Detecting cyberbullying in Spanish texts throughout deep learning techniques

Paúl Cumba-Armijos,
Diego Riofrío-Luzcando*,
Verónica Rodríguez-Arboleda and
Joe Carrión-Jumbo

Digital School,
SEK International University,
Quito, Pichincha, Ecuador
Fax: +593-3994800
Email: pcumba.mti@uisek.edu.ec
Email: diego.riofrio@uisek.edu.ec
Email: veronica.rodriguez@uisek.edu.ec
Email: joe.carrion@uisek.edu.ec
*Corresponding author

Abstract: Recent recollected data suggests that it is possible to automatically detect events that may negatively affect the most vulnerable parts of our society, by using any communication technology like social networks or messaging applications. This research consolidates and prepares a corpus with Spanish bullying expressions taken from Twitter in order to use them as an input to train a convolutional neuronal network through deep learning techniques. As a result of this training, a predictive model was created, which can identify Spanish cyberbullying expressions such as insults, racism, homophobic attacks, and so on.

Keywords: cyberbullying; deep learning; convolutional neuronal network; Spanish; social networks.

Reference to this paper should be made as follows: Cumba-Armijos, P., Riofrío-Luzcando, D., Rodríguez-Arboleda, V. and Carrión-Jumbo, J. (xxxx) 'Detecting cyberbullying in Spanish texts throughout deep learning techniques', *Int. J. Data Mining, Modelling and Management*, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes: Paúl Cumba-Armijos obtained his Masters degree from the SEK International University in 2018, and his degree of Engineer in Computer Systems Engineering from the Polytechnic School of Chimborazo in 2012. He has worked on machine learning techniques for word processing in the field academic and has experience in agile software development techniques in the business field and multi platform development.

Diego Riofrío-Luzcando is the Director of the UISEK Digital School. He has a PhD in Software and Systems (UPM 2017), a Master in Software and Systems (UPM 2012) and a Computer Science and Computing Engineering (National Polytechnic School of Ecuador 2006). His main research interests include virtual education and training environments, smart tutoring systems, data science, and digital transformation.

Verónica Rodríguez-Arboleda is the Coordinator of Basic and Generic Subjects of the UISEK. She is the Teacher of the Master of Cybersecurity at Digital School. She is an accredited researcher at the Senescyt, Master in Business Administration. She has a Master in University Teaching. Computer Science Systems Engineer. Her research interests are information technologies applied to education, virtual environments, cybersecurity.

Joe Carrión-Jumbo received his PhD grade of Computer Science at Universitat Autónoma de Barcelona (UAB) in 2017. He received his Masters degree from the Computer Architecture and Operating Systems (CAOS) at UAB in 2013. He worked for a private company for ten years with four years as manager of the IT Department. He is focused in studies related in areas related to database, data mining and business intelligence.

1 Introduction

By allowing users to share and publish information in the different social networks freely, the problems especially to vulnerable groups such as children and teenagers have increased (David-Ferdon and Hertz, 2007).

This phenomenon is aggravated when minors commit suicide, which has increased in recent years due to the appearance of cyberbullying, mainly because this type of violence does not end when the minor returns to his/her home, unlike the situations of face-to-face violence (Girard, 2017).

According to the Cybersecurity Research Center, about 28% of students from middle and high schools around the USA have been victims of cyberbullying (Patchin, 2016). In Spain, The ANAR Fundation and Mutua Madrileña (ANAR and Mutua Madrileña, 2016) affirm that one in four harassments that occurred in Spanish schools are produced using technological means. It was determined that 30.7% of adolescents suffered from bullying through the internet in Latin America (ESET, 2014). For example, in some cities of Ecuador (Quito, Manta and Guayaquil), a high level of grooming and cyberbullying through social networks has been detected, about 27% of teenagers have suffered marginalisation, 46%harassment, 17% aggression and 10% extortion (Ortega Mora, 2013).

To prevent this type of aggression, it is essential to provide preventive control measures, for example identifying attacks on time (Nandhini and Sheeba, 2015). Data analytics techniques can be used as a prevention tool to detect this type of incidents.

Therefore, this work is the creation of a model using deep learning techniques that identify cyber bullying attacks in Spanish written texts. 83,400 tweets have been used to identify these attacks and compared to texts with no violence at all.

The structure of the remainder of the paper is as follows. Section 2 states relevant works in the fields of sentiment analysis in social networks and automatic cyber bullying prediction. In the Section 3 there will be a description of the proposed model which would detect cyber bullying in texts. Section 4 reports model validation which details the method observed in this study and discusses its results. Section 5 outlines the conclusions of this research and some takes a look at further work within its field.

2 Related work

The related work is divided into three sections.

Section 2.1 briefly presents the main goals, behaviour analysis in social networks, especially how to interpret the sentiments in texts through the use of natural language processing (Spanish). Section 2.2 summarises the reviewed literature about social networks and Cyberbullying published on specialised journals. Finally, Section 2.3 focuses on the use of machine and deep learning to create prediction models that can detect cyberbullying.

2.1 Sentiment analysis in social networks

Historically, the study of human behaviour has been approached with great interest from several disciplines, and this kind of researches is currently being carried out within social networks. Some authors highlight the importance of these services since they allow to extract patterns of a large amount of data to understand the behaviour of the users in order to satisfy their needs in the consumption of information (Hu and Liu, 2012). This work extends this idea by focusing in a real context based on Social networks with specific needs related to a social problem.

For example, some researches are used to understand human sociological behaviour through a probabilistic and statistical model (Zhang et al., 2011), in order to apply this hypothesis text messages have been gathered and classified manually to ensure the quality of the model. Other works use a natural language analysis in order to identify the personality, social situation and interpersonal relationships that are established among individuals in a social network (Pennebaker, 2002). Instead of identifying personal as well as social features, this work purposes classifying expressions for applying this model. There are researches that classify the main aspects that affect human aging based on data mining techniques (Squicciarini et al., 2017). Thus, this work applies techniques to a specific part of psychological human aspects which is the way the humans express their sentiments by analysing the natural language.

Data analysis in social networks can also be used as a mean to improve business, for example to identify rumours in business Twitter accounts (Kaya and Alhajj, 2019), in contrast this work uses the messages of the accounts and the accounts a tried anonymously, or to create customised advertising to the needs of the buyer (Schiaffino and Amandi, 2009). Instead of classifying the user accounts to raise actions, this work classifies messages in an offline way and create models to classify future messages. In this area, sentiments behind comments of users in social networks can be analysed to measure the loyalty and interests of the clients with respect to any brand or product (Neri et al., 2012), which can be obtained by categorising opinions (positive or negative) (Baracho et al., 2012).

Machine learning is one of the techniques that can help analysing data in social networks that can be implemented. For example, to classify the sentiments in positive, negative or neutral depending on the user's opinion about other people, organisations, events, products, services, places, ideas, among others (Beigi et al., 2016; Jayasanka et al., 2013). In works where deep learning is the technique used, the purpose of the models created where the same as those mentioned above, with the difference that the type of neuronal network adopted was the convolutional neural network (Lu et al., 2017; Severyn and Moschitti, 2015).

2.2 Cyberbulling and social networks

In order to identify relevant studies about cyberbullying, a search of indexed journals was carried out. Two search criteria were used, the first one about cyberbullying in specialised data mining journals. The second group of searches on specialised magazines in social networks.

Criterion 1 *Journals on Data Mining:* The journals indexed as Q1, Q2, Q3 and Q4 in the SJR ranking were searched. The title and area of the journal should include the phrases 'data mining'. In total, 12 journals were found, in the Q1 group 8 journals were identified, one in Q2, one in Q3 and two journals in Q4. In the 12 magazines with a direct focus on Data mining, articles on cyberbullying were searched. The result was 4 articles.

Criterion 2 *Magazines on Social Networks:* Magazines Q1 to Q4 that include 'social network' in their title were searched. Thirty four journals were identified among all quartiles. In these journals, articles that include the word cyberbulling in the Title, Abstract and keywords were searched. The result was 67 articles.

In the total of 71 articles, those ones that perform analysis or detection in the content related to cyberbullying were searched. Through a review of the abstract of the articles, 5 articles that propose models of content analysis for cyberbullying detection were identified.

It can be seen in Silva et al. (2018) that they develop an automated model to identify and measure the degree of cyberbullying on Facebook. This research uses the Twitter Social network so it allows us to analyse a different group of users instead of Facebook. In the article, Raisi and Huang (2018) propose a weakly supervised model by means of vocabulary analysis using machine learning techniques which have been evaluated on Twitter, Ask.fm and Instagram networks. This work uses Neural Networks instead and it tries to improve the accuracy for a specific social area which is the cyberbullying behaviour.

The analysis of social networks based on the identification of positive and negative behaviour is studied by Squicciarini et al. (2017), in the article the complexity of the models is analysed due to the high volume of traffic in the networks and their high exchange rate, positive and negative sentiments can be extended to several areas or topics and uses generic terminology. On the other hand, this work aim is to focus on cyberbullying attacks as a main goal.

Some networks that are not currently in use have been studied by Rafiq et al. (2016), in this case the authors analyse the video content of the extinct Vine network by downloading and tagging videos to model and content. The tagging technique is very useful to validate a model, the same technique in order to create a Corpus will be applied. The video tagging strategy to search similar images, audios and videos is also proposed by Soni and Singh (2018), who use a corpus created by Rafiq et al. (2016) to analyse behaviour in text and audios of the videos. The 66-remaining works analyse the presence of cyberbullying in social networks as per in general content.

2.3 Cyberbullying prediction

As mentioned above, machine learning is used to detect sentiments in texts based on a given model. In this way some authors try to identify who?, what?, why?, where? and when? An episode of bullying occurs (Bellmore et al., 2015). Proposed model can be used to the same goal and compare the result with other models. In order to accomplish this goal, 9,764,583 bullying posts were compiled from Twitter to obtain a classification on different categories: Yes or No (is cyberbullying), the role of the author (who), the way bullying is carried out (what), the post typology (why).

Specifically, for the analysis of feelings and the detection of cyberbullying in social networks in Spanish language, machine learning was used in the development of a Bayesian classifier, which was trained with a set of data obtained from the Faculty of Exact Sciences of the University of San Agustín in Buenos Aires (Mercado et al., 2018).

Thus, two researches for the detection of cyberbullying can be highlighted. The objective of the first was to identify global polarities as positive, negative and neutral from texts in Japanese (Ptaszynski et al., 2017), for which the authors had to use morphological analysers to transcribe the texts into English, it is very important to apply this kind of models to other languages like Spanish. And the second research compares results of machine learning techniques with those of deep learning to predict cyberbullying in English, obtaining better results for convolutional neural network models (Agrawal and Awekar, 2018).

Most researches that focus on cyberbullying prediction have been carried out in languages other than Spanish (Arabic, English), being convolutional neural network, the most common method used (Bayari and Bensefia, 2021). This predilection for deep learning techniques may occur due to the fact that it has best-in-class performance on problems in text classification domain (Altınel and Ganiz, 2018).

3 Cyberbullying detection model

3.1 Model architecture

The proposed architecture is structured in phases (Figure 1), which allows to establish a controlled and independent flow of information. Each phase determines one or several processes used for the preparation of the neural network.

As it can be appreciated in 1, the first phase is responsible for collecting the information to train the model, for which 83,400 tweets were obtained. The second phase defines the scheme of data extraction, this process was developed in a Python script which interacts with the social network Twitter through its API.

In the data preparation phase, each tweet was manually classified and labelled with zero (no bullying) or one (bullying), depending on the context and keywords (Del Rey et al., 2002), resulting on 16,247 bullying labeled tweets and 67,153 no bullying labeled tweets. In this phase, data quality has been improved in order to avoid words or symbols that may cause noise in model training. For this, a Python script was also developed in which the following actions were executed:

- Convert text to lowercase.
- Delete all special characters used in each tweet. Only letters remain.
- Remove references of external links (‘http’, ‘https’, ‘www’).
- Eliminate mentions to other users ‘@user’.
- Remove stop words.

Finally, a convolutional neural network was trained using these preprocessed tweets. For this, a Python script was developed by means of Keras 3 module. The convolutional neural networks are an excellent classifier of sentences, because this type of networks can capitalise the distribution of words in a sentence by converting them into a vector matrix (Kim, 2014).

Figure 1 Cyberbullying detection mode architecture (see online version for colours)

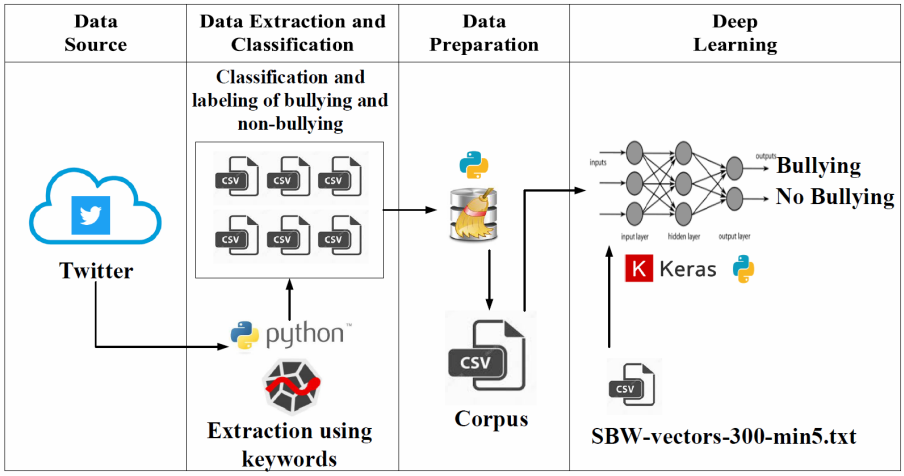


Table 1 Example of keywords for bullying expressions (Spanish)

care asno	care culo	care huevo
eres regalada	homosexual	gay
maricon	mamerto	mojigato
nerd	ojala mueras	orejas elefante
te den culo	te odio	tetas chuecas
care simio	care verga	eres gorda
borrego	maldito	marica
narizon	negro	apestoso
parece cerda	parece marrana	parece marrano
eres deprimente	eres discapacitado	eres patetico

Table 2 Example of keywords for no bullying expressions (Spanish)

gusta	me gustas	linda
gustoso	eres mejor	lindo
pago entradas	animo	enamorarse
hola	amiga	frio
amigo	cuidate	calor
te quiero	gustoso conocerte	tener autoestima
te amo	como estas	vive feliz

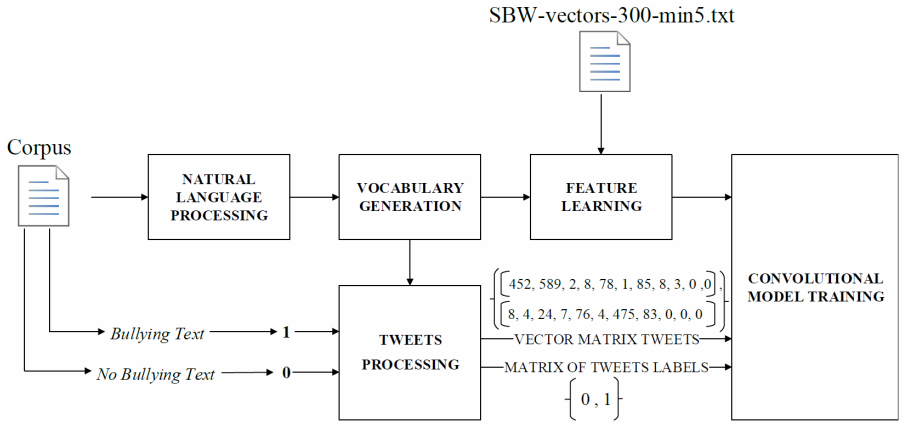
3.2 Data extraction

For the search and extraction of the tweets the labels that young students between 8 and 14 years of age used to denominate the phenomenon of bullying were considered (Del Rey et al., 2002), to search for texts of verbal aggression (bullying), to search for texts of verbal aggression (bullying). Labels such as insults, racist attacks and serious verbal violence (Table 1). At the same time, opposite words to the violent ones were used to find harmless tweets (Table 2).

3.3 Creating the model

A flow with five tasks was defined in order to create the model (Figure 2), based on the traditional machine learning process (Raschka, 2015).

Figure 2 Generation process of the model (see online version for colours)



3.3.1 Natural language processing

The frequency distribution of the words corpus is obtained through the NLTK library.

3.3.2 Vocabulary generation

This vocabulary is created based on the frequency distribution, by implementing the most common words and then an indexed matrix. This matrix starts with the most frequent word assigning the number one (1) as an index, the second most frequent word is assigned the number two (2) as an index, and so on.

3.3.3 Tweets processing

This process decomposes the corpus of the training data set into three vectors.

- Tweets. Stores the values of the indexes, which represent the words of the corpus based on the frequency.
- Labels. Stores the values of the labels: one (1) if is bullying and zero (0) if it is not.
- Ids. Stores the ids of each tweet.

3.3.4 Feature learning

For this, an incrustation of words in Spanish was used, words vectorially represented through a linguistic model consolidated in a file (SBW-vectors-300-min5.txt) from different resources published on the web, and made up of 1.4 billion words in Spanish (Cardellino, 2016). This model establishes analogies between words and their proximity relationships through a vectoral representation, and it was created by using the word2vec algorithm (Mikolov et al., 2013).

3.3.5 Model training

As stated above, a convolutional neural network was used to train the model. This network receives the corpus of preprocessed tweets as input and emits predictions whether a text is bullying or not as output.

4 Model validation

4.1 Method

The validation of the model was carried out in two parts. First, a validation of the consistency of the corpus and then a cross-validation of the model to find the accuracy of its predictions.

4.1.1 Corpus validation

As previously stated, the corpus was structured from a total of 83,400 tweets that describe texts that have bullying and non-bullying signs. Therefore, it is important to establish that the words used to create the corpus are consistent and satisfy the minimum requirements to represent expressions of interpersonal communication.

For this, an analysis of the linguistic distribution of the words in tweets used for the Corpus was carried out, as stated by Zipf (1950). This law expresses that people use a reduced number of words most of the time, while the vast majority of words are used

very rarely in the linguistic communication. This use of words follows a distribution that can be represented by:

$$f(r) = \alpha \frac{1}{r^{\alpha}}; \alpha \approx 1$$

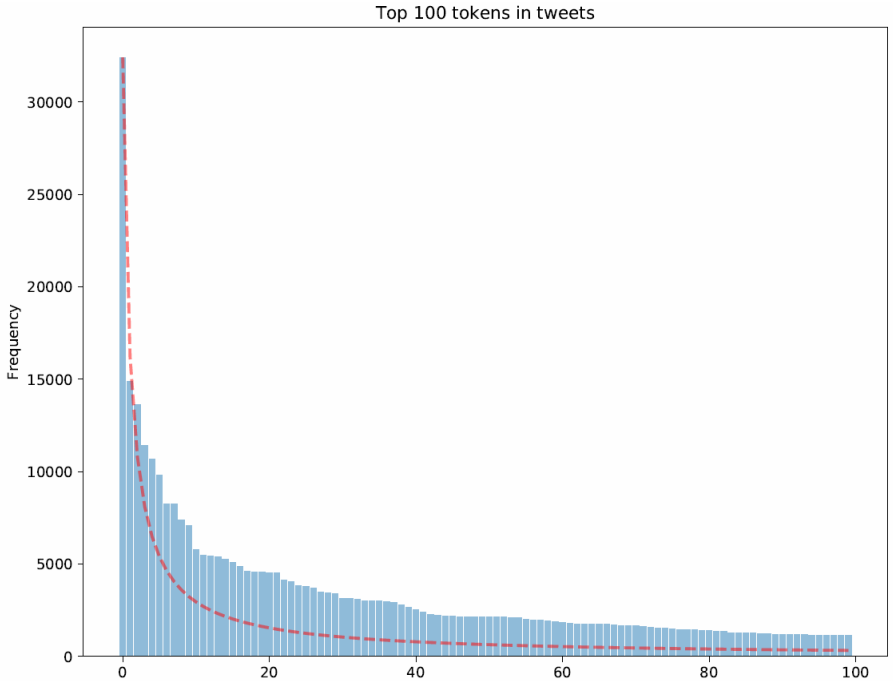
where $f(r)$ represents the frequency that it is most used and α is a number that approaches one. This equation means that the frequency of each word is inversely proportional to its location range according to its frequency, so, the most frequent word will use twice more than the second most frequent word and three times more than the most used words, and so on.

4.1.2 Model cross-validation

For this validation, the total data set was divided into two subsets (training and test sets), after that the validation was carried out in four random iterations. The proportion to perform the data split was 90% for the training set (75,059 tweets) and 10% for the test set (8,341 tweets).

Once the different data sets were obtained, the convolutional network was trained during each iteration with eight epochs.

Figure 3 Top 100 of the frequency distribution of the words in the corpus (see online version for colours)



In each iteration at the end of each epoch a checkpoint was established, storing a model as a result of each epoch. That is to say, eight epochs were executed for each training iteration and a model was stored at each checkpoint, which generated a total of 32 models with their respective accuracy and loss.

These metrics were used to select the best generated model in each training iteration and for each epoch. Thus, in terms of the model that has the highest precision and the least loss.

Once the best model of each iteration was selected, a validation of how well that model predicts the data from the test set was made, with which results of success and failure were obtained for each prediction.

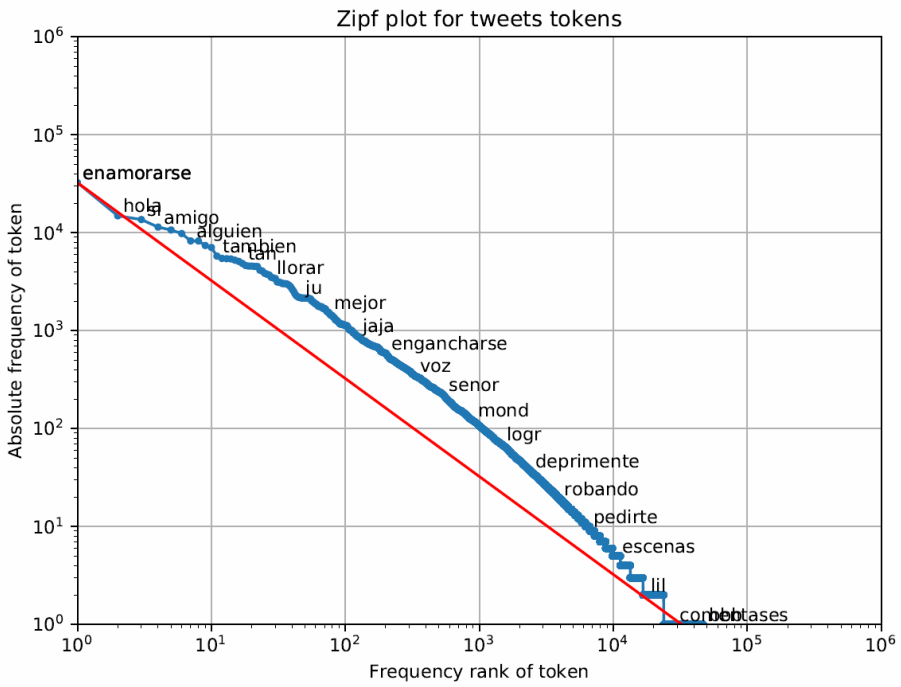
4.2 Results

4.2.1 Corpus validation

Figure 3 shows the distribution according to the frequency range of the words of the bullying and non-bullying tweets used in the corpus, while red dotted line represents the Zipfian distribution.

Another representation to verify that the distribution of the corpus complies with Zipf's law is shown 4, which represents the frequency range of the words used in the corpus vs. its absolute frequency.

Figure 4 Words distribution vs. a Zipf distribution (see online version for colours)



When plotting the frequencies of the words in the plane, an almost linear curve approximates to the red line can be observed, which represents the distribution of Zipf.

4.2.2 Model cross-validation

As stated in the Section 4.1, eight epochs were used for each training iteration of the model and for each iteration the best model was selected. The accuracy and loss of each of these selected models are presented in Table 3.

Table 3 Model selected for each training iteration

Training iteration	*Selected model*	*Accuracy*	*Lost*
1	Epoch 8	99.80%	0.006
2	Epoch 8	99.80%	0.005
3	Epoch 8	99.80%	0.006
4	Epoch 8	99.80%	0.005

With the best selected model, the cross-validation was carried out to prove how well the model predicts a text with and without bullying. Results that are presented in Table 4, where the model hits a 98.85% of the time average prediction.

Table 4 Results of success and failure in the prediction of test set for each iteration

Iteration	*Success percentage*	*Fail percentage*
1	98.96%	1.04%
2	98.84%	1.16%
3	98.92%	1.08%
4	98.67%	1.33%
Average	98.85%	1.15%

4.3 Discussion

Even though in Figure 3 it can be seen that the use of the words in the corpus resembles that distribution, it can also be observed in Figure 4 that there is a deviation of the corpus distribution curve from the expected distribution, which means that in the corpus some words occur with a higher frequency range. This may mean that the tweets used for the corpus have more frequent words than normal expressions. This may be due to the fact that keywords were used to obtain the corpus and these are the most frequent words in the corpus.

In the cross-validation, the results obtained allow us to show that the models in the different validation iterations do not have a high variability among them. Besides that, in each one of the iterations and on average a high percentage of precision is presented.

5 Conclusions and future work

This research presents a contribution to the development of systems that prevent and alert cases of cyberbullying, this is accomplished by taking advantage of the model generated from a corpus with Spanish expressions labelled as bullying and non-bullying.

In the extraction of data from Twitter, the importance of retrieving focused data using keywords is highlighted, which allow us to obtain the expressions that represents concrete situations of interpersonal violence. In addition, this process allowed us to obtain a large amount of information, which was an essential pillar for the preparation of the model.

The training process of the convolutional neuronal network was carried out efficiently using deep learning techniques, since the convolution layers of the neural network facilitated the extraction and the understanding of the characteristics of the linguistic relationship of the words in the expressions that had signs of bullying, thanks to the help of a pre-trained vector matrix with Spanish words. In this process the data preparation that was made to the corpus was of essence, since this guaranteed that the processing of the model was efficient in time and quality of training.

The use of the cross-validation technique allowed us to verify that the models had in average a very high precision (98.19%) when predicting the test data.

Inclusion of data from other sources like other social networks or blogs will be needed in order to generate a larger corpus for a future work with the objective of feeding the neural network with more linguistic characteristics that can be obtained from other expressions used in Spanish, in order to generate a model that can improve the prediction values.

References

Agrawal, S. and Awekar, A. (2018) 'Deep learning for detecting cyberbullying across multiple social media platforms', *European Conference on Information Retrieval*, Springer, pp.141–153.

Altınel, B. and Ganiz, M.C. (2018) 'Semantic text classification: a survey of past and recent advances', *Information Processing & Management*, Vol. 54, No. 6, pp.1129–1153.

ANAR, F. and Mutua Madrileña, F. (2016) *Estudio sobre el ciberbullying según los afectados*; ANAR Fundation, Mutua Madrileña Fundation, Madrid, Tech. Rep. [online] http://www.anar.org/wp-content/uploads/2016/04/Estudio-Bullying-Seg%C3%BAn-los-Afectados-Abril-2016.pdf (accessed 27 October 2018).

Baracho, R.M.A., Silva, G.C. and Ferreira, L.G.F. (2012) 'Sentiment analysis in social networks: a study on vehicles', *ONTOBRAS-MOST*, pp.132–143 [online] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.416.8828&rep=rep1&type=pdf#page=132 (accessed 15 January 2021).

Bayari, R. and Bensefia, A. (2021) 'Text mining techniques for cyberbullying detection: state of the art', *Advances in Science, Technology and Engineering Systems*, Vol. 6, No. 1, pp.783–790.

Beigi, G., Hu, X., Maciejewski, R. and Liu, H. (2016) 'An overview of sentiment analysis in social media and its applications in disaster relief', *Sentiment Analysis and Ontology Engineering*, pp.313–340, Springer, Cham., Switzerland.

Bellmore, A., Calvin, A.J., Xu, J.M. and Zhu, X. (2015) 'The five W's of 'bullying' on Twitter: who, what, why, where, and when', *Computers in Human Behavior*, Vol. 44, pp.305–314 [online] https://www.sciencedirect.com/science/article/abs/pii/S0747563214006621?via%3Dihub.

Cardellino, C. (2016) *Spanish Billion Words Corpus and Embeddings* [online] https://crscardellino.github.io/SBWCE/ (accessed 27 October 2018).

David-Ferdon, C. and Hertz, M.F. (2007) 'Electronic media, violence, and adolescents: an emerging public health problem', *Journal of Adolescent Health*, Vol. 41, No. 6, pp.S1–S5.

Del Rey, R., Genevat, R. and Ruiz, R.O. (2002) 'Etiquetas verbales en el vocabulario de docentes, padres y madres para nominar el fenómeno bullying', *Revista electrónica interuniversitaria de formación del profesorado*, Vol. 5, No. 4, p.11.

ESET (2014) *Reporte Sustentabilidad*, Tech. Rep., [online] http://www.eset-la.com/microsites /responsabilidad-social/pdf/eset-reporte-resumido.pdf (accessed 17 February 2018).

Girard, G. (2017) 'El suicidio en la adolescencia y en la juventud', *Revista de Formación Continuada de La Sociedad Española de Medicina de La Adolescencia*.

Hu, X. and Liu, H. (2012) *Text Analytics in Social Media*, pp.385–414, Mining text data, Springer, Boston, USA.

Jayasanka, R.A.S.C., Madhushani, T., Marcus, E., Aberathne, I.A.A.U. and Premaratne, S. (2013) Sentiment analysis for social media', in *Information Technology Research Symposium*, November, Vol. 11, p.22.

Kaya, M. and Alhajj, R. (2019) *Influence and Behavior Analysis in Social Networks and Social Media*, Springer, Cham., Switzerland.

Kim, Y. (2014) 'Convolutional neural networks for sentence classification', *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 25–29 October, Doha, Qatar, c 2014 Association for Computational Linguistics, pp.1746–1751, arXiv preprint arXiv:1408.5882.

Lu, Y., Sakamoto, K., Shibuki, H. and Mori, T. (2017) 'Are deep learning methods better for twitter sentiment analysis', *Proc. 23rd Annual Meeting of Natural Language Processing*, Japan, pp.787–790.

Mercado, R.N.M., Chuctaya, H.F.C. and Gutierrez, E.G.C. (2018) 'Automatic cyberbullying detection in spanish-language social networks using sentiment analysis techniques', *International Journal of Advanced Computer Science and Applications*, Vol. 9, No. 7, pp.228–235.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013) 'Distributed representations of words and phrases and their compositionality', *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Vol. 2, pp.3111–3119.

Nandhini, B.S. and Sheeba, J. (2015) 'Online social network bullying detection using intelligence techniques', *Procedia Computer Science*, Vol. 45, pp.485–492 [online] https://www.sciencedirect.com/science/article/pii/ S187705091500321X (accessed 01 May 2018).

Neri, F., Aliprandi, C., Capeci, F., Cuadros, M. and By, T. (2012) 'Sentiment analysis on social media', *2012 IEEE/ACMInternational Conference on Advances in Social Networks Analysis and Mining*, IEEE, pp.919–926.

Ortega Mora, A.C. (2013) 'Manifestaciones de la agresión verbal entre adolescentes escolarizados', (Tesis de Licenciatura Educativa, Universidad de Cuenca), Repositorio Institucional – Universidad de Cuenca.

Patchin, J. (2016) *Summary of Our Cyberbullying Research (2004-2016)* [online] https://cyberbullying.org/summary-of-our-cyberbullying-research (accessed 22 February 2018).

Pennebaker, J.W. (2002) 'What our words can say about us: toward a broader language psychology', *Psychological Science Agenda*, Vol. 15, No. 1, pp.8–9.

Ptaszynski, M., Eronen, J.K.K. and Masui, F. (2017) 'Learning deep on cyberbullying is always better than brute force', *IJCAI 2017 3rd Workshop on Linguistic and Cognitive Approaches to Dialogue Agents (LaCATODA 2017)*, Melbourne, Australia, pp.19–25.

Rafiq, R.I., Hosseinmardi, H., Mattson, S.A., Han, R., Lv, Q. and Mishra, S. (2016) 'Analysis and detection of labeled cyberbullying instances in Vine, a video-based social network', *Social Network Analysis and Mining*, Vol. 6, No. 1, p.88.

Raisi, E. and Huang, B. (2018) 'Weakly supervised cyberbullying detection with participant vocabulary consistency', *Social Network Analysis and Mining*, Vol. 8, No. 1, p.38.

Raschka, S. (2015) *Python Machine Learning*, Packt Publishing Ltd., Birmingham.

Schiaffino, S. and Amandi, A. (2009) 'Intelligent user profiling', *Artificial Intelligence An International Perspective*, pp.193–216, Springer, Berlin.

Severyn, A. and Moschitti, A. (2015) 'Twitter sentiment analysis with deep convolutional neural networks', *Proc. 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp.959–962.

Silva, Y.N., Hall, D.L. and Rich, C. (2018) 'BullyBlocker: toward an interdisciplinary approach to identify cyberbullying', *Social Network Analysis and Mining*, Vol. 8, No. 1, p.18.

Soni, D. and Singh, V.K. (2018) 'See no evil, hear no evil: audio-visual-textual cyberbullying detection', *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, CSCW, pp.1–26.

Squicciarini, A., Rajtmajer, S. and Griffin, C. (2017) 'Positive and negative behavioral analysis in social networks', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 7, No. 3, p.e1203.

Zhang, H., Dantu, R. and Cangussu, J.W. (2011) 'Socioscope: human relationship and behavior analysis in social networks', *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 41, No. 6, pp.1122–1143.

Zipf, G.K. (1950) 'Human behavior and the principle of least effort', *Journal of Clinical Psychology*, 1949, Vol. 6, No. 3, pp.306–306, p.573, Addison-Wesley, Cambridge, Mass.