# I2C at PoliticEs 2022: Using Transformers to Identify Political Ideology in Spanish Tweets

Pablo Cáceres Ramos[1], Jacinto Mata Vázquez[2], Victoria Pachón Álvarez[3] and Juan Luis Domínguez Olmedo[4]

[1,2,3,4] *University of Huelva, Spain*

**Abstract**

This paper introduces our approach and systems' description on *PoliticES: Spanish Author Profiling for Political Ideology* from *IberLEF 2022*. Our proposal consists of the development of different classifiers based on transformers for each of the four author profiles. Depending on to the nature of the characteristics to be classified, several pre-processing and optimization techniques have been used. Finally, the submitted system achieved an average macro F1-score of 0.80 in the evaluation phase, obtaining the seventh position out of the twenty participants.

**Keywords**

Gender detection, Profession detection, Political ideology detection, Spanish, Twitter, Deep learning, RoBERTa

## 1. Introduction

Social media has become a platform where more than 4 billion users spend an average time of 2 hours a day expressing freely their thoughts and opinions freely. Gathering relevant information from each individual navigating the Internet has become a challenge due to the exponential growth of social media users. Politics is no stranger to this challenge. Political parties are trying to find new ways of obtaining data to improve their polls before elections. In order to understand citizens more deeply, many models that automatically identify their political profiles based on the opinions expressed on social networks and digital media are being developed.

This paper depicts our approach and systems description on *PoliticES: Spanish Author Profiling for Political Ideology* from *IberLEF 2022* [1] to detect gender, profession, and political ideology in Spanish social networks.

Given the success and popularity of the transformers [2], all the developed models have been based on this technology. For gender extraction, we decided to train three models and build an ensemble to improve the performance of the classifiers. On the other hand, for the extraction of the profession, the tweets of each user were merged to optimize the models. Finally, for the binary and multiclass classification of political ideology, several models based on transformers were fine-tuned.

## 2. Related works

The great increase in the political campaign budgets of the major parties in the US has led to an increasing number of researchers being attracted to this field of research with different proposals and methodologies.

One of these proposals comes hand in hand with the great increase in social networks within the current population, where a lot of users are very involved in the political life on the network, leaving very valuable information for political parties.

IberLEF 2022, September 2022, A Coruña, Spain
pablo.caceres982@alu.uhu.es (P. Cáceres-Ramos); mata@uhu.es (J. Mata-Vázquez); vpachon@uhu.es (V. Pachón-Álvarez); juan.dominguez@dti.uhu.es (J.L. Domínguez-Olmedo)
0000-0002-4372-5189 (P. Cáceres-Ramos); 0000-0001-5329-9622 (J. Mata-Vázquez); 0000-0003-0697-4044 (V. Pachón-Álvarez); 0000-0001-5083-2313 (JLDO)

The first models that came out tried to classify users within the two largest ideological currents. In [3], authors obtained a high percentage of accuracy using SVM to classify the political alignment of Twitter users.

Looking for systems more in line with today’s world, researchers noticed that it wasn’t possible to classify people directly into two large groups, since each person may have some opinions closer to one group while others could differ and be more aligned with the opposite line of thought. Therefore, new models were proposed using a continuous classification into different values, rather than a single discrete classification [4].

Other methodologies used are not so much to use user’s messages, but to use their interactions with other users with the purpose of getting a direct estimate of the political leaning of the population [5].

## 3. Datasets and tasks

The Corpus provided by the organizers is described in [6,7]. For the practice phase, the Corpus contains two datasets: 5000 tweets (100 users) for training and 1000 tweets (20 users) for test. For the evaluation phase, organizers provided a new training dataset composed of 37560 tweets. Finally, a test dataset composed of 12600 tweets was provided.

**Table 1**
Datasets class distribution

| Feature | Practice train dataset | Practice test dataset | Evaluate train dataset |
|---|---|---|---|
| Male | 2650 | 800 | 19320 |
| Female | 2350 | 200 | 15840 |
| Journalist | 672 | 300 | 6661 |
| Politician | 4328 | 700 | 28499 |
| Left | 2750 | 250 | 20760 |
| Right | 2250 | 750 | 14400 |
| Left | 950 | 50 | 9000 |
| Moderate Left | 1800 | 200 | 11760 |
| Moderate Right | 1550 | 500 | 10080 |
| Right | 700 | 250 | 4320 |

Columns of all datasets are structured as follows: *label*, *gender*, *profession*, *ideology_binary*, *ideology_multiclass* and *tweet*.

The task consists of extracting the gender and profession of a Twitter user, as well as their political ideology. Political ideology is considered a binary problem but also a multi-class problem.

## 4. Methodology and experiments

The methodology used to perform the four types of classifications is based on searching for the optimal combination of algorithms and datasets for each one. For this purpose, different transformer-based models, available in the Huggingface [8] transformers library, were fine-tuned. Each type of classification task has its own particularities, so separate pre-processing and adjustments were made to train each model. For example, for the gender classifier and the binary ideology classifier, we used the training set provided in the practice phase. However, for the profession classifier and the multi-class ideology classifier, the training set provided in the evaluation phase was used. In addition, different pre-processing approaches were carried out. For example, for the profession classifier, each user's tweets were merged. However, we used the individual tweets for the rest of the classifiers.

To obtain the best values for the hyperparameters Learning Rate and Weight Decay, we chose the *roberta-base-bne* model with the gender classifier. Table 2 shows the best values obtained to optimize the F1-score. These values were used to fine-tune the other classifiers. Values between 9.00e-05 and 1.00e-05 were searched for the Learning Rate and for the Weight Decay, we used the range between 0.5 and 0.1.

**Table 2**
Fine-tuned hyperparameters obtained by gender classifier

| Learning Rate | Weight Decay | F1 |
|---|---|---|
| 1.896943597947415e-05 | 0.13286869015692937 | 0.71 |
| 2.4142572422467317e-05 | 0.145887175525535 | 0.70 |
| 2.8183970495481102e-05 | 0.23814114973479916 | **0.72** |

All the models were trained with 16 batch size and 512 token length. Early stopping was used to avoid overfitting while training.

## 4.1. Gender classifier

For this classifier, the methodology used in our work is based on creating an ensemble (Figure 1) as a set of pre-trained transformers models. These methods improve the prediction performance of a single model by combining the predictions of a set of models by means of a vote where each model has equal weight.
The models were the following:

- *roberta-base-bne* [9]. A transformer-based masked language model for the Spanish language based on the RoBERTa base model and pre-trained using the National Library of Spain.
- *roberta-large-bne* [10]. A transformer-based masked language model for the Spanish language based on the RoBERTa large model and pre-trained using the National Library of Spain.
- *roberta-base-biomedical-clinical-es* [11]. This model is a RoBERTa-based model trained on a biomedical-clinical corpus in Spanish collected.
- *alberti-bert-base-multilingual-cased* [12]. ALBERTI is a set of two BERT-based multilingual models for poetry.
- *bertin-roberta-base-spanish* [13]. BERTIN is a series of BERT-based models for Spanish. The current model hub points to the best of all RoBERTa-base models trained from scratch on the Spanish portion of mC4 using Flax.

The three models that achieved the best macro average F1-score in test dataset (Table 3) were selected to create the ensemble based on majority voting.

**Table 3**
Results of each model on gender classification

| Models | Accuracy | F1 |
|---|---|---|
| roberta-base-bne | 0.721 | **0.717** |
| roberta-large-bne | 0.715 | **0.712** |
| roberta-base-biomedical-clinical-es | 0.671 | **0.662** |
| bertin-roberta-base-spanish | 0.623 | 0.600 |
| alberti-bert-base-multilingual-cased | 0.589 | 0.588 |
| bertin-roberta-large-spanish | 0.610 | 0.585 |

The ensemble reached a F1-score of 0.744, while the best individual model achieved a F1-score of 0.717.

Before feeding the texts to the classifiers, we performed a simple preprocessing that consisted of remove emoticons. Nevertheless, we considered hashtags must be kept due to the valuable information related to this task they could hold.

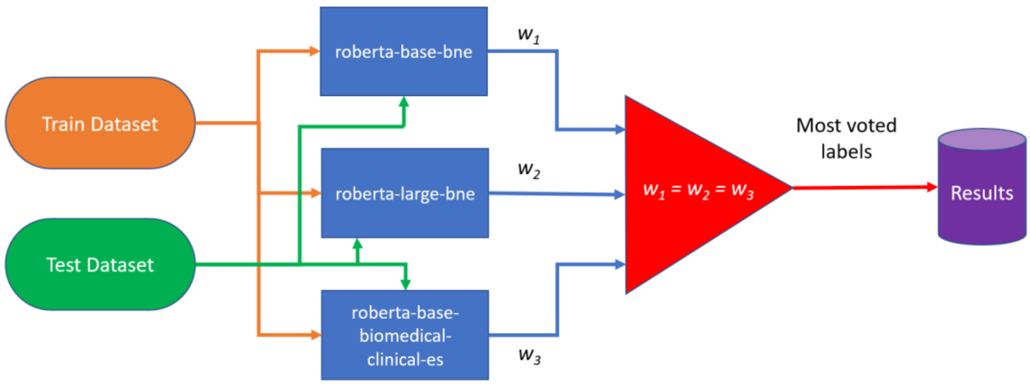


**Figure 1**: Example of ensemble

## 4.2. Profession classifier

For the profession classifier, we merged all the tweets of each user. However, this approach did not perform very well for the other classifiers. By using the *roberta-base-bne* model and the merging of the tweets we improved the results of the classification by 19%, resulting in a final F1-score of 0.87. Due to the length of the tweets once merged, the classifier was trained with the maximum length possible (512 tokens).

In Table 4 we can see the comparative analysis using individual tweets and merged tweets for profession classifier.

**Table 4**
Results of the different methodologies used in the classification of profession

| Methodology | F1 |
|---|---|
| Individual tweets | 0.68 |
| Merged tweets | **0.87** |

## 4.3. Ideology binary classifier

For the political ideology binary classifier, a *roberta-base-bne* model was fine-tuned. The practice training dataset granted better results than the evaluation training dataset when used for this classifier. No preprocessing technique was employed for this model. Finally, the classifier obtained a 0.86 F1-score.

## 4.4. Ideology multiclass classifier

For the multiclass political ideology classification, the model used was the same one that had been previously applied for the binary ideology classification. In this case, we used the provided evaluation training dataset without preprocessing. Best results were obtained from training with 65% of the training dataset (24,414 tweets).

## 5. Results

Table 5 shows the results of our models in the practice and evaluation phases. In the practice phase, a basic setting of the *roberta-base-bne* model was used to train the classifiers. The values achieved in the evaluation phase show the results of applying the methodologies described in section 4.

**Table 5**
F1-score - practice and evaluation systems

| Classifier | Practice Benchmark | Our practice system | Evaluation Benchmark | Our evaluation System |
|---|---|---|---|---|
| Gender | 0.47 | 0.53 | 0.58 | 0.74 |
| Profession | 0.41 | 0.69 | 0.43 | 0.87 |
| Ideology Binary | 0.55 | 0.78 | 0.60 | 0.86 |
| Ideology Multiclass | 0.27 | 0.61 | 0.44 | 0.73 |
| Average | 0.42 | 0.65 | 0.51 | 0.80 |

Our results in the competition among the participants are shown in Table 6.

**Table 6**
Some official results

| Ranking | Team | Average Macro F1 | F1 Gender | F1 Profession | F1 Ideology Binary | F1 Ideology Multiclass |
|---|---|---|---|---|---|---|
| 1 | LosCalis | 0.902262 | 0.902868 | 0.944327 | 0.961623 | 0.800229 |
| 5 | TeamHalBERT | 0.825320 | 0.726020 | 0.897760 | 0.921759 | 0.755739 |
| **7** | **I2C** | **0.799984** | **0.743767** | **0.867561** | **0.862153** | **0.726455** |
| 13 | UNED | 0.740889 | 0.747162 | 0.833309 | 0.818269 | 0.564815 |
| 20 | BASELINE | 0.511228 | 0.576211 | 0.432432 | 0.595665 | 0.440603 |

The confusion matrices of the four classifiers on the test dataset can be found in Figure 2. It can be seen how well the classifier performs when predicting males. Even so, it is not as reliable at predicting females. This may be the result of the large gender imbalance in the training dataset, where male has the highest presence. Looking at the binary ideology, we can see how the classifier tends to predict left more often than right, although this training dataset was balanced. Finally, it is remarkable how the multi-class idelogy classifier fails more with close ideologies, with a tendency towards the left like the binary classifier.

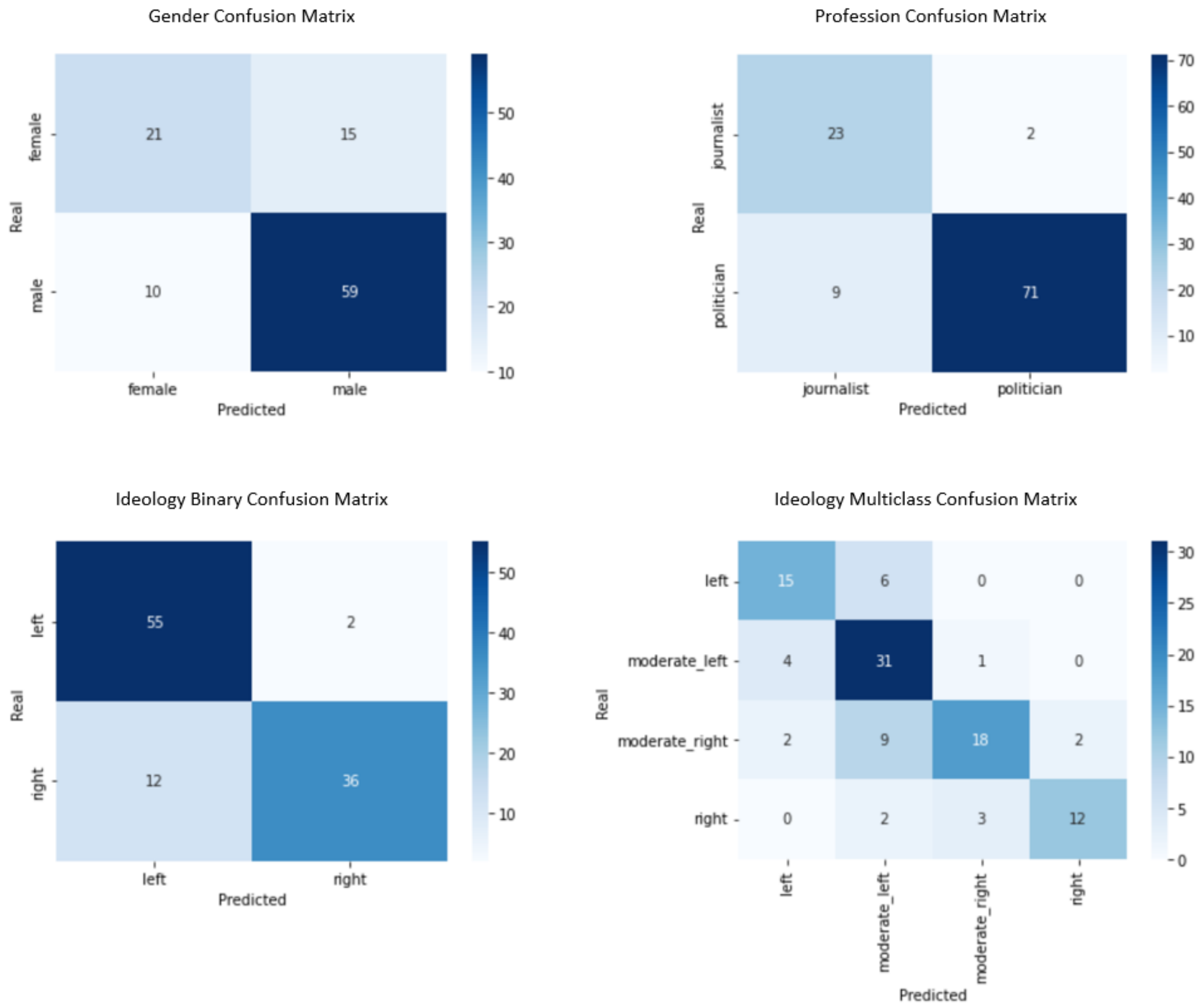


**Figure 2**: Classifiers Confusion Matrices

## 6. Conclusions

In this paper, we presented our proposal for *Spanish Author Profiling for Political Ideology* and the results obtained in the shared task for *IberLEF 2022*. Our approach consisted of fine-tuning transformer-based models. Different approaches have been applied to each classifier in order to achieve the proper results. We proposed an ensemble of models for the gender classifier whereas for the profession classifier, a pre-processing consisting of merging the tweets of each user was carried out. For political ideology classifiers, it has been proposed to use different subsets of tweets to train the models. Our final model achieved a 0.80 macro average F1-score and got the seventh position in the ranking.

In future works we plan to explore other techniques of creating ensemble as well as do more exhaustive hyperparameters search for each classifier and apply this technique to make a real investigation of the predominant political ideologies on Twitter.

## 7. References

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

[2] Gu, Y., Chen, T., Sun, Y., & Wang, B. (2016). Ideology detection for twitter users with heterogeneous types of links. *arXiv preprint arXiv:1612.08207*.
[3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. Advances in neural information processing systems, 30.
[4] M. D. Conover, B. Gon ̧calves, J. Ratkiewicz,A. Flammini, and F. Menczer. Predicting the political alignment of twitter users. In Privacy, Security, Risk and Trust and 2011 IEEE Third Int. Conf. on Social Computing, pages 192–199. IEEE, 201.
[5] F. Wong, C. W. Tan, S. Sen, and M. Chiang. Media, pundits and the us presidential election: Quantifying political leanings from tweets. In Proc. of the Int. Conf. on Weblogs and Social Media, 2013.
[6] García-Díaz, J. A., Almela, Á., Alcaraz-Mármol, G., & Valencia-García, R. (2020). UMUCorpusClassifier: Compilation and evaluation of linguistic corpus for Natural Language Processing tasks. Procesamiento del Lenguaje Natural, 65, 139-142.
[7] García-Díaz, J. A., Colomo-Palacios, R., & Valencia-García, R. (2022). Psychographic traits identification based on political ideology: An author analysis study on Spanish politicians' tweets posted in 2020. Future Generation Computer Systems, 130(1), 59-74.
[8] Hugging Face (2022, March 2). https://huggingface.co/
[9] Gutiérrez-Fandiño, A., Armengol-Estapé, J., Pàmies, M., Llop-Palao, J., Silveira-Ocampo, J., Carrino, C. P., ... & Villegas, M. (2021). Spanish language models. arXiv preprint arXiv:2107.07253.
[10] Gutiérrez Fandiño, A., Armengol Estapé, J., Pàmies, M., Llop Palao, J., Silveira Ocampo, J., Pio Carrino, C., ... & Villegas, M. (2022). Maria: Spanish language models. Procesamiento del Lenguaje Natural, 68.
[11] Carrino, C. P., Armengol-Estapé, J., Gutiérrez-Fandiño, A., Llop-Palao, J., Pàmies, M., Gonzalez-Agirre, A., & Villegas, M. (2021). Biomedical and Clinical Language Models for Spanish: On the Benefits of Domain-Specific Pretraining in a Mid-Resource Scenario. arXiv preprint arXiv:2109.03570.
[12] Hugging Face (2022, March 2). https://huggingface.co/flax-community/alberti-bert-base-multilingual-cased
[13] De la Rosa, J., Ponferrada, E. G., Romero, M., Villegas, P., de Prado Salas, P. G., & Grandury, M. (2022). BERTIN: Efficient Pre-Training of a Spanish Language Model using Perplexity Sampling. Procesamiento del Lenguaje Natural, 68, 13-23.