



Article

# Detecting Aggressiveness in Tweets: A Hybrid Model for Detecting Cyberbullying in the Spanish Language

Manuel Lepe-Faúndez <sup>1</sup>, Alejandra Segura-Navarrete <sup>1</sup>, Christian Vidal-Castro <sup>1,\*</sup> , Claudia Martínez-Araneda <sup>2</sup> and Clemente Rubio-Manzano <sup>1,3</sup> 

<sup>1</sup> Department of Information Systems, University of the Bío-Bío, Avda. Collao 1202, Casilla 5-C, Concepción 4051381, Chile; mlepe@ubiobio.cl (M.L.-F.); asegura@ubiobio.cl (A.S.-N.); clrubio@ubiobio.cl (C.R.-M.)

<sup>2</sup> Computer Science Department, Universidad Católica de la Santísima Concepción, Concepción 4090541, Chile; cmartinez@ucsc.cl

<sup>3</sup> Department of Mathematics, University of Cádiz, 11003 Cádiz, Spain

\* Correspondence: cvidal@ubiobio.cl

**Abstract:** In recent years, the use of social networks has increased exponentially, which has led to a significant increase in cyberbullying. Currently, in the field of Computer Science, research has been made on how to detect aggressiveness in texts, which is a prelude to detecting cyberbullying. In this field, the main work has been done for English language texts, mainly using Machine Learning (ML) approaches, Lexicon approaches to a lesser extent, and very few works using hybrid approaches. In these, Lexicons and Machine Learning algorithms are used, such as counting the number of bad words in a sentence using a Lexicon of bad words, which serves as an input feature for classification algorithms. This research aims at contributing towards detecting aggressiveness in Spanish language texts by creating different models that combine the Lexicons and ML approach. Twenty-two models that combine techniques and algorithms from both approaches are proposed, and for their application, certain hyperparameters are adjusted in the training datasets of the corpora, to obtain the best results in the test datasets. Three Spanish language corpora are used in the evaluation: Chilean, Mexican, and Chilean-Mexican corpora. The results indicate that hybrid models obtain the best results in the 3 corpora, over implemented models that do not use Lexicons. This shows that by mixing approaches, aggressiveness detection improves. Finally, a web application is developed that gives applicability to each model by classifying tweets, allowing evaluating the performance of models with external corpus and receiving feedback on the prediction of each one for future research. In addition, an API is available that can be integrated into technological tools for parental control, online plugins for writing analysis in social networks, and educational tools, among others.

**Keywords:** cyberbullying detect; emotions analysis in Spanish; hybrid approach



**Citation:** Lepe-Faúndez, M.; Segura-Navarrete, A.; Vidal-Castro, C.; Martínez-Araneda, C.; Rubio-Manzano, C. Detecting Aggressiveness in Tweets: A Hybrid Model for Detecting Cyberbullying in the Spanish Language. *Appl. Sci.* **2021**, *11*, 10706. <https://doi.org/10.3390/app112210706>

Academic Editor: Luis Javier García Villalba

Received: 19 September 2021

Accepted: 2 November 2021

Published: 12 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The growing use of social networks has provided a channel to unrestrictedly express feelings and opinions on a mass scale. However, one of the negative aspects is that this has caused an increase in harassment, the so-called cyberbullying, defined as the use of information and communication technologies, like e-mails, text messages from cell phones, social networks, to support the deliberate, repeated, and hostile behavior of an individual or group to harm others, through personal attacks, disclosure of confidential or fake information, among other aspects [1].

According to [2], between 2005 and 2018, there was an increase in cyberbullying cases in Latin America. This study detected the existence of a high percentage of related situations: between 3.5% and 58% of cyber-victims; and between 2.5% and 32% of cyber-aggressors. Those involved are mainly men. In the particular case of Chile, according to [3] study led by PUCV with financing from MINEDUC, and under the agreement with

UNESCO-Santiago, 20% of those surveyed, reported having been treated offensively or in an unfriendly manner by other people. These experiences are 58% in-person cases (without excluding other forms), 28% through a social network, 25% by messaging, and 13% through an online game. A total of 14% of children who use the internet acknowledge having treated another person offensively or rudely in the last year. 76% of these interactions have been face-to-face, 19% through social media, 18% by messaging, and 11% through cell phone messages. Generally, bullying and cyberbullying begin in primary school. This continues in high school, where it reaches its peak. In some cases, these behaviors continue after high school, leading to social isolation, truancy, and low grades [4,5].

Generally, cyberbullying starts with aggressive and repetitive messages from one person or a group of people to another person, where an aggressor and a victim can be seen. Table 1 shows some of the types of cyberbullying [6]. These types of cyberbullying share common elements, like the intention to cause harm and the sending of messages with aggressive and offensive content. Hence, it is important to create detection mechanisms of aggressive messages as a first step for the early detection of cyberbullying.

**Table 1.** Types of cyberbullying [6].

Type of Cyberbullying	Description
Flaming	Sending aggressive, rude, and vulgar messages, targeting one or more people, privately or in an online group
Bullying	Repetitively sending aggressive, rude, and vulgar messages to a person
Cyberstalking	Harassment that includes threats to harm or that is highly intimidating.
Denigration	Sending or publishing harmful, aggressive, fake, or cruel statements about one person to others.
Identity theft	Pretending to be another person and sending or publishing material to make them either look bad or to endanger them.
Outing and trickery	Sending or publishing material about a person that contains sensitive, private, or embarrassing information, including forwarding private messages or images. Tricking people to request embarrassing information that is then made public.
Exclusion	Actions that specifically and intentionally exclude a person from an online group.

Natural Language Processing (NLP) can provide important mechanisms to detect aggression in texts. NLP is an area of research and application that explores how computers can be used to understand and manipulate human expressions in text [7]. NLP addresses different areas, like computing and computer science, linguistics, math, artificial intelligence, and psychology, among others. In recent years, the use of different techniques has been popular to identify the emotions that the author of a comment or message wishes to transmit. Text subjectivity analysis is found as an NLP subcategory. This oversees extracting and classifying the different emotions that the author of a text wants to transmit, and with this obtain valuable information to analyze its content. The analysis of text subjectivity provides different tools that allow detecting aggression in text written on social media. There is consensus in the benefits that the early detection of aggression in messages sent by users provides, allowing taking preventive measures, and thus avoiding the consequences of cyberbullying.

Given that most aggression detection works in texts prior to 2018 have been made for the English language, it seems to be important to focus this work on the analysis of aggression in texts written in the Spanish language. Using the previous work of the Universidad del Bio-Bio research group SoMos, the use of a hybrid approach based on

lexicons and Machine Learning (ML) is proposed. Specifically, a hybrid model is proposed, and its results are compared with models that do not use lexicons in the detection of text aggression.

The rest of this article is organized as follows: the next section presents the background and related works on cyberbullying detection that use machine learning and lexicon-based approaches. Section 3 describes the methodology applied, including a detailed description of the resources used. Section 4 describes the models proposed for classifying aggressive texts. The implementation and performance evaluation of the proposed models are shown in Section 5, using a software specifically created for this purpose. Section 6 presents the discussion of the results achieved. Finally, the conclusions and lines of future works are presented.

### *Background*

Through a revision of the literature, works were studied that propose some aggression detection mechanism. As has already been mentioned, most of these works were applied to texts in English [8–17]. Regarding the approach used, more than 50% of the articles used ML techniques, using different corpora to train classifiers. The most typically used algorithms are Naïve Bayes, Support Vector Machine and Random Forest, due to their widespread use in text classifications. There is a smaller body of works that combine, in one way or another, the ML approach with the use of lexicons, for example, to have predefined lists of bad words that, once detected, are used as features in ML [13,15,17]. In [17] was used exclusively the lexicon-based approach, including 9 bad words chosen by the authors considering their high frequency in situations labeled as Cyberbullying, applying a morphological analysis and information recovery techniques to determine the degree of aggression.

A growing interest can be seen in applying aggression detection in Spanish texts. In [18], 3 corpora are created from Twitter: small corpus (25,304 tweets), medium corpus (229,801 tweets), and big corpus (960,578 tweets). The “Presumed cyberbullying” or “Without cyberbullying” labelling of each tweet is done automatically, bearing in mind the “General insult inventory” [19], as well as adding Ecuadorian insults and the patterns detected in [20]. The model is created using solely ML (Naive Bayes, Support Vector Machine and Logistic Regression) algorithms, and the feature vector is formed using the TF-IDF technique. On evaluating the model with the different corpora, there is an average accuracy of between 80% and 91%, with Support Vector Machine obtains the best result on being applied in the medium corpus, with a peak of 94% accuracy. In addition, a web application is implemented, where the percentage of cyberbullying can be evaluated in real-time in Twitter under 3 scenarios: phrase analysis, analysis of a Twitter profile (bearing in mind the most recent tweets), and trend analysis.

In [21] was present a proposal that analyzes Peruvian phrases. In this work, a Naive Bayes classifier is trained through the NLTK library [22] for Python, using a lexicon [23] and 595 words labeled manually. The Bag of Words method is used to represent the phrases, through a collection of bad words. The model provides the probability that a phrase contains bullying characteristics.

The workshop organized by IberEval has provided important room for aggression detection initiatives in Spanish texts [24]. In this event, the participants faced the challenge of proposing models that allow detecting aggression in a corpus which, in its 2018 version, comprised 10,856 instances (7700 for training and 3153 for evaluation). The corpus gathered tweets of Mexican users, in Spanish. The participants proposed a variety of methodologies, which comprised content-based features (frequencies, scores, POS, specific elements of Twitter, etc.), as well as classical ML (Naïve Bayes, SVM, Logistic regression, etc.) algorithms, and Neural Networks. In [25] was presented the winning model that reached an average F-measure of 0.620 and an accuracy of 0.667. It uses a classifier that utilizes a Support Vector Machine and two lexicons, and through genetic programming, it makes the final prediction. In 2019, the winning work extracts features using Word Embedding

and n-grams, and uses Multilayer Perceptron to classify [26]. The winning team of the 2020 workshop uses a classifier trained to predict aggression, with majority and weighted vote schemes [27]. The training is done by adjusting the model. Table 2 summarizes the works analyzed that make aggression detection in Spanish, indicating the approach, corpus, algorithms, and the features vector reported.

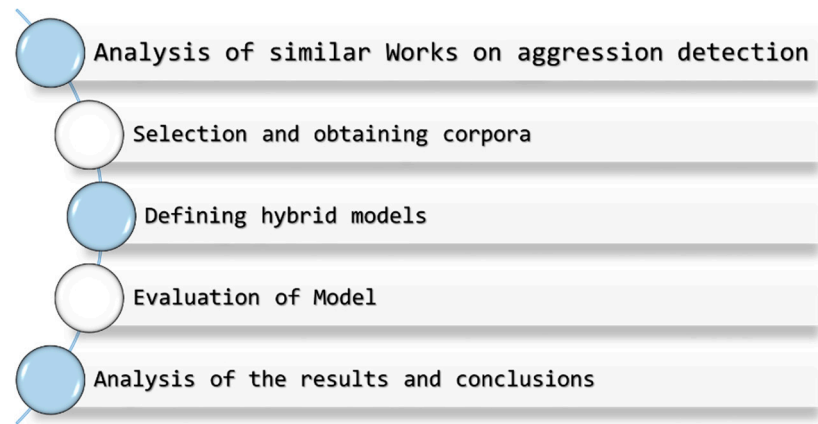
**Table 2.** Summary of aggressive detection models analyzed.

Journals						
Year	Name	Approach	Corpus	Classification Algorithms	Features Vector	Best Result
2018	Automatic cyberbullying detection in spanish-language social networks using sentiment analysis techniques [21].	machine learning & lexicon	100 social network phrases, labeled manually from 0 to 100	naive bayes	bag of words	0.93 accuracy 0.93 accuracy
2019	Presumptive detection of cyberbullying on twitter through natural language processing and machine learning in the spanish language [18].	machine learning	own corpus of 960,578, “presumed cyberbullying” or “without cyberbullying”, automatically labeled	naive bayes, support vector machine & logistic regression	tf-idf	support vector machine with a peak of 0.94 in accuracy.
Workshops						
Year	Name	Place	Approach	Classification Algorithms	Features vector	F-measure
2018	Ingeotec at mex-a3t: author profiling and aggressiveness analysis in twitter using utc and evomsa [25].	1	Combines machine learning models and lexicons	support vector machine	-	0.620
	Attention mechanism for aggressive detection [28].	2	Neural network (bi-lstm) and post-attention lst	neural networks	word embedding & presence of bad words or not according to the lexicon	0.605
2019	Uach at mex-a3t 2019: preliminary results on detecting aggressive tweets by adding author information via an unsupervised strategy [26].	1	Machine learning	multilayer perceptron and support vector machine	word embedding & n-grams.	0.620 (mp)
2020	Transformers and data augmentation for aggressiveness detection in mexican spanish [29].	1 and 2	(1) 20 adjusted betos (2) 20 betos plus data augmentation techniques	neural networks	-	(1) 0.8851 (2) 0.8588
	Detecting aggressiveness in mexican spanish social media content by fine-tuning transformer-based models [30]	3	Fine-tuning pre-trained english, spanish and multilingual-transformer-based models	neural networks	-	0.8538 (1 beto)

## 2. Materials and Methods

Figure 1 describes the methodology used in this work. In the first stage, a revision of the literature was made to get to know the models recently applied for aggression

detection in Spanish texts. Then, the corpora to be used in this work was chosen. After this, the hybrid classification models are defined and implemented, considering the literature revision. In the next stage, the models are evaluated through experiments, measuring their performance using the metrics utilized in the literature. Finally, the results analysis of the evaluation is carried out, reaching conclusions on the hybrid models implemented.



**Figure 1.** Methodology used.

Table 3 corpora are used, that contain tweets in Spanish, manually labeled as “aggressive” and “non-Aggressive”. The first corpus is from tweets by Chilean users, and for this reason, is called the Chilean corpus. It comprises the corpus prepared by [31], with 1470 tweets in the context of aggression against women, and the corpus prepared in [32] which has 1000 tweets. 41% of the total correspond to tweets labeled as “aggressive” and 59% as “non-Aggressive”. The second corpus used was created by [24]. These are tweets in Spanish collated in Mexico City, and were filtered to use only the instances labeled. This corpus of 7332 tweets will be called Mexican corpus, with 28.8% of the tweets labeled as “aggressive” and 71.2% as “non-Aggressive”. The third corpus used is the previous two together, which will be called the Chilean-Mexican corpus. The merger is made with the goal of having a larger corpus with tweets from different countries to test the different models. This last corpus has a total of 9802 tweets, with 31.9% labeled “aggressive” and 68.1% as “non-Aggressive”.

**Table 3.** Number of instances of training and testing of the corpus.

Corpus	Number of Instances		
	Train (70%)	Test (30%)	Total (100%)
Chilean	1729	741	2470
Mexican	5132	2200	7332
Chilean-Mexican	6861	2941	9802

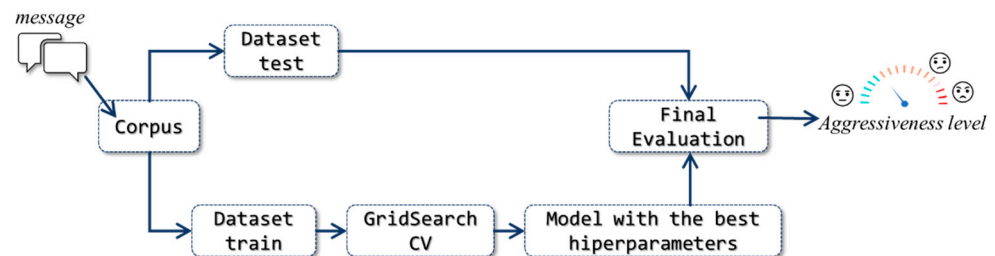
To train the models, 70% of the corpus instances were used, with 30% used to run the performance tests. Table 3 shows the number of instances that were set aside for training and testing in the 3 corpora used.

### 2.1. Proposed Aggressiveness Detection Models

The main feature that differs the approaches is the way of representing the features vector of tweets that receive the ML algorithm as input. All the proposed approaches are implemented with 3 supervised Machine Learning classification algorithms: Support Vector Machine (SVM), Naïve Bayes (NB) and Random Forest (RF).

Hyperparameters and candidate values are defined in each approach, while the definitive values are chosen using the GridSearchCV algorithm, applied to the training datasets of each corpus. GridSearchCV creates an execution matrix where all the possible

combinations of candidate values are evaluated, and the best combination is kept. The cross-validation technique is used to evaluate the performance of each execution, to minimize overadjustment, and the metric used to choose the best combination is the F-measure. Figure 2 shows this process and how the final evaluation of the models is done. The results are presented in the Experiments and Results section.



**Figure 2.** Training and testing process.

### 2.1.1. TF-IDF Approach

This first approach is the simplest and uses the most traditional technique to obtain features of a text, called TF-IDF (Term frequency—Inverse document frequency) ([https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.TfidfVectorizer.html](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html), accessed on 5 November 2021). The purpose of creating this model is to use it as a base to compare the results of the rest of the models that mix Lexicons with ML classifiers.

First, a preprocessing of the text is done, depending on the definitive values of the defined hyperparameters. TF-IDF is used to obtain the features vector of the text, which consists of determining the importance of each word in the phrase depending on the frequency words appear in the corpus. After obtaining the features vector, the different ML classifiers are applied.

### 2.1.2. Lexicon Approach

The second approach implemented used a mixture of the emotions analysis using Lexicons to form the features vector and ML classifiers. The Lexicon used is the one proposed in [33], which consists of an affective Lexicon in Spanish based on an enriched Lexicon, which represents the emotion intensity of each word, as shown in the example in Table 4. This Lexicon considered so-called emotional words.

**Table 4.** Extract from the Intensity Lexicon for the effective anger class.

Word	Intensity
fuss	88
angry	75
fury	75
inspire	10

Initially, a preprocessing of the text is done, filtering special characters (stress marks, punctuation, signs, etc.), as well as eliminating stopwords and the lemmatization of each word depending on the definitive values of defined hyperparameters. Tokenization is done using spacy (<https://spacy.io/models/es>, accessed on 5 November 2021) in Spanish. After preprocessing, the analysis is made with the Lexicons, to form the features vector of each phrase that is comprised by 10 columns, detailed below.

Each one of the first 8 columns represents the sum of intensities of the phrase's words, that appear in the Lexicon which represents the corresponding affective class. The affective classes proposed by [34] are considered (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust).

Column 9 represents the division between the number of bad words (BW) found in the phrase and the number of words this has.

Finally, column 10 represents the number of words in the phrase (NW).

Table 5 shows an example of the features vector obtained for the offensive phrase “Oyyyyy feo culiao insoportable chucha nota esta cagao miedo” (Oi, ugly unbearable fucker, fuck, look they’re fucking scared) to exemplify the process. The angry column has a value of 56, as in the affective class Lexicon, the word “nota” (look) has an intensity of 10 and the word “miedo” (scared) an intensity of 46, therefore, on adding these two intensities, the total is 56. The process was done in the same way for the other columns of the affective class. In the column that represents the result of the division between the number of bad words and the number of words in the phrase, the value is 0.333, since there are 3 bad words found in the defined Lexicon: “feo” (ugly), “culiao” (fucker), “chucha” (fuck), and a total of 9 words in the phrase.

Table 5. Example feature vector Lexicon approach.

Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	BW/NW	BW
56	64	85	64	92	38	68	46	0.333	9

### 2.1.3. TF-IDF Lexicon Approach

For this approach, a mix of the TF-IDF and Lexicon approach is implemented, where the features vector is a concatenation of the TF-IDF vector and the one derived from the Lexicons analysis. Figure 3 shows the process this approach performs. The corpus processing takes two routes to perform each approach, while the preprocessing of the corpus is done in each approach following its previously defined hyperparameters. Finally, these two vectors are concatenated, as shown in Figure 4, to apply the ML algorithms. The size of the TF-IDF vector depends on the vocabulary of each corpus, hence the final size of the vector is subject to the corpus the training is done with.

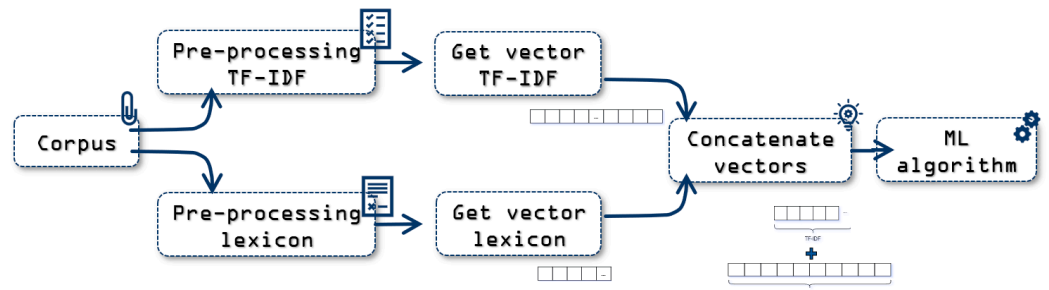


Figure 3. TF-IDF\_Lexicon approach process.

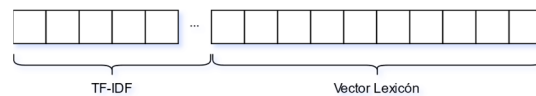


Figure 4. TF-IDF\_Lexicon approach feature vector.

### 2.1.4. Word Embedding Approach

This approach seeks to represent the features vector using the Word Embedding technique. In a similar way as the TF-IDF approach, this is implemented to have a basis for comparison for the rest of the approaches that include Lexicons. Word Embedding is an approach of distribution semantics that represents words of a phrase as real number vectors. This representation has useful grouping properties, as it groups semantically and syntactically similar words. For example, it is expected that the words “dolphin” and “seal” are found to be close, but “Paris” and “dolphin” are not, since there is not a strong relationship between them. Therefore, the words are represented as real value vectors, where each value captures a dimension of the meaning of the word. This means that semantically similar words have similar vectors. In other words, each dimension of the

vectors represents a meaning, and the numerical value in each dimension captures the proximity of the association of the word to said meaning. In [35] was showed the power of Word Embedding. In their work, they establish this tool as being highly effective in different Natural Language Processing tasks, while presenting a neural network architecture that many of the current approaches are based upon.

Firstly, just as in the previous approaches, a preprocessing of the text is done, filtering special characters (stress marks, punctuation, signs, among others), as well as eliminating stopwords and making a lemmatization of each word depending on the definitive values of the defined hyperparameters. Then, the representation of the feature vector of each text is done using the sum of the Word Embedding vectors of each word present in the phrase. In this way, a vector is obtained that represents the entire text. It is worth stating that, after the sum is made, a standardization of the resulting vector is made. Figure 5 shows, as an example, a vectorial representation of the phrase “me gustan los gatos” (I like cats), without standardizing it.

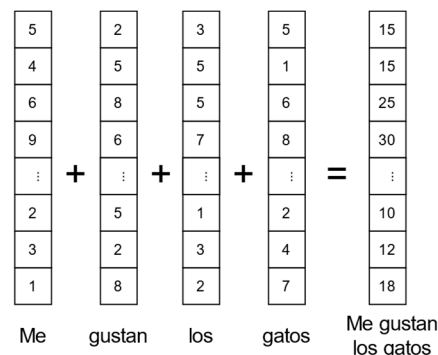


Figure 5. Word Embedding approach feature vector.

A pretrained Word Embedding model is used to obtain the features vector from the phrases. This was implemented with FastText and Skipgram [36] and was trained with 1.4 billion words, using the Spanish Billion Word Corpus [37]. Each vector has 300 dimensions; therefore, each text will be represented with a 300-size vector. This vector is received as input for the classification algorithms implemented.

### 2.1.5. WE\_Lexicon Approach

This approach represents the features vector as a concatenation of the output vectors of Word Embedding and Lexicon approaches. Figure 6 shows the process performed, doing this following its previously defined hyperparameters. Finally, these two vectors are concatenated, as shown in Figure 7, to apply Machine Learning algorithms. The vector size is 310, 300 boxes for the Word Embedding vector, and 10 for the Lexicon analysis vector.

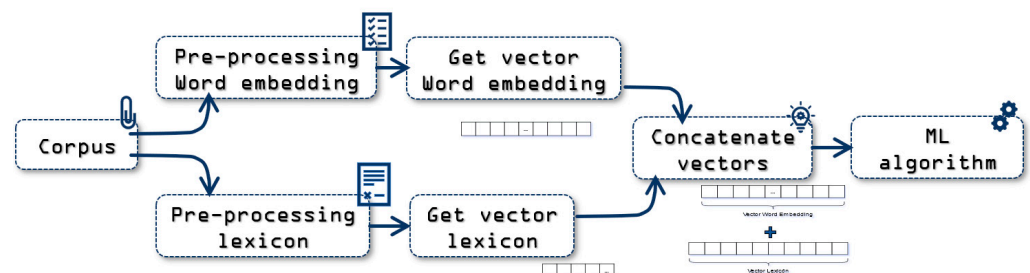


Figure 6. Process approach WE\_Lexicon.



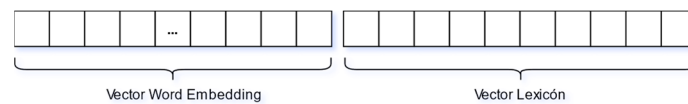


Figure 7. WE\_Lexicon feature vector.

### 2.1.6. WE\_Lexicon TF-IDF Approach

In a similar way as the previous approach, the features vector is presented with the concatenation of the Word Embedding and Lexicon vector, but the TF-IDF vector is also added. Figure 8 shows the process used for this approach. Unlike with the previous approach, now the corpus takes 3 paths to execute the 3 approaches with their preprocessing, following the previously defined hyperparameters. Finally, the 3 vectors are concatenated, as shown in Figure 9, to later apply the Machine Learning classifier. The vector size depends on the corpus and its words. This occurs on having the TF-IDF representation.

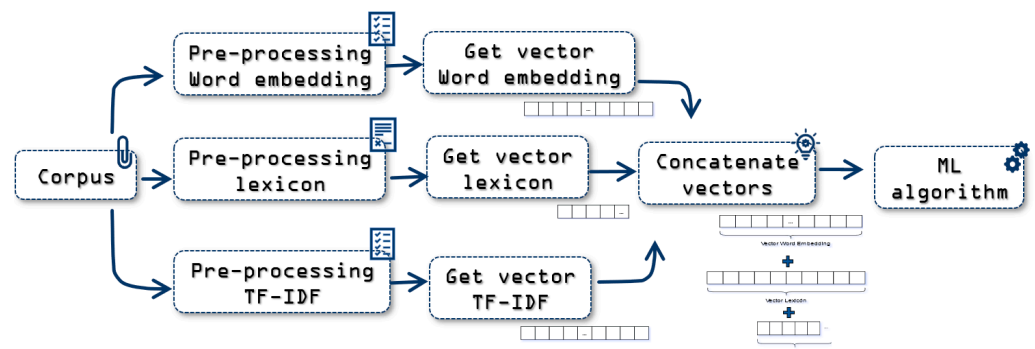


Figure 8. Process approach WE\_Lexicon\_TF-IDF.

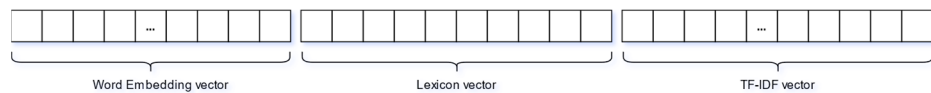


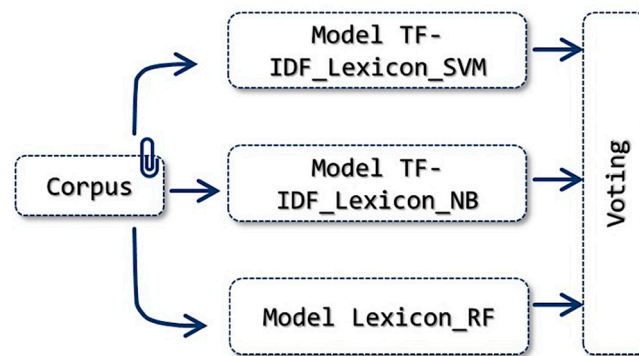
Figure 9. WE\_Lexicón\_TF-IDF feature vector.

### 2.1.7. Ensemble Approach

Below, four models implemented under the “Ensemble Learning” technique are described. Ensemble learning is the process of combining decisions of several Machine Learning models trained to improve overall performance. With the decisions of the different models, a final prediction takes place using different rules such as, for example, the majority vote. The reason behind using Ensemble models is to reduce the prediction generalization error. The prediction error of the model drops when this technique is used, provided that the combined models are diverse and independent. The approach seeks the wisdom of the masses to make a prediction. Although the Ensemble model has multiple base models within it, it acts and behaves as a single model [38]. In the models developed, the final prediction is made using a majority vote, which is implemented with VotingClassifier (<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html>, accessed on 5 November 2021) from the scikit-learn library.

### 2.1.8. TF-IDF\_Lexicon\_E\_Clfs Model

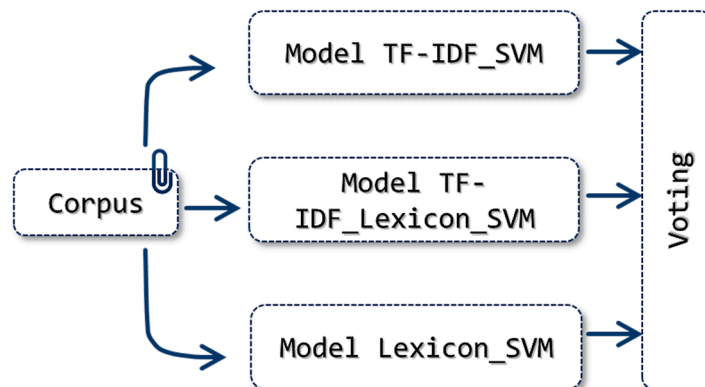
The first model created under this approach combines the three models implemented under the TF-IDF\_Lexicon approach, as shown in Figure 10. Here, the corpus feeds 3 individually trained models, to then make a final prediction about the test corpus using the majority vote technique. This model is created under the hypothesis that the combination of the 3 models that use the TF-IDF\_Lexicon approach will provide better results than each one of them separately, as they use different classifiers. The definitive values of the hyperparameters of each model, found previously using GridSearchCV on the different training datasets of the corpora, are used.



**Figure 10.** TF-IDF\_Lexicon\_E\_Clfs Model.

#### 2.1.9. TF-IDF\_Lexicon\_E\_SVM Model

This model combines those created with the Support Vector Machine classifier in the TF-IDF, Lexicon and TF-IDF\_Lexicon approaches, as shown in Figure 11. Just as in the previous approach, the models are trained separately to then make a final prediction on the test corpus using the majority vote technique. This is implemented, as it is thought that by combining the different ways of obtaining the features vector, the result of the final classification can be improved. The Support Vector Machine classifier is used because it obtained the best performance in the preliminary tests. The different values of each model's hyperparameters, found beforehand using GridSearchCV on the different training datasets of the corpora, are used.



**Figure 11.** TF-IDF\_Lexicon\_E\_SVM Model.

#### 2.1.10. WE\_Lexicon\_TF-IDF\_E\_SVM Model

The third model created combines the models implemented with the Support Vector Machine classifier in the Word Embedding, WE\_Lexicon and WE\_Lexicon\_TF-IDF approaches, as shown in Figure 12. The models are trained separately to then make a final prediction on the test corpus using the majority vote technique. This model, just as in the previous approach, is implemented under the hypothesis that combining the different ways of obtaining the feature vector can improve the final classification result. The hyperparameter values used in each model were found beforehand using GridSearchCV on the different training datasets of the corpora.

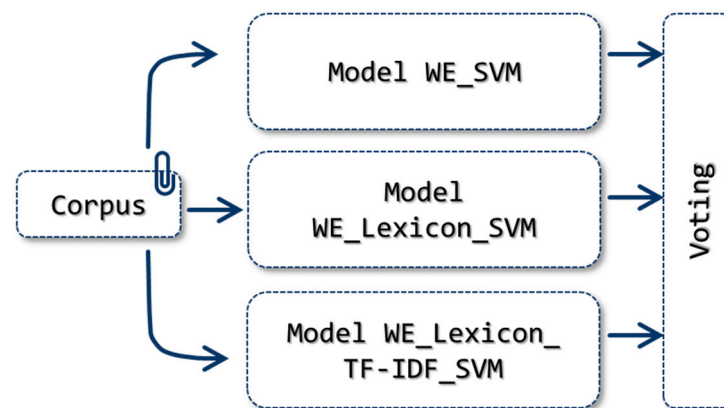


Figure 12. WE\_Lexicon\_TF-IDF\_E\_SVM Model.

#### 2.1.11. E\_SVM\_Approach Model

The final Ensemble model implemented combines all the those implemented with the Support Vector Machine classifier in the different approaches, as shown in Figure 13. Just as in all the previous models, these are trained separately to then make a final prediction on the test corpus using the majority vote technique. The hypothesis behind implementing this is that a greater diversification of the ways of obtaining the features vector can improve the result. It is worth mentioning that this model is more costly in terms of memory and time to train and test the corpus. The hyperparameter values of the models found beforehand in the different corpora are used.

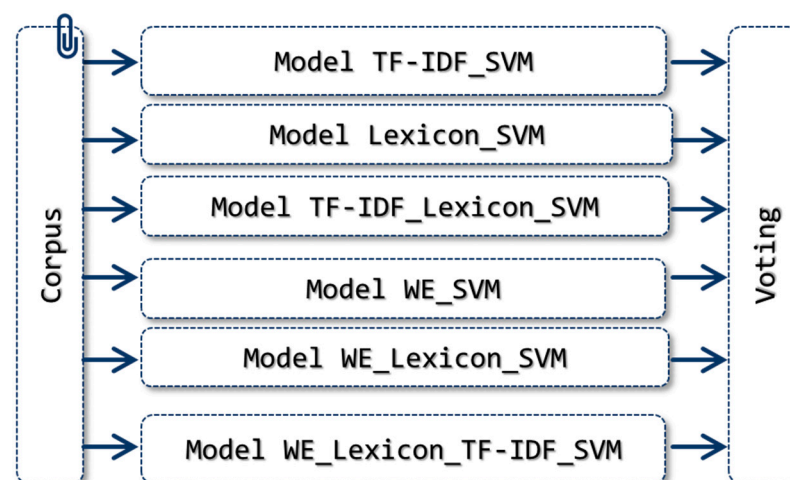


Figure 13. E\_SVM\_Approach Model.

### 3. Implementations and Experimentation

A web application (<http://35.247.212.145/>, accessed on 28 October 2021), which implements the models proposed for the classification of aggressiveness of the different corpora and other new ones written by a user was developed, to show the applicability of the models proposed, and to evaluate their performance through experiments. This application allows classifying comments, receiving feedback of the classification results, and building a base of labeled tweets for future research. The application also allows evaluating the performance of the model chosen with suitably structured test corpora. The performance results are provided using the F-measure metrics: Accuracy, Precision, and Recall. For back-end development, the FastAPI (<https://fastapi.tiangolo.com/>, accessed on 5 November 2021) framework was used, which allows building APIs using the Python programming language. The front-end was implemented using the Vue.js (<https://vuejs.org/>, accessed on 5 November 2021) JavaScript framework, along

with the Vuetify (<https://vuetifyjs.com>, accessed on 5 November 2021) user interface library for Vue.js. Docker (<https://docs.docker.com>, accessed on 5 November 2021), while Docker-compose (<https://docs.docker.com/compose>, accessed on 5 November 2021) was used for the display. The web application code can be downloaded at [https://gitlab.com/ManuelLepeF/lexicon\\_ml\\_agresividad\\_web](https://gitlab.com/ManuelLepeF/lexicon_ml_agresividad_web), accessed on 28 October 2021. Figure 14 shows the user interface of the web application.

**Figure 14.** Web application.

#### *Description of the Experiments*

Using this application, the implemented models were tested on the test datasets of the three corpora: Chilean, Mexican, and Chilean-Mexican. The datasets used in the experiments are available at [https://gitlab.com/ManuelLepeF/lexicon\\_ml\\_agresividad](https://gitlab.com/ManuelLepeF/lexicon_ml_agresividad) (accessed on 5 November 2021). It is worth highlighting that these datasets were not used in the training process, as Figure 2 shows. In this way, the generalization capacity of the models was measured using the F-measure and Accuracy metrics. The hyperparameters used in each model were found using the GridSearchCV technique in the different training datasets of the corpora.

The experiments were made on a server that has the following hardware and software features.

1. Ubuntu 18.04.03 (4.15.0.101-generic kernel)
2. 2 × Intel(R) Xeon(R) Gold 5118 CPU @ 2.30 GHz
3. 62 Gb RAM
4. 48 CPUs

#### **4. Results**

After running the experiments with all the models described above on the 3 corpora, Table 6 shows the results obtained by the models in the F-measure metric in the different corpora used.

As a means of complementing the results, Table 7 shows the results obtained with the *Accuracy* metric in the different corpora used.

**Table 6.** F-measure obtained by the models in the corpus.

Approach	Model	Corpus		
		Chilean	Mexican	Chilean-Mexican
		F-Measure		
TF-IDF	TF-IDF_SVM	0.8671	<b>0.8225</b>	0.8225
	TF-IDF_NB	0.7514	0.7336	0.7351
	TF-IDF_RF	<b>0.8701</b>	0.8069	<b>0.8330</b>
Lexicon	Lexicon_SVM	0.8535	0.5735	0.6528
	Lexicon_NB	0.8321	<b>0.6408</b>	0.6865
	Lexicon_RF	<b>0.8627</b>	0.6347	<b>0.7016</b>
TF-IDF_Lexicon	TF-IDF_Lexicon_SVM	0.8648	<b>0.8330</b>	<b>0.8372</b>
	TF-IDF_Lexicon_NB	0.7829	0.7420	0.7543
	TF-IDF_Lexicon_RF	<b>0.8839</b>	0.7960	0.8231
WordEmbedding	WordEmbedding_SVM	<b>0.8547</b>	<b>0.7831</b>	<b>0.7900</b>
	WordEmbedding_NB	0.8253	0.7504	0.7633
	WordEmbedding_RF	0.8170	0.7296	0.7252
WE_Lexicon	WE_Lexicon_SVM	<b>0.8908</b>	<b>0.7874</b>	<b>0.8086</b>
	WE_Lexicon_NB	0.8495	0.7551	0.7696
	WE_Lexicon_RF	0.8833	0.7107	0.7257
WE_Lexicon_TF-IDF	TF-IDF_SVM	<b>0.8731</b>	<b>0.8394</b>	<b>0.8507</b>
	TF-IDF_NB	0.7842	0.7420	0.7537
	TF-IDF_RF	0.8501	0.7061	0.7033
Ensemble	TF-IDF_Lexicon_E_Clfs	0.8828	0.8191	<b>0.8399</b>
	TF-IDF_Lexicon_E_SVM	0.8716	<b>0.8308</b>	0.8356
	WE_Lexicon_TF-IDF_E_SVM	<b>0.8851</b>	0.7879	0.8219
	E_SVM_Approaches	0.8804	0.7868	0.8146

**Table 7.** Accuracy obtained by the models in the corpus.

Approach	Model	Corpus		
		Chilean	Mexican	Chilean-Mexican
		Accuracy		
TF-IDF	TF-IDF_SVM	0.8690	<b>0.8281</b>	<b>0.8473</b>
	TF-IDF_NB	0.7530	0.7522	0.7500
	TF-IDF_RF	<b>0.8717</b>	0.8204	0.8418
Lexicon	Lexicon_SVM	0.8556	0.6977	0.7154
	Lexicon_NB	0.8367	0.6822	0.7113
	Lexicon_RF	<b>0.8636</b>	<b>0.7000</b>	<b>0.7211</b>
TF-IDF_Lexicon	TF-IDF_Lexicon_SVM	0.8663	<b>0.8395</b>	<b>0.8439</b>
	TF-IDF_Lexicon_NB	0.7813	0.7486	0.7558
	TF-IDF_Lexicon_RF	<b>0.8852</b>	0.8122	0.8340
WordEmbedding	WordEmbedding_SVM	<b>0.8569</b>	<b>0.7972</b>	<b>0.8021</b>
	WordEmbedding_NB	0.8259	0.745	0.7599
	WordEmbedding_RF	0.8218	0.7713	0.7616
WE_Lexicon	WE_Lexicon_SVM	<b>0.8920</b>	<b>0.8027</b>	<b>0.8184</b>
	WE_Lexicon_NB	0.8502	0.7495	0.7664
	WE_Lexicon_RF	0.8852	0.7590	0.7626
WE_Lexicon_TF-IDF	WE_Lexicon_TF-IDF_SVM	0.8744	<b>0.8431</b>	<b>0.8548</b>
	WE_Lexicon_TF-IDF_NB	0.7827	0.7486	0.7548
	WE_Lexicon_TF-IDF_RF	0.8542	0.7590	0.7528
Ensemble	TF-IDF_Lexicon_E_Clfs	0.8839	0.8309	<b>0.8480</b>
	TF-IDF_Lexicon_E_SVM	0.8731	<b>0.8386</b>	0.8435
	WE_Lexicon_TF-IDF_E_SVM	<b>0.8866</b>	0.8022	0.8310
	E_SVM_Approach	0.8825	0.8081	0.8289

## 5. Discussion

Graphs are presented for each metric used as a means of visually comparing the performance of the models of each approach in the different corpora. The models are

presented in different colors. Figure 15 shows the performance of the models in the F-measure metric in the 3 corpora used. It is seen that the model that obtains the best performance in this metric in the Chilean corpus is WE\_Lexicon\_SVM, with 0.8908. For the Mexican and the Chilean-Mexican corpora, it is the WE\_Lexicon\_TF-IDF\_SVM, with 0.8394 and 0.8507, respectively.

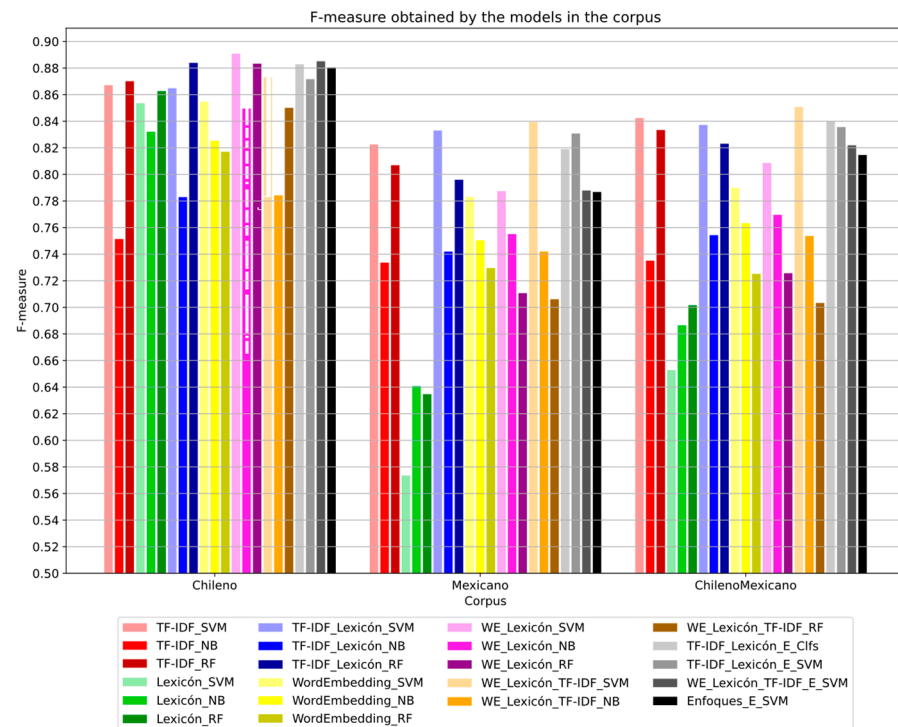


Figure 15. Comparison of F-measure obtained by the models.

On the other hand, Figure 16 shows that the model with the best performance in the Accuracy metric for the Chilean corpus is WE\_Lexicon\_SVM, with 0.892. For the Mexican and Chilean-Mexican corpora, it is the WE\_Lexicon\_TF-IDF-SVM, model with 0.8431 and 0.8548, respectively.

In general terms, it is seen that the models, in the different metrics, have a similar behavior. Meanwhile, the models generally have a better performance in the Chilean corpus, followed by the Chilean-Mexican one, and finally, the Mexican one. In the graphs, it is seen that the models with a hybrid approach have a better performance compared to the approaches that do not use Lexicons in the Chilean corpus, followed by the Mexican one, and finally, the Chilean-Mexican one. As can be seen in Table 8, the Chilean corpus processed with the 8 hybrid models outperforms the best model that does not use Lexicons. In the case of the Mexican corpus, there are 3 hybrid models that obtain better performance than the best model that does not use Lexicons, as seen in Table 9. Finally, in Table 10, it is seen that only one model outperforms the best hybrid model that does not use Lexicons in the Chilean-Mexican corpus.

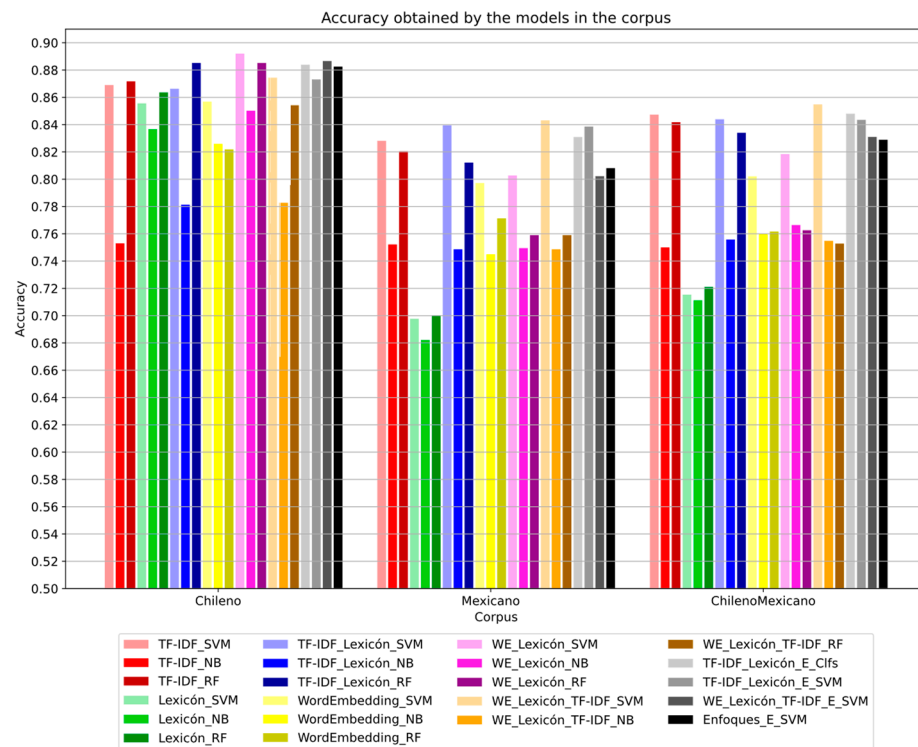


Figure 16. Comparison of Accuracy obtained by the models.

Table 8. Hybrid models outperforming the best non-Lexicon model in the Chilean corpus.

Model	F-Measure
WE_Lexicon_SVM	0.8908
WE_Lexicon_TF-IDF_E_SVM	0.8851
TF-IDF_Lexicon_RF	0.8839
WE_Lexicon_RF	0.8833
TF-IDF_Lexicon_E_Clf	0.8828
E_SVM_Approach	0.8804
WE_Lexicon_TF-IDF_SVM	0.8731
TFIDF_Lexicon_E_SVM	0.8716
TF-IDF_RF (Does not use Lexicons)	0.8701

Table 9. Hybrid models outperforming the best non-Lexicon model in the Mexican corpus.

Model	F-Measure
WE_Lexicon_TF-IDF_SVM	0.8394
TF-IDF_Lexicon_SVM	0.8330
TFIDF_Lexicon_E_SVM	0.8308
TF-IDF_SVM (Does not use Lexicons)	0.8225

Table 10. Hybrid models outperforming the best non-Lexicon model in the Chilean-Mexican corpus.

Model	F-Measure
WE_Lexicon_TF-IDF_SVM	0.8507
TF-IDF_SVM (Does not use Lexicons)	0.8424

As a summary, Table 11 shows the models that obtain the best results according to the F-measure and Accuracy metrics for each corpus used. It is seen that for the Chilean corpus, this is the WE\_Lexicon\_SVM model, and for the Mexican and Chilean-Mexican corpora, it is the WE\_Lexicon\_TF-IDF\_SVM. These three models use Word Embedding and

Lexicons to extract the features of the texts. This shows that the best results are obtained by incorporating this technique. On the other hand, it is seen that the best result is obtained in the Chilean corpus, achieving a value of 0.89 of F-measure and Accuracy, followed by the Chilean-Mexican corpus, and finally, the Mexican. This can be explained due to the lack of specific Mexican words in the different Lexicons used, especially the bad words Lexicon.

Finally, it is seen that the models with the best results used Support Vector Machine as a Machine Learning classifier. With this, it is reasserted that this seems to be a good algorithm to perform the text classification of the three algorithms tested.

**Table 11.** Models with the best performance in the corpus.

	Model	F-Measure	Accuracy
Chilean	WE_Lexicon_SVM	0.8908	0.8920
Mexican	WE_Lexicon_TF-IDF_SVM	0.8394	0.8431
Chilean-Mexican	WE_Lexicon_TF-IDF_SVM	0.8507	0.8548

Table 12 shows the results that obtain the best models of each corpus in the F-measure metric (Table 11), and the models of the base approaches that obtain the best results. It is seen that the broadest difference is found in the Chilean corpus, followed by the Mexican one, and finally, the Chilean-Mexican one. It can also be seen that the difference is broader with the models that use the Word Embedding-based approach.

**Table 12.** F-measure comparison of the best models with the best base models.

Corpus	Best Model	Best Model TF-IDF Approach	Best Model Word Embedding Approach
		F-Measure	
Chilean	0.8908	0.8701	0.8547
Mexican	0.8394	0.8225	0.7831
Chilean-Mexican	0.8507	0.8424	0.7900

## 6. Conclusions

This article presented several hybrid models, whose idea is using the Lexicon and Machine Learning approach to analyze emotions in user comments, specifically to detect aggression in texts written in Spanish. 5 approaches are proposed to create different models: Lexicon, TF\_IDF\_Lexicon, WE\_Lexicon, WE\_Lexicon\_TF-IDF, and the Ensemble approach, which differentiate mainly in the way of extracting the feature vector from the text. The 2 TF-IDF and Word Embedding approaches are also implemented, which do not use Lexicons, to compare them with the other models.

In each one of the models created, the best hyperparameters are sought from the training dataset of each corpus using GridSearchCV, to then perform experimentation on the test datasets and, through this, compare the results obtained in each model and select the best models in each one of the corpora. The models that obtained the best results use approaches that mix Word Embedding, Lexicons, and ML classifiers, outperforming the base models. The results indicate that hybrid models obtain the best results in the 3 corpora, over the models implemented that do not use Lexicons. This shows that, by mixing the approaches, the aggressiveness detection improves. It is worth highlighting that hybrid models have a better performance in the Chilean corpus, because the Lexicons have a better coverage or coincidence with Spanish words used in Chile, than what occurs with the Spanish used in Mexico.

On the other hand, all the models that obtain better results in the corpora use the Support Vector Machine as a classifier. Using the experiments that were run, it can be reasserted that this is one of the best algorithms to perform aggressiveness classification compared to the other algorithms used.



Finally, a web application was created, that allows showing the applicability of the proposed models, allowing classifying tweets or comments, evaluating the models implemented, and receiving user feedback on the prediction of the models, that allows generating a database for future research. It is worth mentioning that the backend of the web application is implemented as an API, meaning it can be used by external services.

In future work, incorporating Mexican words into the different Lexicons used is considered, especially in the bad words one, to check whether the performance of the models implemented on the Mexican corpus improves. Likewise, using different dictionary type Lexicons is considered, as these include more words than the Lexicon used in this work. The intention is also to implement the management of quantifiers, negations, and emojis in text preprocessing, as this work does not consider these. It is also considered important to incorporate other Ensemble models in the experimentation, using different ML classifiers. While it is felt that it is important to incorporate models based on neural networks in the future, to classify and mix these results with Lexicon-based models.

**Author Contributions:** Conceptualization, M.L.-F., A.S.-N. and C.V.-C.; Data curation, M.L.-F.; Formal analysis, M.L.-F., A.S.-N. and C.V.-C.; Investigation, M.L.-F., A.S.-N., C.V.-C. and C.M.-A.; Methodology, M.L.-F. and A.S.-N.; Software, M.L.-F.; Supervision, A.S.-N. and C.V.-C.; Validation, M.L.-F., C.M.-A. and C.R.-M.; Visualization, M.L.-F. and A.S.-N.; Writing—original draft, M.L.-F. and C.V.-C.; Writing—review & editing, A.S.-N., C.M.-A. and C.R.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** This work has been done in collaboration with the research group SOMOS (SOftware-MOdeling-Science), funded by the Research Agency of the Bío-Bío University, and the Engineering Faculty and Computer Science Department of the Universidad Católica de la Santísima Concepción, Chile.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Nocentini, A.; Calmaestra, J.; Schultze-Krumbholz, A.; Scheithauer, H.; Ortega, R.; Menesini, E. Cyberbullying: Labels, Behaviours and Definition in Three European Countries. *Aust. J. Guid. Couns.* **2010**, *20*, 129–142. [\[CrossRef\]](#)
2. Garaigordobil, M.; Mollo-Torrico, J.P.; Larrain, E. Prevalencia de Bullying y Cyberbullying En Latinoamérica: Una Revisión. *Rev. Iberoam. Psicol.* **2019**, *11*, 1–18. [\[CrossRef\]](#)
3. Cádiz, P.C.; Claro, M.; Peña, D.L.; Antezana, L.; Maldonado, L. *Implementación de Estudio de Usos, Oportunidades y Riesgos en el uso de TIC por Parte de Niños, Niñas y Adolescentes en Chile*; Pontificia Universidad Católica de Chile; Mineduc y OREALC/UNESCO: Santiago, Chile, 2017.
4. Hicks, J.; Jennings, L.; Jennings, S.; Berry, S.; Green, D.-A. Middle School Bullying: Student Reported Perceptions and Prevalence. *J. Child Adolesc. Couns.* **2018**, *4*, 195–208. [\[CrossRef\]](#)
5. Campbell, M.; Morgan, N. Adults Perceptions of Bullying in Early Childhood. In *Child and Adolescent Wellbeing and Violence Prevention in Schools*; Slee, P.T., Skrzypiec, G., Cefai, C., Eds.; Routledge: Oxfordshire, UK, 2017; pp. 101–108. [\[CrossRef\]](#)
6. Willard, N.E. *Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress*; Research Press: Champaign, IL, USA, 2007.
7. Scherer, K.R. Emotion as a Multicomponent Process: A Model and Some Cross-Cultural Data. *Personal. Soc. Psychol. Rev.* **1984**, *5*, 37–63.
8. Serhrouchni, A. Multilingual Cyberbullying Detection System. In Proceedings of the 2019 IEEE International Conference on Electro Information Technology (EIT), Brookings, SD, USA, 20–22 May 2019; pp. 40–44. [\[CrossRef\]](#)
9. Chatzakou, D.; Kourtellis, N.; Blackburn, J.; De Cristofaro, E.; Stringhini, G.; Vakali, A. Mean Birds: Detecting Aggression and Bullying on Twitter. In Proceedings of the 2017 ACM on Web Science Conference, Troy, NY, USA, 25–28 June 2017; pp. 13–22.
10. Gordeev, D. Automatic Detection of Verbal Aggression for Russian and American Imageboards. *Procedia—Soc. Behav. Sci.* **2016**, *236*, 71–75. [\[CrossRef\]](#)

11. Sharma, H.K.; Kshitiz, K. NLP and Machine Learning Techniques for Detecting Insulting Comments on Social Networking Platforms. In Proceedings of the 2018 International Conference on Advances in Computing and Communication Engineering, ICACCE 2018, Paris, France, 22–23 June 2018; pp. 265–272. [CrossRef]
12. Pawar, R.; Agrawal, Y.; Joshi, A.; Gorrepati, R.; Raje, R.R. Cyberbullying Detection System with Multiple Server Configurations. In Proceedings of the 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, USA, 3–5 May 2018; IEEE Computer Society: Washington, DC, USA, 2018; pp. 90–95. [CrossRef]
13. Balakrishnan, V.; Khan, S.; Fernandez, T.; Arabnia, H.R. Cyberbullying Detection on Twitter Using Big Five and Dark Triad Features. *Personal. Individ. Differ.* **2019**, *141*, 252–257. [CrossRef]
14. Al-Garadi, M.A.; Varathan, K.D.; Ravana, S.D. Cybercrime Detection in Online Communications: The Experimental Case of Cyberbullying Detection in the Twitter Network. *Comput. Hum. Behav.* **2016**, *63*, 433–443. [CrossRef]
15. Del Bosque, L.P.; Garza, S.E. Aggressive Text Detection for Cyberbullying. In Proceedings of the Mexican International Conference on Artificial Intelligence, Tuxtla Gutierrez, Mexico, 16–22 November 2014.
16. Murnion, S.; Buchanan, W.J.; Smales, A.; Russell, G. Machine Learning and Semantic Analysis of In-Game Chat for Cyberbullying. *Comput. Secur.* **2018**, *76*, 197–213. [CrossRef]
17. Ptaszynski, M.; Masui, F.; Nitta, T.; Hatakeyama, S.; Kimura, Y.; Rzepka, R.; Araki, K. Sustainable Cyberbullying Detection with Category-Maximized Relevance of Harmful Phrases and Double-Filtered Automatic Optimization. *Int. J. Child-Comput. Interact.* **2016**, *8*, 15–30. [CrossRef]
18. Leon-Paredes, G.A.; Palomeque-Leon, W.F.; Gallegos-Segovia, P.L.; Vintimilla-Tapia, P.E.; Bravo-Torres, J.F.; Barbosa-Santillan, L.I.; Paredes-Pinos, M.M. Presumptive Detection of Cyberbullying on Twitter through Natural Language Processing and Machine Learning in the Spanish Language. In Proceedings of the 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Valparaiso, Chile, 13–27 November 2019; pp. 1–7. [CrossRef]
19. Celdrán, P. *El Gran Libro de Los Insultos: Tesoro Crítico, Etimológico e Histórico de Los Insultos Españoles*; La Esfera de los Libros: Madrid, Spain, 2009.
20. Tapia, F.; Aguinaga, C.; Lujé, R. Detection of Behavior Patterns through Social Networks like Twitter, Using Data Mining Techniques as a Method to Detect Cyberbullying. In Proceedings of the 7th International Conference on Software Process Improvement (CIMPS), Guadalajara, Mexico, 17–19 October 2018; pp. 111–118. [CrossRef]
21. Montufar Mercado, R.; Chacca Chuctaya, H.; Castro Gutierrez, E. Automatic Cyberbullying Detection in Spanish-language Social Networks using Sentiment Analysis Techniques. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2018**, *9*. [CrossRef]
22. Loper, E.; Bird, S. NLTK: The Natural Language Toolkit. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*; Association for Computational Linguistics: Sydney, Australia, 2006; pp. 69–72.
23. Ríos, M.; Gravano, A. Spanish DAL: A Spanish Dictionary of Affect in Language. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, GA, USA, 14 June 2013; pp. 21–28.
24. Álvarez-Carmona, M.; Guzmán-Falcón, E.; Montes-y-Gómez, M.; Escalante, H.J.; Villaseñor-Pineda, L.; Reyes-Meza, V.; Rico-Sulayes, A. Overview of MEX-A3T at IberEval 2018: Authorship and Aggressiveness Analysis in Mexican Spanish Tweets. *CEUR Workshop Proc.* **2018**, *2150*, 74–96.
25. Graff, M.; Miranda-Jiménez, S.; Tellez, E.S.; Moctezuma, D.; Salgado, V.; Ortiz-Bejar, J.; Sánchez, C.N. INGEOTEC at MEX-A3T: Author Profiling and Aggressiveness Analysis in Twitter Using MTC and EvoMSA. *CEUR Workshop Proc.* **2018**, *2150*, 128–133.
26. Casavantes, M.; López, R.; González, L.C. UACH at MEX-A3T 2019: Preliminary Results on Detecting Aggressive Tweets by Adding Author Information via an Unsupervised Strategy. In Proceedings of the IberLEF@SEPLN, Bilbao, Spain, 24 September 2019; Volume 2421, pp. 537–543.
27. Cañete, J.; Chaperon, G.; Fuentes, R.; Ho, J.-H.; Kang, H.; Perez, J. Spanish Pre-Trained BERT Model. PML4DC, ICLR 2020. pp. 1–10. Available online: <https://users.dcc.uchile.cl/~ljperez/papers/pml4dc2020.pdf> (accessed on 28 October 2021).
28. Enrique Muñoz Cuza, C.; Liz De la Peña Sarracén, G.; Rosso, P. Attention Mechanism for Aggressive Detection. *CEUR Workshop Proc.* **2018**, *2150*, 114–118.
29. Guzman-Silverio, M.; Balderas-Paredes, Á.; López-Monroy, A.P. Transformers and Data Augmentation for Aggressiveness Detection in Mexican Spanish. *CEUR Workshop Proc.* **2020**, *2664*, 293–302.
30. Tanase, M.A.; Zaharia, G.E.; Cercel, D.C.; Dascalu, M. Detecting Aggressiveness in Mexican Spanish Social Media Content by Fine-Tuning Transformer-Based Models. *CEUR Workshop Proc.* **2020**, *2664*, 236–245.
31. Riquelme, R. Detección de Violencia Verbal Hacia Las Mujeres En Redes Sociales Mediante Técnicas de Aprendizaje Automático. [Memoria de Título]. 2019. Available online: [http://repobib.ubiobio.cl/jspui/bitstream/123456789/2692/1/Riquelme\\_Silva\\_Ricardo.pdf](http://repobib.ubiobio.cl/jspui/bitstream/123456789/2692/1/Riquelme_Silva_Ricardo.pdf) (accessed on 28 October 2021).
32. Lepe, M. Modelos Híbridos Basados En Lexicones y Machine Learning Para La Detección de Agresividad Sobre Textos En Idioma Español. [Memoria de grado]. 2021. Available online: [http://www.mcc.ubiobio.cl/docs/tesis/manuel\\_lepe-tesis\(manuellepe\).pdf](http://www.mcc.ubiobio.cl/docs/tesis/manuel_lepe-tesis(manuellepe).pdf) (accessed on 28 October 2021).
33. Segura Navarrete, A.; Martínez-Araneda, C.; Vidal-Castro, C.; Rubio-Manzano, C. A Novel Approach to the Creation of a Labelling Lexicon for Improving Emotion Analysis in Text. *Electronic Library*. 2021. Available online: [https://www.researchgate.net/publication/349119080\\_A\\_novel\\_approach\\_to\\_the\\_creation\\_of\\_a\\_labelling\\_lexicon\\_for\\_improving\\_emotion\\_analysis\\_in\\_text](https://www.researchgate.net/publication/349119080_A_novel_approach_to_the_creation_of_a_labelling_lexicon_for_improving_emotion_analysis_in_text) (accessed on 28 October 2021).

34. Plutchik, R. A General Psychoevolutionary Theory of Emotion. In *Emotion: Theory, Research, and Experience: Vol. 1. Theories of Emotion*; Plutchik, R., Kellerman, H., Eds.; Academic Press: New York, NY, USA, 1980; pp. 3–33.
35. Collobert, R.; Weston, J. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In Proceedings of the 25th International Conference on Machine Learning, New York, NY, USA, 5–9 July 2008; pp. 160–167. [[CrossRef](#)]
36. Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching Word Vectors with Subword Information. *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 135–146. [[CrossRef](#)]
37. Cardellino, C. Spanish Billion Words Corpus and Embeddings. 2016. Available online: <https://crscardellino.github.io/SBWCE/> (accessed on 28 October 2021).
38. Kotu, V.; Deshpande, B. Data Mining Process. In *Predictive Analytics and Data Mining*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 17–36. [[CrossRef](#)]