URJC-Team at PoliticIT: Political Ideology Detection in Italian Texts Using Transformers Architectures*

Miguel Ángel Rodríguez-García[1,*,†]

[1]Universidad Rey Juan Carlos, Spain

Abstract

Over the years, psychologists have examined human personalities to understand their behaviours. This analysis has demonstrated the existence of several factors that are highly correlated to their conduct that supply clues about thought patterns. Consequently, in recent years the automatic prediction of personality traits has received considerable attention from the community. Thus, EVALITA proposes an evaluation campaign of Natural Language Processing, which suggests the PoliticIT task, which aims to recognise Twitter users' political polarity by analysing their comments. In particular, this task is composed of two subtasks focused on binary and multiclass classification problems. In this work, it is proposed a system which utilises Deep Learning architectures to address these subtasks. Three different pre-trained versions of the BERT transformer model are employed for each task. The outcomes of each generated model reached acceptable scores on the binary classification problems, but its performance dropped slightly on the multiclassification problem. In binary classification, 0.77 and 0.72 were achieved in gender and ideology tasks, respectively, and 0.56 in the multi-class.

Keywords

Natural Language Processing, Transformers architecture, Political Ideology Detection, Deep Learning,

1. Introduction

Personality is defined through people's behaviour, motivation, emotion and features of their thought patterns [1]. Besides, it accompanies us, impacting our daily lives significantly and affecting our decisions, satisfaction with life, well-being, happiness, preferences and desires [2]. The capability to automatically infer personality traits has an incredible amount of practical applications since it helps us understand human nature deeply, enabling us to discover how humans behave and think in determined situations [3].

Social Media has revolutionized the way that people interact, providing an open and direct path for communicating with each other anytime and anywhere and for other varied purposes like enjoyment or simply information access [4, 5, 6]. Hence, with the increasing popularity of this interactive way, Social Media has become a convenient resource for studying deeply human behaviour, considering personality traits and linguistic behaviour [7]. In this context, the target of this work was beyond this analysis, trying to understand the connection between the conduct of linguistic humans and their political ideology polarity.

This article describes the approach submitted to the challenge proposed in the EVALITA campaign for recognizing the political polarity in Italian texts [8, 9, 10]. It relates a system based on Deep Learning architectures like Encoders-Decoders and their families of masked-language models. Specifically, the proposed solution for the challenge is based on Transformers, where three different Bidirectional Encoder Representations from Transformers (BERT) are employed on each subtask. The selection was taken as a consequence of a literature review, where it was studied the highest performance of these models on natural language classification problems.

The rest of the paper is organised as follows. Section 2 a brief overview of the related work. Section 3 details the datasets' content delivered for each subtask and the proposed system's architecture. Section 4 analyses the results achieved on each subtask in the challenge. Finally, Section 5 compiles the findings gained facing this challenge.

2. Related work

Several studies discovered the party ideology in humans is strongly related to their personal traits [11, 12]. Its analysis is relevant since it provides more detailed information about how humans behave and thinks in determined situations [3]. Hence, hereafter, it is provided with a cutting-edge context about the proposed solutions in political ideology recognition. The study begins with the system proposed by Baly et al., in [13], where they employed two models based on two deep learning architectures: Long Short-Term Memory networks (LSTMs) and Bidirectional Encoder Representations from Transformers (BERT) to predict the political polarity of news arti-

EVALITA 2023: 8th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, Sep 7 – 8, Parma, IT

*Corresponding author.

†These authors contributed equally.

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

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

cles. Then, in the experiments, they fine-tuned the hyperparameters of both models, playing with the length of the inputs, sizes of the architecture, learning rates, and batch sizes, among others. As a result, they concluded that the BERT overperformed the LSTM architecture in the experiments accomplished. In the same way, Iyyer et al., in [14] applied a Recursive Neural Network (RNN) model for identifying political polarity signals at the sentence level. For the experiments, they created a new dataset that combines Convote, an initial dataset created, and a filtered version of the Ideological Books Corpus (IBC), a collection of resources whose authors had well-known political leanings. In the results, they created a baseline by selecting Machine Learning techniques like Logistic regression to compare with more complex strategies based on Deep Learning like RNN. The results show how RNN clearly outperform traditional methods. Finally, in [15], Ozturk and Ozcan studied the application of varied Machine Learning Strategies such as Transformers, Long Short Term Memory (LSTM), and Support Vector Machines, among others, on the ideology classification problem. After testing different settings, they closed that a couple of Transformers architectures employed in the task achieved greater precision than others. Therefore, given the notable performance of the Transformers architectures, it was decided to select this choice to deal with the subtasks proposed.

3. Material and methods

This section describes the methods developed to face the tasks proposed and the distribution of the datasets delivered.

3.1. Data

The dataset was composed of tweets harvested from Twitter user accounts, whose political affiliation could be easily deduced since they have a clear link to the party they belong to [16, 17]. The collected tweets were grouped into clusters to avoid ethical and privacy issues and preprocessed to anonymize users' names and remove those tweets, including web content, news, and links. Furthermore, they were labelled, considering gender: male and female, and political spectrum from two perspectives: i) binary: right and left, and ii) multiclass: right, moderate right, left, and moderate left. As a result, the dataset was composed of a set of clusters, which contained 80 tweets each. The delivered dataset was organized into 80%-20% for training and testing in the proposed challenge. Table 1 depicts the distribution of the datasets delivered for practice and evaluation.

The distribution analysis shows the imbalance of the dataset on each task. The highest is on the practice dataset, in which the label `'left'` only has 11 clusters against the label `'right'`, which has 37. For its part, in the dataset delivered for evaluation, the less imbalanced dataset is in `'gender'`, where labels `'male'` and `'female'` differ in 2 clusters. On the contrary, the more noticeable difference is again on `'ideology_multiclass'`, where there is decompensation between the four types of labels included.

3.2. Method

The system proposed for the challenge is composed of a set of three modules: i) the cleaner, which is responsible for removing useless information from tweets; ii) the classifier, which represents the main module in the system since it carries out categorization tasks; iii) the evaluator, which aims at quantifying the performance of the system. The classifier is composed of three different versions of the BERT model [18]. Each one was pre-trained differently by using dissimilar datasets. The version dedicated to the subtask gender was trained by employing text obtained from Wikipedia dump and OPUS corpora collection. [1] The model responsible for identifying the political polarity in the binary classification problem was the XXL version of before, which was trained by employing the OSCAR corpus. [2] Finally, for the multiclassification subtask, it was utilized AlBERTo, a language model trained to understand the users' jargon in social networks [19]. [3] Figure 1 depicts the modular architecture proposed.

The system' architecture has been configured as a pipeline, and it works as follows: two inputs are defined to receive training and testing datasets. When the training dataset is provided, the cleaner module removes the elements in the tweets that do not provide valuable information as links, symbols, and emojis. Then, depending on the faced classification task, a different model is trained in the classifier module. Next, the evaluator employes the test dataset to quantify the precision of the model. As a result, it provides three different outcomes, the resulting CSV file, the main classification metrics and the confusion matrix.

4. Results and Discussion

This section details the outcomes obtained on each subtask delivered in the challenge. The metrics selected to assess the performance of the proposed systems in the challenge are precision, recall, and F1-score. Table 2 scrutinizes the results achieved by the proposed architecture on each classification task.

[1] https://huggingface.co/dbmdz/bert-base-italian-cased
[2] https://huggingface.co/dbmdz/bert-base-italian-xxl-cased
[3] https://huggingface.co/m-polignano-uniba/bert_uncased_L-12_H-768_A-12_italian_alb3rt0

		Practise		Evaluation	
		Train	Test	Train	Test
gender	male	23	23	810	318
	female	25	25	488	135
Total		48	48	1298	453
ideology_binary	right	37	37	578	205
	left	11	11	720	248
Total		48	48	1298	453
ideology_multiclass	right	10	10	131	51
	moderate_right	27	27	447	154
	left	1	1	558	148
	moderate_left	10	10	162	100
Total		48	48	1298	453

Table 1
The distribution of the delivered dataset in clusters.

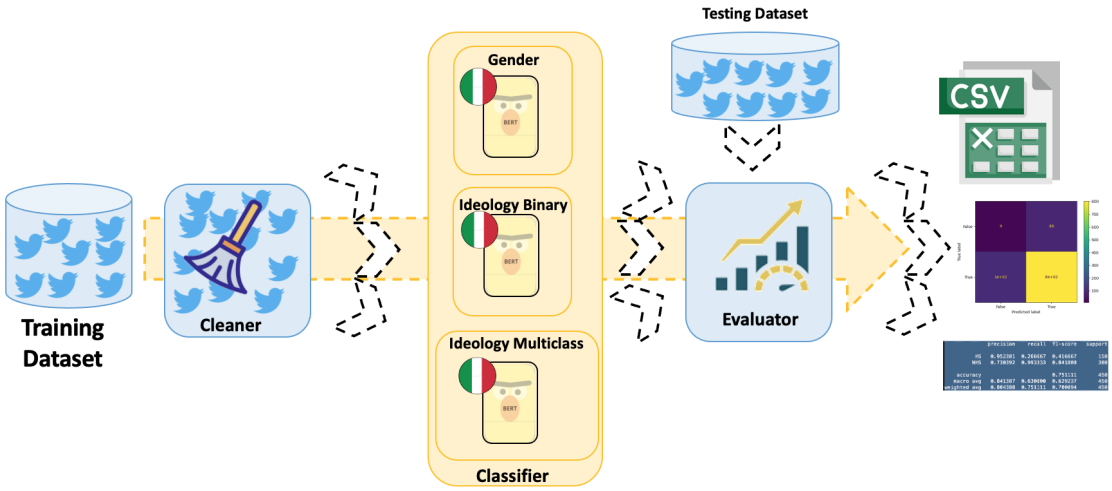
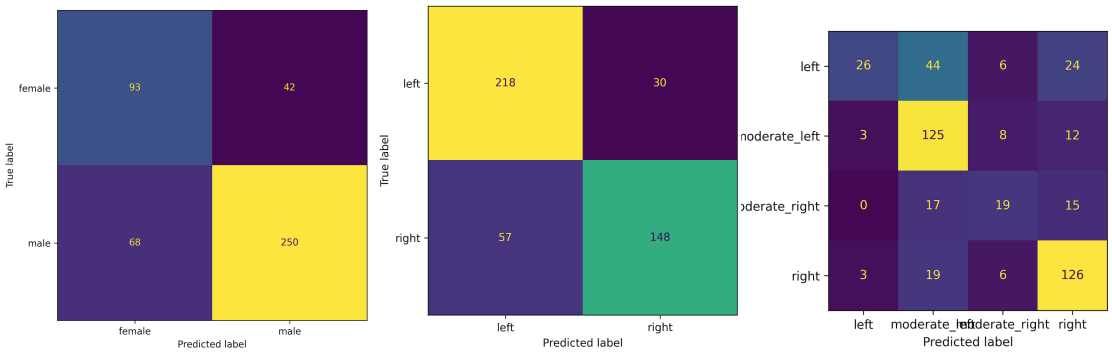


Figure 1: Architecture of the proposed system.

		Evaluation		
Task	Label	Precision	Recall	F1-score
gender	female	0.51	0.63	0.56
	male	0.78	0.72	0.75
	Macro AVG	0.65	0.68	0.66
ideology_binary	right	0.76	0.85	0.8
	left	0.8	0.69	0.74
	Macro AVG	0.78	0.77	0.77
ideology_multiclass	right	0.75	0.87	0.81
	moderate_right	0.54	0.42	0.47
	left	0.85	0.31	0.45
	moderate_left	0.65	0.89	0.75
	Macro AVG	0.7	0.62	0.62

Table 2
Results on the official test set.



(a) Gender Task. (b) Ideology Binary Task. (c) Ideology Multiclass Task.

Figure 2: Confusion matrix computed on each task.

The table was divided into three zones, one per each subtask delivered. The best results were achieved in classifying the political polarity, primarily on tagging tweets as a `'right'` in binary and multiclass classification problems, where both employed models obtained an F1-score of 0.8 and 0.81, respectively. If we look at the dataset distribution in Table 1, both cases did not contain the highest number of samples. Therefore, from this behaviour, it can be deduced that, even when the dataset is non-balanced and the amount of training data is not abundant, both models could recognise the correct label with a high percentage. Moreover, these high results suggest that the samples found in the dataset and used during the training phase contained more discriminate features, improving the models' discrimination capacity. Conversely, the low results obtained on `'ideology_multiclass'`, where the precision dropped to 0.54 and 0.65 in `'moderate_right'` and `'moderate_left'`, respectively, reveals that despite the highest number of training data, the outcomes were not extremely high, indicating the characteristics extracted from tweets were not discriminant enough to teach the model how to recognise these labels during the classification tasks. Finally, to have a deeper understanding of the models' behaviour, Figure 2 shows the confusion matrices of each classification task.

Starting for the gender task, the confusion matrix reveals that the model had great difficulties in characterizing tweets written by males since it misclassified 68 against 42 mistakes made in female classification. For the political polarity identification, in the binary classification problem, the models made more mistakes in recognizing right than left. In the multilabelling task, the errors were in the label `'left'` when the model tried to differentiate this against `'moderate_left'` and `'right'`, making 44 and 24 mistakes, respectively. From this amount of elevated errors, it can be deduced that the text annotated by these labels, `'right'` and `'moderate_left'`, contains similar features to `'left'` since the model can not differ between them. Consequently, a more exhaustive study of the dataset is required to identify precisely the most discriminant features between these labels and reduce the ratio of mistakes made.

At the close of this section, it is compared the results obtained by the architecture proposed against the three first participants and the baseline delivered. Table 3 shows an excerpt of the official leader board, which depicts the F1-score obtained by these proposals.

With regard to the baseline, which combines the Bag of Words (BoW) and the logistic regression, it achieves better results than our proposal on the binary task of ideology, reaching 0.81 against 0.77. However, it is obtained much better results in the other tasks, doubling the performance in the multi-classification task. This imbalance between the two approaches reveals the influence of the choice of features on the behaviour of the techniques since it is compared two ways of extracting features, a set of isolated words versus embeddings, in other words, words versus semantics. The higher results obtained by the baseline show that the selected features are highly discriminative and influence the highest results. The lower results show exactly the opposite. There are similar features assigned to these classification tags that cause the approach to make mistakes, which has a negative impact on the results obtained. On the other hand, the results obtained by the first three proposals in the ranking are very far from the results obtained by the other proposals. Only in the multiclassification task, the proposed architecture can approximate the performance of the third classified opponent, which means there is still work to be done in fine-tuning the architecture and extracting features.

Team Name	F1Gender	F1Ideology Binary	F1Ideology Multiclass
TuebingenPoliticIT	0.79	0.93	0.75
INFOTEC-LaBD	0.82	0.86	0.72
extremITA	0.77	0.92	0.62
...	...	...	...
NLP_URJC	0.66	0.77	0.62
UMU_TEAM	0.53	0.81	0.36

Table 3
The official leader board.

5. Conclusions

This article describes the proposed system for EVALITA, an evaluation campaign focused on Natural Language Processing and Speech challenges for the Italian language in 2023, promoted by the Italian Association of Computational Linguistics (AILC). In particular, this work faced the PoliticIT challenge, which motivates the development of smart systems capable of recognizing the ideological polarity in italian texts. The challenge included three subtasks that address three different classification problems, two binary, where it is demanded to categorize gender and political polarity and one multiclass, where the political polarity was extended from two to four different classes. To deal with these tasks, a system based on the Transformers models. The system includes three versions of the BERT model pre-trained using three datasets. The best performance obtained by the system was on the binary political polarity classification subtask. Not too far are the remaining tasks, in which the system gains relatively similar scores. In spite of the results obtained being quite elevated, there is still a significant range of improvement.

In future work, several lines would be interesting to explore for extending the architecture presented here. Firstly, it would be interesting to use augmentation techniques to study their effects on the Transformers models selected. A more detailed dataset study would be interesting to address, primarily to identify discriminant features that can boost the system's performance.

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] Y. Mehta, N. Majumder, A. Gelbukh, E. Cambria, Recent trends in deep learning based personality detection, Artificial Intelligence Review 53 (2020) 2313–2339.

[2] G. Durand, J. Lobbestael, The relationship between psychopathic personality, well-being, and adaptive traits in undergraduates, Personality and Individual Differences 204 (2023) 112050.

[3] P. Sharma, D. Jain, A. Dhull, Psychology approach: Through ai, ml and dl, Psychology 7 (2021).

[4] G. Appel, L. Grewal, R. Hadi, A. T. Stephen, The future of social media in marketing, Journal of the Academy of Marketing science 48 (2020) 79–95.

[5] F. Monti, F. Frasca, D. Eynard, D. Mannion, M. M. Bronstein, Fake news detection on social media using geometric deep learning, arXiv preprint arXiv:1902.06673 (2019).

[6] L. Xiao, J. Mou, Social media fatigue-technological antecedents and the moderating roles of personality traits: The case of wechat, Computers in Human Behavior 101 (2019) 297–310.

[7] G. Mavis, I. H. Toroslu, P. Karagoz, Personality analysis using classification on turkish tweets, International Journal of Cognitive Informatics and Natural Intelligence (IJCINI) 15 (2021) 1–18.

[8] D. Russo, S. M. Jiménez-Zafra, J. A. García-Díaz, T. Caselli, M. Guerini, L. A. Ureña-López, R. Valencia-García, 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).

[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] D. Russo, S. M. Jiménez Zafra, J. A. García-Díaz, T. Caselli, M. Guerini, L. A. Ureña-López, R. Valencia-García, PoliticIT at EVALITA 2023: Overview of the Political Ideology Detection in Italian Texts Task, in: M. Lai, S. Menini, M. Polignano, V. Russo, R. Sprugnoli, G. Venturi (Eds.), Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop, 2023.

[11] 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.

[12] A. Furnham, M. Fenton-O'Creevy, Personality and political orientation, Personality and Individual Differences 129 (2018) 88–91.

[13] R. Baly, G. D. S. Martino, J. Glass, P. Nakov, We can detect your bias: Predicting the political ideology of news articles, arXiv preprint arXiv:2010.05338 (2020).

[14] 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.

[15] O. Ozturk, A. Özcan, Ideology detection using transformer-based machine learning models (2022) 1–23. doi:10.13140/RG.2.2.12303.51362.

[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.

[17] J. A. García-Díaz, S. M. Jiménez-Zafra, M.-T. M. 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 69 (2022) 265–272.

[18] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. arXiv:1810.04805.

[19] M. Polignano, P. Basile, M. De Gemmis, G. Semeraro, V. Basile, et al., Alberto: Italian bert language understanding model for nlp challenging tasks based on tweets, in: CEUR Workshop Proceedings, volume 2481, CEUR, 2019, pp. 1–6.