# SQYQP@Vaxxstance: Stance Detection for the Antivaxxers Movement

Jesús Calleja and Ariane Méndez

University of the Basque Country UPV/EHU, Spain
{jcalleja009,amendez026}@ikasle.ehu.eus

**Abstract.** In this paper an approach regarding the Stance Detection task for the Antivaxxers Movement is described. Systems for stance classification in Basque and Spanish are built using different textual and contextual information, features and dataset sizes. In addition, different word embeddings are tried when training models, with the aim of identifying how they perform in varying language settings. The results show that contextual models work better, specially in Basque, where a big performance improvement is achieved by adding contextual data.

**Keywords:** Stance detection · Multilingualism · Text categorization · Deep learning · Feature engineering · Word embeddings.



## 1 Introduction

The anti vaccine movement in the world has grown up recently because of the new COVID vaccines that are being authorized to use for the general public in the shortest time the world has ever seen. This rapidness has turned on the alarms on the society and more and more sceptic people are appearing. The measurement of this trend can be made through social media, where millions of people post their opinions online, so that everyone can read them. The objective of this project is to, given a sample of the comments about vaccines that are being written on Twitter, build a model that is able to classify opinions shared on social media about vaccines.

Such tasks, where opinions posted on social media are to be classified according to the posture they express regarding a certain topic, take place within the Stance Detection field [3, 5, 7, 8], which has appeared relatively recently in the NLP area. The growth of social media, the ability to process a great size of data and the improvement of text processing techniques has allowed the scientific community and enterprises to take advantage of this new and ever growing information in order to understand the opinion of the public about certain topics.

*IberLEF 2021, September 2021, Málaga, Spain.*

In this particular scenario, our objective is to classify tweets that talk about vaccines in three categories: in FAVOR of the vaccines, AGAINST the vaccines or NONE of them where either the user is neutral or the stance is unclear. We also aim to analyze whether contextual information, such as Twitter user follower connections, is useful for achieving better predictions than those made by sole use of tweets.

In order to complete the task, two main types of models are built in this work: textual models and contextual models. On the one hand, textual models do not make use of anything other than the content of the tweets. In short, different types of embeddings are used (depending on the language) and fed to an LSTM to get the general representation of the text. This representation is later fed to a linear layer that outputs the final stance for each tweet. See Section 4.1 for further details on this approach. On the other hand, contextual models use the confidences for each label (obtained from the textual models) and the mean distance from a user that wrote a certain tweet to other users who wrote FAVOR, AGAINST and NONE tweets. The final features are inserted into a Multilayer Perceptron. This procedure is explained in detail in Section 4.2.

This work is especially attractive because it is performed in two different languages, Spanish and Basque, the last one being low-resourced. This can show the challenges of trying to build a classifier (or any other model) with few data in different languages and push new techniques to embrace this problem into the state of the art.

## 2 Related Work

Stance Detection is a task that has gained popularity due to the recent ascent of social media. For this reason, the works that can be found in the field are quite new. The most important ones are gathered in past tasks [3, 7, 8] where different issues and techniques were proposed.

To start with, in SemEval-2016 Task 6 [7], a dataset [6] about different social issues was given to the participants with the objective of researching new ways to deal with Stance Detection. As an additional subtask, methods for weakly supervised classification were discussed. The most important findings discovered in this task were that the task was still novel, as the methods used to classify the tweets were not very sophisticated. Some of the participants used external knowledge, like lexicons or pre-trained word embeddings from Google News or directly from Twitter. At this time, the use of neural networks was reduced to recurrent neural networks, at most.

In the MultiStanceCat task from IberLEF 2018 [8], the topic was the Independence Referendum in Catalunya. The data made available was formed by tweets about the topic. Participants used different approaches; however, most of them strived towards pre-processing the data, i.e. getting contextual information like mentions and hashtags in the tweet and the usage of Support Vector Machines (SVMs). The team named "Casacufans" also used the images that went along with the tweets to try to get more information, employing a Convolutional

Neural Network (CNN) to do so. Results showed that by using contextual information classifiers were able to detect the stance of the tweets significantly better. Regrettably, as reported by the organizers of the task, the members of said team did not provide a working note explaining their approach in detail [8].

The last task that was part of the Evaluation of NLP and Speech Tools for Italian (EVALITA) campaign in 2020, SardiStance [3], officially added the subtask of using contextual information. In this case, in addition to the tweets about the Sardine Movement, contextual information such as followers and other additional data about users was also provided. For this scenario, pre-trained language models based on the Transformer architecture [9] (i.e. AlBERTo, GilBERTo, UmBERTo) were used to get a better representation of the textual information of the tweets. In the contextual subtask, participants used different features from Twitter and other sources with different types of information (psychological, emotional). From this task, the proposal in [4] was specially inspiring, where, as contextual information, the followers' network was used in order to get the distance to/from against, favor and neutral users.

In the task were this work is based, apart from textual and contextual information, new subtasks were available, like the open track were any type of information can be used and data augmentation and exploitation of information in other languages (cross-language models) is encouraged.

## 3 Data

The data used for this work is provided in the VaxxStance Shared Task [1]. A set of training data is given for each language. In each set, two types of data can be found: textual data and contextual data.

Textual data consists of a list of instances with the following elements: a tweet, its id, the username who wrote the tweet and the stance it represents, that is, whether the tweet is in FAVOR of vaccines, AGAINST vaccines or NONE in cases where either the user is neutral, the stance is not clear or it is not specified. The exact number of tweets of each stance for the two languages can be seen in Table 1, as well as the number of accounts from which those tweets were fetched.

**Table 1.** Number of tweets and users who wrote them for each language.

| | **FAVOR** | **AGAINST** | **NONE** | **Total** | **Users** |
|---|---|---|---|---|---|
| **Basque** | 327 | 219 | 524 | 1070 | 149 |
| **Spanish** | 937 | 475 | 591 | 2003 | 1261 |

For contextual information, 5 collections of data are made available. The first one focuses on user information, providing, for each user, the user id, the number of tweets posted by the account, the number of accounts followed by the user, the number of users who follow the account, the time of the user registration on Twitter and whether (and which) emojis are present on the user's Twitter bio. The second data file provides specific information about tweets, such as the tweet id, the id of the user who wrote it, the number of retweets and favourites, the device or operating system from which it was posted and the time of creation. The third and fourth groups of context data give information about user relationships. To be more precise, the number of times a source user retweets a target user is stored. The fourth file, which also describes these relationships between users, collects all the retweets from the timelines of the users contained in the training set. This is the only context file that is only provided for Basque because of the low number of retweets of the tweets that appear in the training data. Finally, the last set of context data ties two users if the source user follows the target user.

In this work, only the last context data mentioned is used. Using this information, a network of the users in the training set with the distance between them (minimum number of sequential user follower jumps) is created.

## 4 Methodology

The participation method chosen for the Shared Task is the Close Track, which means that no extra data can be used apart from that mentioned in Section 3. Furthermore, in this setting two systems for each language need to be trained: one that simply uses the textual information and a second one that, in addition, makes use of the contextual data.

### 4.1 Using textual data

Flair [2] is the tool chosen to train the model that is fed with textual information. Before training, however, some data pre-processing is needed. The proposed data-cleaning method consists in removing URLs, hashtags and user ("@") symbols from all the tweets. This way, all characters that are not words or do not have a meaning are removed. Any column in the textual information datasets that does not correspond to the tweets or the labels is also discarded.

After being pre-processed, the data is used to train a Stance Detection model for each language. To do so, different word embeddings are tried to see which ones perform better in the current task and whether using some of them supposes drastic changes depending on the language. These embeddings are fed to an LSTM to obtain a text representation of the data, which is later fed to a linear layer that outputs the label for each data sample.

It is important to consider that, at the time of making these experiments, data against which the quality of the created models could be tested was not provided. Therefore, data was partitioned into a training and a development set. In the case of Spanish, 90% of the tweets (1802 instances) are used for training and 10% (201 instances) for development. For Basque, since the corpus is smaller, the data is divided as follows: 95% of the tweets for training (1016 instances) and 5% of the tweets for development (54 instances).

For both languages, different models are trained using static embeddings and Transformer embeddings, with 20 epochs each. For static embeddings various combinatios of character embeddings, word embeddings and Flair embeddings are used. As for Tranformer embeddings, Multilingual cased BERT embeddings are employed, which are trained on cased text in the top 104 languages with the largest Wikipedias, as well as XML-RoBERTa embeddings, which are trained on 2.5 TB of newly created clean CommonCrawl data in 100 languages. These multilingual Transformer embeddings are expected to work extremely well in high-resource languages like English and Spanish, but their performance tends to be worse when working with low-resource languages [4].

Table 2 shows a summary of all the different language and embedding combinations tried and the F1 scores obtained by each of those models.

**Table 2.** F1 score obtained by the different systems trained for the textual setting. First column indicates the embedding-type used in each experiment.

| | **Basque** | **Spanish** |
|---|---|---|
| **Word Embeddings + Character Embeddings** | 0.7037 | 0.7015 |
| **Flair Embeddings + Character Embeddings** | **0.7222** | 0.7164 |
| **Multilingual BERT Embeddings** | 0.6253 | **0.7612** |
| **XML-RoBERTa Embeddings** | 0.5294 | 0.5611 |

The best model for Spanish obtains a F1 score of 0.7612 and is built by using BERT multilingual embeddings at document level. The best model for Basque obtains a F1 score of 0.7222 and is built combining Flair embeddings trained for Basque and character embeddings. These are the models used to predict stances when adding contextual information (see Section 4.2).

### 4.2 Adding contextual data

As mentioned in Section 2, usage of contextual data, that is, additional information surrounding the tweet (user information, interaction information), can improve the performance of the classifiers when it comes to grasp the stance of the tweet. This makes sense, as there can be cases where the intention of the tweet could be unclear when the opinion of the user is transmitted subtly. Attempts to overcome that problem can be made by giving the model information about who the author of the tweet is. As it will be seen later, this improves the quality of the classification model.

This work follows the approach by [4], where a network formed by the authors of the tweets and the connections between them is composed. To build it, the *friend_train* and *friend_test* CSV files for each language are used, which contain connections between two users. These connections are present whenever the first user is following the other one on Twitter. Thus, the network is formed by nodes representing users and edges representing follows.

Once the network is built, a subgraph formed only by the users that appear in the training set and the test set, separately, is created. However, the distance between them is needed, so the minimum distance between every possible pair of users in each set is computed.

From the resulting final graph, the mean squared distance between a user and the rest of the users that are in favor, against or neutral is calculated. For that, Equation 1 is used, where $|T|$ is the total number of users with a stance and $d^2_{n\to i}$ is the squared distance between users $n$ and $i$. As a note, the distance between two not-connecting users is arbitrarily set to 100.

$$d_T(n) = \frac{\sum_{i=1}^{|T|} \frac{1}{d^2_{n\to i}}}{|T|} \tag{1}$$

Three features for each user are obtained from the aforementioned computation. Each of them is the mean squared distance to users with favor, against and neutral stance. This features are added to the output of the best textual model for each language. The output is formed by the confidence that the model has for a sample for each stance.

The final input to the contextual model is the confidence of the textual model for each stance in a sample plus the mean squared distance of the author of the sample to users for each stance. This information is fed to a Multilayer Perceptron. The model is built with three hidden layers with 100, 200 and 100 neurons each, with the maximum iterations set to $20,000$. The rest of the parameters are the default ones.

The results of the model can be seen in the Contextual columns of Table 3.

## 5 Results

Table 3 shows all the results and allows us to compare the performance in both language settings. To be more precise, the table shows the F1 scores for the AGAINST and FAVOR stances and the F1 MACRO obtained when testing the textual and contextual models against the test set.

The contextual model shows an overall improvement from the textual model, proving that extra information can be helpful for boosting performance of systems that are built for the purpose of identifying stances. In fact, the contextual model improves its results in every stance in both languages significantly, as it goes up in a range from 2.45 to 26.36. It is specially helpful in the case of Basque: FAVOR and AGAINST labels are rightly predicted more often.

The difference in performance between the textual and the contextual systems is more noticeable in Basque compared to Spanish, as the gap between the scores obtained for both models in Basque is significantly higher than the gap in Spanish. The same occurs regarding individual stances: performance on the AGAINST stance is greatly improved using contextual information compared to the FAVOR stance. It is noteworthy that these two settings where the contextual model performs better, the Basque scenario and the AGAINST class, are also the settings with less resources: the training dataset contains double the amount of tweets for Spanish than for Basque (in addition to Basque being a low-resourced language in itself) and in both languages the number of tweets in favor of the vaccine is higher. This further proves the importance of employing contextual data, as it shows that additional information can give a big boost to systems in certain settings where they do not perform as good as they can in other languages or stances.

Regarding the textual setting, an observation worth mentioning is the performance of the different embeddings in each of the languages. The best results for classifying in Spanish are obtained using Multilingual BERT, hence proving our hypothesis of achieving great performance with high-resource languages. The contrary is also proven by checking the results for the Basque systems: the lowest scores are obtained using Transformer embeddings and the highest using static embeddings specifically trained for this language.

**Table 3.** Result comparison between the textual and contextual settings for both languages. F1 scores of AGAINST and FAVOR stances are shown, as well as the corresponding F1 macro.

| | **Basque** | | | **Spanish** | | |
|---|---|---|---|---|---|---|
| | **AGAINST** | **FAVOR** | **F1 MACRO** | **AGAINST** | **FAVOR** | **F1 MACRO** |
| **Textual** | 0.3881 | 0.4631 | 0.4256 | 0.5714 | 0.7761 | 0.6738 |
| **Contextual** | **0.6517** | **0.5294** | **0.5906** | **0.6627** | **0.8006** | **0.7317** |

## 6 Conclusion and Future Work

We have explored different language models to classify the tweets using textual information. The results we got were interesting. For the case of Basque, the best model was the result of mixing static and dynamic word embeddings. However, in the Spanish model, the best one was the multilingual BERT. This could be due to the under-representation of Basque in multilingual models. Spanish is the second most spoken language in the world, and, accordingly, it has a lot of linguistic resources available. Basque however, is a small language spoken by less than a million of speakers.

The contextual information was harder to get, but was worth using, as the best models got even better. The importance of using additional external data was proven, specially for certain settings where less data exists, in this case the Basque language and the AGAINST stance. As future work, the way of retrieving and calculating contextual features could be further analyzed, so that the models' performance improves.

## References


1. Agerri, R., Centeno, R., Espinosa, M., Fernández de Landa, J., Rodrigo, Á.: VaxxStance@IberLEF 2021: Going Beyond Text in Crosslingual Stance Detection. Procesamiento del Lenguaje Natural 67(0) (2021)
2. Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., Vollgraf, R.: Flair: An easy-to-use framework for state-of-the-art nlp. In: NAACL 2019, 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations). pp. 54–59 (2019)
3. Cignarella, A.T., Lai, M., Bosco, C., Patti, V., Paolo, R., et al.: Sardistance@ evalita2020: Overview of the task on stance detection in italian tweets. In: EVALITA 2020 Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. pp. 1–10. CEUR (2020)
4. Espinosa, M.S., Agerri, R., Rodrigo, Á., Centeno, R.: Deepreading @ sardistance: Combining textual, social and emotional features. In Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020) (2017)
5. Küçük, D., Can, F.: Stance detection: A survey. ACM Computing Surveys (CSUR) **53**(1), 1–37 (2020)
6. Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: A dataset for detecting stance in tweets. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). pp. 3945–3952 (2016)
7. Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: Semeval-2016 task 6: Detecting stance in tweets. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). pp. 31–41 (2016)
8. Taulé, M., Pardo, F.M.R., Martí, M.A., Rosso, P.: Overview of the task on multimodal stance detection in tweets on catalan #1oct referendum. In: IberEval@ SEPLN. pp. 149–166 (2018)
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017), https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf