to improve the adjustment of the weights in the training, but it was observed that this did not contribute to an improvement, because the precision obtained was the same. Thus, only one layer and the function *Softmax* were added.

For all cases, multilingual and Spanish-language models, pre-trained models available on the *Hugging Face* website were used. This training was carried out with the same data set and parameters as in our Spanish model. Parameters include the percentage of data for training and testing, 60% and 40%, respectively, the number of training epochs is 3, the maximum token length is 70, a dropout of 40%, a Batch Size of 16, an optimizer Adam and a learning rate of $2e-5$. The development was done in Python 3, the PyTorch, and Transformers libraries for BERT programming, and the training was done on a GPU in the Google Colaboratory programming environment.

Table 1 shows the values obtained with the three precision measures, Precision, *F1* and Recall that were applied to the BERT models trained with the same set of data and parameters. The pre-trained models exclusively in Spanish are the first two in the table, they yielded a higher precision than most of the multilanguage models, and in some cases the difference is wide. The third model that appears in the table, *nlptown/bert-base-multilingual-uncased-sentiment*, is multilanguage with which an acceptable precision of 0.96 was obtained for training and 0.81 in tests, which is considered similar to that obtained with the best model in Spanish. Other multilanguage models, although they had a lower precision, are also acceptable, since they are in values of 0.79 and 0.80 in the tests.

To measure the precision of the models, the Recall, *F1* and Precision metrics were used. True Positives are the number of texts that the model correctly classified as positive. False Positives are texts wrongly classified as positive. True Negatives is the number of texts correctly classified as negative, while False Negatives represent those classified as negative when they are actually positive. These concepts are used to define the following formulas:

Table 1 Evaluation of training models

Pre-trained BERT model	Precision Training/tests	<i>F1</i> Training/tests	Recall Training/tests
dccuchile/bert-base-spanish-wwm-uncased	0.97/0.81	0.97/0.81	0.97/0.81
dccuchile/bert-base-spanish-wwm-cased	0.96/0.81	0.96/0.81	0.96/0.81
nlptown/bert-base-multilingual-uncased-sentiment	0.96/0.81	0.96/0.81	0.96/0.81
bert-base-multilingual-cased	0.26/0.26	0.50/0.50	0.34/0.34
distilbert-base-multilingual-cased	0.26/0.26	0.50/0.50	0.34/0.34
Davlan/bert-base-multilingual-cased-ner-hrl	0.87/0.77	0.87/0.77	0.87/0.77
Babelscape/wikineural-multilingual-ner	0.95/0.80	0.95/0.79	0.95/0.79
ambersoad/bert-multilingual-passage-reranking-msmarco	0.90/0.79	0.91/0.79	0.91/0.79
Davlan/distilbert-base-multilingual-cased-ner-hrl	0.26/0.26	0.50/0.50	0.34/0.34
nlpodyssey/bert-multilingual-uncased-geo-countries-headlines	0.96/0.80	0.96/0.80	0.96/0.80
Recognai/distilbert-base-es-multilingual-cased	0.26/0.26	0.50/0.50	0.34/0.34
gunghio/distilbert-base-multilingual-cased-fine-tuned-conll2003-ner	0.26/0.26	0.50/0.50	0.34/0.34

The lines that are in bold are the pre-trained models exclusively in Spanish

Table 2 Model training accuracy

	Precision	Recall	f1-score	Support
0	0.98	0.96	0.97	3052
1	0.96	0.98	0.97	3319
Accuracy			0.97	6371
Macro avg	0.97	0.97	0.97	6371
Weighted avg	0.97	0.97	0.97	6371

Table 3 Model accuracy with test data

	Precision	Recall	f1-score	Support
0	0.81	0.79	0.80	2037
1	0.81	0.83	0.82	2211
Accuracy			0.81	4248
Macro avg	0.81	0.81	0.81	4248
Weighted avg	0.81	0.81	0.81	4248

$$\text{recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

Tables 2 and 3 show the details of the precision that was obtained with the BERT model in Spanish, *dccuchile/bert-base-spanish-wwm-uncased*.

Figure 3 shows the evolution of the model during training. It can be observed that as this training progresses, the accuracy increases until you reach the reported value of 0.97. The progress in the accuracy of the model, considering the test data, was presented to a lesser, but an acceptable measure of 0.81. Table 4 shows the results.

Experiments were carried out varying the epochs from 3 to 8. It was observed that the precision in the training no longer increased considerably, and it was decided to end the process as soon as it yielded a difference of 0.15 between the training and the tests. The second epoch produced a precision of 0.7975 for the tests, which does not vary much from the value obtained in the third, 0.8109. It was decided to take three epochs, because it represents an acceptable average number for all BERT models, since they were compared with each other, considering equal parameters for all.

Training and testing were done with 10,619 tweets retrieved and pre-processed, 60% for training and 40% for testing. Tweets are from August, September, and October 2021. The model finished with an accuracy of 0.9728 for the training data set and 0.8109 for the test data set. This measurement is greater than those obtained with other

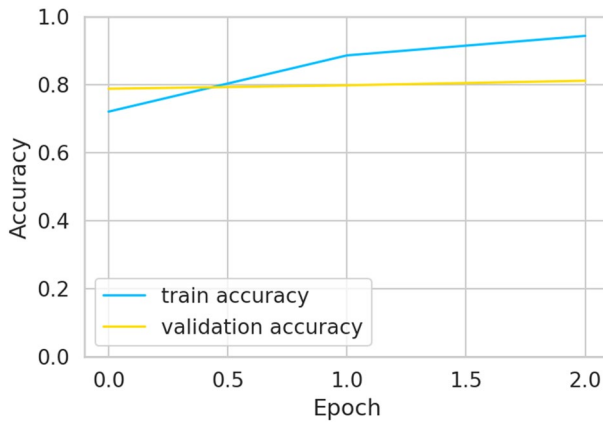


Fig. 3 Model training and testing accuracy

Table 4 Precision by Epochs for training and testing model Spanish-uncased

	Epoch 1	Epoch 2	Epoch 3
Training	0.7199	0.8854	0.9728
Test	0.7874	0.7975	0.8109

Table 5 Precision values, $F1$ -score and training recall

	BERT	SVM	Logistic regression	Decision trees	Naive Bayes
Accuracy	0.97	0.77	0.77	0.76	0.92
$F1$ -score	0.97	0.52	0.52	0.52	0.92
Recall	0.97	0.39	0.38	0.38	0.92

Table 6 Accuracy values, $F1$ -score, and test recall

	BERT	SVM	Logistic regression	Decision trees	Naive Bayes
Accuracy	0.81	0.77	0.77	0.76	0.77
$F1$ -score	0.81	0.52	0.52	0.52	0.77
Recall	0.81	0.38	0.38	0.38	0.77

classification methods, such as Decision Trees, Support-Vector Machines, Naive Bayes, and Logistic Regression. Tables 5 and 6 show training and test results of the accuracy obtained with the same data set in different classification models. These models were trained with the same data set and preprocessing described above. In the programming for the training of these models, the Sklearn library for Python was used, in the Google

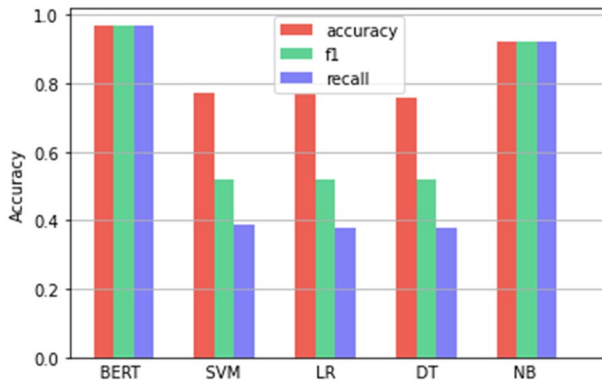


Fig. 4 Accuracy of models in training

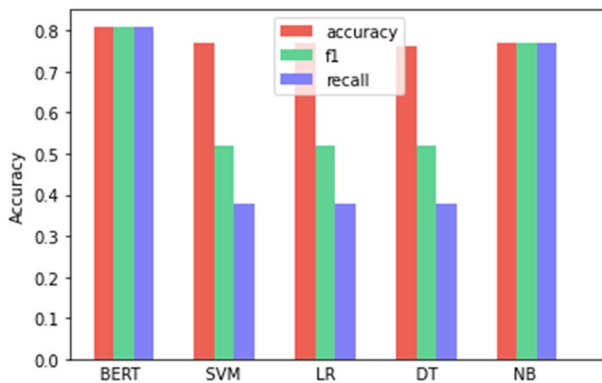


Fig. 5 Accuracy of models in test

Colaboratory programming environment. The BERT model included in the tables for comparison with other models is the one with the best precision in Spanish.

Figures 4 and 5 graphically show the results of training and testing of the BERT, Support Vector Machines (SVM), Logistic Regression (LR), Decision Trees (DT) and Naive Bayes (NB) models. Here, it can be observed that the BERT model in Spanish obtained higher precision in both cases, training, and tests. The data returned by the Naive Bayes model are also acceptable. For the BERT model in Spanish-uncased and Naive Bayes, the three precision measures are congruent, the other models show a variation between Precision, Recall and F1.

4 Results

We applied the machine-learning model BERT developed here, based on *dccuchile/bert-base-spanish-wwm-uncased*, to analyze new tweets between the dates of December 1 and 31, 2021, with the results shown below, 33,776 tweets

were evaluated, and all were submitted to the learning model for the sentiment classification. The results showed 13,452 positive and 20,324 negative tweets, representing 39.8% and 60.2% of the total tweets, respectively. The nature of the topic investigated is complicated for Twitter users, this can be perceived in the fear they demonstrate for the pandemic, since the months of the beginning of this phenomenon; however, in December 2021, the Omicron variant was entering Mexico, which caused uncertainty in people about this public health event. This could be the explanation for why most tweets are negative on this topic. However, it is interpreted as a reasonable balance between texts labeled as positive and negative.

Figure 6a shows the score of tweets with the model developed for classifying texts in Spanish. The orange line is that of the negative tweets, so we can notice that the number of tweets labeled as negative is higher. Among the factors that predominated for the negative scores, were traits of anxiety, depression, and concern about the pandemic situation, which seemed to get longer and longer. The positive ones were characterized by comments about the end of the pandemic, economic recovery, and the possibility of going out again and resuming activities as they were held before the onset of this phenomenon.

Figure 7 shows the positive and negative scores per day, from 1 to 31 December 2021. The first shows the days from 1 to 15, while the second, the days from 16 to 31. Here, it is clear that texts classified as negative are predominant.

Below are the broken-down quantitative results produced by the BERT model selected for application in the sentiment analysis in December 2021. The model classified them with the correct precision, according to that shown in the training. These results are used to reach a conclusion about the sentiment shown by the Twitter community in Mexico, on issues related to COVID-19.

The conversations in the collected data were measured to find out what the interactions between Twitter users about COVID-19 topics are like. This also includes measuring the sentiment of the reaction in *replies*, *retweets* and *likes*. This measurement is important, because a tweet can be positive or negative, but it can provoke a series of reactions, in favor or against, through replies and *retweets*.

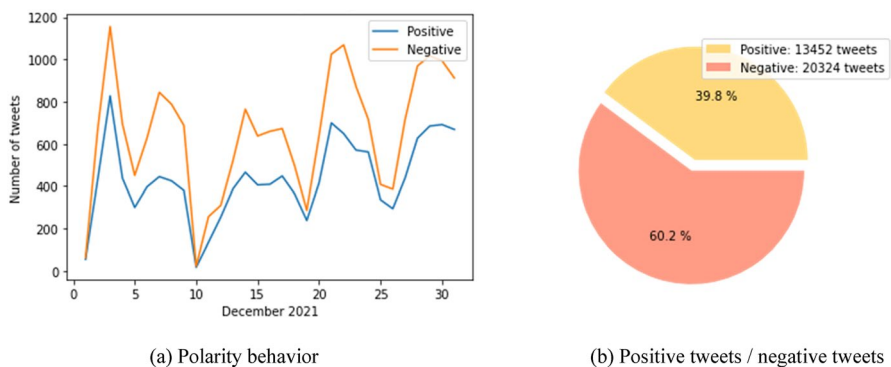


Fig. 6 COVID-19 tweet polarity

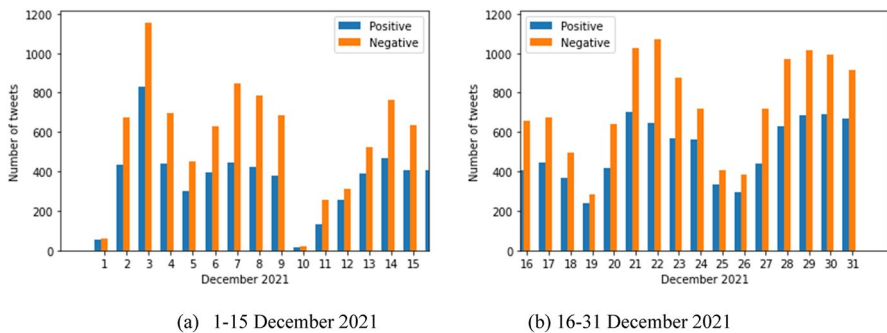


Fig. 7 Tweet score by day, 1–31 December 2021

The graph in Fig. 8 shows the cumulative *likes* in tweet ranges that contain multiples of 1000 *likes*. The ranges are made up of tweets that contain between 1–1000 *likes*, 1001–2000, and so on. It can be seen that the first ranges are the ones that accumulate the most *likes*. The total number of *likes* that show a positive sentiment is 303,268 and 382,939 negative ones, in a total number of tweets of 33,776. It was graphed in ranges of tweets due to the enormous amount of data that was evaluated.

It is considered that the *likes* in a publication ratify the idea that is being expressed, that is, the user agrees with the content and sentiment that the text expresses. The negative *likes* exceeded the positive ones in the sample for the month of December 2021, the total number of positives is 303,268 and 382,939 negatives.

The retweets were also evaluated according to the polarity assigned by the model to the Twitter texts. The retweets can be commented on, that is, assign a text to them, and they can represent a similar polarity to the original tweet, or an opposite polarity. Graph 9 shows the number of positive and negative *retweets* reported by the model from the input data set. Again, as in the previous case, of the *likes*, the negative ones are greater than the positive ones, with a total of 135,884 and 78,544, respectively. In Fig. 9a, you can see the distribution of positive and negative *retweets* in ranges of tweets that contain from 1 to 1000 *retweets*. In the range from 0.0 to 1.0,

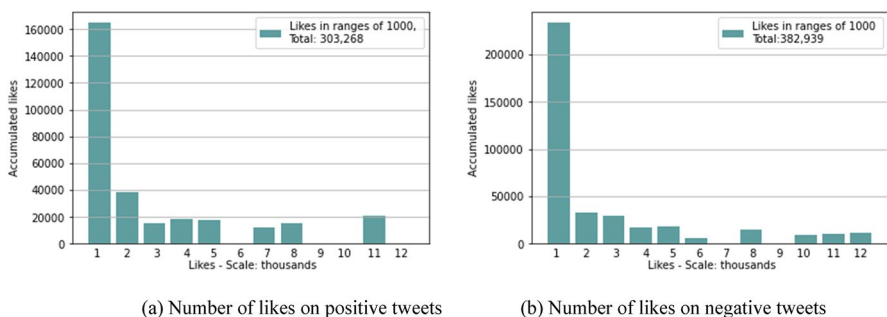


Fig. 8 Positive y negative likes

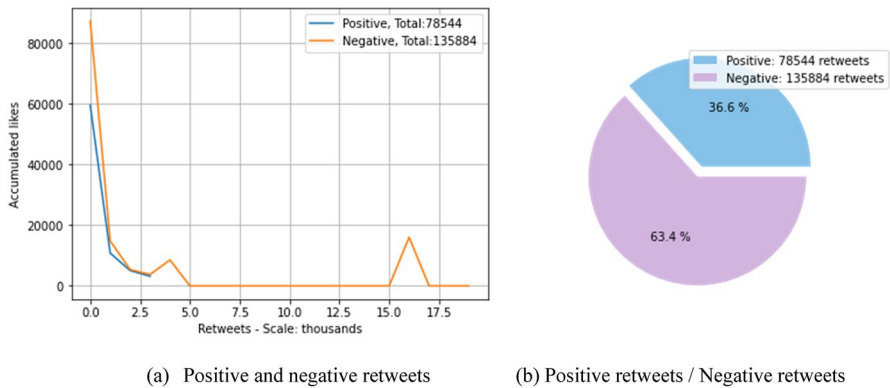


Fig. 9 Retweet polarity

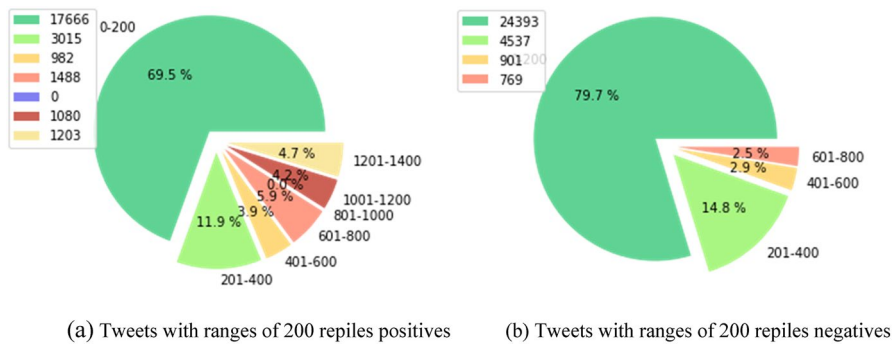


Fig. 10 Polarity in replies

appears the accumulated number of tweets that have between 1 and 1000 *retweets*, in the range from 1.0 to 2.0 are those that have from 1001 to 2000 *retweets*, and so on with the rest. In the graph, it can be seen that at all times the negative *retweets* surpassed the positive ones.

Replies are another interesting indicator that allows you to measure the impact of a publication, through comments, and can be positive or negative, unlike *likes*. The graphs in Fig. 10 show groups or ranges of replicas, positive and negative, respectively, in the total set of *tweets*. For example, in Fig. 10a, the *tweets* that contain from 1 to 200 positive replies are grouped in the first block and together they represent 69.5% of the total replies and so with the other ranges. The total of positive replies is 25,434 and the negative is 30,600 (Fig. 11). The model obtained several negative replies greater than the positive ones. The first blocks in both Fig. 10a, b indicate that most replies were concentrated there, 17,666 and 24,393, respectively. The block marked with 0 in Fig. 10a indicates that the range of 800–1000 replicas is empty.

Fig. 11 Positive replies/negative replies

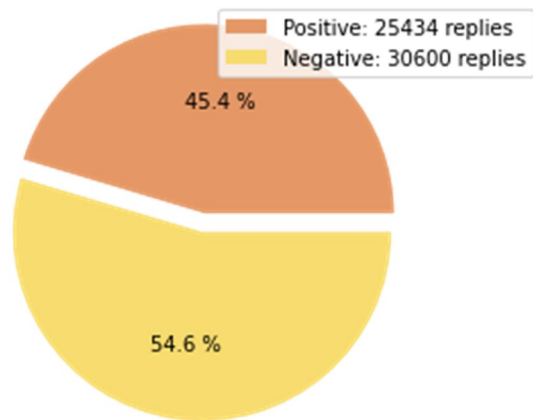


Figure 12 shows the comparison between positive and negative replies now in ranges of 1000. In the first range, where tweets containing between 1 and 1000 replies are considered, most of the replies are concentrated there. For the positive and negative replicas, it was found that the accumulated ones are in the first block. This means that users reply to tweets that do not contain many replicas, according to the data obtained. In addition, it can be noticed that the positive replies are more distributed throughout the graph, which means that there are more tweets in those higher ranges but with few responses. In Fig. 12, you can see the comparison of the behavior of the positive and negative replicas. In the range of tweets containing between 2001 and 3000 replies, the positive ones surpass the negative ones, and then in the ranges from 3001 to 6000, there are only tweets with replies classified as positive, although in low numbers.

To make a comparison of the behavior of these three indicators, *likes*, *retweets* and *replies*, the following graphs in Fig. 13 are shown, where the distribution of each one can be seen in blocks of 2000. As in the previous cases, the blocks are

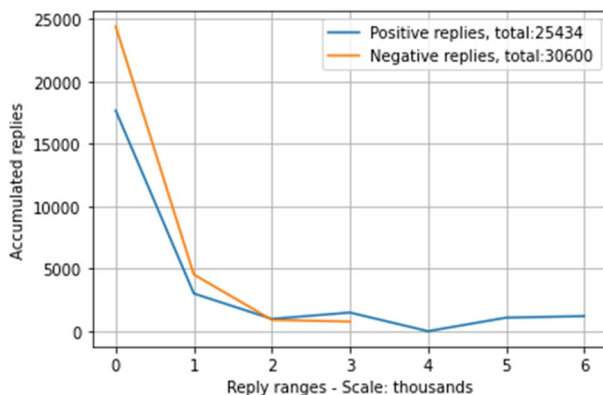


Fig. 12 Number of replies positive and negative in tweets with ranges of 1000

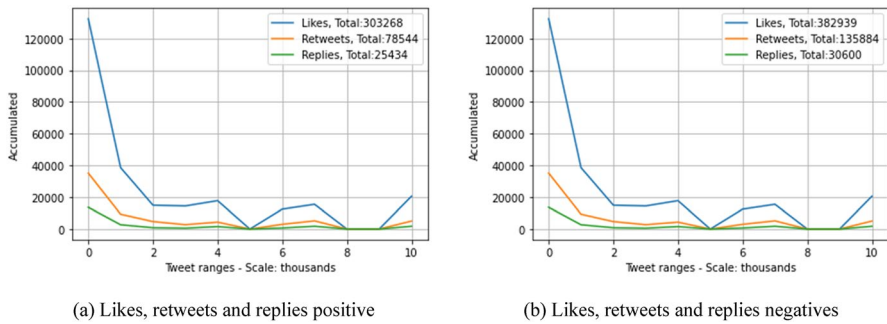


Fig. 13 Comparison of likes, retweets and replies

defined by the tweets that have a few *likes*, *retweets*, and *replies* between 1–2000, 2001–4000 and so on. The largest number of these three indicators is between the range of 0–2000, that is, most of the tweets analyzed contain between 1 and 2000 *likes*, *retweets*, and *replies*. The totals can be seen in graph 13 and the number of tweets that the model classified with negative sentiments is higher than the positive ones.

5 Discussion

The Deep Learning model trained exclusively in Spanish based on the pre-trained model *dccuchile/bert-base-spanish-wwm-uncased* had a performance according to the precision shown in the training. With this model, it was possible to classify the aforementioned set of tweets from December 2021. The model was not aware of this data set and the results quantitatively analyzed show consistency with the tests carried out in the training stage.

The model was trained specifically for the topic of COVID-19 and returned a higher number of negative tweets than positive ones, 20,324 negatives versus 13,452 positives. The reason why more negative than positive tweets were obtained is due to the complex issue of COVID-19 which caused major health problems in the world population and in Mexico. The conversations on Twitter on this topic show concern about infections and deaths, among others.

The analysis of the replicas yielded very interesting data, since they are not necessarily in the same line of semantics and sentiment as the tweet to which they belong. This means that there are tweets that are answered or commented on by people who have totally opposite ideas, for example, positive tweets can have a greater number of negative replies and vice versa.

Retweets, like replicas, can present an inverse behavior, although to a lesser extent than these. However, *retweets* are more similar to *likes* in that they are posts that primarily show approval of the original post, although there are exceptions. In general, for *likes*, *tweets* and *retweets*, the model yielded more negative than positive texts, which is considered consistent due to the nature of the topic addressed.

Regarding the comparison of BERT models that are made in this work, it was shown that multi language models can have a high performance; however, they can have some deficiencies in the translation from Spanish to another language and vice versa. The model selected to carry out the sentiment analysis about COVID-19 in Twitter Mexico shows an acceptable accuracy of 81% in the tests carried out during the training. The two best multi language models showed interesting results in precision, since for the training they had 0.95 and 0.96 and for the tests 0.79, 0.80, and 0.81, the others had lower values.

The comparison of the performance of SVM, Logistic Regression, Decision Trees and Naive Bayes models with the BERT models trained with an exclusive base in Spanish, showed that the accuracy of these is higher. The only one that had a precision close to the BERT model was Naive Bayes; it can be considered similar to the model in Spanish.

6 Conclusions

The deep learning model using a network of bidirectional transformers, with a pre-trained model, presented acceptable results, for Spanish-language tweets, about COVID-19. The new training of the pre-trained model, with the fine-tuning for the classification of tweets, provides an interesting precision that allows high reliability in the result. In addition to the set of tests that were used together with the training, we tested the model with 33,776 new tweets that it did not know. The result is correct, where most of the texts were classified as negative, which is normal for the type of topic investigated, COVID-19. The model obtained an accuracy of 97% in training and 81% in tests, so it is concluded that the training is adequate to assign correct scores, positive or negative, in COVID-19 texts, in the Spanish language.

It was demonstrated that the pre-trained model in Spanish, allows the new model to know the form of writing in this language, even with the adaptations applied. These adjustments were made to allow the new model to know how to classify tweets into positive or negative, according to the mood they express.

We compared the result of our model with other models generated from the same data set for training and testing. The model presented here reported greater accuracy than the others, with 97.0% and 81.0%.

Measuring the sentiments of a particular population about the COVID-19 pandemic, in this case, Mexico is important to know and make decisions about issues related to this public health phenomenon. In Mexico, data sources are limited, for example, in the number of deaths, hospitalized people, or in the number of tests to detect SARS-CoV-2. In the latter case, the lack of diagnostic tests was evident, due to the rapid entry into Mexico of the Omicron variant, where it is estimated that millions of people became infected in just a few weeks. We applied the model, in the month of December 2021, with the beginning of the fourth wave of infections caused by Omicron and the results it yielded are a reliable indicator of the mood of the Twitter community in Mexico.

6.1 Future Work

The applications of pre-trained models in Spanish will continue to be researched. A study will be carried out on classifiers with BERT, based on the Spanish language, for the polarity of opinions in social networks. This polarity is a variant of sentiment analysis, but it is not tied to specific words in the text that it can use in lexical-based labeling. This is a classification problem too; however, opinions about two opposing positions can be expressed in words that demonstrate positive or negative sentiment. For example, in the case of measuring polarity on political issues, where there are two clearly opposed positions, opinions such as "this is the best government in history" and others such as "the opposition is demonstrating an excellent performance" can be found. In the lexicon-based sentiment analysis, these two sentences would be classified as positive; however, they demonstrate contrary positions.

In this case, a Deep Learning model that learns to classify contrary positions regardless of the words in the text is required, that is, the model must be able to recognize the semantics of the input texts. To do this, it is considered that a set of training and test data as close as possible to the characteristics of the sample language should be worked on.

Another topic for future work, related to the project is how the model understands the semantics of the ironic, sarcastic, and double-meaning comments in the Spanish language, which are used in Mexico. This is because texts with these characteristics can be interpreted as positive when in fact, they are negative, or vice versa. In addition, the extraction of tweets by topics and the relationship that exists between them, for example, in the general topic of COVID-19, we find many tweets that refer to the characteristics of infections or symptoms, care for infected patients, vaccines, post-COVID-19 sequels, hospitalizations, doctors, medicines and others. The sentiments of each of these topics are not necessarily similar; some have a clearly positive tendency, while others are negative.

Funding No funding was received to assist with the preparation of this manuscript. No funding was received for conducting this study. No funds, grants, or other support were received.

Data availability The data associated with this publication is available with the authors.

Declarations

Conflict of Interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Ahmad, M., Aftab, S., Bashir, M.S., Hameed, N.: Sentiment analysis using SVM: a systematic literature review. *Int. J. Adv. Comput. Sci. Appl.* (2018). <https://doi.org/10.1456/IJACSA.2018.090226>

2. El Alaoui, I., Gahi, Y., Messoussi, R., et al.: A novel adaptable approach for sentiment analysis on big social data. *J. Big Data* (2018). <https://doi.org/10.1186/s40537-018-0120-0>
3. Albalawi, Y., Buckley, J., Nikolov, N.S.: Investigating the impact of pre-processing techniques and pre-trained word embeddings in detecting Arabic health information on social media. *J. Big Data* (2021). <https://doi.org/10.1186/s40537-021-00488-w>
4. Alsayat, A.: Improving sentiment analysis for social media applications using an ensemble deep learning language model. *Arab. J. Sci. Eng.* (2022). <https://doi.org/10.1007/s13369-021-06227-w>
5. Canales, L., Strapparava, C., Boldrini, E., Martinez-Barco, P.: Intensional learning to efficiently build up automatically annotated emotion corpora. *IEEE Trans. Affect. Comput.* (2020). <https://doi.org/10.1109/TAFFC.2017.2764470>
6. Dashtipour, K., Poria, S., Hussain, A., et al.: Multilingual sentiment analysis: state of the art and independent comparison of techniques. *Cogn. Comput.* **8**, 757–771 (2016). <https://doi.org/10.1007/s12559-016-9415-7>
7. de la Rosa, J., Pérez, Á., de Sisto, M., et al.: Transformers analyzing poetry: multilingual metrical pattern prediction with transformer-based language models. *Neural Comput. Appl.* (2021). <https://doi.org/10.1007/s00521-021-06692-2>
8. Devika, M.D., Sunitha, C., Ganesh, A.: Sentiment analysis: a comparative study on different approaches. *Procedia Comput. Sci.* (2016). <https://doi.org/10.1016/j.procs.2016.05.124>
9. Eirinaki, M., Pisal, S., Singh, J.: Feature-based opinion mining and ranking. *J. Comput. Syst. Sci.* (2012). <https://doi.org/10.1016/j.jcss.2011.10.007>
10. González, J.A., Hurtado, L.F., Pla, F.: TWiLBert: pre-trained deep bidirectional transformers for Spanish Twitter. *Neurocomputing* **426**, 58–69 (2021). <https://doi.org/10.1016/j.neucom.2020.09.078>
11. Gopalakrishnan, V., Ramaswamy, C.: Patient opinion mining to analyze drugs satisfaction using supervised learning. *J. Appl. Res. Technol.* **15**(4), 311–319 (2019). <https://doi.org/10.1016/j.jart.2017.02.005>
12. Grissette, H., Nfaoui, E.H.: Affective concept-based encoding of patient narratives via sentic computing and neural networks. *Cogn. Comput.* (2022). <https://doi.org/10.1007/s12559-021-09903-z>
13. Hasni, S., Faiz, S.: Word embeddings and deep learning for location prediction: tracking Coronavirus from British and American tweets. *Soc. Netw. Anal. Min.* (2021). <https://doi.org/10.1007/s13278-021-00777-5>
14. Havey, N.F.: Partisan public health: how does political ideology influence support for COVID-19 related misinformation? *J. Comput. Soc. Sci.* **3**, 319–342 (2020). <https://doi.org/10.1007/s42001-020-00089-2>
15. Jelodar, H., Wang, Y., Orji, R., Huang, S.: Deep sentiment classification and topic discovery on novel Coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE J. Biomed. Health Inform.* (2020). <https://doi.org/10.1109/JBHI.2020.3001216>
16. Jiménez-Zafra, S.M., Martín-Valdivia, M.T., Molina-González, M.D., Ureña-López, L.A.: How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for the medical domain. *Artif. Intell. Med.* (2019). <https://doi.org/10.1016/j.artmed.2018.03.007>
17. Kaliyar, R.K., Goswami, A., Narang, P.: FakeBERT: fake news detection in social media with a BERT-based deep learning approach. *Multimed. Tools Appl.* (2021). <https://doi.org/10.1007/s11042-020-10183-2>
18. Kaur, H., Ahsaan, S.U., Alankar, B., et al.: A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets. *Inf. Syst. Front.* (2021). <https://doi.org/10.1007/s10796-021-10135-7>
19. Kowsari, K., JafariMeimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., Brown, D.: Text classification algorithms: a survey. *Information* **10**, 150 (2019). <https://doi.org/10.3390/info10040150>
20. Kula, S., Kozik, R., Choraś, M.: Implementation of the BERT-derived architectures to tackle disinformation challenges. *Neural Comput. Appl.* (2021). <https://doi.org/10.1007/s00521-021-06276-0>
21. Kumar, A., Gupta, P., Balan, R., et al.: BERT based semi-supervised hybrid approach for aspect and sentiment classification. *Neural Process. Lett.* (2021). <https://doi.org/10.1007/s11063-021-10596-6>
22. Kwon, D., Kim, S., Wei, C.H., Leaman, R., Lu, Z.: EzTag: tagging biomedical concepts via interactive learning. *Nucleic Acids Res.* (2018). <https://doi.org/10.1093/nar/gky428>
23. Li, Z., Chen, X., Xie, H., et al.: EmoChannel-SA: exploring emotional dependency towards classification task with self-attention mechanism. *World Wide Web* **24**, 2049–2070 (2021). <https://doi.org/10.1007/s11280-021-00957-5>

24. Li, X., Fu, X., Xu, G., Yang, Y., Wang, J., Jin, L., Liu, Q., Xiang, T.: Enhancing BERT representation with context-aware embedding for aspect-based sentiment analysis. *IEEE Access* (2020). <https://doi.org/10.1109/ACCESS.2020.2978511>
25. Lighthart, A., Catal, C., Tekinerdogan, B.: Systematic reviews in sentiment analysis: a tertiary study. *Artif. Intell. Rev.* (2021). <https://doi.org/10.1007/s10462-021-09973-3>
26. Liu, Y., Whitfield, C., Zhang, T., et al.: Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning. *Health. Inf. Sci. Syst.* (2021). <https://doi.org/10.1007/s13755-021-00158-4>
27. Mishra, S., Prasad, S., Mishra, S.: Exploring multi-task multi-lingual learning of transformer models for hate speech and offensive speech identification in social media. *SN Comput. Sci.* **2**, 72 (2021). <https://doi.org/10.1007/s42979-021-00455-5>
28. Nandwani, P., Verma, R.: A review on sentiment analysis and emotion detection from text. *Soc. Netw. Anal. Min.* (2021). <https://doi.org/10.1007/s13278-021-00776-6>
29. Nazir, A., Rao, Y., Wu, L., Sun, L.: Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. *IEEE Trans. Affect. Comput.* (2020). <https://doi.org/10.1109/TAFFC.2020.2970399>
30. Ngoc, P.V., Ngoc, C.V.T., Ngoc, T.V.T., et al.: A C4.5 algorithm for English emotional classification. *Evol. Syst.* **10**, 425–451 (2019). <https://doi.org/10.1007/s12530-017-9180-1>
31. Palani, B., Elango, S., Viswanathan, K.V.: CB-Fake: a multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. *Multimed. Tools Appl.* (2022). <https://doi.org/10.1007/s11042-021-11782-3>
32. Ruz, G.A., Henríquez, P.A., Mascareño, A.: Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Gener. Comput. Syst.* (2020). <https://doi.org/10.1016/j.future.2020.01.005>
33. Sarracén, G.L.D., Rosso, P.: Offensive keyword extraction based on the attention mechanism of BERT and the eigenvector centrality using a graph representation. *Pers. Ubiquitous Comput.* **27**, 45–57 (2021). <https://doi.org/10.1007/s00779-021-01605-5>
34. Schiller, B., Daxenberger, J., Gurevych, I.: Stance detection benchmark: how robust is your stance detection? *Künstl. Intell.* **35**, 329–341 (2021). <https://doi.org/10.1007/s13218-021-00714-w>
35. Singh, M., Jakhar, A.K., Pandey, S.: Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Soc. Netw. Anal. Min.* (2021). <https://doi.org/10.1007/s13278-021-00737-z>
36. Sourì, A., Hosseinpour, S., Rahmani, A.M.: Personality classification based on profiles of social networks users and the five-factor model of personality. *HCIS* **8**, 24 (2018). <https://doi.org/10.1186/s13673-018-0147-4>
37. Trabelsi, M., Chen, Z., Davison, B.D., et al.: Neural ranking models for document retrieval. *Inf. Retr. J.* **24**, 400–444 (2021). <https://doi.org/10.1007/s10791-021-09398-0>
38. Vashisht, G., Sinha, Y.N.: Sentimental study of CAA by location-based tweets. *Int. J. Inf. Technol.* **13**, 1555–1561 (2021). <https://doi.org/10.1007/s41870-020-00604-8>
39. Viji, D., Revathy, S.: A hybrid approach of Weighted Fine-Tuned BERT extraction with deep Siamese Bi-LSTM model for semantic text similarity identification. *Multimed. Tools Appl.* (2022). <https://doi.org/10.1007/s11042-021-11771-6>
40. Villatoro-Tello, E., Parida, S., Kumar, S., et al.: Applying attention-based models for detecting cognitive processes and mental health conditions. *Cogn. Comput.* (2021). <https://doi.org/10.1007/s12559-021-09901-1>
41. Wang, Y., Chu, Y.M., Thaljaoui, A., et al.: A multi-feature hybrid classification data mining technique for human-emotion. *BioData Min.* (2021). <https://doi.org/10.1186/s13040-021-00254-x>
42. Wang, Z., Cui, X., Gao, L., et al.: A hybrid model of sentimental entity recognition on mobile social media. *EURASIP J. Wirel. Commun. Netw.* (2016). <https://doi.org/10.1186/s13638-016-0745-7>
43. Wang, J., Zhang, G., Wang, W., et al.: Cloud-based intelligent self-diagnosis and department recommendation service using Chinese medical BERT. *J. Cloud Comput.* **10**, 4 (2021). <https://doi.org/10.1186/s13677-020-00218-2>
44. Zerva, C., Nghiem, M.Q., Nguyen, N.T.H., et al.: Cited text span identification for scientific summarisation using pre-trained encoders. *Scientometrics* **125**, 3109–3137 (2020). <https://doi.org/10.1007/s11192-020-03455-z>
45. Zhang, Z., Han, J., Deng, J., Xu, X., Ringeval, F., Schuller, B.: Leveraging unlabeled data for emotion recognition with enhanced collaborative semi-supervised learning. *IEEE Access* (2018). <https://doi.org/10.1109/ACCESS.2018.2821192>

46. Zhou, J., Yang, S., Xiao, C., Chen, F.: Examination of community sentiment dynamics due to COVID-19 pandemic: a case study from a state in Australia. *SN Comput. Sci.* **2**, 201 (2021). <https://doi.org/10.1007/s42979-021-00596-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Salvador Contreras Hernández¹  · María Patricia Tzili Cruz¹  ·
José Martín Espínola Sánchez¹  · Angélica Pérez Tzili² 

María Patricia Tzili Cruz
patricia.tzili@upvm.edu.mx

José Martín Espínola Sánchez
jose.espinola@upvm.edu.mx

Angélica Pérez Tzili
angy_tzili@hotmail.com

¹ Department of Informatics, Universidad Politécnica del Valle de México, 54910 Tultitlán Estado de México, Mexico

² School of Medicine, Instituto Politécnico Nacional, 07738 Mexico City, Mexico