UMUTeam at PoliticES-IberLEF2023: Evaluating Transformers for Detecting Political Ideology in Spanish Texts

Ronghao Pan[1,*,†], Camilo Caparrós-Laiz[1,*,†] and Ángela Almela[2,*,†]

[1]*Facultad de Informática, Universidad de Murcia, Campus de Espinardo, 30100, Spain*
[2]*Facultad de Letras, Universidad de Murcia, Campus de La Merced, 30001, Murcia, Spain*

Abstract

This paper describes the participation of the UMUTeam in the PoliticES shared task organized at IberLEF 2023 within the SEPLN conference. It is an automatic document classification task on clusters of texts, which consists of extracting the self-assigned gender and profession as demographic traits, and political ideology as a psychographic trait from a set of texts written in Spanish by several authors that share those traits. For this task, we have fine-tuned the MarIA, a transformer model proposed for Spanish, to create classification models for each feature. After several submissions for these tasks, our team ranked seventh out of 12 participants, with an average F1 score of 69.225% of all classification models.

Keywords

Natural Language Processing, Transformers, Politic ideology detection, Large Language Model, Multiclass classification

1. Introduction

In general, individuals tend to be reluctant to heed the advice and instructions of politicians who do not coincide with their ideology. In extreme cases, individuals may show a strong bias towards a particular political party while roundly rejecting others' ideology. In this line, [1] analyzed the relationship between personality traits and political ideology by collecting data from 21 countries. The author observed a correlation between political ideology and the big five personality traits. Therefore, political ideology is a psychographic characteristic that helps to understand individual and social behavior, as it encompasses moral and ethical values and inherent attitudes, evaluations, biases, and prejudices [2]. Furthermore, it can enhance micro-targeting efforts, enabling public authorities and local governments to adopt more effective communication strategies during crises [3].

With the aim of detecting political ideology information from Spanish texts, the shared-task *PoliticES* [4] has been organized as part of the IberLEF 2023 [5] workshop within the framework of the SEPLN 2023 conference. The organizers proposed an automatic document classification task on clusters of texts, consisting of the extraction of self-assigned gender and profession as

IberLEF 2023, September 2023, Jaén, Spain
ronghao.pan@um.es (R. Pan); camilo.caparrosl@um.es (C. Caparrós-Laiz); angelalm@um.es (Á. Almela)
0009-0008-7317-7145 (R. Pan); 0000-0002-5191-7500 (C. Caparrós-Laiz); 0000-0002-1327-8410 (Á. Almela)

demographic traits and political ideology as a psychographic trait from a set of texts written in Spanish by several authors that shared those traits.

This work presents the participation of the UMUTeam in this shared task, which is based on exploring the fine-tuning of different pre-trained and Transformers-based Large Language Models (LLMs) to detect gender, profession, and political ideology. The rest of the paper is organized as follows. Section 2 presents the task and dataset provided. Section 3 describes the methodology of our proposed system for addressing the task. Section 4 shows the results obtained. Finally, Section 5 concludes the paper with some findings and possible future work.

2. Task description

The shared task *PoliticES* 2023, organized at IberLEF workshop, aims to extract political ideology information from a set of texts written in Spanish by several authors. Specifically, the organizers propose a task of identifying demographic traits (gender and profession) and political ideology as psychographic traits from a given cluster of texts (document classification). In the case of the identification of political ideology, the organizers have defined it as a binary problem and as a multiclass problem.

The dataset provided is an extension of the PoliCorpus 2020 dataset [3] and the corpus used for the PoliticES 2022 shared task [6]. It was gathered between 2020 and 2022 from Twitter accounts of politicians, political journalists, and celebrities in Spain using UMUCorpusClassifier [7]. The users of the dataset are labeled with their gender (male, female), profession (politician, journalist, celebrity), and political spectrum on two axes: binary (left, right) and multiclass (left, moderate left, moderate right, right). Regarding the tweets collected from each user, the organizers removed discarded retweets and tweets that contain headlines from news sites and removed tweets written in languages other than Spanish. Moreover, they anonymized them by replacing all mentions with *@user*, except for real users. Furthermore, other entities, such as political party references, are also replaced with *@political_party* token. The final dataset is composed of approximately 2,800 different clusters. The tweets that belong to each cluster are selected favoring diversity, including texts from different dates and topics. For this shared task, the dataset is divided into training and test sets (80%-20%).

The training set consists of 180,000 tweets from 2,250 clusters and the test set consists of 547 clusters with a total of 43,760 tweets. The distribution of demographic and psychographic traits of each cluster is shown in Table 1.

3. Methodology

This task involves identifying gender and profession (demographic traits) and political ideology (psychographic traits) of users in a given set of texts. The pipeline used to participate in this task can be described as follows. First, the dataset has been processed, and emoji features have also been added. Second, the training dataset was split into training and validation at the tweet level. Third, a classification model is created for each of the features using the fine-tuning approach. Fourth, having the classification models at the sentence level, two strategies have been evaluated to identify the demographic and psychographic traits of the users (at document

Table 1
The distribution of demographic and psychographic traits of each cluster.

Trait		Training	Test	Total
Gender	**Male**	1,493	381	**1,874**
	Female	757	166	**923**
Profession	**Politician**	752	186	**938**
	Journalist	1,385	305	**1,690**
	Celebrity	113	56	**169**
Binary ideology	**Left**	1,255	327	**1,582**
	Right	995	220	**1,215**
Multiclass ideology	**Moderate left**	825	210	**1,035**
	Left	430	117	**547**
	Moderate right	728	153	**881**
	Right	267	67	**334**

level): (1) **mode**, which consists of predicting each user's tweet individually and selecting the most repeated label among the results obtained with the classifier, and (2) **highest probability**, which selects the label with the highest probability. The system architecture is depicted in Figure 1

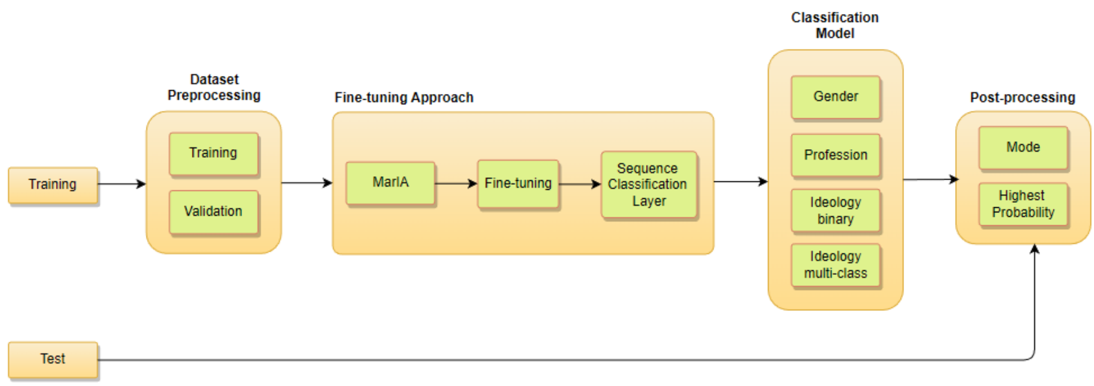


Figure 1: Overall system architecture.

3.1. Dataset preprocessing

As mentioned above, tweets from the same user will have the same demographic and psychographic traits, Therefore, to carry out the problem of identifying these traits, we have created a sentence-level classification model for each of them with all the tweets and the distribution of demographic and psychographic traits of the dataset as shown in Table 2. In this case, we have divided the training set into two subsets (80%-20%): training and validation. The customized validation split is created using stratified sampling, in order to maintain a balance between the labels.

Table 2
The distribution of demographic and psychographic traits of the training and validation split.

Trait		Training	Validation
Gender	**Male**	95,552	23,888
	Female	48,448	12,112
Profession	**Politician**	48,128	12,032
	Journalist	88,640	22,160
	Celebrity	7,232	1,808
Binary ideology	**Left**	80,320	20,080
	Right	63,680	15,920
Multiclass ideology	**Moderate left**	52,800	6,880
	Left	27,520	13,200
	Moderate right	46,592	4,272
	Right	17,088	11,648

3.2. Fine-tuning approach

We utilized the fine-tuning approach of a transformer-based masked language model for Spanish called MarIA [8] to carry out the identification of different features. MarIA is based on the RoBERTa base model and has been pre-trained using the largest Spanish corpus known to date, with a total of 570 GB of clean and deduplicated text processed. The fine-tuning process involves adapting and adding a classification layer to the model to perform the training of the complete model. In this way, the model takes advantage of MarIA's pre-trained linguistic knowledge and adapts it specifically for a particular classification task, which can significantly improve performance on that task. The model has been fine-tuned with a training batch size of 16, 6 epochs, a learning rate of 2e-5, and a decay of weights of 0.01.

4. Results

This section describes the systems submitted by our team in each run and the overall results obtained in this shared task. It should be noted that each participating team was allowed to submit ten runs.

We sent two runs for this task. The results, and a brief description of each, are depicted in Table 3. The first run is based on the mode strategy, which consists of selecting the most frequent label obtained in the user's text set for each feature, achieving an average F-score of 69.225%. The second run, which entailed the highest probability strategy, achieved an average F-score of 55.889%. It can be observed that the macro F1-score obtained with the mode strategy applied in different classification models for both demographic and psychographic features performed better than the highest probability strategy. In addition, the binary classification models (gender and binary ideology) outperformed the multiclass models (profession and ideology multiclass).

The official leaderboard for this task is depicted in Table 4. We achieved the seventh position in the ranking with an average F-score of 69.225%. The teams *ELiRF VRAIN* and *HiTZ-IXA* achieved the best position, outperformed our best run with an average F-score of 81.131% and 79.348%, respectively.

Table 3
Results of each feature. For each strategy, the macro precision (M-P), macro recall (M-R), and macro F1-score (M-F1) are reported.

Strategy	Avg. F1	Gender			Profession			Ideology binary			Ideology multiclass		
		M-P	M-R	M-F1	M-P	M-R	M-F1	M-P	M-R	M-F1	M-P	M-R	M-F1
Mode	**0.69225**	0.77837	0.66614	**0.68363**	0.91832	0.60625	**0.61636**	0.90809	0.84467	**0.86002**	0.76738	0.58239	**0.60901**
Highest Prob.	**0.55889**	0.758522	0.56140	**0.53508**	0.76068	0.44681	**0.44273**	0.80514	0.66657	**0.66131**	0.64008	0.58348	**0.59644**

Table 4
Official leaderboard of PoliticES shared task.

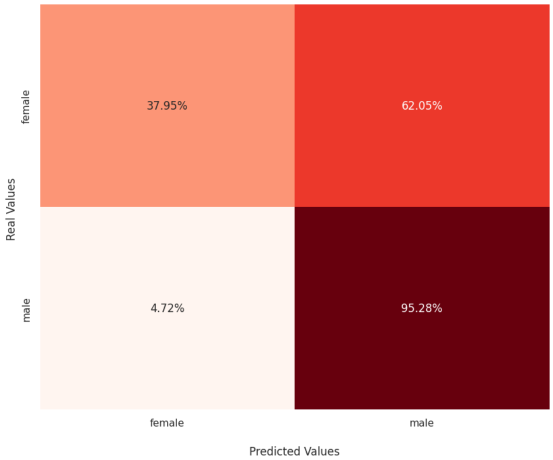
#	Team Name	Avg. F1	F1 Gender	F1 Profession	F1 Binary Ideology	F1 Multiclass Ideology
1	ELiRF VRAIN	0.811319	0.829633	0.827618	0.896715	0.691309
2	HiTZ-IXA	0.793477	0.795627	0.860824	0.877588	0.639871
3	ESCOM-IPN	0.785280	0.769522	0.785898	0.894368	0.691334
4	INGEOTEC	0.777584	0.711549	0.837945	0.891394	0.669448
-	-	-	-	-		
7	**UMUTeam**	**0.692255**	**0.683632**	**0.616362**	**0.860017**	**0.609011**
-	-	-	-	-		
12	miwytt	0.538552	0.611954	0.543934	0.634900	0.363420

In order to perform the error analysis and check what kind of wrong predictions our system makes, a normalized confusion matrix with truth labels have been used, which consists of a table showing the distribution of the predictions of a model with respect to the truth label of the data. The confusion matrix of the system using the mode strategy is shown in Figure 2. It can be noticed that our model tends to confuse the female gender with the male gender, with a percentage of 37.95%. As for the prediction of users' professions, the model usually confuses celebrities with journalists with a percentage of 92.86% because there are few celebrity texts in the training set (see Table 2). Regarding the identification of political ideology, it can be observed that the model has a good accuracy in the identification of left-wing with a percentage of 99.39%. However, it tends to confuse right-wing and left-wing political ideology, with a percentage of 30.45%. In the case of the identification of political ideology at different levels (multiclass classification), it is observed that by having more moderate left cases in the training set, the model has an accuracy of 90% in the identification of this class. However, the model tends to confuse left-wing ideology with moderate left (52.99%) and right-wing with moderate right ideology (58.21%).

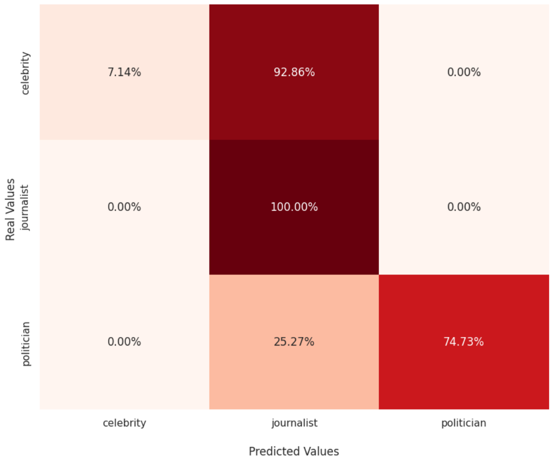
5. Conclusion

These working notes summarize the participation of the UMUTeam in the PoliticES shared task (IberLEF 2023). We achieved a 7/12 on the mean of all F1-score (69.225%) for the demographic and psychographic feature identification models. For this, we used the MarIA fine-tuning approach with the processed dataset and emoji features.

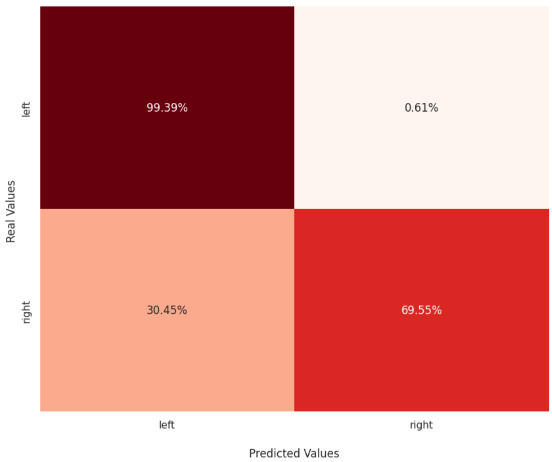
As future work, we are planning to improve our pipeline using an expanded LLMs model with political speech, i.e., fine-tuning a Masked Language Model (MLM) model with political text



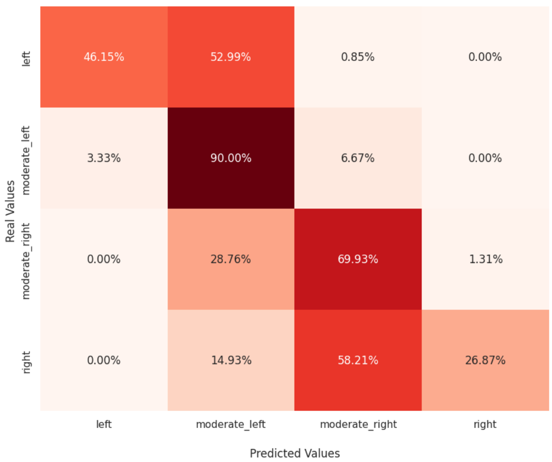
(a) Gender



(b) Profession



(c) Binary ideology



(d) Multiclass ideology

Figure 2: Confusion matrix of the system with mode strategy.

and later fine-tuning this model for detecting political ideology. In addition, we are planning to fine-tune other pre-trained Spanish models to see if they improve MarIA's performance. The source code is available via GitHub.

Acknowledgments

This work is part of the research projects AIInFunds (PDC2021-121112-I00) and LT-SWM (TED2021-131167B-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. This work is also part of the research project LaTe4PSP (PID2019-107652RB-I00/AEI/ 10.13039/501100011033) funded by MCIN/AEI/10.13039/501100011033.

References

[1] M. Fatke, Personality traits and political ideology: A first global assessment, Political Psychology 38 (2017) 881–899. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/pops.12347. doi:https://doi.org/10.1111/pops.12347. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/pops.12347.

[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] 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.

[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 at IberLEF 2023: Political ideology detection in Spanish texts, Procesamiento del Lenguaje Natural 71 (2023).

[5] S. M. Jiménez-Zafra, F. Rangel, M. Montes-y Gómez, Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org, 2023.

[6] 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 68, 265-272 (2022-09).

[7] J. A. G.-D. y Ángela Almela y Gema Alcaraz-Mármol y Rafael Valencia-García, Umucorpusclassifier: Compilation and evaluation of linguistic corpus for natural language processing tasks, Procesamiento del Lenguaje Natural 65 (2020) 139–142. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6292.

[8] A. G. Fandiño, J. A. Estapé, M. Pàmies, J. L. Palao, J. S. Ocampo, C. P. Carrino, C. A. Oller, C. R. Penagos, A. G. Agirre, M. Villegas, Maria: Spanish language models, Procesamiento del Lenguaje Natural 68 (2022). URL: https://upcommons.upc.edu/handle/2117/367156#.YyMTB4X9A-0.mendeley. doi:10.26342/2022-68-3.