URJC-Team at PoliticES-IberLEF2023: Political Ideology Detection Using Hybrid Architecture

Miguel Ángel Rodríguez-García[1,*]

[1]Universidad Rey Juan Carlos, Spain

Abstract

The emergence of social media has promoted the proliferation of information sources, where increasingly users irrigate these environments with their opinions, naturally associated with a particular political ideology. Hence, there is a strong correlation between personal characteristics and social behaviour. Various authors in the literature agree to characterize political ideology as a psychographic trait due to this strong correlation. Consequently, its identification help the research community to have a deeper understanding of humans and their social behaviours. In this work, it is described the system submitted to IberLEF's challenge of detecting political ideology in the Spanish language. The system includes a hybrid approach which combines two traditional strategies with two transformers models to deal with the four classification tasks delivered. As traditional approaches, it was used a Logistic Regression model (LR), and as a more complex architecture, it was employed the pre-trained models RoBERTa and BETO. The system achieves acceptable results, reaching the mid-table of the table positions on the leaderboard.

Keywords

Deep Learning, Transformers, Natural Language Processing, Text Classification, Ideology Detection, Political Ideology

1. Introduction

The proliferation of opinions in social media has grown exponentially in the last few years [1]. Users employ these technologies as a communication channel to virtually share information with thousands of users without regulation or quality control [2]. This situation has made users become direct content creators capable of circulating opinions much faster, motivating the research community to utilise these information sources for analysing human behaviour [3]. In literature, several works propose intelligent systems to infer numerous personal characteristics such as age, gender, or psychological traits through analysing the language used [4, 5, 6]. In this sense, Political ideology is considered a psychological trait whose understanding can reveal personal features to understand better humans and their social behaviours [7].

Generally, politics has a tremendous impact on our society, influencing the decisions taken in our daily lives [8]. Its automatic extraction can be addressed in Artificial Intelligence by different models, from traditional statistical models to complex neural architectures or even more based on Deep Learning [9, 10]. In this work, it is addressed the shared task PoliticES

IberLEF 2023, September 2023, Jaén, Spain

*Corresponding author.

miguel.rodriguez@urjc.es (M. Á. Rodríguez-García)

0000-0001-6244-6532 (M. Á. Rodríguez-García)

located on IberLEF 2023 track [11, 12], which mainly consists of extracting political ideology from two perspectives, binary a multiclass classification problem. Both challenges have been addressed by a hybrid system which combines traditional Machine Learning models and Deep Learning architectures.

The remainder of the manuscript is organised as follows. Section 2 an analysis of the related work. Section 3 details the distribution of the datasets delivered for each task and describes the hybrid method proposed for the challenge. Section 4 presents and discusses the results achieved in the challenge. Finally, Section 5 summarises the findings harvested facing this challenge.

2. Related work

Political ideology detection is a relevant challenge due to the strong correlation with personality traits such as morals, attitudes, and ethical values, among others [7]. In literature, several approaches have been proposed to identify political ideology. For instance, Ansari et al., in [13], use classical machine learning techniques to identify the inclination of political opinions on social media. They harvested 6060 tweets linked to Delhi Elections 2020, which were converted into lowercase and cleaned, removing determined undesired data like numerals, and URLs, among others. Then, each tweet was assigned a label considering its sentiment tone. In experiments, they employed Support Vector Machine (SVM), Random Forest, Logistic Regression and Naïve Bayes, where the SVM outperforms others. Using the same social media, but a more complex architecture, Kabir and Madria in [14] proposed a Deep Learning model based on a pre-trained BERT-base for ideology detection and polarization analysis on COVID-19-related tweets. The steps followed were similar to the previous ones. First, they collected 831 million tweets, which were reduced to 512 million after filtering and pre-processing. Then, three annotators annotated 10K tweets with ten different emotion types. Next for the emotion classification problem, they developed a Deep Neural Network with adversarial sample generation and learning for classifying tweets into these different emotion types. In contrast, a Deep Neural Network architecture based on the BERT-base, was used for political ideology detection. They also conducted hyperparameter tuning to achieve the best performance. Not only Twitter is selected as a primary data source, Tasnim et al., [15] used other social media environments like Facebook, Youtube and Instagram to determine the political ideology in Bengali text by employing Skip-gram and CBOW models. They pre-processed the comments, removed the irreverent data, and used Word2Vec as a frequency analyser to filtered out and only considering embeddings repeated higher than a given threshold. Finally, the models were employed in the experiments, and the results showed that Skip-gram outperformed the CBOW model and two more selected approaches.

After the study, it is worth stressing that the two types of methods analysed, Machine Learning and Deep Learning, both reached good results in the targeted classification problem. Hence, this was the first step which boosted the construction of the hybrid model proposed for the IberLEF challenge.

3. Material and methods

This section details the datasets delivered in the challenge in the practising and evaluating phase and the architecture designed to address the tasks proposed.

3.1. Data

The dataset PoliticEs was compiled from Twitter accounts of politicians, political journalists, and celebrities in Spain from 2020 and 2022 [16]. Several characteristics were utilised to select users' accounts. For instance, in the case of politicians, it was considered their affiliation to political parties; for political journalists, it was considered the inclination of the newspapers where they usually write; for celebrities, it was considered the kind of political parties supported. Once collected, the tweets were pre-processed to discard those that share content and contain mentions of new sites, among others. Then, the tweets were organized into clusters of 80 tweets posted by different users. The resulting dataset was composed of approximately 2800 clusters. Next, the authors labelled by considering the following labels: gender, profession, binary and multiclass political spectrum. Table 1 shows the dataset distribution delivered in the practising and evaluating phase.

		Practice		Evaluation	
		Train	Test	Train	Test
gender	male	119440	43760	9800	30480
	female	60560		4600	13280
Total		180000	43760	14400	43760
profession	politician	60160	43760	4800	14880
	journalist	110800		8640	24400
	celebrity	9040		960	4480
Total		180000	43760	14400	43760
ideology_binary	right	79600	43760	6400	17600
	left	100400		8000	26160
Total		180000	43760	14400	43760
ideology_multiclass	right	21360	43760	1680	5360
	moderate_right	58240		4720	12240
	left	34400		2760	9360
	moderate_left	66000		5240	16800
Total		180000	43760	14400	43760

Table 1
Distribution of the datasets delivered in practice and evaluation phase

Table 1 shows, for each classification task, the number of samples delivered for each label. Thus, a more detailed picture of the dataset distribution is shown. It is worth mentioning that in the practise phase, the provided test datasets did not provide labelled tweets. Hence, the test column has the same value in all tasks. Other features to highlight are the unbalanced amount of samples provided on each task, above all, on the "gender" task, where the label "male" almost duplicated the opposite, and the slight amount of samples provided for the label "celebrity", which represents in practice and evaluation almost a 5% of the number of examples supplied for this task.

3.2. Method

The PoliticEs challenge of the IberLef evaluation campaign delivers as a primary task extracting political ideology on text from two perspectives, binary and multiclass classification problems. Apart from that, demographic traits like gender and profession have to be distinguished too. Hence, in summary, the challenge proposes four classification subtasks. In response, it is proposed a system that combines Machine Learning and Deep Learning techniques to deal with these subtasks. In particular, i) for the gender classification subtask, the transformer model BETO (https://huggingface.co/dccuchile/bert-base-spanish-wwm-cased) was applied; for the profession classification subtask was employed a traditional Machine Learning model, Logistic Regression; iii) for classifying the binary political ideology was utilised the Logistic Regression too; and, finally, for the multiclass version was employed another transformer architecture, RoBERTa (https://huggingface.co/PlanTL-GOB-ES/roberta-large-bne). Figure 1 depicts the system architecture prepared for the subtasks in the challenge.

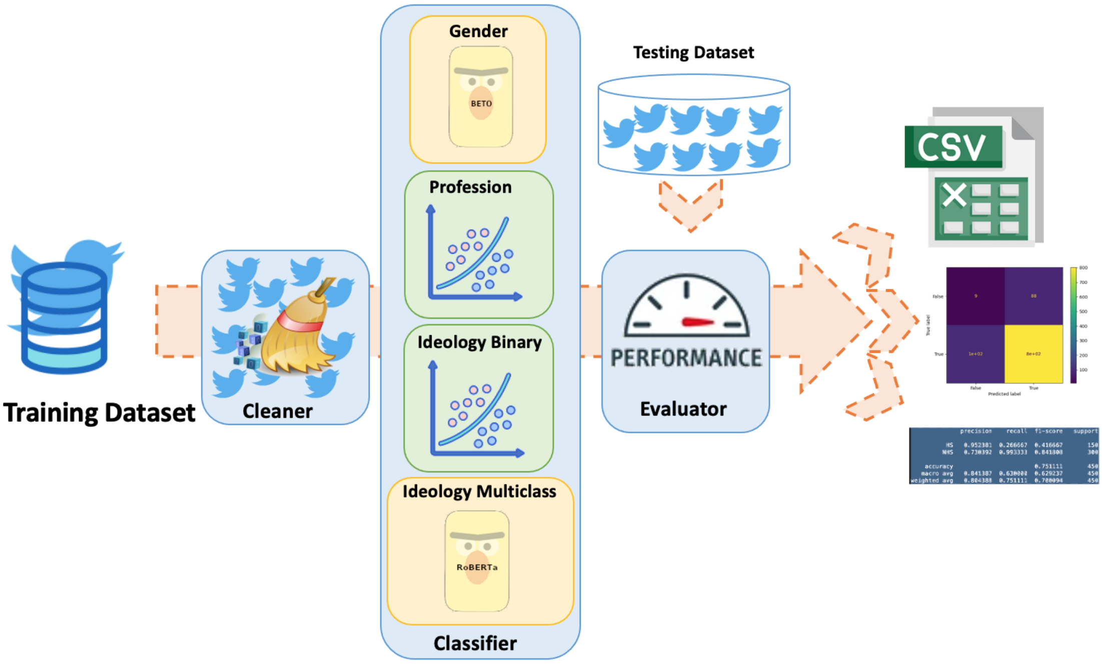


Figure 1: Architecture of the proposed system.

As we can see in Figure 1, the system has been configured as a sequence of three modules: the cleaner is responsible for cleaning the tweet's content, removing metadata and special symbols associated with comments delivered in social networks like links, hashtags, emojis, among others. Next, the classifier module aims at categorizing the tweets by applying four models, one per each different task. Finally the evaluator model is in charge of quantifying the performance of each model. Thus, the system works as follows, when a training dataset is given as input to the system, each tweet is firstly pre-processed by the cleaner module, where the superfluous information is removed. Then, each tweet with its four labels, one per each classification task, is used for training each model in isolation. Finally, the evaluator module utilises the trained models and the testing dataset to assess their performance. As a result, it provides the stats about the achieved performance, a confusion matrix for each task, and a comma-separated file similar to the testing dataset, including the predicted labels.

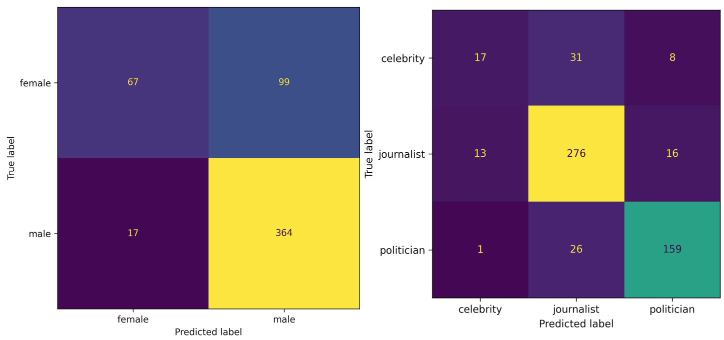
4. Results and Discussion

In this section, it is analysed the performance of the architecture proposed. It has to be underlining that, to facilitate the experiments' reproducibility, hereafter, it is revealed the parameters employed by each model during the evaluation. For the transformer architectures it was established 512, 32 and 2e-5 as embedding lengths, batch size and learning rates parameters, respectively. On the other hand, the Logistic Regression models were configured by setting 100, 100 and One-vs-Rest as a C parameter, max iterations and heuristic method. The standardised metrics selected to evaluate the performance were Precision, Recall and F1-measure. Table 2 concentrates on the results achieved on each challenge's classification task.

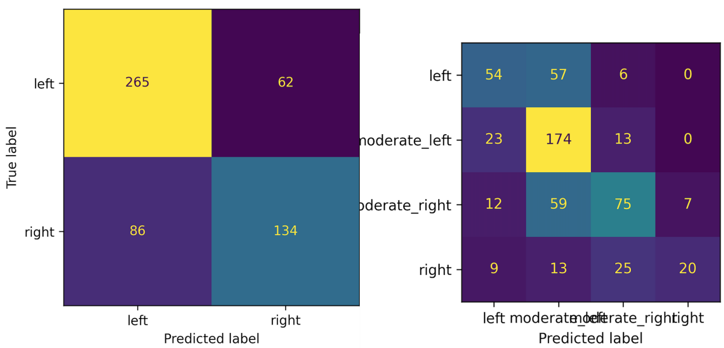
		Evaluation		
Task	Label	Precision	Recall	F1-score
gender	female	0.6	0.66	0.63
	male	0.84	0.81	0.83
	Macro AVG	0.72	0.74	0.73
profession	celebrity	0.55	0.3	0.39
	journalist	0.83	0.9	0.86
	politician	0.87	0.85	0.86
	Macro AVG	0.75	0.68	0.7
ideology_binary	right	0.68	0.61	0.64
	left	0.75	0.81	0.78
	Macro AVG	0.72	0.71	0.71
ideology_multiclass	right	0.5	0.65	0.56
	moderate_right	0.53	0.27	0.36
	left	0.68	0.43	0.53
	moderate_left	0.57	0.66	0.61
	Macro AVG	0.57	0.5	0.52

Table 2
Results obtained on the official test set.

As we can see in Table 2, the highest results were achieved in the profession task, specifically, on classifying tweets as journalist and politician, where the traditional classification model reaches the precision and recall values of 0.83 and 0.9 on journalist and 0.83 and 0.85 on politician classification. Conversely, the worst precision is obtained by the 'ideology_multiclass' task, where RoBERTa is attained 0.5, categorizing tweets as "right". Besides, the worst recall is reached by the Logistic Regression model in the profession classification problem, where it is gained at 0.3. At the overall performance, the best average is obtained by the pre-trained transformer BETO model in gender classification, reaching 0.72, 0.74, and 0.73 on precision, recall and F1-score, respectively. In contrast, the other transformer model employed, RoBERTa, achieves the worst result on the multiclass classification problem, achieving 0.57 on precision, 0.5 on recall, and 0.52 F1-score. To examine deeply the outcomes accumulated, Figure 2 shows the confusion matrix of each experiment conducted.



(a) Gender Task. (b) Profession Task.



(c) Ideology Binary Task. (d) Ideology Multiclass Task.

Figure 2: Confusion matrix computed on each task.

The study of the confusion matrix gives us a more detailed picture of the system's behaviour when a classification task is carried out. Thus, it can be noticed that the biggest mistake made by the system was on the Gender task, where the BETO model was unable to differentiate females from males, reaching 99 samples misclassified. In contrast, if we look at the multiclass problem's outcomes, RoBERTa could precisely distinguish between left and right since it made no mistake. However, its performance dropped with moderate ideologies where, a high number of samples were mistakenly classified. From these outcomes, it could be inferred that the model could not capture discriminating features that precisely represent both types of moderate ideologies against the extreme. On the binary task, although the outcomes were better than the multiclass task, it is also observed significant amounts of samples misclassified, 62 and 86 from predicting left to right and vice-versa. Therefore, it is believed that one of the reasons for these low performances may be due to the distribution of the dataset, which indicates that more samples are required to train the models better.

5. Conclusions

This manuscript describes the proposed system submitted for the shared task PoliticES allocated on the shared evaluation campaign of Natural Language Processing systems, IberLEF 2023. The approach implements a hybrid strategy comprised of traditional Machine Learning techniques and Deep Learning architectures. In particular, each delivered task is completed by a different model. Thus, the profession and ideology binary are addressed by a Machine Learning model, namely, Logistic Regression, and the gender and multiclass ideology tasks are accomplished by transformers models, namely, BETO and RoBERTa. The best performance was achieved by the Logistic Regression on the profession task. In contrast, the BETO transformer model obtained the lowest results on the multiclass classification problem, which exposes an issue that could be on the model configuration, data distribution, or feature selection.

As future work, various lines would be interesting to explore. For instance, it would be interesting to try augmentation techniques to analyse their effect on each model's performance. Besides, to change the models for building a more complex architecture based on Deep Learning. Finally, it would also be significant to try other text feature extractors methods to analyse their impacts on the models.

6. Acknowledgments

This research has been partially supported by grants: PID2021-125709OA-C22, funded by MCIN/AEI/10.13039/501100011033, and "ERDF A way of making Europe"; P2018/TCS-4566, funded by Comunidad de Madrid and European Regional Development Fund, and "Programa para la Recualificación del Sistema Universitario Español 2021-2023".

References

[1] H.-C. Soong, N. B. A. Jalil, R. K. Ayyasamy, R. Akbar, The essential of sentiment analysis and opinion mining in social media: Introduction and survey of the recent approaches and techniques, in: 2019 IEEE 9th symposium on computer applications & industrial electronics (ISCAIE), IEEE, 2019, pp. 272–277.

[2] R. Baly, G. Karadzhov, A. Saleh, J. Glass, P. Nakov, Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media, arXiv preprint arXiv:1904.00542 (2019).

[3] V. Christienne Grace Regodon, Personality detection from text, based on the mbti model (2020).

[4] R. Hirt, N. Kühl, G. Satzger, Cognitive computing for customer profiling: meta classification for gender prediction, Electronic Markets 29 (2019) 93–106.
[5] A. Z. Klein, A. Magge, G. Gonzalez-Hernandez, Reportage: Automatically extracting the exact age of twitter users based on self-reports in tweets, PloS one 17 (2022) e0262087.
[6] Y. Mehta, S. Fatehi, A. Kazameini, C. Stachl, E. Cambria, S. Eetemadi, Bottom-up and top-down: Predicting personality with psycholinguistic and language model features, in: 2020 IEEE International Conference on Data Mining (ICDM), IEEE, 2020, pp. 1184–1189.
[7] B. N. Bakker, Y. Lelkes, A. Malka, Reconsidering the link between self-reported personality traits and political preferences, American Political Science Review 115 (2021) 1482–1498.
[8] B. Massumi, Politics of affect, John Wiley & Sons, 2015.
[9] M. Iyyer, P. Enns, J. Boyd-Graber, P. Resnik, Political ideology detection using recursive neural networks, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014, pp. 1113–1122.
[10] K. M. Alzhrani, Political ideology detection of news articles using deep neural networks, Intelligent Automation & Soft Computing 33 (2022) 483–500.
[11] J. A. García-Díaz, S. María 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).
[12] 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.
[13] M. Z. Ansari, A. F. Siddiqui, M. Anas, Inferring political preferences from twitter, in: Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 3, Springer, 2021, pp. 581–589.
[14] M. Y. Kabir, S. Madria, A deep learning approach for ideology detection and polarization analysis using covid-19 tweets, in: Conceptual Modeling: 41st International Conference, ER 2022, Hyderabad, India, October 17–20, 2022, Proceedings, Springer, 2022, pp. 209–223.
[15] Z. Tasnim, S. Ahmed, A. Rahman, J. F. Sorna, M. Rahman, Political ideology prediction from bengali text using word embedding models, in: 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE, 2021, pp. 724–727.
[16] 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.