

# UMUTeam at Dipromats 2023: Propaganda Detection in Spanish and English Combining Linguistic Features with Contextual Sentence Embeddings

José Antonio García-Díaz<sup>1</sup>, Rafael Valencia-García<sup>1</sup>

<sup>1</sup>Facultad de Informática, Universidad de Murcia, Campus de Espinardo, 30100, Spain

## Abstract

These notes summarise the UMUTeam's contribution to the Dipromats joint task of IberLEF 2023, which deals with the fine-grained detection of propaganda techniques in the political domain, using texts written in English and Spanish. Our contribution is based on the combination of linguistic features and sentence embeddings extracted for several large language models using ensemble learning and knowledge integration. We rank third in the binary classification subtask, first in the multi-classification subtask, and second in the multi-label classification subtask.

## Keywords

Propaganda Identification, Feature Engineering, Transformers, Knowledge Integration, Ensemble Learning, Natural Language Processing

## 1. Introduction

According to [1], propaganda is *an evolving set of techniques and mechanisms that facilitate the spread of ideas and actions*. Therefore, propaganda uses rhetorical figures to enhance replication. The techniques used in propaganda have been analysed in [2]. However, identifying propaganda has not attracted much interest from journalists, fact-checkers or researchers. This lack of attention reduces efforts to combat misleading and manipulative information. Propaganda's deceptive intent can be more elusive and cunning than disinformation, as its content need not necessarily be untrue, and its effects may only become apparent through systematic, long-term observation.

The Dipromats 2023 shared task [3] is a Natural Language Processing challenge focused on identifying propaganda techniques by analysing the language used by official authorities on social networks. To this end, the organisers of the challenge have compiled a dataset of micro-posting messages on Twitter in Spanish and English from diplomatic profiles of China, Russia, the United States and the European Union. Specifically, two subtasks are proposed. The first subtask is a binary classification in which participants have to decide when a text contains propaganda. The second subtask is to categorise the techniques used to spread propaganda. This categorisation is done in two ways: a multi-classification approach and a multi-label approach

---


IberLEF 2023, September 2023, Jaen, Spain

✉ joseantonio.garcia8@um.es (J. A. García-Díaz); valencia@um.es (R. Valencia-García)

🆔 0000-0002-3651-2660 (J. A. García-Díaz); 0000-0003-2457-1791 (R. Valencia-García)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Dataset

According to the organisers of Dipromats 2023, the dataset is a cleaned and filtered version of more than one million tweets in different languages, collected between 1 January 2020 and 11 March 2021. The selected accounts are from governments, embassies or consulates, among others. The dataset is divided into tweets written in Spanish and English. The final dataset consists of 9,501 Spanish tweets and 14,747 English tweets. In addition, the organisers divided the data using temporal criteria to decide on the training and test sets.

For labelling, the organisers of the Dipromats 2023 shared task have used the taxonomy proposed in [2] but including other techniques. They have also grouped the techniques into four main categories: (1) Appeal to Commonality, which includes techniques related to patriotism based on fallacious reasoning and emotions; (2) Discrediting the Opponent, with techniques that show hostility towards the political opponent using fallacies and evoking negative emotions; (3) loaded language, which refers to the usage of hyperbolic language, metaphors and expressions with strong emotional implications; and (4) appeal to authority, which includes appeals to false authority and band wagoning, which refers to the attempt to persuade the audience to take an action because someone else is taking the same action.

Table 1 shows the statistics of the Spanish and English partitions of the Dipromats 2023 task. As can be seen, the dataset is very unbalanced. Furthermore, there are no instances of documents marked as *appeal to false authority* in the English partition and for *bandwagoning* in the Spanish partition.

**Table 1**

Dataset statistics for the Spanish and English partitions of the Dipromats 2023 joint task

Category	Label	Spanish				English			
		train	val	test	total	train	val	test	total
1	ad populum	36	23		59	52	20		72
1	flag waving	179	55		234	423	122		545
2	absurdity appeal	8	11		19	17	13		30
2	demonization	26	15		41	22	22		44
2	doubt	18	9		27	51	25		76
2	fear appeals	40	21		61	34	23		57
2	name calling	51	39		90	152	61		213
2	propaganda slinging	85	39		124	82	32		114
2	scapegoating	2	2		4	7	5		12
2	undiplomatic assertiveness	314	116		430	427	136		563
3	loaded language	292	97		389	704	209		913
4	appeal to false authority	3	3		6	-	-	-	-
4	bandwagoning	-	-	-	-	1	1		2
total		1054	430		1484	1972	669		2641

### 3. Methodology

We focus this problem on language, as there are powerful LLMs available specifically for both Spanish and English. However, we only focus on the third subtask, the fine-grained propaganda characterisation, because it also solves the first task concerning the binary propaganda identification. In this sense, we reduce the number of models we need to train, thus saving time and resources.

In short, our methodology for solving this task is a typical machine learning pipeline that consists of cleaning the dataset, extracting features from the documents, and training and evaluating some neural network models with a custom split before sending our final runs.

During the data cleansing phase, we convert numbers into fixed tokens to give the model some generation. For the same reason, we remove mentions and hyperlinks among other proper elements from social networks. Our last step in the data cleaning stage is to find and expand acronyms and proper language from the message text.

In the feature extraction phase, we extract linguistic features (LFs) from UMUTextStats [4] and sentence embeddings for several LLMs after fine tuning. It is worth mentioning that some of the LLMs are used exclusively on Spanish, English or both, depending on whether they are pre-trained for a specific language or if they are multilingual. The LLMs evaluated are the following:

- **BERT and BETO.** Bidirectional Encoder Representations from Transformers (BERT) [5] is an LLM that uses the Transformer architecture to learn contextual word embeddings that better capture the semantics and relationships between words in a sentence. BETO [6] is the Spanish version of BERT, trained on data from Spanish wikis, subtitles, speeches from the Spanish parliament, among others.
- **ALBERT and ALBETO.** ALBERT is a lighter and more efficient version of BERT [7]. ALBERT uses parameter factorisation and shared pre-training, allowing it to be more efficient in its use of computational resources without significantly compromising the performance of the model. There is also a version of ALBERT trained on Spanish data [8].
- **DistilBERT [9] and DistilBETO [8].** These are versions of BERT and BETO based on distillation, which is another way of constructing lightweight LLMs.
- **RoBERTa and Maria.** The RoBERTa architecture is an improved version of BERT in the pre-training approach and some technical aspects [10]. MarIA [11] is a model based of the RoBERTa architecture, but trained with Spanish data.
- **BERTIN.** This is a Spanish LLM [12] based on the RoBERTa architecture. Unlike MarIA, BERTIN is trained on the Spanish part of the mC4 dataset during an event sponsored by Google Cloud.
- **multilingual BERT.** This is a multilingual version of BERT [5]. It has the same architecture, but is loaded with data from more than 100 languages.
- **XLM.** This is a multilingual LLM [13] that can transfer knowledge from one language to another, allowing models trained in one language to be used for tasks in other languages without the need for additional training.

- **XLM-Twitter**. It is an alternative version of XLM that it is based on RoBERTA, but trained on almost 198 million of tweets written in different languages [14].
- **MDeBERTA**. This is the multilingual version of DeBERTA, an LLM that uses a disentangled attention. This LLM is currently in its third version [15].
- **Legal-BERT**. This is an English LLM model based on BERT but trained for the legal domain [16]. This means that it contains about 12 GB of texts on legislation, court cases or contracts extracted from public sources. It should be noted that this model is lighter than BERT.

For each LLM, we obtain its sentence embeddings, since a fixed representation of the data simplifies the task of combining the LLM with the linguistic features. In order to know the best configuration for each LLM, we train 10 models for each LLM for Spanish and English, evaluating different learning rates, training epochs, batch sizes, warm-up steps and weight decay. This step is carried out using RayTune [17] with Distributed Asynchronous Hyperparameter Optimisation (HyperOptSearch) with the Tree of Parzen Estimators (TPE) algorithm [18] and the ASHA scheduler (because it favours parallelism). The table 2 shows the best configuration found for each LLM. It can be observed that all the models require a larger number of training epochs, between 4 and 5, with a few exceptions (AIBETO in Spanish and multilingual BERT and DeBERTA in English). Even LegalBERT, being the model with texts more related to this shared task, needed 5 epochs to obtain its better result in this experiment.

**Table 2**

Hyperparameter tuning of the LLM before obtaining the sentence embeddings

LLM	learning rate	train epochs	batch size	warmup steps	weight decay
Spanish					
AIBETO	3.3e-05	2	8	0	0.026
BERTIN	1e-05	5	8	0	0.28
BETO	2.4e-05	5	8	250	0.14
DistilBETO	4.7e-05	4	8	0	0.22
MarIA	3e-05	4	8	500	0.25
mBERT	4e-05	5	8	500	0.1
mDeBERTA	4.8e-05	4	16	0	0.02
XLM	2.2e-05	5	8	1000	0.14
XLM-Twitter	4.4e-05	5	16	0	0.23
English					
AIBERT	3e-05	4	8	500	0.038
BERT	2.8e-05	5	8	250	0.18
DistilBERT	3.9e-05	5	8	0	0.3
LegalBERT	4.2e-05	5	8	1000	0.043
mBERT	3.4e-05	3	16	250	0.13
mDeBERTA	3e-05	3	16	1000	0.079
XLM	2.8e-05	4	16	0	0.019
XLM-Twitter	3.5e-05	4	8	0	0.091

We then obtained the contextual sentence embeddings from the classification token, as suggested in [19]. This fixed representation of each document in the corpus allows us to more easily combine these embeddings with each other or with external features.

With these sentence embeddings we train another neural network model, but using Keras and simple neural networks. This process allows a fair comparison with the LFs and the training of a multi-input neural network based on Knowledge Integration (KI), which combines all feature sets at once. In this stage we test different numbers of hidden layers and neurons arranged in different shapes, including the linear function between layers. The learning rate, batch size and dropout mechanism are also evaluated. The results are shown in the table 3. As these features are already fine-tuned in the previous step, we can observe that most of the architectures are simple, being mostly shallow neural networks with one or two hidden layers at most. The most notable difference is the number of neurons, as we obtained 1024 neurons in one layer in Spanish, but only 16 in English using Knowledge Integration.

Please note that the mBERT model is not included in the Spanish experiments due to a human error during our participation. As this model is not part of the KI strategy, we have decided not to include it in the results.

**Table 3**

Best hyperparameters for the deep learning models for Spanish (left) and English (right)

	hidden layers	neurons	dropout	lr	batch size	activation
feature-set	Spanish					
LF	1	48	False	0.001	64	linear
AIBETO	1	128	False	0.001	64	sigmoid
BERTIN	2	37	0.1	0.001	64	tanh
BETO	1	512	False	0.001	32	tanh
DistilBETO	1	256	0.1	0.001	64	relu
MarIA	1	128	False	0.001	64	tanh
mDeBERTA	1	128	0.2	0.01	64	linear
XLM	2	64	0.2	0.01	32	linear
XLM-Twitter	2	64	0.2	0.01	32	linear
KI	1	1024	False	0.001	64	linear
feature-set	English					
LF	2	12	False	0.001	32	linear
AIBERT	2	128	False	0.001	64	linear
BERT	2	512	0.1	0.001	64	sigmoid
DistilBERT	2	512	0.2	0.001	64	tanh
LegalBERT	2	512	0.1	0.001	64	linear
mBERT	2	512	0.1	0.001	64	sigmoid
mDeBERTA	1	128	False	0.001	64	linear
XLM	1	37	0.1	0.001	64	sigmoid
XLM-Twitter	1	37	0.1	0.001	64	sigmoid
KI	1	16	0.2	0.01	32	linear

Finally, we build the ensemble learning models based on combining the outputs of the models

trained with the sentence embeddings for each LLM and the LFs. We use three strategies to combine these outputs. The first is called highest probability, as we choose the maximum probability for each label. The second strategy is based on averaging the probabilities of each model in the ensemble, and the last strategy consists in the mode of each label in the predictions.

## 4. Results and discussion

First, we report our results using our custom validation split. Note that these results focus on the multi-label task. The results are shown in Table 4 for Spanish (left) and English (right). It can be seen that the best results were obtained with individual models rather than with strategies for combining the results. This is the case of BETO for Spanish and ALBERT in English. Usually, the combination of features have reported better performance in other tasks such as Sentiment Analysis [20], hate speech detection [21], satire identification [22] or author profiling [23]. As for the results obtained by the LFs, they are more limited in English. This is to be expected, since UMUTextStats focuses on Spanish although the features based on stylometry and morphosyntax are also suitable for English. The ensemble learning strategies achieve competitive results. The highest probability strategy is usually the model that achieves better recall and good precision for this task.

**Table 4**

Results of the custom validation split of the Dipromats 2023 shared task for Spanish (left) and English (right)

Spanish				English			
feature-set	precision	recall	f1-score	feature-set	precision	recall	f1-score
LF	40.541	20.316	18.525	LF	16.098	11.263	11.811
ALBETO	39.186	16.940	19.064	ALBERT	47.333	38.441	<b>40.806</b>
BERTIL	16.444	18.583	15.295	BERT	51.590	35.224	39.537
BETO	<b>54.319</b>	28.802	<b>34.466</b>	DISTILBERT	50.154	36.250	39.959
DISTILBETO	50.079	25.314	28.810	LEGALBERT	41.092	32.772	35.488
MARIA	53.382	28.373	34.077	MBERT	50.906	26.988	30.711
MDEBERTA	17.495	15.361	13.316	MDEBERTA	41.551	26.597	27.861
XLM	13.037	11.496	9.832	XLM	24.639	17.946	19.836
KI	40.138	26.708	30.219	KI	45.724	38.582	40.470
EL (HIGHEST)	30.561	<b>46.612</b>	31.991	EL (HIGHEST)	32.807	<b>60.277</b>	40.397
EL (MEAN)	48.092	16.006	18.042	EL (MEAN)	54.675	24.432	30.494
EL (MODE)	48.344	16.574	19.056	EL (MODE)	<b>55.516</b>	24.428	30.384

For the Dipromats 2023 shared task, participants are allowed to submit 5 runs. We decided to send three runs based on the ensemble learning strategies, as they give very competitive results, and reserve runs as small internal baselines, one based on the linguistic features (run 04) and another based on BETO and BERT (run 05).

Next, we report the results of the official leader board. The metric used to compare the systems is the ICM-Hard [24]. According to the organisers, there is a baseline based on RoBERTa (MarIA

for Spanish) for Task 1, and for Task 2 they trained the same models, but instead of using a multi-label fashion, they trained all the labels separately (including the negative class). However, we suspect that these baselines will appear in the task overview, as the results obtained with these baselines seem to be simple heuristics based on less frequent labels.

The Table 5 shows the official leaderboard of the Dipromats 2023 Joint Task. We have published only one run per competitor, as we believe this is the fairest leaderboard. In this sense, we are in second place in the binary classification task, with an ICM hard of 0.1165. We obtained this result with our third run, based on ensemble learning with the mode of predictions. The results obtained by all the participants are similar. They have an average macro F1 score of 77.108% and a standard deviation of 1.9 (F1 score results are not shown in this table). For the second task (multi-classification) we get the best result with an ICM-Hard of -0.0037 with our third run and -0.005 with our first run. These results are followed by VRain-ELiRF (ICM-Hard of -0.0117). For the third task, the multi-label classification, the best result is achieved by the VRain-ELiRF team (ICM-Hard of -0.1232), followed by us (ICM-Hard of -0.1318) with our fifth run. It should be noted that our best result was obtained in the fifth run, based on BERT for the English and BETO for the Spanish.

**Table 5**

Official leaderboard for the three subtasks of Dipromats 2023 shared task

Subtask 1		Subtask 2		Subtask 3	
Team	ICM-Hard	Team	ICM-Hard	Team	ICM-Hard
Marco Casavantes	0.1576	UMUTEAM	-0.0037	VRain-ELiRF	-0.1232
Fran Rodrigo	0.1185	VRain-ELiRF	-0.0117	UMUTEAM	-0.1318
UMUteam	0.1043	Fran Rodrigo	-0.0217	Fran Rodrigo	-0.1768
VRain-ELiRF	0.1069	Marco Casavantes	-0.1487	Marco Casavantes	-0.6844
Mario Graff	0.0102	baseline	0.04	baseline	-12.5753
baseline	-2.0089	-	-9.5803	-	-

Next, Table 6 depicts the results obtained for each run on the test set. The first three runs (01, 02 and 03) are based on the ensemble learning strategy of the LLMs and LFs, but use different strategies for combining the models. The first run is based on the average probabilities, the second run is based on the highest probability and the third run is based on the mode. We use the fourth and fifth runs as internal baselines. The fourth run is the linguistic features separately and the fifth run is the features from BERT for the English texts and BETO for the Spanish texts.

As noted when reviewing the rankings, it is noticeable that the fifth run, based on a fine-tuned BERT and a fine-tuned BETO, outperforms the other runs. Looking at the results per run, ensemble learning based on averaging probabilities (run 01) and mode (run 03) achieve similar results. The second run, based on the highest probability, achieves limited results in task 2, especially in the Spanish split. The fourth run, based on linguistic features only, gave the most limited results in all tasks. These results are not surprising, as LFs in isolation are not able to capture the same patterns as state-of-the-art LLMs. However, the fact that combining the features with ensemble learning does not always improve the results draws our attention and further work should be done to perform an error analysis.

**Table 6**

Results per run using the test set. Runs 01, 02 and 03 are based on ensemble learning strategies (average probabilities, highest strategy and mode). The fourth run is a deep learning model with LFs and the fifth run is based on BERT and BETO depending on whether the text is in English or Spanish. The results are presented using the ICM-Hard

run	Subtask 1		Subtask 2		Subtask 3	
	Spanish	English	Spanish	English	Spanish	English
01	0.1316	0.101	-0.0192	-0.0193	-0.1838	-0.1289
02	-0.790	-0.2648	-1.8017	-0.7525	-2.1964	-1.1581
03	0.1323	0.1025	-0.018	-0.0183	-0.1831	-0.1328
04	-0.6874	-0.4029	-1.3937	-0.8811	-1.7334	-1.16915
05	0.096	0.111	-0.0718	0.003	-0.181	-0.1086

## 5. Conclusions

In this paper we have presented our approach to solving the Dipromats 2023 shared task. We focus on propaganda characterisation in a multi-label way, as the models trained for this task can also solve the propaganda identification and propaganda characterisation task using a multi-classification approach. Our approach evaluates linguistic features and sentence embeddings from multiple LLMs, including models specific to English, Spanish and other multilingual models. We achieve competitive results in all tasks and we are very satisfied with the results. We think that this is a very relevant task and we expect to participate in similar tasks in the future, as fighting propaganda and misinformation is a relevant challenge in our daily lives.

In terms of further work, we should compare our results in Task 1 if we had trained a model focused on propaganda identification. In addition, it draws our attention to the fact that the results of models based on BERT and BETO outperform more sophisticated approaches that have been effective in other joint tasks. Accordingly, we will perform a detailed error analysis for each propaganda technique.

## Acknowledgments

This work is part of the research projects AIInFunds (PDC2021-121112-I00) and LT-SWM (TED2021-131167B-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. This work is also part of the research project LaTe4PSP (PID2019-107652RB-I00/AEI/ 10.13039/501100011033) funded by MCIN/AEI/10.13039/501100011033.

## References

- [1] C. Sparkes-Vian, Digital propaganda: The tyranny of ignorance, *Critical sociology* 45 (2019) 393–409.



- [2] G. Da San Martino, S. Yu, A. Barrón-Cedeno, R. Petrov, P. Nakov, Fine-grained analysis of propaganda in news article, in: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), 2019, pp. 5636–5646.
- [3] Pablo Moral, Guillermo Marco, Julio Gonzalo, Jorge Carrillo-de-Albornoz, Iván Gonzalo-Verdugo, Overview of DIPROMATS 2023: automatic detection and characterization of propaganda techniques in messages from diplomats and authorities of world powers, *Procesamiento del Lenguaje Natural* 71 (2023).
- [4] J. A. García-Díaz, P. J. Vivancos-Vicente, A. Almela, R. Valencia-García, Umutextstats: A linguistic feature extraction tool for spanish, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 6035–6044.
- [5] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: <https://aclanthology.org/N19-1423>. doi:10.18653/v1/N19-1423.
- [6] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, *Pml4dc at iclr 2020* (2020) 1–10.
- [7] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, ALBERT: A lite BERT for self-supervised learning of language representations, *CoRR abs/1909.11942* (2019). URL: <http://arxiv.org/abs/1909.11942>. arXiv:1909.11942.
- [8] J. Cañete, S. Donoso, F. Bravo-Marquez, A. Carvallo, V. Araujo, Albeto and distilbeto: Lightweight spanish language models, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 4291–4298.
- [9] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, *arXiv preprint arXiv:1910.01108* (2019).
- [10] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, *CoRR abs/1907.11692* (2019). URL: <http://arxiv.org/abs/1907.11692>. arXiv:1907.11692.
- [11] A. G. Fandiño, J. A. Estapé, M. Pàmies, J. L. Palao, J. S. Ocampo, C. P. Carrino, C. A. Oller, C. R. Penagos, A. G. Agirre, M. Villegas, Maria: Spanish language models, *Procesamiento del Lenguaje Natural* 68 (2022). URL: <https://upcommons.upc.edu/handle/2117/367156#.YyMTB4X9A-0.mendeley>. doi:10.26342/2022-68-3.
- [12] J. de la Rosa, E. G. Ponferrada, M. Romero, P. Villegas, P. González de Prado Salas, M. Grandury, BERTIN: efficient pre-training of a spanish language model using perplexity sampling, *Proces. del Leng. Natural* 68 (2022) 13–23. URL: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6403>.
- [13] A. Conneau, G. Lample, Cross-lingual language model pretraining, in: Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019, pp. 7059–7069.
- [14] F. Barbieri, L. E. Anke, J. Camacho-Collados, Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 258–266.

- [15] P. He, J. Gao, W. Chen, Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing, arXiv preprint arXiv:2111.09543 (2021).
- [16] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, I. Androutsopoulos, Legal-bert: The muppets straight out of law school, in: Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 2898–2904.
- [17] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, I. Stoica, Tune: A research platform for distributed model selection and training, arXiv preprint arXiv:1807.05118 (2018).
- [18] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, Advances in neural information processing systems 24 (2011).
- [19] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019, Association for Computational Linguistics, 2019, pp. 3980–3990. URL: <https://doi.org/10.18653/v1/D19-1410>. doi:10.18653/v1/D19-1410.
- [20] J. A. García-Díaz, F. García-Sánchez, R. Valencia-García, Smart analysis of economics sentiment in spanish based on linguistic features and transformers, IEEE Access 11 (2023) 14211–14224.
- [21] J. A. García-Díaz, S. M. Jiménez-Zafra, M. A. García-Cumbreras, R. Valencia-García, Evaluating feature combination strategies for hate-speech detection in spanish using linguistic features and transformers, Complex & Intelligent Systems (2022) 1–22.
- [22] J. A. García-Díaz, R. Valencia-García, Compilation and evaluation of the spanish satcorp2021 for satire identification using linguistic features and transformers, Complex & Intelligent Systems 8 (2022) 1723–1736.
- [23] 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.
- [24] E. Amigo, A. Delgado, Evaluating extreme hierarchical multi-label classification, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 5809–5819. URL: <https://aclanthology.org/2022.acl-long.399>. doi:10.18653/v1/2022.acl-long.399.