UMUTeam at PoliticIT-EVALITA2023: Evaluating Transformer Model for Detecting Political Ideology in Italian Texts

Ronghao Pan[1], Ángela Almela[2] and Francisco García-Sánchez[1]

[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 PoliticIT shared task organized at EVALITA 2023. It is an automatic document classification task on clusters of texts, which consists of extracting self-assigned gender as a demographic trait, and ideology as a psychographic trait through a set of texts written in Italian by several authors sharing these traits. For this task, we used the fine-tuning approach of a pre-trained transformer-based masked language model for Italian called *dbmdz/bert-base-italian-cased* to carry out the identification of different features. After several submissions for these tasks, our team ranked sixth out of 7 participants, with an average F1 score of 70.426% of all classification models. However, our binary political ideology classification model obtained the fourth-best result with an F1 score of 86.63%.

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

1. Introduction

Political ideology is considered to be a psychographic trait that can be used to understand individual and social behavior, including moral and ethical values, as well as inherent attitudes, appraisals, biases, and prejudices [1]. The relationship between personality traits and political ideology was demonstrated in [2], in which the authors collected data from 21 countries and concluded that political ideology is related to the big five personality traits and that the results of this relationship vary across countries, especially as a function of the level of prosperity.

Political ideology has a great influence on society and can guide our daily decisions, just like other psychographic traits such as personality. However, these decisions are made both consciously and unconsciously, as the social and cultural environment influences our ideology. In general, however, people tend to be more reluctant to follow advice and directions from politicians who do not coincide with their ideology [3]. In some extreme cases, people may be strongly biased in favor of one political party and, at the same time, strongly disagree with others, which may lead to irrational decisions and put people's lives at risk by ignoring specific recommendations of the authorities.

Thus, the aim of PoliticIT at EVALITA 2023 [4] is to extract information about users' political ideology through text clusters in Italian. For this purpose, the task is divided into several objectives such as extracting self-assigned gender as a demographic trait, and ideology as a psychographic trait through a set of texts written in Italian by several authors sharing these traits. This task builds on a previous task called PoliticES presented at IberLEF 2022 [5] where the dataset was an extension of the PoliCorpus 2020 dataset [3].

Transformer models are a neural network architecture used in various fields of NLP and other machine learning domains. The idea of a Transformer network originated from a paper called *Attention is all you need* [6], developed by researchers at Google to address the language translation problem. Unlike recurrent neural networks (RNNs) and convolutional neural networks (CNNs), which were widely used in NLP before the advent of Transformers, this model relies on attention mechanisms to process sequences more efficiently and capture long-range relationships between words in a sentence or the order of words in a sequence [7]. Pretrained models based on Transformers are language models generated through massive training on large amounts of unlabeled text, aiming to learn language patterns and features from the training data. This enables them to capture semantic, syntactic, and contextual information, developing a deep understanding of language in general. For example, BERT (Bidirectional Encoder Representations from Transformers) is a Transformer-based model that achieved remarkable accuracy and advanced the state of the art in various Natural Language Processing (NLP) tasks [8]. Another advantage of pre-trained models is the ability to add spe-

EVALITA 2023: 8th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, Sep 7 – 8, Parma, IT
ronghao.pan@um.es (R. Pan); angelalm@um.es (Á. Almela); frgarcia@um.es (F. García-Sánchez)
0009-0008-7317-7145 (R. Pan); 0000-0002-1327-8410 (Á. Almela); 0000-0003-2667-5359 (F. García-Sánchez)

cific layers to adapt them to particular tasks. In other words, the process of taking a pre-trained model on vast amounts of data and continuing the training on a smaller and task-specific dataset is known as fine-tuning. The advantage of fine-tuning is that it allows leveraging the general language knowledge acquired during pretraining and applying it to specific tasks with relatively small amounts of training data.

This paper presents the UMUTeam's participation in this shared task, based on the fine-tuning of a pre-trained Italian Transformer model to detect the gender and political ideology of users through their texts written in Italian (document-level classification). 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 *PoliticIT* 2023, organized at EVALITA [9], is a document classification problem that aims to extract self-assigned gender as a demographic trait, and political ideology as a psychographic trait from a given text cluster [4]. The organizers propose political ideology as a binary classification problem and a multi-class classification problem. The organizers created text clusters by mixing some of these extracted tweets to avoid ethical and privacy concerns about author profiling on Twitter. Consequently, all clusters are composed of texts written by different users sharing all the features under evaluation. The users of the dataset are labeled with their gender (male, female), and political spectrum on two axes: binary (left, right) and multi-class (left, moderate left, moderate right, right). Regarding the tweets collected from each user, the organizers discarded retweets and tweets that contain headlines from news sites and removed tweets written in languages other than Italian. Moreover, they anonymized them by replacing all mentions of the politicians and other Twitter accounts mentions with *@user*. Furthermore, other entities, such as political party references, are also replaced with *@political_party* token. 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%). For these different classification problems, we have used the same approach, which consists in fine-tuning a pre-trained Italian model based on Transformer for different demographic and psychographic trait classification tasks at the tweet level. Thus, the training set consists of 103,840 tweets from 1,298 clusters and the test set consists of 453 users with a total of 36,240 tweets. The distribution of demographic and psychographic features of the cluster is shown in Table 1.

Trait		Training	Test	Total
Gender	**Male**	810	318	**1128**
	Female	488	135	**623**
Ideology binary	**Left**	720	248	**968**
	Right	578	205	**783**
Ideology multi-class	**Moderate left**	558	148	**706**
	Left	162	100	**262**
	Moderate right	131	154	**285**
	Right	447	51	**498**

Table 1
The distribution of demographic and psychographic traits of the clusters.

3. Methodology

This task involves identifying the gender (demographic trait) and political ideology (psychographic trait) of users in a given set of texts (document-level classification). The task addresses gender traits as a binary classification problem and political ideology as both a binary and a multi-class classification problem. The approach used to solve the challenge of classifying both demographic and psychographic features at the document level consists in creating a phrase-level classification model by fine-tuning a pre-trained Italian model based on BERT [8] called *dbmdz/bert-base-italian-cased* [10]. The system architecture is depicted in Figure 1 and the pipeline used to participate in this task can be described as follows. First, the dataset has been processed, and the emoji in the text have been transformed into text through the *emoji* library[1]. Second, the training set has been divided into two parts: training and validation. Third, classification models are created for each feature using the approach explained above. 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 level): (1) **mode**, which consists in 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. Next, some more details about the preprocessing and fine-tuning stages are provided.

3.1. Dataset preprocessing

Tweets from the user have the same demographic and psychographic features, so to carry out the task of identifying these features, we created a sentence-level classification model for each feature with all tweets from all users as the training set. The distribution of demographic

[1]https://github.com/kyokomi/emoji

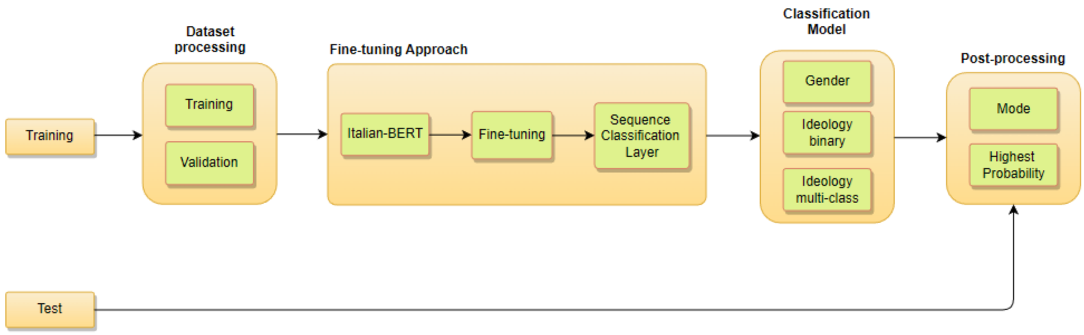


Figure 1: Overall system architecture.

and psychographic features in the dataset is shown in Table 2. In this case, the organizers have only provided the training set, so we have divided the training set into two subsets (80-20%): training and validation. The custom validation split is created using stratified sampling, in order to maintain a balance between labels.

Trait		Training	Validation
Gender	**Male**	51840	12960
	Female	31232	7808
Ideology binary	**Left**	46080	11520
	Right	36992	9248
Ideology multi-class	**Moderate left**	35712	8928
	Left	10368	2592
	Moderate right	8384	2096
	Right	28608	7152

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

3.2. Fine-tuning approach

We utilized the fine-tuning approach of a transformer-based masked language model for Italian called *dbmdz/bert-base-italian-cased*[2] to carry out the identification of different features. Based on the BERT base model, *dbmdz/bert-base-italian-cased* has been pre-trained using a recent Wikipedia dump and various texts from the OPUS corpora collection [11]. The final training corpus has a size of 13GB and 2,050,057,573 tokens. The fine-tuning process consists of adapting and adding a classification layer to the model to perform the training of the full model. In this way, the model takes advantage of the pre-trained linguistic knowledge of *dbmdz/bert-base-italian-cased* and adapts it specifically for a particular classification task, which can significantly improve performance on that task. To do this, we have used the "transformers" library[3] from HuggingFace, which includes a classification layer for pre-trained models. This classification layer is responsible for mapping the contextualized representations of input tokens to the specific output of the corresponding task. For sequence classification, the output layer usually consists of a classification head, which comprises one or more fully connected layers followed by a softmax activation. Finally, the model has been fine-tuned with a training batch size of 16, 6 epochs, a learning rate of 2e-5, and a weight decay of 0.01.

4. Results

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

We submitted a total of 2 runs for this task. The results and a brief description of each are shown in Table 3. The first run is based on running the different classification models on the user texts and then using the "mode" strategy, which consists in selecting the most frequent label obtained in the set of user texts for each feature. This run obtained an average F1 score of 0.70426. The second run has the same structure as the first one but with a different post-processing strategy; in this case, the "high probability" strategy has been used, which consists of making a decision based on the highest probability of all the predictions obtained on the user texts. In this case, it performed worse than with the "mode" strategy, with an average F1 score of 0.49232.

To perform the error analysis and check in which case the models give wrong predictions, a normalized confusion matrix with truth labels has been used, which consists of a table showing the distribution of the predictions of a model with respect to the truth label of the

[2] https://github.com/dbmdz/berts

[3] https://huggingface.co/docs/transformers/index

Strategy	Avg. F1	Gender			Ideology binary			Ideology multi-class		
		M-P	M-R	M-F1	M-P	M-R	M-F1	M-P	M-R	M-F1
Mode	**0.70426**	0.78736	0.69179	**0.71271**	0.87796	0.86244	**0.86638**	0.85372	0.55402	**0.53369**
Highest Prob.	**0.49323**	0.63904	0.62732	**0.63168**	0.68540	0.58355	**0.53658**	0.39979	0.36964	**0.31144**

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

#	Team Name	Avg. F1	F1 Gender	F1 Ideology Binary	F1 Ideology Multi-class
1	Tübingen	0.824057	0.792469	0.928223	0.751477
2	INFOTEC-LaBD	0.800788	0.824287	0.860230	0.717849
3	extremITA	0.771867	0.769002	0.925572	0.621027
4	INGEOTEC	0.762001	0.732259	0.848525	0.705220
5	Teeeech	0.751067	0.752146	0.887720	0.613335
6	**UMUTeam**	**0.704262**	**0.712714**	**0.866382**	**0.533688**
7	NLP_URJC	0.684715	0.661327	0.770840	0.621979

Table 4
Official leaderboard PoliticIT shared task.

data. The confusion matrix of the system using the mode strategy is shown in Figure 2. It can be observed that our model tends to confuse the female gender with the male gender and right-wing with left-wing political ideology, and this may be due to the fact that there are fewer examples in the training set (see Table 2), which makes the model predict the wrong case. In the multi-class classification model of political ideology in our model, it fails mainly in the left-wing prediction, which tends to predict moderate left with a probability of 58%, and the moderate right-wing prediction, which tends to predict moderate left with a probability of 74.51%.

Figure 3 shows the confusion matrices of the system with the highest probability strategy. It can be observed that this strategy tends to confuse right and left ideology in the identification of binary ideology, so in Figure 3b it is shown that there are 78.05% cases that have been mispredicted. As for the identification of multiclass ideology, the model performs better than the mode strategy in identifying left and moderate right with a probability of 40% and 52.94%, respectively. However, it tends to misidentify moderate left and right ideology poorly, with hit probabilities of 9.45% and 45.45% and a difference of 89.86% and 40.91% with respect to the mode strategy (see Figure 2c and 3c).

The official leaderboard for this task is depicted in Table 4. We achieved the sixth position in the ranking with an average F1 score of 0.7042. However, our binary political ideology classification model obtained the fourth-best result with an F1 score of 0.8663.

5. Conclusion

This paper summarizes the participation of UMUTeam in the PoliticIT shared task (EVALITA 2023). We achieved a 6/7 on the mean of all F1 scores (70.42%) for the demographic and psychographic feature identification models. For this, we used the fine-tuning approach of a pre-trained transformer-based masked language model for Italian called *dbmdz/bert-base-italian-cased*[4] to carry out the identification of different features.

As future work, we are planning to improve our pipeline using an expanded transformer-based masked language model with political speech, i.e., extend a Masked Language Model (MLM) model with political text and later fine-tuning this model for detecting political ideology. In addition, we are planning to fine-tune other pre-trained Italian models to see if they improve the *dbmdz/bert-base-italian-cased* performance.

Acknowledgments

This work is part of the research projects LaTe4PoliticES (PID2022-138099OB-I00) funded by MCIN/AEI/10.13039/501100011033 and the European Fund for Regional Development (FEDER)-a way to make Europe and LaTe4PSP (PID2019-107652RB-I00/AEI/ 10.13039/501100011033) funded by MCIN/AEI/10.13039/501100011033. This work is also part of the research projects AIInFunds (PDC2021-

[4]https://github.com/dbmdz/berts

121112-I00) and LT-SWM (TED2021-131167B-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

References

[1] 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. doi:https://doi.org/10.1111/j.1540-5907.2011.00568.x.

[2] M. Fatke, Personality traits and political ideology: A first global assessment, Political Psychology 38 (2017) 881–899. doi:https://doi.org/10.1111/pops.12347.

[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. doi:https://doi.org/10.1016/j.future.2021.12.011.

[4] Russo, Daniel and Jiménez Zafra, Salud María and García Díaz, José Antonio and Caselli, Tommaso and Guerini, Marco and Ureña López, Luis Alfonso, and Valencia García, Rafael, Overview of PoliticIT2023@EVALITA: Political Ideology Detection in Italian Texts, 8th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2023) (2023).

[5] García-Díaz, José Antonio AND Jiménez Zafra, Salud M. AND Martín Valdivia, María Teresa AND García-Sánchez, Francisco AND Ureña López, Luis Alfonso AND Valencia García, Rafael, Overview of politices 2022: Spanish author profiling for political ideology, Procesamiento del Lenguaje Natural 69 (2022a) 265–272.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, CoRR abs/1706.03762 (2017). URL: http://arxiv.org/abs/1706.03762. arXiv:1706.03762.

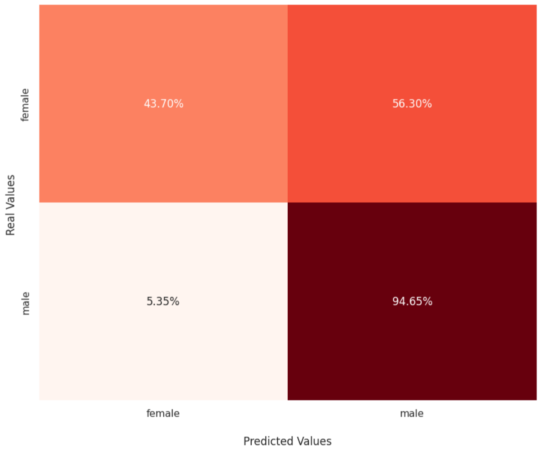
[7] T. Lin, Y. Wang, X. Liu, X. Qiu, A survey of transformers, AI Open 3 (2022) 111–132. URL: https://www.sciencedirect.com/science/article/pii/S2666651022000146. doi:https://doi.org/10.1016/j.aiopen.2022.10.001.

[8] 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.

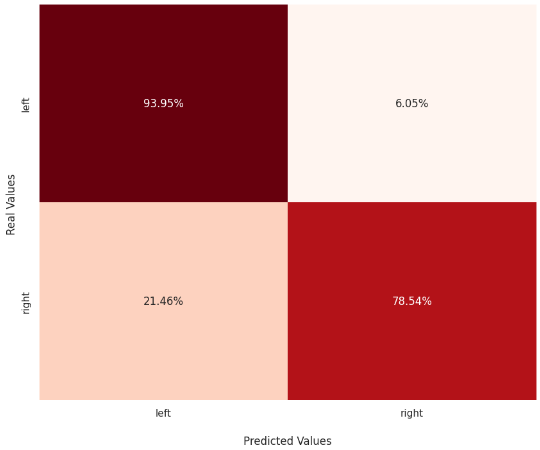
[9] M. Lai, S. Menini, M. Polignano, V. Russo, R. Sprugnoli, G. Venturi, Evalita 2023: Overview of the 8th evaluation campaign of natural language processing and speech tools for italian, in: Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR.org, Parma, Italy, 2023.

[10] S. Schweter, Italian bert and electra models, 2020. URL: https://doi.org/10.5281/zenodo.4263142. doi:10.5281/zenodo.4263142.

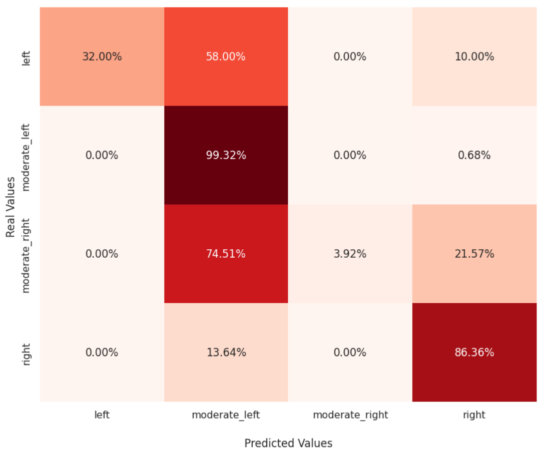
[11] J. Tiedemann, Parallel data, tools and interfaces in OPUS, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 2214–2218. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf.



(a) Gender

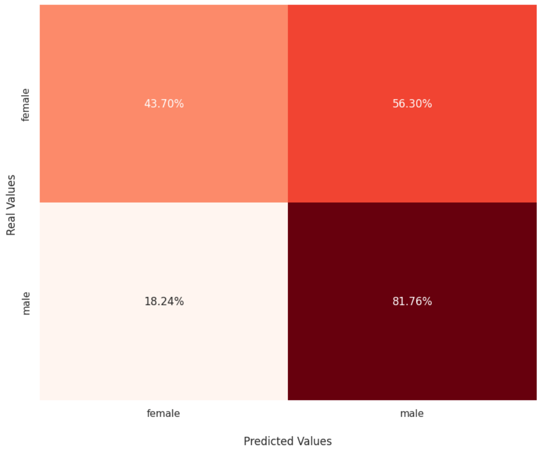


(b) Ideology binary

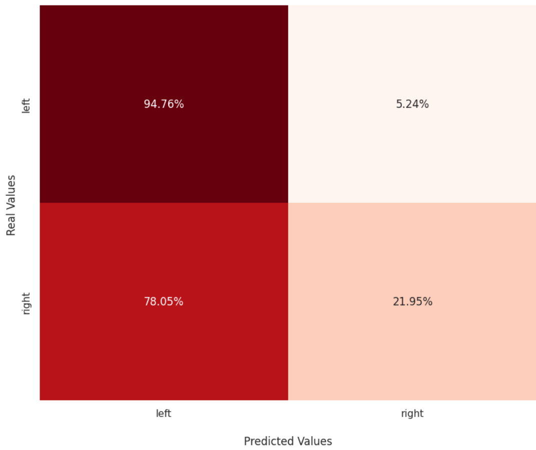


(c) Ideology multi-class

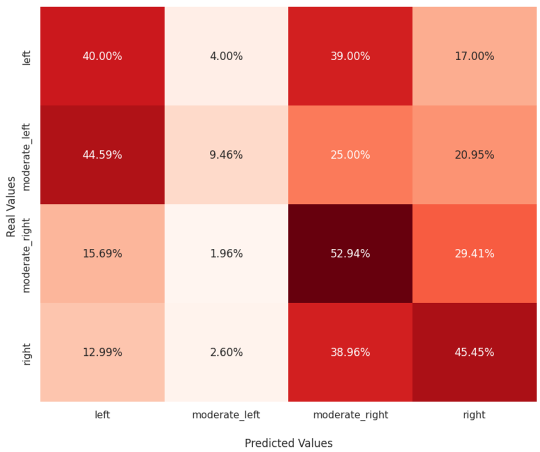
Figure 2: Confusion matrix of the system with mode strategy.



(a) Gender



(b) Ideology binary



(c) Ideology multi-class

Figure 3: Confusion matrix of the system with the highest probability strategy.