

Overview of PoliticEs 2022: Spanish Author Profiling for Political Ideology

Resumen de la tarea PoliticEs 2022: Perfilado del Autor Español por su Ideología Política

José Antonio García-Díaz¹, Salud María Jiménez-Zafra²,
María-Teresa Martín Valdivia², Francisco García-Sánchez¹,
L. Alfonso Ureña-López², Rafael Valencia-García¹

¹Facultad de Informática, Universidad de Murcia, Campus de Espinardo, 30100, Spain

²Computer Science Department, SINAI, CEATIC, Universidad de Jaén, 23071, Spain
{joseantonio.garcia8,frgarcia,valencia}@um.es
{sjzafra,maite,laurena}@ujaen.es

Abstract: This paper presents the PoliticEs 2022 shared task, organized at IberLEF 2022 workshop, within the framework of the 38th International Conference of the Spanish Society for Natural Language Processing. This task aims to extract the political ideology from a given user's set of tweets. Specifically, it focused on the identification of the gender and the profession, as demographic traits, and the political ideology from a binary and multi-class perspective, as a psychographic trait. The PoliticEs task attracted 63 teams that registered through CodaLab. Finally, 20 submitted results and 14 presented working notes describing their systems. Most of the teams proposed transformer-based approaches, although some of them also used traditional machine learning algorithms or even a combination of both approaches.

Keywords: Author profiling, political ideology, author analysis, demographic and psychographic traits.

Resumen: Este artículo presenta la tarea PoliticEs 2022, organizada en el taller IberLEF 2022, en el marco de la 38 edición del Congreso Internacional de la Sociedad Española para el Procesamiento del Lenguaje Natural. Esta tarea tiene como objetivo extraer la ideología política de un usuario a partir de un conjunto de tuits publicados por él. En concreto, se centró en la identificación del género y la profesión, como rasgos demográficos, y la ideología política desde una perspectiva binaria y multiclase, como rasgo psicográfico. La tarea PoliticEs atrajo a 63 equipos que se inscribieron a través de CodaLab. Finalmente, 20 enviaron resultados y 14 presentaron artículos describiendo sus sistemas. La mayoría de los equipos propusieron enfoques basados en transformers, aunque algunos de ellos también utilizaron algoritmos tradicionales de aprendizaje automático o incluso una combinación de ambos enfoques.

Palabras clave: Perfilado de usuarios, ideología política, análisis de autores, rasgos demográficos y psicográficos.

1 Introduction

Political ideology is a psychographic trait that can be used to understand individual and social behaviour, including moral and ethical values as well as inherent attitudes, appraisals, biases, and prejudices (Verhulst, Eaves, and Hatemi, 2012). The relationship between personality traits and political ideology was demonstrated in Fatke (2017). The author gathered data from 21 countries and found a correlation between political ideology and the big five personality traits.

For instance, he found that conscientiousness is strongly correlated with the right wing, whereas openness to experience and agreeability were notably more correlated to the left wing. Moreover, our political ideology has a great influence in our daily lives. For example, Baumgaertner, Carlisle, and Justwan (2018) found a correlation between political ideology and the attitude of citizens to vaccination campaigns of infectious diseases.

The PoliticEs shared task organized at IberLEF 2022 (Montes-y Gómez et al., 2022)

aims to extract political ideology information from texts. For this, an author profiling task is proposed. It is focused on the identification of the gender, the profession, and the political spectrum from a binary and multi-class perspective.

In recent years, several shared tasks have been organized on author analysis under the PAN workshop series (Bevendorff et al., 2021). The novelty of the PoliticEs task is that, to the best of our knowledge, none of these previous tasks have focused on political ideology.

The rest of the paper is organized as follows. Section 2 describes the PoliticEs shared task. Section 3 presents the dataset provided in the competition. Section 4 summarizes the participant approaches. Section 5 shows the results and a discussion thereof. Finally, Section 6 concludes the paper with some insights and future works.

2 Task description

The PoliticEs shared task consists of extracting the gender and the profession as demographic traits, and the political ideology as a psychographic trait from a given user's set of tweets. Political ideology is considered as a binary (pib) and as a multiclass problem (pim). The possible categories of each trait are as follows:

- gender: male, female.
- profession: political, journalist.
- pib: left, right.
- pim: left, moderate left, right, moderate right.

The challenges involved in this shared task are:

1. Extracting political ideology from a text collection. To the best of our knowledge, this is the first Spanish shared task focused on this.
2. Multi-class classification. The author profiling task should be addressed from a binary and multi-class perspective with four different classes.

The competition was organized through CodaLab and is accessible at the following link: <https://codalab.lisn.upsaclay.fr/competitions/1948>. It was divided into

3 phases: Practice, Evaluation and Post-evaluation. In the Practice phase, the participants were provided with a subset of the training data to familiarize with the training data format, and with a notebook with a baseline based on Bag of Words (BoW) to have a starting point for system development. Later, they were provided with the full training set to develop their approaches. For this, they were allowed to make a maximum of 100 submissions in CodaLab. It should be mentioned that in the Evaluation phase, the test partition was provided for the participants to label it using the developed systems. This partition was used to evaluate the teams. They were allowed to make a maximum of 10 submissions through CodaLab, from which each team had to select the best one for ranking. The ranking was determined using the arithmetic mean of the macro f1-score of the gender, profession, binary political ideology, and multi-class political ideology.

3 Dataset

The dataset for this shared task is an extension of the Spanish PoliCorpus 2022 (García-Díaz, Colomo-Palacios, and Valencia-García, 2022), which consists of a set of tweets from the timelines of the Twitter accounts of politicians and journalists in Spain. The politicians are members of the government, congress and senate of Spain along with mayors, presidents of the autonomous communities, former politicians, and collaborators whereas the journalist accounts belong to journalists associated to political press, from Spanish newspapers such as ABC, El País, ElDiario, El Mundo or La Razón among others. The dataset was compiled using the UMUCorpusClassifier tool (García-Díaz et al., 2020).

The users of the dataset are labelled with their gender (male, female), profession (political, journalist), and their political spectrum on a binary axis (left, right) and a multi-class axis (left, moderate left, moderate right, right).

Regarding the tweets collected from each user, we discarded retweets and tweets that contains headlines from news sites. We also removed tweets written in languages other than Spanish. Moreover, we anonymised them by replacing all mentions with the token @user, except for the real users, that were encoded with the token @user and a correlative

number. We did this to hinder the author’s traits identification.

The final dataset is composed of around 400 different users with at least 120 tweets. For the shared task, training and test sets were released (80%-20%). We released the dataset in two splits: training and testing. However, in the first stages of the competition, we released an early birds dataset composed by a subset of 5,000 tweets from the training dataset. It is worth noting that the accounts from training and testing are completely independent in order to prevent automatic classifiers learn to identify the authors rather than the traits. The number of users per set and trait are shown in Table 1.

4 Participant approaches

The PoliticEs shared task attracted 63 teams that registered in CodaLab, of which 20 submitted results and 14 presented working notes describing their systems. The following is a brief summary of the participants’ proposals:

- **(1st) LosCalis (Carrasco and Rosillo, 2022)**. This system is based on transformers (Vaswani et al., 2017). It combines BETO (Cañete et al., 2020) and MarIA (Gutiérrez Fandiño et al., 2022), and employs both architectures for document level characteristics extraction together with a Multi-Layer Perceptron for labels decoding.
- **(2nd) NLP-CIMAT (Villa-Cueva et al., 2022)**. The authors propose PolitiBETO, a pretrained BETO model (Cañete et al., 2020) in the political domain, based on the use of domain adaptation and ensemble learning. They 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.
- **(3rd) Alejandro Mosquera (Mosquera, 2022)**. He explores the use of L2-regularized logistic regression model based on word and character n-grams features along with readability features. This work is notable for the analysis of adversarial attacks on the author profiling challenge.
- **(4th) CIMAT 2021 (Santibáñez-Cortés et al., 2022)**. This team defines different classification models per each trait. Specifically, they use fine-tuned BERT (Kenton and Toutanova, 2019) models for the gender and profession, XGBoost for binary ideology, and Logistic Regression for multiclass ideology.
- **(5th) HalBERT (Holgado and Sinha, 2022)**. The authors evaluate multiple feature sets, and deep learning and machine learning models. They also explore data augmentation and ensemble learning techniques. They find that GloVe embedding features and term-frequency based features, like TF-IDF, can be very helpful and can provide comparative results to deep learning approaches.
- **(7th) I2C (Ramos et al., 2022)**. Their proposal is based on the used of transformers (Vaswani et al., 2017). For gender extraction, they build an ensemble as a set of pre-trained transformers models (RoBERTa (Liu et al., 2019), ALBERTI¹ and BERTIN (De la Rosa et al., 2022)). For the identification of the profession, the tweets of each user are merged to optimize the models. Finally, for the binary and multi-class classification of political ideology, the ROBERTA model was fine-tuned.
- **(8th) TeamMX (Ochoa-Hernández and Alemán, 2022)**. The authors analyze several methods for feature selection and machine learning (Random Forest and SVM) and deep learning (Multi-Layer Perceptron) classifiers. Finally, for determining the gender, they select the best 200 Pearson’s correlation words, using TF-IDF and SVM classifier. For the identification of the profession, they use transition point analysis with lemmas using TF-IDF and Random Forest classifier. For the binary classification of the ideology, they select the set based study using TF-IDF and SVM classifier. For the multi-class classification of the ideology they use an average analysis with lemmas using frequency and Multi-Layer Perceptron classifier.

¹<https://huggingface.co/flax-community/alberti-bert-base-multilingual-cased>

Trait		Training	Test	Total
Gender	<i>Male</i>	177	69	246
	<i>Female</i>	136	36	172
Profession	<i>Politician</i>	251	80	331
	<i>Journalist</i>	61	26	87
Binary ideology	<i>Left</i>	178	57	235
	<i>Right</i>	135	48	183
Multiclass ideology	<i>Moderate left</i>	102	36	138
	<i>Left</i>	76	21	97
	<i>Moderate right</i>	94	31	125
	<i>Right</i>	41	17	58

Table 1: Corpus statistics per trait.

- **(9th) UniRetro (Manea and Dinu, 2022)**. The authors propose two approaches: the first one based on using TF-IDF on SentencePiece pretrained and custom tokens obtained by Named Entity Encapsulation, and the second one consisting of fine-tuning BETO (Cañete et al., 2020) and DistilBETO (Cañete et al., 2022).
- **(13th) UNED (Rodrigo, Fabregat, and Centeno, 2022)**. This team explores two approaches. The first is based on approximate nearest neighbours, which obtains low scores for individual results but a great score when combining several outputs. The second uses some fine-tuned BERT systems, obtaining the best results.
- **(14th) THANGCIC (Ta et al., 2022)**. They present a system based on multilingual BERT (Kenton and Toutanova, 2019), fine-tuned for sentiment analysis, which has been trained with product reviews written in different languages.
- **(15th) URJC-Team (Rodríguez-García, Montalvo Herranz, and Martínez Unanue, 2022)**. This team explores two machine learning algorithms (Logistic Regression and SVM) using a pre-processing module that cleans the tweets and a feature extractor module that combines character and word features with two different settings, with and without stopwords. Finally, they select SVM classifier with stopwords, which provides the best results.
- **(16th) SINAI (Espin-Riofrio, Ortiz-Zambrano, and Montejo-**

Ráez, 2022). The authors propose a voting classifier model that leverages the use of several classical classifiers (Logistic Regression, Random Forest, Decision Trees, Multi-Layer Perceptron, and Gradient Tree Boosting) using as features the combination of stylometry measures with embeddings obtained from MarIA (Gutiérrez Fandiño et al., 2022), a Spanish RoBERTa model for text representation.

- **(19th) UC3MDeep (García-Ochoa Martín-Forero, Massotti López, and Segura-Bedmar, 2022)**. The authors explore several machine learning approaches (K-Nearest Neighbours, Random Forest and Logistic Regression) with different configurations. They obtain the best scores using Logistic Regression without penalty and with a saga solver.
- **INFOTEC-LaBD (Cabrera, Tellez, and Miranda, 2022)**. The proposal of these authors is based on a low-dimensional stacking model approach, which was designed to create both transparent and competitive user profiling models. The results of this team were late in the challenge, due to a confusion with the deadline.

5 Results and discussion

The official leaderboard of the PoliticEs shared task is shown in Table 2. It can be seen the results of the 19 participants that submitted results in time, plus the results of the baseline provided as a notebook, plus the results of the INFOTEC-LaBD team, which submitted a few hours late due to a mistake.

Team	Average Macro-F1	F1-gender	F1-profession	F1-ideology (binary)	F1-ideology (m-class)
LosCalis	0.90226 (01)	0.90287 (01)	0.94433 (01)	0.96162 (01)	0.80023 (04)
NLP-CIMAT	0.89096 (02)	0.78484 (06)	0.92125 (03)	0.96148 (02)	0.89628 (01)
Alejandro Mosquera	0.88918 (03)	0.82671 (03)	0.93345 (02)	0.95152 (03)	0.84504 (03)
CIMAT_2021	0.87976 (04)	0.83683 (02)	0.89500 (05)	0.94167 (04)	0.84553 (02)
HalBERT	0.82532 (05)	0.72602 (13)	0.89776 (04)	0.92176 (05)	0.75574 (06)
Bernardo	0.81996 (06)	0.79178 (04)	0.84982 (08)	0.91315 (06)	0.72511 (08)
I2C	0.79998 (07)	0.74377 (11)	0.86756 (07)	0.86215 (09)	0.72646 (07)
TeamMX	0.79849 (08)	0.78222 (07)	0.82681 (11)	0.82143 (11)	0.76349 (05)
UniRetro	0.78694 (09)	0.73798 (12)	0.88346 (06)	0.90220 (07)	0.62412 (12)
joseluisUS	0.78164 (10)	0.79178 (04)	0.79532 (13)	0.91315 (06)	0.62631 (11)
ErHulio	0.77040 (11)	0.75278 (09)	0.71208 (16)	0.89207 (08)	0.72466 (09)
AzaelCC	0.75180 (12)	0.78050 (08)	0.89500 (05)	0.79935 (14)	0.53234 (18)
UNED	0.74089 (13)	0.74716 (10)	0.83331 (09)	0.81827 (12)	0.56482 (15)
THANGCIC	0.72724 (14)	0.69146 (15)	0.81471 (12)	0.75769 (16)	0.64511 (10)
URJC-Team	0.72192 (15)	0.65987 (16)	0.83298 (10)	0.80811 (13)	0.58672 (13)
SINAI	0.72147 (16)	0.78571 (05)	0.75395 (15)	0.78469 (15)	0.56154 (16)
UC3M-DEEPLNLP-2	0.64315 (17)	0.69388 (14)	0.47324 (17)	0.82917 (10)	0.57629 (14)
probatzen	0.61084 (18)	0.59167 (18)	0.77987 (14)	0.67453 (18)	0.39729 (20)
UC3MDeep	0.58644 (19)	0.64892 (17)	0.40341 (19)	0.74638 (17)	0.54704 (17)
BASELINE	0.51123 (20)	0.57621 (19)	0.43243 (18)	0.59567 (19)	0.44060 (19)
INFOTEC-LaBD	0.72426 (15)	0.71275 (14)	0.61111 (17)	0.95152 (03)	0.72426 (10)

Table 2: PoliticEs official leaderboard (ranking per metric is shown between parenthesis).

The system that obtained the overall highest performance was LosCalis, with an average macro-f1 of 0.90226, combining BETO and MarIA for document level characteristics extraction together with a Multi-Layer Perceptron classifier for labels decoding. It was followed by NLP-CIMAT and Alejandro Mosquera with an average macro-f1 of 0.89096 and 0.88918, respectively. The NLP-CIMAT team proposed PolitiBETO, based on domain adaptation and ensemble learning. Alejandro Mosquera used word and character n-grams features along with readability features with a L2-regularized logistic regression classifier.

Regarding the results per trait, on the one hand, in relation to the demographic traits, gender has been the most difficult for the participants to classify and, on the other hand, with respect to the psychographic trait, political ideology, the multi-class classification has been the most complex.

Concerning the approaches used, most of the teams propose approaches based on transformers (BETO, MarIA, RoBERTa, ALBERTI, BERTIN, DistilBERT, and multilingual BERT), mainly fine-tuning the pre-trained models. Some of them also use traditional machine learning algorithms, being SVM and Logistic Regression the most frequent. There are teams that define differ-

ent models for the identification of each trait, although most use a single model for all of them. Some of them also combine different approaches through ensemble learning and only one team explores data augmentation techniques.

6 Conclusions

This paper presents the first edition of the PoliticEs task at IberLEF 2022. It is an author profiling task for political ideology in Spanish. So far, several tasks on authorship analysis have been organized in the PAN workshop series (Bevendorff et al., 2021), but none of them focuses on political ideology. Political ideology is a psychographic trait that can be used to understand individual and social behavior. Because of its relevance, we intend to promote author profiling research for political ideology in Spanish through the organization of this shared task.

We are very pleased with the impact of the PoliticEs task, as 63 teams registered for it through CodaLab, the platform on which the competition was organized, which is accessible at the following link: <https://codalab.lisn.upsaclay.fr/competitions/1948>. Finally, of all the registered teams, 20 submitted results and 14 presented working notes to describe their systems, which are summarized in this paper.

As expected, approaches based on transformers are the trend solutions presented by participating teams, but some of them also used traditional machine learning systems or even a combination of them. Finally, it should be mentioned that gender and political ideology multi-class have been the traits most difficult to classify for the participants.

As future work, we plan to extend the dataset by including more users who are neither politicians nor journalists. For this, we ask users to voluntarily sent their tweets at the same time they define their political spectrum. Another idea is to include more subtasks concerning author analysis. For example, we are planing to add a subtask related to stance detection, in order to determine which authors are in favor of certain topics and which users are against. We can use this information to define clusters of users and to observe whether there is a relationship between the topics and the political ideology.

Acknowledgements

This work was supported by Project LaTe4PSP (PID2019-107652RB-I00) funded by MCIN/AEI/10.13039/501100011033, Project AllInFunds (PDC2021-121112-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, Project LIVING-LANG (RTI2018-094653-B-C21) funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe, and Big Hug project (P20_00956, PAIDI 2020) and WeLee project (1380939, FEDER Andalucía 2014-2020) funded by the Andalusian Regional Government. In addition, José Antonio García-Díaz has been supported by Banco Santander and University of Murcia through the industrial doctorate programme, and Salud María Jiménez-Zafra has been partially supported by a grant from Fondo Social Europeo and Administración de la Junta de Andalucía (DOC_01073).

References

Baumgaertner, B., J. E. Carlisle, and F. Justwan. 2018. The influence of political ideology and trust on willingness to vaccinate. *PloS one*, 13(1):e0191728.

Bevendorff, J., B. Chulvi, G. L. D. L. Peña Sarracén, M. Kestemont, E. Man-

javacas, I. Markov, M. Mayerl, M. Potthast, F. Rangel, P. Rosso, et al. 2021. Overview of PAN 2021: authorship verification, profiling hate speech spreaders on twitter, and style change detection. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 419–431. Springer.

- Cabrera, H., E. S. Tellez, and S. Miranda. 2022. INFOTEC-LaBD at PoliticES 2022: Low-dimensional Stacking Model for Political Ideology Profiling. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Cañete, J., G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez. 2020. Spanish pre-trained bert model and evaluation data. *Pml4dc at iclr*, 2020:1–10.
- Cañete, J., S. Donoso, F. Bravo-Marquez, A. Carvallo, and V. Araujo. 2022. Albetó and distilbetó: Lightweight spanish language models. *arXiv preprint arXiv:2204.09145*.
- Carrasco, S. S. and R. C. Rosillo. 2022. LosCalis at PoliticEs 2022: Political Author Profiling using BETO and MarIA. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- De la Rosa, J., E. G. Ponferrada, M. Romero, P. Villegas, P. G. de Prado Salas, and M. Grandury. 2022. Bertin: Efficient pre-training of a spanish language model using perplexity sampling. *Procesamiento del Lenguaje Natural*, 68:13–23.
- Espin-Riofrio, C., J. Ortiz-Zambrano, and A. Montejo-Ráez. 2022. SINAI at PoliticEs 2022: Exploring Relative Frequency of Words in Stylometrics for Profile Discovery. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Fatke, M. 2017. Personality traits and political ideology: A first global assessment. *Political Psychology*, 38(5):881–899.
- García-Díaz, J. A., Á. Almela, G. Alcaraz-Mármol, and R. Valencia-García. 2020. UMUCorpusClassifier: Compilation and

- evaluation of linguistic corpus for Natural Language Processing tasks. *Procesamiento del Lenguaje Natural*, 65(0):139–142.
- García-Díaz, J. A., R. Colomo-Palacios, and R. Valencia-García. 2022. Psychographic traits identification based on political ideology: An author analysis study on spanish politicians’ tweets posted in 2020. *Future Generation Computer Systems*, 130:59–74.
- García-Ochoa Martín-Forero, Á., A. Massotti López, and I. Segura-Bedmar. 2022. UC3MDeep at PoliticEs 2022: Exploring Traditional Machine Learning Algorithms for Political Ideology Detection. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Gutiérrez Fandiño, A., J. Armengol Estapé, M. Pàmies, J. Llop Palao, J. Silveira Ocampo, C. Pio Carrino, C. Armentano Oller, C. Rodríguez Penagos, A. Gonzalez Agirre, and M. Villegas. 2022. MarIA: Spanish language models. *Procesamiento del Lenguaje Natural*, 68.
- Holgado, C. G. and A. Sinha. 2022. HalBERT at PoliticEs 2022: Are Machine Learning Algorithms better for Author Profiling? In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Kenton, J. D. M.-W. C. and L. K. Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.
- Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Manea, A.-A. and L. P. Dinu. 2022. UniRetro at PoliticEs@IberLef 2022: Political Ideology Profiling using Language Models. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Montes-y Gómez, M., J. Gonzalo, F. Rangel, M. Casavantes, M. Á. Álvarez-Carmona, G. Bel-Enguix, H. Jair Escalante, L. Freitas, A. Miranda-Escalada, F. Rodríguez-Sánchez, A. Rosá, M. A. Sobrevilla-Cabezudo, M. Taulé, and R. Valencia-García, editors. 2022. *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*.
- Mosquera, A. 2022. Alejandro Mosquera at PoliticEs 2022: Towards Robust Spanish Author Profiling and Lessons Learned from Adversarial Attacks. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Ochoa-Hernández, J. L. and Y. Alemán. 2022. TeamMX at PoliticEs 2022: Analysis of Feature Sets in Spanish Author Profiling for Political Ideology. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Ramos, P. C., J. M. Vázquez, V. P. Álvarez, and J. L. D. Olmedo. 2022. I2C at PoliticEs 2022: Using Transformers to Identify Political Ideology in Spanish Tweets. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Rodrigo, Á., H. Fabregat, and R. Centeno. 2022. UNED at PoliticEs 2022: Testing Approximate Nearest Neighbors and Spanish Language Models for Author Profiling in Political Ideology. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Rodríguez-García, M. Á., S. Montalvo Heranz, and R. Martínez Unanue. 2022. URJC-Team at PoliticEs 2022: Political Ideology Prediction using Linear Classifiers. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. *CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain*.
- Santibáñez-Cortés, E., A. Carrillo-Cabrera, Y. A. Castillo-Castillo,

- D. Moctezuma, and V. Muñiz-Sánchez. 2022. CIMAT_2021 at PoliticEs 2022: Ensemble Based Classification Algorithms for Author Profiling in Spanish Language. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain.
- Ta, H. T., A. B. S. Rahman, L. Najjar, and A. Gelbukh. 2022. THANGCIC at PoliticEs 2022: Term-based BERT for Extracting Political Ideology from Spanish Author Profiling. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain.
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Verhulst, B., L. J. Eaves, and P. K. Hatemi. 2012. Correlation not causation: The relationship between personality traits and political ideologies. *American journal of political science*, 56(1):34–51.
- Villa-Cueva, E., I. González-Franco, F. Sanchez-Vega, and A. P. López-Monroy. 2022. NLP-CIMAT at PoliticEs 2022: PolitiBETO, a Domain-Adapted Transformer for Multi-class Political Author Profiling. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. CEUR Workshop Proceedings, CEUR-WS, A Coruña, Spain.