UniRetro at PoliticEs@IberLef 2022: Political Ideology Profiling using Language Models

Andrei-Alexandru Manea[1], Liviu Petrişor Dinu[1,2]

[1]*Faculty of Mathematics and Computer Science, University of Bucharest, Romania*
[2]*Human Language Technologies Research Center, University of Bucharest, Romania*

Abstract

In this paper, we describe the participation of UniRetro's team at IberLef 2022 on "PoliticEs: Spanish Author Profiling for Political Ideology" task. The main goal is to process social media posts in Spanish in order to extract the user's political ideology, gender, and profession. Our first approach for this task consists of using TF-IDF applied on SentencePiece pretrained and custom tokens obtained by Named Entity Encapsulation. In our second approach, we finetuned deep pretrained models on Spanish using handcrafted class weigths. We obtained 9th place in the general leader board. We also performed an importance analysis in order to prove that the deep models were capable of better generalization.

Keywords

IberLef, UniRetro, BETO, DistilBETO, SentencePiece, TF-IDF

1. Introduction

In the book "Political Ideology: An Introduction" of Heywood [1], the author defines the linear spectrum as a horizontal axis with left and right orientations, as illustrated in the Figure 1.

Figure 1: The linear political spectrum containing Communism as the extreme-left regime and Fascism as the extreme-right regime

Many left activists speak about environmental concerns, human rights and support social equality. However, right-wing people use their voice against taxes and consider normal the current social orders and hierarchies. Their speech is important because people are usually more enthusiastic to follow advice and directions from politicians who share their ideology. They also may copy their attitude, lifestyle, and other traits.

Political ideology is a psychographic trait that can be used to understand individual and social behavior, including moral and ethical values as well as inherent attitudes, appraisals, biases,

IberLEF 2022, September 2022, A Coruña, Spain
andrei.manea.dev@gmail.com (A. Manea); liviu.p.dinu@gmail.com (L. P. Dinu)
© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3202/politices-paper8.pdf

and prejudices Verhulst et al. [2]. For instance, an experiment made Baumgaertner et al. [3] for analyzing the socio-political characteristics of attitudes about vaccination for infectious diseases indicated that conservative respondents are less likely to express pro-vaccination beliefs than other individuals. However, openness to experience and agreeability were notably more correlated to the left-wing.

For gaining more popularity, some politicians state that they value the "will of the people". This is called populism and populist politicians use to state that the public institutions are ruled by a self-serving minority who seems to manipulate the virtuous majority of common people. The ideology of the elite is being criticized and the listening group receives it as an "us vs them" question.

In this paper, we noticed in the first models, that the decisive features for detecting the left orientation are words denoting the right orientation and vice-versa. The paper is organized as follows: Section 2 contains general observations of related articles, Section 3 has the description of the task provided by the organizers. The 4th Section contains the dataset source description alongside a general overview. The Section of experiments (5) contains a general description of the experiments and 3 subsections regarding the results obtained using TF-IDF features (Subsection 5.1), Pretrained Embeddings (Subsection 5.2) and the best results compared to the evaluation leaderboard (Subsection 5.3). The following Section (6) contains observations extracted from feature importance analyses of the best models using TF-IDF features on gender and binary political ideology classification. Eventually, Section 7 contains the conclusions and future work of the entire project.

2. Related Work

The political orientation of a certain individual may be determined by a variety of stances. One of the recent competitions was SemEval-2016, Task 6 "Detecting Stance in Tweets" Mohammad et al. [4]. Given a tweet and a target entity (person, organization, etc.), automatic natural language systems must determine whether the tweeter is in favor, against, or neutral with the specific target. The first task was to determine the stance for several targets from the training set, whereas in the second task, other targets were employed. This proves there is a need for stance models to work on any given target.

From a political perspective, other stance models were released Kawintiranon and Singh [5], regarding the user position against Donald Trump and Joe Biden in the US election 2020. Instead of using the classical Masked Language Model (MLM) objective for self-supervised learning, knowledge enhanced version was employed. This presumes that significant tokens generated using the log-odds ratio are incorporated into the learning process and used to improve the classification task.

Another previous article uses a set of opinions extracted from the text and builds a knowledge graph Chen et al. [6]. However, these predefined opinions have to cover a great part of the dataset and the graphs are not able to extend for searching the opinions regarding the new topics rising on social media.

There are also several text augmentation techniques presented in Li and Caragea [7] for integrating the target through the inner text. In our binary task of detecting political ideology, although we performed tweet classification, we may consider that our target is the abstract center of the political spectrum. Similar targets may be derived for the multiclass objective.

There is a target-free approach in English, in a Stanford project for classifying congressional debate transcripts from Ideological Books Corpus (IBC), Simoes and Castanos [8]. They used BERT models to find a correlation between the texts and the "de jure" political standing of the authors.

3. Task Description

The challenges involved were to extract:

1. gender, with label ∈ {female, male}
2. profession, with label ∈ {politician, journalist}
3. political ideology in binary format, with label ∈ {left, right}
4. political ideology in multi-class format, with label ∈ {left, moderate_left, moderate_right, right}

Each user contains submissions from Twitter [9], a submission consists of a post or comment. More details regarding how data was collected are in the next section. We performed the training and inference at tweet level and for each user, the dominant label was submitted as the answer. For evaluation, the F1 macro score was used for each task and a general metric representing the mean of the 4 values.

4. Dataset

To the best of the knowledge of the organizers, this is the first Spanish shared task focused on this topic.

Using the UMUCorpusClassifier of García-Díaz et al. [10], the tweets were collected during 2020 and 2021 from the accounts of politicians and political journalists in Spain. For each user, the political affiliation can be guessed according to the party to which politicians belong, or the editorial line of the newspapers where journalists write. The politicians may be members of the Government, Congress and Senate, mayors of the important cities, presidents of the autonomous communities (Catalonia), former politicians, and other collaborators associated with political parties.

Journalists were selected from different Spanish news media, such as ABC, El País, ElDiario, El Mundo or La Razón among others. They are relevant people within political journalism and some of them publish opinion articles and participate in television gatherings, so their political orientation can be guessed with certain guarantees.

Other mentioned links and hyperlinks to the news were removed. The dataset is considered an extension of the PoliCorpus 2020 García-Díaz et al. [11] and more information can be found in García-Díaz et al. [12].

We received the final sets through the Codalab [13] platform, the labels for the test set being released after the evaluation stage. In the Table 1 there are general observations of each set. It is easy to see that on the training set, the most unbalanced tasks are profession and multi-class ideology prediction. The distributions of the other 2 were acceptable, as we can see in the Table 1.

Table 1
General observation of the complete dataset for each task and set

	training			development			testing		
Label	Num. tweets	Num. users	Avg. tweets/user	Num. tweets	Num. users	Avg. tweets/user	Num. tweets	Num. users	Avg. tweets/user
female	16320	136	5561.9	2350	47	2301.3	4320	36	5486.5
male	21240	177	5846.3	2650	53	2424.4	8280	69	5689.6
journalist	7381	62	5309.1	672	14	2152.3	3000	25	5340.4
politician	30179	252	5801.8	4328	87	2373.8	9600	80	5707.3
left	21360	178	5612.8	2750	55	2276.6	6840	57	6628.9
right	16200	135	5867.7	2250	45	2476.5	5760	48	5609.3
left	9120	76	5952.1	950	19	2460.9	2520	21	5961.9
moderate_left	12240	102	5360.0	1800	36	2179.3	4320	36	5434.6
moderate_right	11280	94	5830.5	1550	31	2446.8	3720	31	5333.6
right	4920	41	5953.0	700	14	2542.3	2040	17	6112.1

The training was performed at tweet level. The tweet was cleaned from punctuation signs, except for the question and the exclamation mark. The emojis and other unusual characters were removed as well. An exception has been made for emojis indicating a number that was converting into numbers and other symbols being translated in Spanish: e.g. "≤" became "más bajo o igual" ("lower or equal").

The name of the political parties and the user id in the dataset were anonymized by replacing them with the tokens "[POLITICAL_PARTY]" and "@user"; as a processing step, the mask first was translated as "partido político", in order to be tokenized as usual words. The tweets from the training and developed set were shuffled and split into our training and validation set, obtaining a similar distribution to the given sets. The set was lazy loaded at training and validation time using dataloaders from Pytorch library Paszke et al. [14].

5. Experiments

The system for performing experiments consists of importing a pretrained tokenizer, a pretrained model and finetuning them on the training set. All the tokenizers and the deep models are imported from HuggingFace API, whereas the other models which were using TF-IDF features are imported from the Scikit-learn library [15]. In both cases, most of the hyperparameters of the learning process were fixed and we have been focused more on data-driven approaches.

Inspired by BERT-base models adapted for Spanish [16], we decided to use the tokenizer "dccuchile/bert-base-spanish-wwm-uncased" from HuggingFace library and apply it to clean tweets. It is a SentencePiece tool [17] that performs subword tokenization and detokenization in an unsupervised manner. We used BETO tokenizer, which uses tokens from Spanish vocabulary which was used for training BETO (Spanish BERT) in the self-supervising manner of the Masked Language Model (MLM). Here is an example of tokenization for sentence "Unos PGE 2019 hubieran facilitado todo pero en breve estará resuelta para bien la partida 2018", we obtain ['unos', 'p', '##ge', '2019', 'hubieran', 'facilitado', 'todo', 'pero', 'en', 'breve', 'estará', 'resuel', '##ta', 'para', 'bien', 'la', 'partida', '201', '##8']. The sentence translated into English: "Some PGE 2019 would have facilitated everything but soon the 2018 game will be resolved for good" and it refers to the planning of the general state budget (PGE).

The tokens that start with # are subwords that are word possible inflections. It is easy to see that it performs tokenization even on named entities. Our dataset contains a lot of institution names, and political entities, so we tried different methods for overcoming these issues. We implemented two types of features: 1) TF-IDF-based and 2) using pretrained embeddings.

5.1. TF-IDF Features

In this subsection, the aforementioned tokenizer was applied in 3 main approaches which are listed below:

1. clean text with few added tokens
2. clean text with few added tokens and MixUp input
3. a linguistic signature of the clean text

The tokens obtained from the training set were turned into TF-IDF vectors, using tools from the Scikit-learn library and then fed into the models: SVM and RandomForest. The training was performed locally on a personal computer, ending with the model weights being saved and their performance was also measured on the TF-IDF features of the validation set.

We observed that there are some named entities that were poorly tokenized; e.g. "donald trump", became ["donald", "tru", "##mp"]. In the first approach, we mapped similar named references into a new token. We called this named entity encapsulation and chose only the following 6 personalities:

- "C's" -> "C_s"
- "el Rey", "Majestad El Rey" -> "el_Rey"
- "Donald Trump", "Trump" -> "D_Trump"
- "Biden", "Joe Biden" -> "J_Biden"
- "PedroSanchez", "Pedro Sánchez", "Sánchez" -> "P_Sanchez"
- "Adolfo Suàrez", "Adolfo Suárez" -> "A_Suarez"

The second approach contains a MixUp technique Hongyi Zhang [18], which consists of generating 30% of the train set size, by MixUp augmentation. This consists of randomly uniform sampling pair of distinct indices, i, j and obtaining a new input and label:

$$x_k = \lambda x_i + (1 - \lambda)x_j \tag{1}$$

$$y_k = round(\lambda y_i + (1 - \lambda)y_j) \tag{2}$$

In our method. $\lambda = 0.8$, x_i is an input vector and y_i its corresponding label. the function *round* is meant to approximate and cast the label to an integer value. This approach came up with a slight improvement, as we can see in Table 2 on the best model predicting the binary ideology. However, on the gender task, it brought more bias, which can be observed by the difference between the F1 score macro versus weighted. So we considered it a bit risky and was not extended for the other tasks.

Table 2
Several tests were performed on our validation set with the light models. For SVMs, the kernel name is appended in the model name column

task	model name	approach	Accuracy	F1 Weighted	F1 Macro
gender	svm_rbf	1	70.46%	**70.41%**	**69.80%**
gender	svm_linear	1	64.66%	64.78%	64.28%
gender	RandomForest	1	**70.88%**	70.32%	69.40%
gender	svm_rbf	3	56.06%	56.07%	55.26%
profession	svm_rbf	1	**86.01%**	**85.86%**	**77.09%**
profession	svm_linear	1	77.96%	79.70%	70.94%
profession	svm_rbf	2	78.14%	73.35%	47.36%
ideology_binary	svm_rbf	1	79.37%	79.29%	78.81%
ideology_binary	svm_linear	1	75.09%	75.15%	74.72%
ideology_binary	svm_rbf	2	**79.54%**	**79.44%**	**78.94%**
ideology_multiclass	svm_rbf	1	**65.04%**	**64.94%**	**63.92%**
ideology_multiclass	RandomForest	1	59.23%	58.17%	55.82%
ideology_middle	svm_rbf	1	**77.97%**	**78.07%**	**76.62%**
ideology_middle	svm_linear	1	72.92%	73.29%	71.92%
ideology_combined	svm_rbf	1	63.55%	63.58%	62.00%

The linguistic features used in the third approach consists of extracting the words that contain at least one of the following morphological property: gender, number, or person. Each word is also replaced with a signature containing the part of speech and the mentioned properties. If one of the features is missing, it is replaced with “unk”. For instance, “Dicen” becomes “VERB:unk:plur:3”. Our expectation was to work for the gender classification task, but the Table 2indicates that these features are not sufficient.

For almost every approach and each task we applied TF-IDF features on the obtained tokens and start the training with Support Vector Machines and Random Forest Classifier.

In the same Table, we may notice the arrival of a new task (ideology_middle). It was meant to simplify the multiclass ideology by creating a new label which represents if the user orientation is moderate (1) or not (0). Using the best model from this artificial task and the best from ideology_binary, we derive a new solution. The multiclass decision is derived from the ideology_middle and ideology_binary. For instance, if a tweet receives the pair (0, 1), it will be encoded as moderate_left. This trial did not improve the SVM that performs one versus all classification, but it would have been an interesting solution if at least one of the models had been deep and had higher performance.

5.2. Pretrained Embeddings

In this section, we could not provide the named entity encapsulation, previously mentioned. This is due to the fact that we must come up with new word embeddings trained from this or another context. The training process was performed by feeding the training data to the model, computing the cross-entropy (weighted or not) and minimizing it with the optimizer AdamW (Loshchilov and Hutter [19]), from the Pytorch library. Most of the hyperparameters of the optimizer and others involved in the training process were fixed $learning\_rate = 2 \times 10^{-5}$, $batch\_size = 16$, $\varepsilon = 10^{-8}$, except for the class weights which are further mentioned.

After a training epoch, the model was tested on the validation set (validation epoch) and if the mean loss is lower than the previous best, the model weights are locally saved. This process was performed in a Jupyter Notebook deployed on the Kaggle platform. At the end of the experiment, the stored results and logs were downloaded and visualised.

In the first attempt, we finetuned the BETO previously mentioned Cañete et al. [16] for ideology_binary. The original model class was overridden in order to edit the last fully connected layer: the original model contains a layer mapping vector of dimension $d = 768$ to the number of classes. We replaced this classifier with a sequence of 2 layers with the following number of neurons: $768 > 256 > num\_classes$, with a dropout probability of 20% and a GELU activation function. Its results were far lower than those of light models, as the Table 3 shows.

Table 3
Several tests were performed on our validation set with the deep models: BETO and DistilBETO

task	model name	approach	Accuracy	F1 Weighted	F1 Macro
gender	DistilBETO	raw	72.39%	72.16%	71.47%
profession	DistilBETO	raw	87.26%	86.58%	77.47%
profession	DistilBETO	weighted	84.52%	85.23%	77.57%
ideology_binary	BETO	raw	56.90%	41.30%	-
ideology_binary	BETO	biased	68.77%	66.49%	-
ideology_binary	DistilBETO	raw	80.31%	80.36%	80.05%
ideology_multiclass	DistilBETO	raw	65.56%	65.41%	63.53%
ideology_multiclass	DistilBETO	weighted	63.87%	63.94%	62.73%

Inspired by Ben-Zaken et al. [20] we retried the procedure but only trained the biases. This time, the performance was similar to the light models but did not prove to be better. This approach is named "biased" in Table 3.

The best solution came from DistilBETO Cañete et al. [21], a DistilBERT model trained on Spanish corpora. Being a model with a fewer number of parameters than BETO, we performed the training, having all the gradients unfrozen. The first results are called the "raw approach" in Table 3.

This time, the performance was significantly better. In the final round of training, we decided to use custom class weights in the training. There are a lot of ways to compute the class weights. If we consider each class id $c \in C$, and N_c the number of samples in the training set, corresponding to the mentioned label, we choose an empirical suitable value $M = 10^3$, and the class weights becomes $w_c = \frac{M}{N_c}$. The weights were used in the loss function, which was categorical cross-entropy. For profession task, the value was $M = 10^3$, whereas for multiclass ideology, was another value $M \in [0.1, 0.9]$. We find this interval suitable for the current optimizer, AdamW and the learning rate. These were the “weighted” approaches.

5.3. Results

For submitting the results, in the first attempt, we selected each user and merged all the tweets in one string, and then send the decoded labels. This performed badly because the models were trained on individual tweets. So, the second submission contains the result obtained by applying inference at tweet level and choosing the most dominant class.

The results were much better, but far away from the results obtained on the validation set. This means that the light models had overfitted some specific words in accomplishing the task. However, the deep models worked better by obtaining a higher performance at user level (Table 4) than at tweet level (Table 3). This is because they are capable of analyzing an entire context and not only the frequency of most common keywords. The third submission was the first containing the deep models trained on raw tweets on ideology and with custom weights on gender task.

The last submission (UniRetro_3) comes up with a new model for the profession task trained with class weights. In Table 4 its values are present in the public leaderboard as well. On the other results, the number mentioned in the round brackets is the hypothetical corresponding place if the final submission was based on that version.

Table 4
Submission leader board containing our submissions compared to the top 3 and the baseline. Each result contains the ranking for the corresponding task in the final leather board

submission_id	Avg. F1 Macro	Gender F1	Profession F1	Ideology Binary F1	Ideology Multiclass F1
LosCalis	0.902262 (1)	0.902868 (1)	0.944327 (1)	0.961623 (1)	0.800229 (4)
NLP-CIMAT-GTO	0.890961 (2)	0.784836 (6)	0.921250 (3)	0.961482 (2)	0.896275 (1)
Alejandro Mosquera	0.889182 (3)	0.826714 (3)	0.933452 (2)	0.951519 (3)	0.845044 (3)
UniRetro_3	0.786940 (9)	0.737976 (12)	0.883463 (6)	0.902198 (7)	0.624124 (12)
UniRetro_2	0.733122 (14)	0.737976 (12)	0.668188 (17)	0.902198 (7)	0.624124 (12)
UniRetro_1	0.684379 (17)	0.693878 (15)	0.668188 (17)	0.796037 (15)	0.579415 (14)
UniRetro_0	0.581635 (20)	0.629944 (18)	0.432432 (19)	0.702381 (18)	0.561781 (16)
Baseline	0.511228 (20)	0.576211 (19)	0.432432 (18)	0.595665 (19)	0.440603 (19)

6. Interpretation

Here we analyze the feature’s impotence, finding relevant patterns in profiling the authors. Although the best models in this area were SVM with RBF kernel, it is worth mentioning that those with linear kernel have a similar performance. They also have the advantage of being easy to interpret.

For the binary classification tasks, we saved the vocabulary and extracted the weight for each word, from SVM parameters. The words corresponding to the top 15 lowest weights will determine if the sample belongs to class 0, and the top 15 highest for class 1.

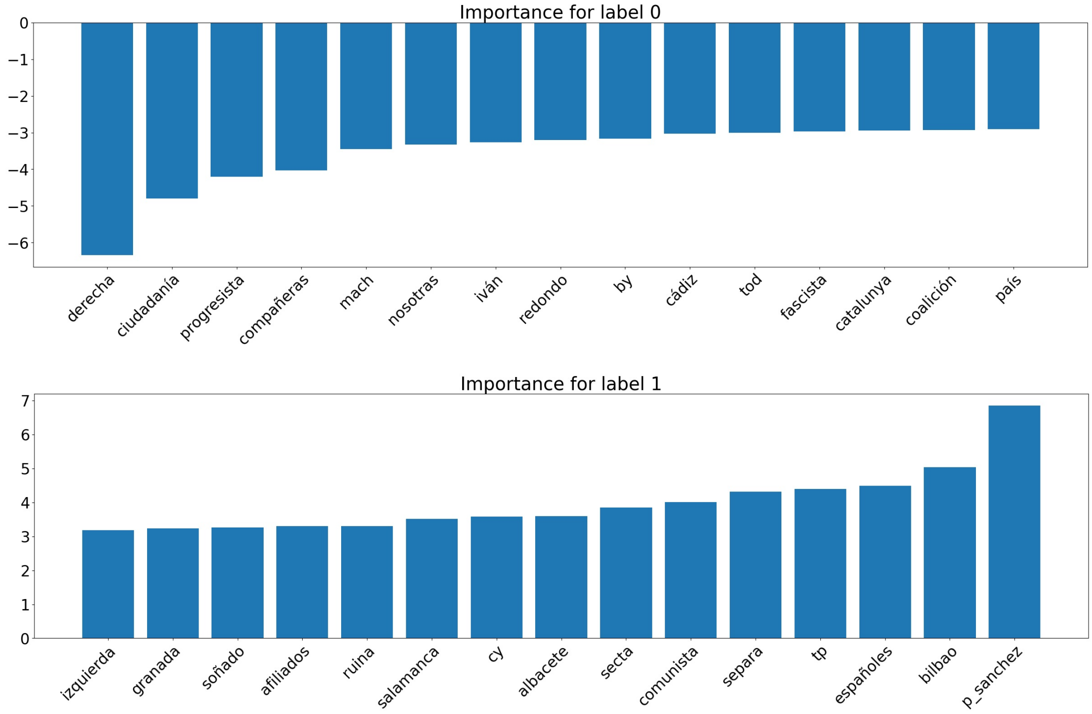

Figure 2: Feature (word) importance for best Linear SVC model predicting the binary ideology; label 0 represents the left orientation and 1 the right orientation

Figure 2 contains the importance of the best model for predicting the binary ideology. It is interesting that texts with the left orientation uses words like "derecha" (right), "progresista" (progressive), "fascista" (fascist), referring to a right-wing vocabulary. At the same time, texts with the right orientation contain "izquierda" (left), "comunista" (communist), "p_sanchez" which stands for Pedro Sánchez, former Prime Minister of Spain and former Secretary-General of Spanish Socialist Workers' Party. This experiment summarizes the populism pattern, presented in Section 1.

Another interesting behavior we noticed, was in the gender classification task. Figure 3 illustrates that the decisive words for both classes were the gender of the noun proud ("orgulloso" and "orgullosa"). The next features representing texts written by women also contain "mujeres" (women), "feminista" (feminist), whereas the male side contains more toponyms (Bilbao, Cantabria, etc.). This demonstrates that the linguistic features we tried to use in the previous experiments, will not be sufficient for gender classification. However, the topic of discussion, the named references may be more relevant in this task.

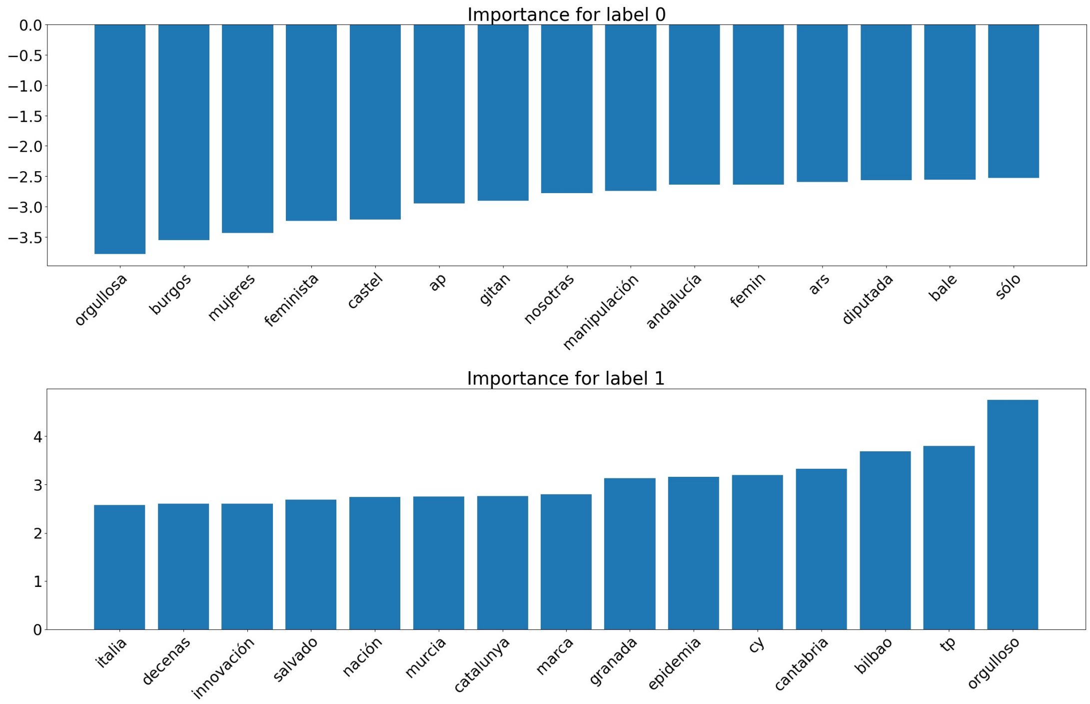

Figure 3: Feature (word) importance for best Linear SVC model predicting the gender; label 0 represents the female writers and 1 the male writers

7. Conclusion and Future Work

In this paper, we presented the contributions of the UniRetro team in the IberLef, PoliticEs 2022 "Spanish Author Profiling for Political Ideology" shared task. We have used a variety of models and techniques including Transformers, Knowledge Distillation, and other Machine Learning models.

We described 2 solutions, each one containing 3 versions of data processing. Different from previous methods, we found a relevant interpretation for which the TF-IDF features may not be sufficient for solving this task and tried to use several techniques like MixUp, named entity encapsulation. Therefore deep models proved to work better in generalizing the problem. Our best ranking was 6th place for profession classification using DistilBETO with handcrafted class weights.

For future work, we aim to adapt the MixUp technique for deep models as well. For avoiding the tokenization of named entities, we had an attempt of isolating them into special tokens, as we did with the 5 personalities. However, the NER model failed to deliver properly the named institution and had a lot of false positives. We also tried with a POS tagger and had a similar issue. So, there is a need for a better Spanish model trained for Named Entity Recognition. Another limitation we encountered was trying to finetune a large Albeto model. This may be overcome by deploying the training in a special cloud and searching for a smooth learning rate accordingly to our dataset, class weights, and other hyperparameters.

References

[1] A. Heywood, Political ideologies: An introduction, Macmillan International Higher Education, 2017.

[2] B. Verhulst, L. J. Eaves, P. K. Hatemi, Correlation not causation: The relationship between personality traits and political ideologies, American Journal of Political Science 56 (2012) 34–51. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5907.2011.00568.x. doi:https://doi.org/10.1111/j.1540-5907.2011.00568.x. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5907.2011.00568.x.

[3] B. O. Baumgaertner, J. E. Carlisle, F. Justwan, The influence of political ideology and trust on willingness to vaccinate, PLoS ONE 13 (2018).

[4] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, C. Cherry, SemEval-2016 task 6: Detecting stance in tweets, in: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, San Diego, California, 2016, pp. 31–41. URL: https://aclanthology.org/S16-1003. doi:10.18653/v1/S16-1003.

[5] K. Kawintiranon, L. Singh, Knowledge enhanced masked language model for stance detection, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 4725–4735. URL: https://aclanthology.org/2021.naacl-main.376. doi:10.18653/v1/2021.naacl-main.376.

[6] W. Chen, X. Zhang, T. Wang, B. Yang, Y. Li, Opinion-aware knowledge graph for political ideology detection., in: IJCAI, volume 17, 2017, pp. 3647–3653.

[7] Y. Li, C. Caragea, Target-aware data augmentation for stance detection, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 1850–1860. URL: https://aclanthology.org/2021.naacl-main.148. doi:10.18653/v1/2021.naacl-main.148.

[8] A. A. Simoes, M. Castanos, Fine-tuned bert for the detection of political ideology, 2020.

[9] [Online], 2006, Twitter Webpage, URL: https://twitter.com/.

[10] J. A. García-Díaz, Á. Almela, G. Alcaraz-Mármol, R. Valencia-García, Umucorpusclassifier: Compilation and evaluation of linguistic corpus for natural language processing tasks, Procesamiento del Lenguaje Natural 65 (2020) 139–142.

[11] J. A. García-Díaz, R. Colomo-Palacios, R. Valencia-García, Psychographic traits identification based on political ideology: An author analysis study on spanish politicians' tweets posted in 2020, Future Generation Computer Systems 130 (2022) 59–74. URL: https://www.sciencedirect.com/science/article/pii/S0167739X21004921. doi:https://doi.org/10.1016/j.future.2021.12.011.

[12] J. A. García-Díaz, S. M. Jiménez-Zafra, M. T. Martín-Valdivia, F. García-Sánchez, L. A. Ureña-López, R. Valencia-García, Overview of PoliticES 2022: Spanish Author Profiling for Political Ideology, Procesamiento del Lenguaje Natural 69 (2022).

[13] A. Pavao, I. Guyon, A.-C. Letournel, X. Baró, H. Escalante, S. Escalera, T. Thomas, Z. Xu, Codalab competitions: An open source platform to organize scientific challenges, Technical

report (2022). URL: https://hal.inria.fr/hal-03629462v1.
[14] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
[15] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, G. Varoquaux, API design for machine learning software: experiences from the scikit-learn project, in: ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 2013, pp. 108–122.
[16] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, in: PML4DC at ICLR 2020, 2020.
[17] T. Kudo, J. Richardson, SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 66–71. URL: https://aclanthology.org/D18-2012. doi:10.18653/v1/D18-2012.
[18] Y. N. D. D. L.-P. Hongyi Zhang, Moustapha Cisse, mixup: Beyond empirical risk minimization, International Conference on Learning Representations (2018). URL: https://openreview.net/forum?id=r1Ddp1-Rb.
[19] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, in: International Conference on Learning Representations, 2018.
[20] E. Ben-Zaken, S. Ravfogel, Y. Goldberg, Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models, 2021.
[21] J. Cañete, S. Donoso, F. Bravo-Marquez, A. Carvallo, V. Araujo, Albeto and distilbeto: Lightweight spanish language models (2022).