Modeling User Behaviour through Stance Detection

Murali Kondragunta and **Marina Duhart**
University of the Basque Country UPV/EHU
{mkondragunta001,mduhart001}@ikasle.ehu.eus

Abstract

In this paper, we describe our work done toward the VaxxStance@IberLEF detection challenge 2021. After benchmarking the language models in a textual setting, we experimented with the exploitation of contextual information by training a classification model on top of contextual features along with the textual model features. We observed that users' network graph is a key feature in identifying ideological bubbles online and improving the model's performance.

1 Introduction

Given the current present pandemic circumstances, the anti-vaccination movement has become a noteworthy topic. As a result, identifying it has developed a particular research interest. Agerri et al. (2021) introduced a shared task called VaxxStance shared task at IberLEF 2021, where the objective is to identify the stance of a tweet in Spanish and Basque languages.[1] *Stance* can be defined as the overall position held by a person towards an object or statement. Hence, when referring to stance detection and classification we aim to settle the user's attitude in a text towards a given topic. Thus, the shared task aims to build a system that helps in determining if a tweet expresses a `FAVOR`, `AGAINST` or `NEUTRAL` (`NONE`) opinion regarding vaccines' controversy.

Of the three tracks provided in VaxxStance shared task, we chose the Close Track, where the evaluation is language-specific and only the provided data must be used. It has two evaluation settings:

1. Textual: Only the tweets provided must be leveraged.

2. Contextual: Along with the text, user-based Twitter information like followers, number of statuses posted, etc. can be used by the participants.

To that extent, we build a system for stance detection in each language. Following (Espinosa et al., 2020), our workflow is divided into two phases, one for each evaluation setting. In the first phase, large pre-trained language models are fine-tuned and benchmarked with textual information as input to identify the best performing language model. In the second phase, we evaluate the impact of contextual features. This involves user information such as number of posts, number of followers, user connections, etc. Later, we consider predictions from the best performing model as textual features and concatenate them with contextual features to train a simple machine learning model, Support Vector Machine (SVM) (Cortes and Vapnik, 1995).

2 Related Work

Information extraction has been a key research topic in Natural Language Processing (NLP). With the rise of social media, extracting valuable insights from online sources has been of major interest in academia, industry and politics. This led to various sub-fields of information extraction like opinion mining, fake news detection and emotion analysis. ALDayel and Magdy (2021) explain that in contrast to sentiment analysis, *stance detection mainly focuses on identifying a person's standpoint or view toward an object of evaluation, either to be in favour of (supporting) or against (opposing) the topic*. This is assessed by a combination of features besides solely textual content, such as contextual and network features also known as users' connections and interactions (friends, followers, emojis, hashtags, etc.). In the last decade, there has been a steady increase in the number of publications on stance detection (ALDayel and Magdy, 2021). SemEval 2016 stance detection task (Mohammad et al., 2016) marked a sudden increase in efforts

[1] https://vaxxstance.github.io/

Feature
Number of posts made
Number of followers
Number of friends
Number of emojis in profile bio
Time difference between profile created and tweet posted
Network feature: Connections between the users

Table 1: Types of features

Spanish			
Train (2003)		Test (694)	
FAVOR	937	FAVOR	359
AGAINST	475	AGAINST	140
NONE	591	NONE	195
Basque			
Train (1070)		Test (312)	
FAVOR	327	FAVOR	85
AGAINST	219	AGAINST	92
NONE	524	NONE	135

Table 2: Number of training and testing instances per each language in VaxxStance dataset

directed towards stance detection, reflected in the sudden jump in the number of publications.

Since COVID-19 has been a dominant topic online for the last 2 years, there have been multiple efforts toward stance detection (Ceslov, 2021; Cotfas et al., 2021; Mubarak et al., 2022) and anti-vaccine fake news detection (Hayawi et al., 2022). (Agerri et al., 2021) introduced a shared task to build a system for stance detection in Spanish and Basque. As the Italian stance detection dataset (Cignarella et al., 2020) is similar to the VaxxStance task, we follow the work of one of the submissions called `DeepReading @ SardiStance` (Espinosa et al., 2020) in building our system.

3 Data

As mentioned in section 1, the VaxxStance dataset has data related to two evaluation settings i.e textual and contextual. In the Spanish dataset, training instances are 2003 and testing instances are 694. Whereas, in the Basque dataset, training instances are 1070 and testing instances are 312. The stance distributions are shown in Table 2.

We plotted hashtags per stance to see if they are a distinguishing factor for stance detection. From Figure 1, we observed that the above hypothesis stands true.

4 Systems Description - Methodology

4.1 Task A: Only Textual features

Considering the success of fine-tuning language models on downstream tasks, we benchmark mBERT (Devlin et al., 2018) and IXAmBERT (Otegi et al., 2020) on the textual data. While mBERT is pretrained on 104 languages, ixamBERT is only trained on 3 languages namely, English, Spanish and Basque. With this experiment, we would like to know if the downstream task benefits more from pretraining on a large number of languages or only from languages of interest.

4.1.1 Data Preprocessing

As most language models are trained on raw text available online, we believe that any kind of traditional preprocessing steps like stop-word removal, stemming and lemmatization would hinder the learning process. So, in the textual track, hyperlink removal is the only preprocessing step we perform.

4.2 Task B: Leveraging Contextual Information

Table 1 shows the user attributes used in this work. We formulate that Twitter users, in general, tend to interact within ideological bubbles. This is a trend that can be explained by the social phenomenon of *homophily*, according to which users tend to associate and bond with others similar to them. This phenomenon shapes users' behavioural data, and it can be valuable to learn and infer stances (ALDayel and Magdy, 2021). In order to test this hypothesis, we test network-specific features in isolation. So, we categorised the contextual features into two subgroups based on network features.

1. Network features → connections between users.

2. Contextual features → rest of the features, except network features. Throughout the paper, these features are referred to as *contextual features*.

4.2.1 Network features

As previously mentioned, studying connections and distance between users will contribute to stance detection. To identify how close or distant a user is to a stance, Espinosa et al. (2020) build a network graph and calculate the distance between the current user and users of different stances using the following formula,

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

where $|T|$ stands for total the number of users of a particular stance and $d^2_{n\rightarrow i}$ refers to the distance between the current node n and particular stance node i. We use the same formula (Eqn. 1) for calculating network features. Here, distance refers to the number of hops made to reach the target node. This way, the mean distance with different stances would give us 3 network features.

4.2.2 Contextual features

Numeric features like `# of posts`, `# of emojis in profile bio` & `# of followers` are considered without any change. Besides previous features, we have also considered `profile creation time` & `tweet posted time`. We conjectured that the time difference between them can lead us to an interesting insight: that fake profiles created for spreading fake news would present a shorter time lapse between the account creation and the time in which tweets are posted. The time difference is mentioned in number of days. Anything less than a day is counted as 0 days. This is a hypothesis that we will like to discuss and study in future work.

5 Results

5.1 Task A

As mentioned in subsection 4.1, we fine-tune and compare two language models, mBERT and IXAmBERT. With a validation split of 0.2, we choose our learning rate as 3e-5 and the number of epochs as 20.

From Table 3 and 4, we observed that IXAmBERT performs better than mBERT. Therefore, we consider it the best language model for extracting textual features i.e. prediction probabilities for each stance. IXAmBERT's success also proves that the downstream task benefits more from language models pretrained on languages of interest than on a large number of languages.

5.2 Task B

For all the experiments in Task B, SVM is selected as the base machine learning model and hyperparameters are selected based on the evaluation set.

Spanish			
Modelname	**AGAINST**	**FAVOR**	**Average**
ixamBERT	0.72	0.85	0.79
mBERT	0.66	0.83	0.75

Table 3: Spanish Language: Language models performance on textual features. Average indicates the average of FAVOR and AGAINST stance F1-scores.

Basque			
Modelname	**AGAINST**	**FAVOR**	**Average**
ixamBERT	0.7	0.62	0.66
mBERT	0.54	0.52	0.53

Table 4: Basque Language: Language models performance on textual Features. Average indicates the average of FAVOR and AGAINST stance F1-scores.

5.2.1 Network Features: Distance-based Heuristic for Prediction

To confirm the hypothesis about users' ideological bubbles, we start with a simple baseline where we predict the stance of a tweet solely based on the distance between the target user who posted it and users whose posts have been classified as `FAVOR`, `AGAINST` and `NONE`. Stance with the maximum score, according to the metric (Eqn. 1), is picked as the prediction. An F1-score of 0.93 from table 5 shows that users who are against the vaccination are closely connected. However, tweets that were expected to be predicted as `FAVOR` are being predicted as `NONE`. An outcome that might be related to the complexity of identifying stance in texts.

While the hypothesis stands true for Spanish, it doesn't apply to Basque (Table 8). This can be attributed to the fact that the number of basque training instances is limited to constructing clusters/bubbles.

5.2.2 Network Features: Machine Learning model

When we trained an SVM with network features (only) as the input, there has been an improvement to the simple heuristic method. Column 2 in the Table 5 and Table 8 show the improvement.

5.2.3 Network features: Adding contextual and textual features

+ Contextual Features: From Table 5 and Table 8, we can infer that there hasn't been any improvement when the contextual features are concatenated with the network features. Moreover, the model's

Stance	Simple Baseline	SVM	(+ Contextual features)	(+ Textual features)	(+Contextual + Textual features)
FAVOR	0.56	0.66	0.65	0.88	0.88
AGAINST	0.93	0.94	0.94	0.89	0.87
NONE	0.61	0.63	0.59	0.77	0.79
Average	0.745	0.8	0.795	**0.885**	0.875

Table 5: Spanish Language: Performance on network features. A plus sign (+) indicates the concatenation of network features with other features as an input to the SVM model. Average indicates the average of FAVOR and AGAINST stance F1-scores.

Stance	SVM	(+ Network features)	(+ Textual features)	(+ Network + Textual features)
FAVOR	0.72	0.65	0.86	0.88
AGAINST	0.26	0.94	0.76	0.87
NONE	0.17	0.59	0.79	0.79
Average	0.49	0.795	0.81	**0.875**

Table 6: Spanish Language: Performance on contextual features. A plus sign (+) indicates the concatenation of contextual features with other features as an input to the SVM model. Average indicates the average of FAVOR and AGAINST stance F1-scores.

Stance	SVM	(+ Network features)	(+ Textual features)	(+Network + Textual features)
FAVOR	0.17	0.58	0.62	0.63
AGAINST	0.62	0.83	0.7	0.8
NONE	0.53	0.55	0.76	0.72
Average	0.395	0.705	0.66	**0.715**

Table 7: Basque Language: Performance on contextual features. A plus sign (+) indicates the concatenation of contextual features with other features as an input to the SVM model. Average indicates the average of FAVOR and AGAINST stance F1-scores.

performance was drastically reduced in the Basque dataset.

+ Textual Features: In the case of Spanish data, table 5 shows that textual features complement the network features in improving the model's performance. Whereas in Basque data, similar to contextual features, textual features affect the model's performance albeit lesser than contextual features.

5.3 Contextual features

From table 6 and table 7, we observed that contextual features alone are ineffective for stance detection. However, when combined with contextual and textual features, there has been a performance improvement.

6 Discussion

We have observed that network features independently play an important role in stance detection. Combining them along with textual features has shown better scores in the model's performance, in the case of the Spanish dataset. In contrast, for the Basque dataset, network features obtained better results by themselves than concatenated with other features. These results have led us to conclude that there is a need for further analysis to understand why models obtain better results for Spanish than for Basque.

7 Conclusions

In this work, we explored the importance of different types of features and evaluated our hypothesis on ideological bubbles. We showed that network features alone contribute a significant amount of information, followed by textual features from fine-tuned language model. We found contextual features i.e., user attributes, to be the least informative among all features.

From figure 1 and 2, we observed that hashtags are indeed a good indicator of stance detection. In future work, we want to find hashtag splitters for both Spanish and Basque languages. We would also like to include more contextual features and test their impact.

References

Rodrigo Agerri, Roberto Centeno, María Espinosa, Joseba Fernandez de Landa, and Alvaro Rodrigo. 2021. Vaxxstance@ iberlef 2021: Overview of the task on going beyond text in cross-lingual stance detection. *Procesamiento del Lenguaje Natural*, 67:173–181.

Abeer ALDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management*, 58(4):102597.

Rodica Ceslov. 2021. Detecting stance on covid-19 vaccine in a polarized media. *CUNY Academic Works*.

Alessandra Cignarella, Mirko Lai, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020. Sardistance @

Stance	Simple Baseline	SVM	(+ Contextual features)	(+ Textual features)	(+Contextual + Textual features)
FAVOR	0.68	0.66	0.58	0.68	0.63
AGAINST	0.64	0.9	0.83	0.8	0.8
NONE	0.73	0.56	0.55	0.75	0.72
Average	0.66	**0.78**	0.705	0.74	0.715

Table 8: Basque Language: Performance on network features. A plus sign (+) indicates the concatenation of network features with other features as an input to the SVM model. Average indicates the average of FAVOR and AGAINST stance F1-scores.

evalita2020: Overview of the task on stance detection in italian tweets.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Liviu-Adrian Cotfas, Camelia Delcea, Ioan Roxin, Corina Ioanas, Gherai Dana Simona, and Federico Tajariol. 2021. The longest month: Analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. *IEEE Access*, PP:1–1.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

María Espinosa, Rodrigo Agerri, Alvaro Rodrigo, and Roberto Centeno. 2020. Deepreading @ sardistance: Combining textual, social and emotional features.

K. Hayawi, S. Shahriar, M.A. Serhani, I. Taleb, and S.S. Mathew. 2022. Anti-vax: a novel twitter dataset for covid-19 vaccine misinformation detection. *Public Health*, 203:23–30.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 31–41.

Hamdy Mubarak, Sabit Hassan, Shammur Absar Chowdhury, and Firoj Alam. 2022. Arcovidvac: Analyzing arabic tweets about COVID-19 vaccination. *CoRR*, abs/2201.06496.

Arantxa Otegi, Aitor Agirre, Jon Ander Campos, Aitor Soroa, and Eneko Agirre. 2020. Conversational question answering in low resource scenarios: A dataset and case study for basque. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 436–442.

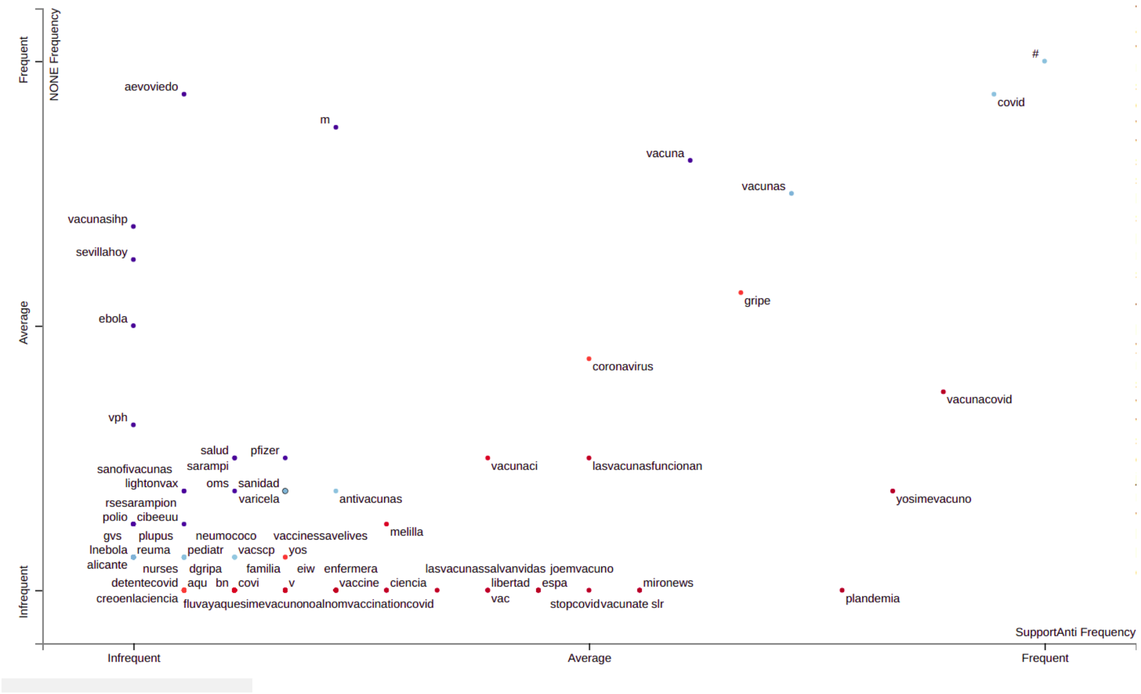
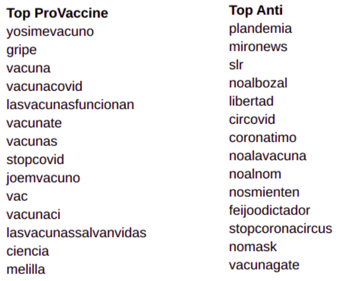


Figure 1: Spanish hashtags according to the stance. Click here for an interactive version.



Top ProVaccine
yosimevacuno
gripe
vacuna
vacunacovid
lasvacunasfuncionan
vacunate
vacunas
stopcovid
joemvacuno
vac
vacunaci
lasvacunassalvanvidas
ciencia
melilla

Top Anti
plandemia
mironews
slr
noalbozal
libertad
circovid
coronatimo
noalavacuna
noalnom
nosmienten
feijoodictador
stopcoronacircus
nomask
vacunagate

(a) hashtags with pro-vaccine stance

(b) hashtags with anti-vaccine stance

Figure 2: Top Spanish hashtags per each stance