NLP-CIMAT at PoliticEs 2022: PolitiBETO, a Domain-Adapted Transformer for Multi-class Political Author Profiling

Emilio Villa-Cueva[1,*,†], Ivan González-Franco[1,4,†], Fernando Sanchez-Vega[1,2,3,†] and Adrián Pastor López-Monroy[1,†]

[1]*Mathematics Research Center (CIMAT), Jalisco S/N Valenciana, 36023 Guanajuato, GTO México.*

[2]*Consejo Nacional de Ciencia y Tecnología (CONACYT), Av. de los Insurgentes Sur 1582, Benito Juárez, 03940, CDMX, México.*

[3]*El Colegio de México (COLMEX), Picacho Ajusco 20, Tlalpan, 14110, CDMX, México.*

[4]*University of Guanajuato, Loma del Bosque 103, Lomas del Campestre. León, Gto., México*

Abstract

This work is an overview of the NLP-CIMAT-GTO team submission for the shared task of PoliticEs at IberLEF 2022, which consisted of Political Author Profiling. In this paper we propose PolitiBETO, a model pretrained in a domain of political tweets. We achieve this by proposing a learning framework based on two ideas: Domain Adaptation and Ensemble Learning of Transformers. In summary, our pipeline consists of first adapting the domain of a pretrained BETO to a political language domain. Subsequently, we compose an ensemble using several instances of pretrained adapted BETO models, which predicts the test data at a tweet level. These predictions are then merged through a majority vote to determine the labels of a given author based on their tweets. Our proposal obtained a first place for the multi-class political ideology label by an important margin of more than 6%. Also, it scored first in terms of average performance of the two political-ideology labels and second overall.

Keywords

Political ideology detection, Transformer Ensemble, Domain Adaptation

1. Introduction

Political ideology can be an important driving factor in how people act in society and has been shown to be related to personality traits [1]. The political compass of a person can also relate to other factors such as openness to experiences, consciousness, or attitude towards vaccination campaigns [2, 3]. Because of this, identifying the political ideology of groups of people can become a useful tool for a variety of applications.

IberLEF 2022, September 2022, A Coruña, Spain

*Corresponding author.
†These authors contributed equally.
evillacueva@gmail.com (E. Villa-Cueva); ivan.gonzalez@cimat.mx (I. González-Franco); fernando.sanchez@cimat.mx (F. Sanchez-Vega); pastor.lopez@cimat.mx (A. P. López-Monroy)
0000-0003-4998-7929 (E. Villa-Cueva); 0000-0002-8533-2818 (F. Sanchez-Vega); 0000-0003-1018-4221 (A. P. López-Monroy)

PoliticEs [4] is a shared task at IberLEF2022 that consisted in profiling Spanish journalists and politicians on Twitter, focused on determining gender, profession, but more importantly, the political ideology. The ideology profiling task has two aspects: (1) Binary, which consists in classifying whether an author's ideology is *left* or *right*, and (2) multiclass, that aims to categorize this in a more specific sense by distinguishing between left, moderate-left, moderate-right, and right. The most interesting targets of the competition are the last two ideology labels, which make this task the first of its kind in Spanish.

The authors of the tweets selected for the task were rather relevant selected people from different government levels (politicians) and different newspapers (journalists). Thus, the dataset is composed of tweets from around 400 authors with at least 120 tweets for each. In this work, we outline the strategy of the NLP-CIMAT-GTO team proposal, which was primarily aimed at the identification of multiclass and binary political ideology.

Our proposed approach is inspired in three insights: (a) A Two-stage Domain Adaptation of a Transformer language model to a general political domain, (b) transfer learning from a general political domain to a highly specialized domain (the PoliticEs task), and (c) maximizing the classification performance of the base Transformer by mitigating the well-known performance variance of models based in the BERT architecture [5, 6, 7]. Based on these insights, our system follows three steps. For the first step, we present PolitiBETO, a Transformer-based model pretrained in a specific political-tweets language domain, which we accomplish by adapting BETO [8] using a two-stage TF-based Domain Adaptation. In the second step, we fine-tune several instances of this model using a single-task and a multi-task configuration and we subsequently use these trained instances as building blocks for our third step, an ensemble that we employ to predict the tweets in the test set using two distinct schemes: a direct inference of each tweet and an inference over packages of 3 tweets concatenated. Finally, we determine the overall predictions for each author by using a majority vote over the predictions of their tweets. Our system, which is mostly tailored to the task of political author profiling obtained the highest score by an important margin for the multi-class ideology label, the first place in average performance for the ideology labels and second place for the average performance of all the classes.

This paper is structured as follows: In Section 2 we introduce a few ideas that are useful for the description of our pipeline, which we carefully illustrate in Section 3. Section 4 describes the PoliticEs dataset as well as the data used for the Domain Adaptation. In Section 5 we show the evaluation results of our proposed framework. Section 6 discusses the ethical implications of our strategy. Finally, Section 7 contains the conclusion of this work.

2. Related Work

2.1. Pretraining Language Models for specific domains.

The versatility of BERT [9] allows it to be fine-tuned for a wide variety of tasks in NLP, but it comes at the cost of a general domain pretraining that may be slightly different from the domain of the objective task. The most obvious example is language. For example, for Spanish tasks, BETO [8] —a BERT model pretrained in Spanish— will be a significantly better model than BERT presumably because the similarity of the pretraining corpus and downstream tasks is greater.

Interestingly, the latter can be taken a step further by pretraining BERT in a specific language domain, which has shown to outperform general language domain pretraining. SciBERT [10] and BioBERT [11] are two of the most notable works that have been pretrained for scientific and biomedical domains respectively, surpassing vanilla BERT when evaluated for specific tasks in these fields. In a more recent work, BERTweet, a BERT model pretrained in around 850M (English) tweets achieve an outstanding performance in Twitter-related tasks [12].

Notwithstanding that pretraining language models provides a huge advantage, the pretraining process of models such as BERT is a highly expensive and complex, task that is potentially prohibitive without huge computational resources. Furthermore, the task of collecting a corpus of suitable size to pretrain BERT can be a difficult task depending on the desired domain. In this work we relieved that by conducting a simple Domain Adaptation such as the one proposed by [13], which takes an already pretrained BERT and resumes the training for the Masked Language Modelling task by using a significantly smaller corpus. It is composed of 50% of source-domain data (the data it was initially pretrained on) and 50% target-domain data (data in the desired language domain), resulting in a model with improved performance in the new domain by using significantly fewer resources.

2.2. High Quality Fine tuning and Stability of BERT-based models.

Despite its outstanding performance for many NLP tasks, several studies have shown that BERT is quite sensitive to the initialization weights of its classification layer, as well as to the data ordering during the fine-tuning stage [5, 6, 7]. This means that some random seeds will achieve significantly higher performance metrics than others, resulting in a high variance in the performance of the models.

One of the solutions for this is to use ensembles of several instances of a model. For example, the authors of [14, 15] introduced this approach in previous competitions obtaining promising results. They found that, in practice, using these ensembles not only reduced the variance in performance metrics of the model but also improved it by some degree. Moreover, they empirically noted that as more models were used to build the ensemble, the performance also improved, but this improvement seemed to reduce for more than 10 or 15 models.

The key idea of the system is fine-tuning several instances of the same BERT model for a given task, then predicting the test set by aggregating the outputs from the softmax layer of those instances. The classification for each input is done by carrying out a weighted voting procedure that selects the class with the highest value. (See Figure 1). Although this instability is most present in datasets smaller than PoliticEs, we believe that for this task our model will also benefit from this approach, so we implement an ensemble of this kind in our pipeline.

3. A Robust Pipeline for Political Ideology Identification

In this section, we describe the three steps that compose the procedure we use for the task of Political Author Profiling, illustrated in Figure 2.

In Section 3.1 we present the first step, that consist in a two-stage Domain Adaptation procedure. Section 3.2 explain the second step, in which we fine-tune the models to use in the

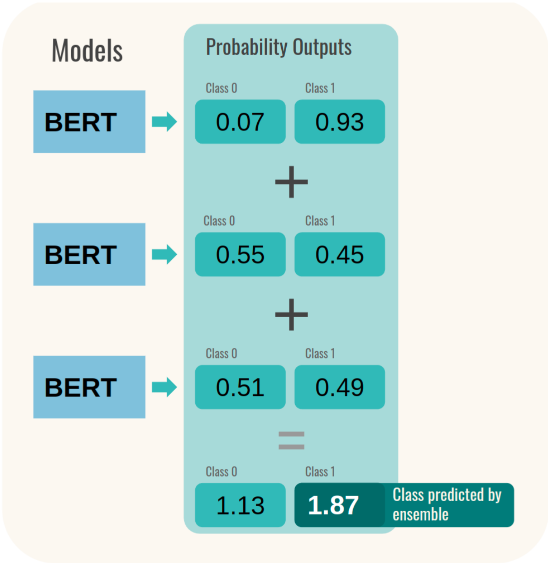

Figure 1: An example of a traditional weighted voting system for an ensemble that uses 3 instances of BERT.

Transformer Ensemble. Section 3.3 contains the procedure for the Ensemble prediction and Author Profiling.

3.1. A Two-Stage Domain Adaptation (PolitiBETO)

BETO [8] is a BERT model that has been pretrained using corpora of general-domain Spanish and generally performs effectively for a variety of Spanish NLP tasks. However, the language domain of the PoliticEs corpus —political tweets in Spanish— is quite specific. We propose a Two-Stage Domain Adaptation on BETO with the intuitive idea to teach BETO the structure of language in Twitter and further specialize it for a general political language of newspapers. Both Domain Adaptations are carried following the procedure suggested for AdaptaBERT [13]. It consists in continuing BETO's pretraining through the *Masked Language Modelling* task for more epochs. The process uses a corpus composed of 50% of source-domain data (in the same language domain of BETO original pretraining) and 50% of target-domain data (about social media and politics in our case). In this paper, we propose a slightly different, but effective two-step Domain Adaptation, resulting in the two following adaptations;

1. A first step Domain Adaptation to a social media domain, using tweets in Spanish obtained from the *TwitterSentimentDataset* corpus, available at [16]. The entire corpus —which contains contains around 250k tweets— was used for the Domain Adaptation. The resulting dataset was composed with 50% of tweet-domain ($250k$) and 50% general-domain ($250k$) data.
2. A second step Domain Adaptation to a political domain employing news articles related to Spain politics, for the political nature of the task. The process for acquiring these tweets is described in section 4.2. This dataset resulted in 25% tweet-domain ($250k$), 25%

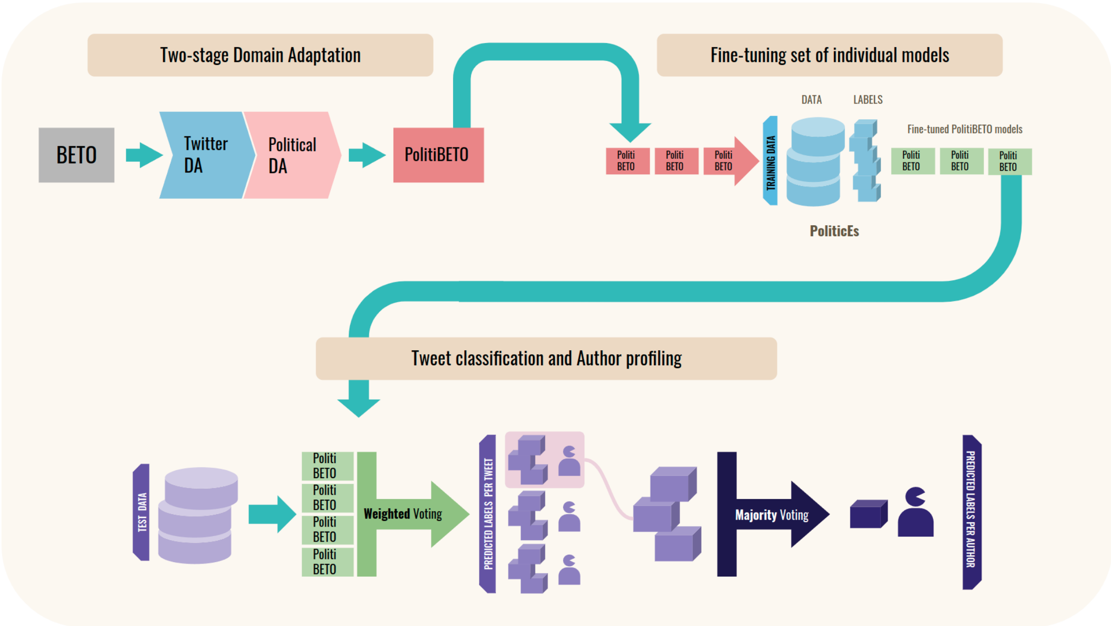

Figure 2: Pipeline of the proposed method

general-domain ($250k$), and 50% political-domain ($500k$) data.

We discuss the details of our Domain Adaptation implementation in Section 4.3. This procedure resulted in a pretrained model tailored to solve NLP tasks in a setting of political tweets in Spanish, which from here onward we will refer to as *PolitiBETO*. For the following steps of our pipeline, we used PolitiBETO as our base model. A difference between our approach and AdaptaBERT is that we condition BETO for the domain of political tweets by using a simple Term-Frequency bounded Domain Adaptation. It follows the same procedure as AdaptaBERT, with the main difference that we bound the possible masked tokens to those whose Term-Frequency falls within the quantiles $Q(0.55)$ and $Q(1.0)$.

We make PolitiBETO's pretrained weights available at https://huggingface.co/nlp-cimat/politibeto, to be fine-tuned either for the PoliticEs corpus or other similar tasks.

3.2. Fine-tuning models to build an ensemble.

The next step is to train the different instances of PolitiBETO to use in an ensemble configuration, we follow a standard fine-tuning procedure experimenting with three different configurations:

1. Single-task binary fine-tuning: we fine-tune a model for each binary label. This means, one for *gender*, the other for *profession*, and another for *ideology-binary*.
2. Multi-task binary fine-tuning: our model is trained to predict three binary categories at the same time: *gender*, *profession*, and *ideology-binary*.
3. Single-task multi-class fine-tuning: for classifying the label *ideology-multiclass*.

The reason we train two different settings for the binary labels is that we hypothesize that some labels could benefit from being trained in a multi-task setting. Other labels, on the other hand, could perform better when trained individually. The fine-tuning is made at a tweet level using the entire set of tweets, without any kind of distinction among authors. For each configuration, we train 15 model instances that will be eventually used in an ensemble configuration. A simple diagram of these configurations is depicted in the left panel of Figure 3.

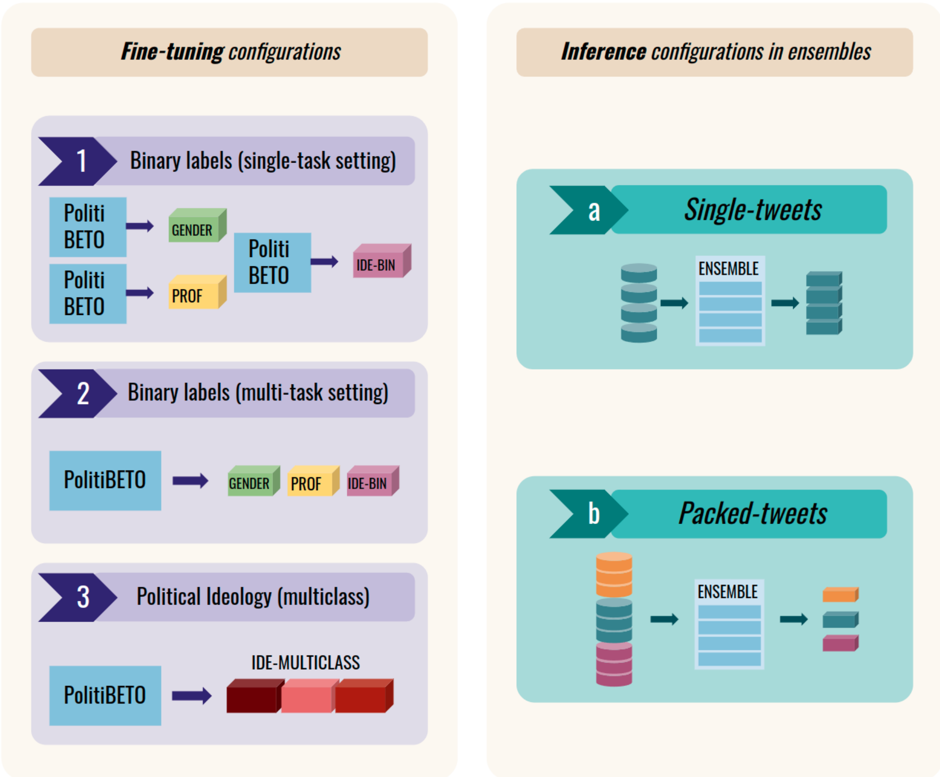

Figure 3: Illustration of the fine-tuning configurations and Ensemble approaches used.

3.3. Inference per tweet using Ensembles and Author Profiling

We approach the Author Profiling problem by first predicting the test dataset at a tweet level. Instead of directly predicting labels per author, we use portions of their tweet history to make the predictions. These portions can be either a single tweet or a package of tweets. We experiment with both approaches in this step (See right panel of Figure 3):

a) *Single-tweets* approach: Predict each tweet individually, ending up with a set of N predictions.
b) *Packed-tweets* approach: Predict *packages* of k concatenated tweets that belong to the same author (and thus share the same labels) and have been randomly sampled. This reduces the size of predicted labels to around N/k.

The packed-tweets technique feeds our model more information attempting to improve its odds of selecting the correct categories. Using these two approaches, we predict the test set using the different configurations of the models trained in Section 3.2, and merge their predictions per configuration using an Ensemble of 15 PolitiBETO models configured as described in Section 2.2, this is to reduce the variance that might come from PolitiBETO and to improve its performance as much as possible (See right panel of Figure 3). Their individual predictions are aggregated using weighted voting and we end up with a set of predicted labels for each Tweet (or package of tweets).

Finally, to determine the predicted labels for each author, we perform a straightforward majority vote using the predicted labels that belong to their tweets (or tweet packages). This simply assigns each author the most common value of their tweets for each label.

4. Dataset and Experimental Settings

4.1. The PoliticEs dataset.

The PoliticEs dataset —an extension to the PoliCorpus [17]— has more than 45 thousand tweets from political journalists and politicians and has been labeled for the task of author profiling. In the case of politicians, users were selected from members (or former members) of political parties or different ranks of the government, whose political ideology could be figured from the party they belong to. Users of political journalists, on the other hand, were selected from different Spanish news media whose political affiliation can be guessed according to the editorial line where they write.

In total, around 400 users with more than 120 tweets each were assigned the following labels:

- Gender: *male, female.*
- Profession: *politician, journalist.*
- Ideology-binary: *left, right.*
- Ideology-multiclass: *left, moderate-left, moderate-right, right.*

For this shared task, this data was split into training and test sets composed of 80% and 20% of the data respectively. Of course, Twitter accounts were anonymized by replacing any mention with @user, and authors that are in the training set were not in the test set to avoid author identification instead of author profiling.

4.2. Political Corpus in Spanish for Domain Adaptation

To obtain corpora with contemporary topics in Spanish politics for the political Domain Adaptation discussed in section 3.1, we selected the newspaper *El País*. Compared to other newspapers such as *ABC* and *El Mundo*, studies have concluded that *El País* tends to be more sober and analytical when reporting news [18]. The same study suggested that it is less compromised with a fixed posture and manages to be more objective. However, others have pointed out that in previous years it may have been more inclined towards a leftist stance[19]; and more recently towards a more rightist posture [20, 21]. Because of this, the resulting Domain Adaptation corpora may not contain a perfectly balanced amount of articles with different currents. In any case, it is still a meaningful source of target-domain data for the Domain Adaptation.

We selected *El Pais*'s articles out of those issued from January 2012 to March 2022. Besides its title and content paragraphs, each article contains the Section where it belongs and a set of tags.

To choose those that were useful for our purpose, we filtered the publications from the selected timespan with a condition: either their section is *España*, or their tags include the substring "españ" (using a case-insensitive comparison). Some of the possible tags could be, for example, *España*, *Español*, *Legislación española*, *Literatura española* and *Federación española*. After filtering with those conditions, the final dataset summed to around 312 thousand articles, with the following characteristics:

a) Their most frequent sections are (in descending order): *Política*, *Cataluña*, *Madrid*, *Comunidad valenciana*, *Pais Vasco*, *Economía*, *Opinión* and *Cultura*.
b) They have the tags (in descending order): *España*, *Política*, *Administración pública*, *Justicia*, *Partidos políticos*, *Sociedad*, *Cataluña*, *Economía* and *Cultura*.

Nevertheless, due to time constraints, we only used a subset of around 500k paragraphs (along with their respective Title and Tags) randomly sampled from the 312k articles of the final dataset for the Domain Adaptation procedure.

4.3. Implementation Details

For the Domain Adaptation procedure introduced in Section 3.1, we train BETO for the Masked Language Modelling task for three epochs using a batch size of 128 and a learning rate of 5×10^{-5}.

To train our adapted models for the target task, we employed a standard preprocessing for the tweets, as shown in Figure 4. For the fine-tuning of the binary labels, we train models for three epochs and models using a learning rate of 1×10^{-5}. And for the multi-class label, we train for five epochs using a learning rate of 4×10^{-5}. For the experiments with RoBERTuito [22] in Section 5.1 we used a learning rate of 1×10^{-5} for the multi-class setting as well because a larger learning rate failed to converge. We use a batch size of 16 for both cases. For the *packed* inference, we set $k = 3$ due to the maximum sentence size BERT can receive as an input. Everything was trained using an NVIDIA Titan RTX GPU.

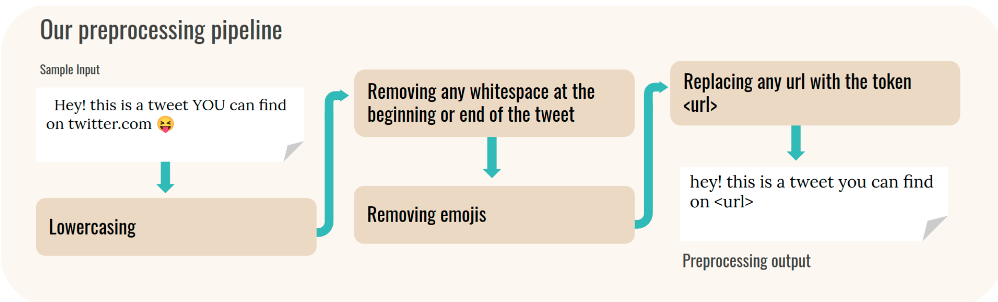

Figure 4: Our preprocessing system.

5. Evaluation

In this section, we exhibit the results of our system. First, we show preliminary evaluations of the effectiveness of Domain Adaptation and Ensembles. Secondly, we indicate the performance metrics obtained in the test set by using the configurations discussed in Section 3.

5.1. A preliminary evaluation of the use of Two-Stage Domain Adaptation and Ensembles for the task Ideology Identification.

To properly assess whether our system was effective for the shared task, we evaluated how our two main ideas —Domain Adaptation and Ensemble learning— accomplished the classification of political ideology, which was our main concern for the Author Profiling task. We assembled an *in-house* training and evaluation data partition using a proportion of 85% and 15% of the training dataset of PoliticEs. This partition allowed us to run as many experiments as we needed to assess how the Domain Adaptation and Ensemble strategies improved BETO for the task of binary and multi-class ideology identification. For this part we entirely removed the authors from the dataset, focusing on anonymously predicting the correct labels using a single-task fine-tuning.

Our experiments, shown in Table 1, indicate that both aspects of our approach improve the model's performance in the task. The main improvement comes from Domain Adaptation, which yields a notable performance boost over the vanilla BETO, leveraging the knowledge of the two domains. We also compare PolitiBETO with RoBERTuito [22], a model based on the RoBERTa architecture [23] that has been pretrained from scratch using around 500M tweets and has been designed for tasks in a Spanish social media domain (such as PoliticEs). We observe that even if RoBERTuito performs significantly better than BETO in most tasks, PolitiBETO still achieves higher performance metrics for all the labels due to its highly specific language domain. These preliminary runs demonstrate that PolitiBETO outperforms a general domain model such as BETO, and a specific social media domain model, namely RoBERTuito, for tasks related to political social media. However, we also note that the second step of the Domain Adaptation (with the political corpora) slightly reduces the model's ability to predict gender.

Furthermore, we observed that the Domain Adaptation procedure also increased the variance in individual model results. This justifies, even more, the necessity of an ensemble to achieve reliable results whilst improving performance. The ensemble learning results show that there is indeed an important improvement, especially for multi-class ideology.

5.2. Evaluation of submissions on the PoliticEs test set.

In this subsection, the results obtained for the test set of the PoliticEs task of author profiling are displayed. First, the results for *multiclass-ideology* are shown in Table 2, we see that for this label, packing the tweets strongly improves performance by more than 36% (See right panel of Figure 3 for an illustration of the packed and single-tweet approaches).

Binary-classification results are exhibited in Table 3. We can see that for this case the best approach varies depending on the label. For *binary-ideology* the best approach is a single-task model trained with packed tweets, which is —as expected— the same approach followed for

Table 1
Preliminary experiments on in-house data partitions to estimate the effectiveness of the proposed strategies. BETO** refers to the first step of the Domain Adaptation only for a Tweet Domain. We trained five models for each configuration to compute their average individual performance and then build their respective ensembles.

	F1-macro of single models (Average of 5 runs)			
Model	Ideology - Binary	Ideology - Multi	Gender	Profession
BETO	76.44 ± 0.188	62.373 ± 0.234	66.314 ± 0.268	77.552 ± 0.381
RoBERTuito	78.43 ± 0.331	62.27 ± 1.244	67.676 ± 0.245	77.713 ± 0.254
BETO**	77.873 ± 0.194	62.797 ± 0.36	**68.373 ± 0.16**	77.497 ± 0.474
PolitiBETO	**78.908 ± 0.242**	**64.656 ± 0.422**	67.347 ± 0.141	**78.129 ± 0.4**
	F1-macro of ensembles of 5 instances			
Model	Ideology - Binary	Ideology - Multi	Gender	Profession
BETO	77.24	65.08	67.29	77.97
RoBERTuito	79.10	63.75	67.78	77.82
BETO**	78.67	65.13	**68.60**	77.95
PolitiBETO	**79.42**	**66.42**	68.24	**78.13**

Table 2
PolitiBETO ensembles for multi-class ideology in PoliticEs test set.

Tweet-inference method	F1 - macro
Single-tweets	65.7
Packed-tweets	**89.63**

multiclass-ideology.

Table 3
PolitiBETO ensembles for binary labels.

Configuration	F1-macro		
	Binary Ideology	Profession	Gender
multi-task+single-tweets	90.22	77.51	**78.48**
single-task+single-tweets	89.21	**92.13**	75.53
single-task+packed-tweets	**96.15**	69.44	57.08

Nevertheless, packing tweets seems to harm the performance of the other labels. Both gender and profession labels worked significantly better without packing the tweets. For the former, we see that the multi-task approach seems to deliver the best metrics. For the latter, a single-task strategy performs the best.

We selected the best performing approaches for each label for our chosen final submission, resulting in a combination of different ensembles of PolitiBETO models (shown in Table 4). Our final submission obtained first place in the label of multiclass-ideology identification by an important margin of more than 6%, a first place in average performance of the political-ideology labels and a second place overall in the competition by average performance of all the labels.

Table 4
Our submitted results. * Indicates a first place and ** a second place.

	F1 macro					
	Multi-class Ideology	Binary Ideology	Profession	Gender	Average Ideology	Average all-labels
Baseline	44.06	59.57	43.24	57.62	51.81	51.12
Participant's average	67.07	85.58	80.60	75.14	76.33	77.10
Our selected submission	**89.63***	**96.15****	**92.13**	**78.48**	**92.89***	**89.10****

6. Ethical issues

There are a few ethical aspects of this work that must be taken into consideration. First, it is important to note that the purpose of a system tailored to identify the political inclinations of users is intended to be employed for merely informative or research-related purposes. Applications such as targeted advertising are not the purpose of this work.

Secondly, the use of ensembles does improve our pipeline's performance, but to the detriment of using significantly more computational resources than using a single model. In our particular case, it uses around fifteen times more resources since we train 15 PolitiBETO models. Regardless, we counteract this with the resources we save from using a Domain Adaptation in the first step instead of pretraining an entire model from scratch, resulting in a significant decrease in our pretraining carbon footprint.

7. Conclusion

In this work, we proposed PolitiBETO and a powerful framework suitable for the shared task of PoliticEs 2022. PolitiBETO is built by means of a two-stage Domain Adaptation that first learns language in a Twitter domain and then in a political domain. Our framework mainly consisted of an Ensemble configuration of several instances of a PolitiBETO model. Our experiments demonstrate that the proposed strategy is successful for the task of binary and multi-class political ideology classification, which we can harness for the task of author profiling by voting over many tweets of the same author, obtaining outstanding results for this task as well.

The focus of our work was in political specialization. For future work, we are interested in extending our framework beyond two stages. For example, three or four stages that makes possible to better learn categories such as gender and profession.

Our pipeline demonstrated to be highly effective for the task of political author profiling, obtaining exceptional results for this shared task, where it ranked first for multi-class political ideology identification and scored a second-place by its average F1 macro over all the labels.

Acknowledgments

The authors thank *Consejo Nacional de Ciencia y Tecnología* (CONACYT), *Centro de Investigación en Matemáticas* (CIMAT) and *Instituto Nacional de Astrofísica, Óptica y Electrónica* (INAOE) for the computer resources provided through the INAOE Supercomputing Laboratory's Deep Learning Platform for Language Technologies (*Laboratorio de Supercómputo: Plataforma de Aprendizaje Profundo*) with the project *"Identification of Aggressive and Offensive text through* *specialized BERT's ensembles"* and CIMAT Bajio Supercomputing Laboratory (#300832). Sanchez-Vega would like to thank CONACYT for its support through the Program *"Investigadoras e Investigadores por México"* by the project *"Desarrollo de Inteligencia Artificial aplicada a la prevención de violencia y salud mental."* (ID. 11989, No. 1311) and the COLMEX Interdisciplinary Data Science Program (Open Society Grant).

References

[1] M. Fatke, Personality traits and political ideology: A first global assessment, Political Psychology 38 (2017) 881–899. URL: http://www.jstor.org/stable/45095184.

[2] B. Baumgaertner, J. E. Carlisle, F. Justwan, The influence of political ideology and trust on willingness to vaccinate, PLOS ONE 13 (2018) e0191728. URL: https://doi.org/10.1371/journal.pone.0191728.

[3] 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 (2011) 34–51. URL: https://doi.org/10.1111/j.1540-5907.2011.00568.x.

[4] 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).

[5] J. Dodge, G. Ilharco, R. Schwartz, A. Farhadi, H. Hajishirzi, N. A. Smith, Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping, CoRR abs/2002.06305 (2020). URL: https://arxiv.org/abs/2002.06305. arXiv:2002.06305.

[6] M. Mosbach, M. Andriushchenko, D. Klakow, On the stability of fine-tuning BERT: misconceptions, explanations, and strong baselines, CoRR abs/2006.04884 (2020). URL: https://arxiv.org/abs/2006.04884. arXiv:2006.04884.

[7] T. Zhang, F. Wu, A. Katiyar, K. Q. Weinberger, Y. Artzi, Revisiting few-sample BERT fine-tuning, CoRR abs/2006.05987 (2020). URL: https://arxiv.org/abs/2006.05987. arXiv:2006.05987.

[8] 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.

[9] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.

[10] I. Beltagy, K. Lo, A. Cohan, SciBERT: A pretrained language model for scientific text, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3615–3620. URL: https://aclanthology.org/D19-1371. doi:10.18653/v1/D19-1371.

[11] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics 36 (2019) 1234–1240. URL: https://doi.org/10.1093/bioinformatics/btz682. doi:10.1093/bioinformatics/btz682.

[12] D. Q. Nguyen, T. Vu, A. Tuan Nguyen, BERTweet: A pre-trained language model for

English tweets, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 9–14. URL: https://aclanthology.org/2020.emnlp-demos.2. doi:10.18653/v1/2020.emnlp-demos.2.

[13] X. Han, J. Eisenstein, Unsupervised domain adaptation of contextualized embeddings: A case study in early modern english, CoRR abs/1904.02817 (2019). URL: http://arxiv.org/abs/1904.02817. arXiv:1904.02817.

[14] M. Guzman-Silverio, Á. Balderas-Paredes, A. P. López-Monroy, Transformers and data augmentation for aggressiveness detection in mexican spanish, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) co-located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, September 23th, 2020, volume 2664 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 293–302. URL: http://ceur-ws.org/Vol-2664/mexa3t_paper9.pdf.

[15] V. Gómez-Espinosa, V. Muñiz-Sánchez, A. P. López-Monroy, Transformers pipeline for offensiveness detection in mexican spanish social media, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021), CEUR Workshop Proceedings, CEUR-WS.org, 2021.

[16] D. Garnacho, TwitterSentimentDataset: Dataset de Sentimientos en Español, https://github.com/garnachod/TwitterSentimentDataset, 2015.

[17] 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.

[18] M. J. R. Crespo, El país, abc y el mundo: Tres manchetas, tres enfoques de las noticias, 1999.

[19] M. Vázquez Bermúdez, Los medios toman partido, Ámbitos. Revista Internacional de Comunicación (2006). URL: https://www.redalyc.org/articulo.oa?id=16801513.

[20] A. Torrús, "El País" cambia de director y la redacción teme un giro a la derecha, Público (2014). URL: https://www.publico.es/actualidad/pais-cambia-director-y-redaccion.html.

[21] I. Sáenz-de Ugarte, Por qué El País quiere irse más a la derecha, El Diario.es (2014). URL: https://www.eldiario.es/rastreador/testamento-pais-periodicos_132_5004225.html.

[22] J. M. Pérez, D. A. Furman, L. A. Alemany, F. Luque, Robertuito: a pre-trained language model for social media text in spanish, CoRR abs/2111.09453 (2021). URL: https://arxiv.org/abs/2111.09453. arXiv:2111.09453.

[23] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, CoRR abs/1907.11692 (2019). URL: http://arxiv.org/abs/1907.11692. arXiv:1907.11692.