A Spanish dataset for Targeted Sentiment Analysis of political headlines

Juan Manuel Pérez[1] [a], Emilio Recart Zapata[2,3] [a], Tomás Alves Salgueiro[2], Damián Furman[1], and Pablo Nicolás Fernández Larrosa[2]

[1] Instituto de Ciencias de la Computación, CONICET, UBA
{jmperez, dfurman}@dc.uba.ar
[2] Instituto de Fisiología, Biología Molecular y Neurociencia, CONICET, UBA,
{talvessalgueiro, fernandezlarrosa}@fbmc.fcen.uba.ar
[3] Facultad de Psicología, UBA
recartemilio@psi.uba.ar

Abstract. Subjective texts have been extensively studied due to their potential to influence behaviors. While most research has focused on user-generated texts in social networks, other types of texts, such as news headlines expressing opinions on certain topics, can also influence judgment criteria during political decisions. In this paper, we address the task of Targeted Sentiment Analysis for news headlines related to the 2019 Argentinean Presidential Elections, published by major news outlets. To facilitate research in this area, we present a polarity dataset comprising 1,976 headlines that mention candidates at the target level. Our experiments using state-of-the-art classification algorithms based on pre-trained language models demonstrate the usefulness of target information for this task. We also provide public access to our data and models to foster further research.

1 Introduction

Extracting opinions from subjective texts has attracted a lot of interest since the eclosion of Internet and Social Networks, given the unprecedented availability of opinion-rich resources [24]. Most works for opinion mining are directed towards user-generated texts from social networks; however, some other texts –such as news headlines– also convey subjective content about certain topics or entities.

Particularly, it is of interest to analyze the role of the media and campaigns in the formation of judgment criteria during political decisions [16]. The rise of social networks marked a great dynamism in the management and transfer of large volumes of data, playing a fundamental role in the transmission of information in political and candidate campaigns at a massive level [29].

Numerous studies analyzed the role of social media, print media, campaigns or fake news in the formation of judgement criteria and political decisions [35, 5, 6, 33, 29], by affecting subjective variables (such as trust) [1, 14, 18, 26, 17, 27,

[a] These authors contributed equally to this work

Received January 2023 Accepted April 2023; Published May 2023

29, 15, 12]. The current work is part of a research project to evaluate the cognitive processes underlying the presidential elections, and assessing the impact of greater exposure to positive content in news headlines associated with the candidates. Our previous work has shown how this type of information can, either by mere repetition or by association with strong emotional content, modulate the decision-making process [2] For this purpose, we are interested in analyzing headlines mentioning candidates by the main national written media.

In this paper, we present an approach to the task of **Targeted Sentiment Analysis** for the domain of newspaper headlines. To the best of our knowledge, no Spanish dataset is available for this task. To bridge this gap, we present a novel dataset of headlines mentioning candidates in the 2019 elections in Argentina, having annotations at target level instead of assigning a single polarity to the whole sentence. Preliminary experiments with state-of-the-art techniques suggest that classifiers that consume both the headline and the target improve their performance for the task over those that only consume the headline, giving indications that both sources of information are useful. We make our dataset and the pre-trained models available for further research.

Our contributions are the following:

- We present a novel dataset of Spanish news headlines for the task of targeted sentiment analysis, with annotations at target level.
- We show that state-of-the-art classifiers can leverage the target information to improve their performance.
- We make our code, dataset and pre-trained models available for further research.

The rest of the paper is organized as follows. Section 2 reviews previous work on sentiment analysis and targeted sentiment analysis. Section 3 describes briefly the context which we are studying, namely the 2019 Argentina presidential elections. Section 4 describes the dataset and the annotation process. Section 5 presents the experimental setup for the classification experiments and Section 6 its results. Finally, Section 8 concludes the paper and discusses future work.

2 Previous work

Sentiment analysis and opinion mining are widely used applications of natural language processing (NLP), and have been the focus of several workshops, including SemEval and overviews [32, 9]. In general, sentiment analysis involves predicting the polarity of a text, which can be either positive, negative, or neutral, or a Likert-scale rating ranging from negative to positive [34]. Aspect-based sentiment analysis (ABSA) is a variant of this task, where the polarity prediction is performed on a specific aspect of the text, allowing for the extraction of multiple opinions [25]. Similarly, targeted sentiment analysis focuses on analyzing the polarity for a given entity mentioned in the text [22]. However, despite the availability of many resources for the Spanish language [21, 9], there are currently no datasets available for this language in this particular subtask.

Pre-trained language models based on transformers [37] and transfer learning are state-of-the-art for most NLP tasks, and BERT [7] and its variants [19] are among the most popular classification techniques today, showing top performances on language-understanding benchmarks such as GLUE [38]. There are several pre-trained models available for the Spanish language, such as BETO [4], RoBERTa$_{es}$ [11], and others [30, 28]. [23] provide a good overview of pre-trained models across various domains and languages.

3 2019 Argentina Presidential Elections

The Argentine Presidential Elections consisted of a two-stage election. The first step, called PASO (which took place in 11th August 2019), aims to filter the main presidential formulas that will run in the General Election (27th October 2019). Only candidates who obtain more than 1.5% of the vote in the PASO advance to the General Election. The candidates who advanced to the 2019 General Election were: M. Macri (which we note MM), A. Fernández [AF], R. Lavagna [RL], N. Del Caño [NDC], J. Gómez Centurión [JGC] and J. L. Espert [JLE]. MM was the official candidate (moderate right), being at the time of the election president of Argentina elected in 2015. AF was the candidate of the majority opposition, an alliance of mostly Peronist sectors (although it also includes non-Peronist sectors), which governed Argentina during 2003-2015. NDC represents the minority left-wing opposition, while JGC and JLE represent the minority radical right-wing opposition. RL represents a moderate right-wing alternative.

4 Data

To generate our dataset, we first collected a total of 22510 newspaper articles published between 21 September and 27 October 2019 (during the General Election period) from the main national news outlets: Ambito, Clarin, El Cronista, INFOBAE, La Nacion, Pagina 12, Perfil, Popular, La Izquierda Diario, Prensa Obrera, Tiempo Argentino. We used the *newspaper3k* [4] python library to collect the articles.

For the purposes of the present work, we selected a sample from the headlines mentioning at least one of the contending parties or candidates in the national election. Table 1 list the entities, keywords and phrases searched for in the headlines. Mentions of other candidates of the same political force (e.g. state governor) were not included in this analysis.

Three annotators were hired to label the headlines. For each pair of headline and target (a political party or candidate), each annotator was required to assigned a polarity to the pair. Annotation guidelines were provided, consisting of instructions and examples of how to perform the task, and also interviews were held with the participants to consolidate the labeling criteria. To diminish

[4] https://github.com/codelucas/newspaper

Table 1. Keywords for Candidate/Presidential Formulas Analysis

Candidate Target	President	Vice-president	Political Force	Other References
AF	Alberto Fernandez; Alberto	Cristina Fernandez de Kirchner; Cristina; Cristina Fernandez; Cristina Kirchner; CFK	Frente de Todos	Kirchnerismo
MM	Mauricio Macri; Macri; Mauricio	Miguel Angel Pichetto; Pichetto	Juntos por el Cambio	Macrismo, PRO
RL	Roberto Lavagna; Lavagna	Juan Manuel Urtubey; Urtubey	Consenso Federal	–
NDC	Nicolas Del Caño; Del Caño	Romina Del Pla; Del Pla	Frente de Izquierda y de los Trabajadores-Unidad; Frente de Izquierda; FIT	–
JGC	Jorge Gomez Centurion; Centurion	Cynthia Hotton; Hotton	Frente NOS; NOS	–
JLE	Jose Luis Espert; Espert	Luis Rosales	Unite por la Libertad y la Dignidad; UNITE	–

political biases as much as possible, we masked the targets to the annotators (e.g. Mauricio Macri was shown as `[TARGET]`).

Each instance was annotated by the three workers as the task is subjective. The agreement was measured using Krippendorff's Alpha [13] and turned out to be $\alpha = 0.62$ (moderate/substantial agreement). A majority voting scheme was used to aggregate the labels, discarding those targets for which the three raters assigned different polarities.

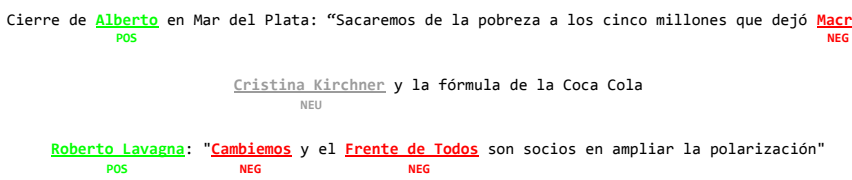
The resulting dataset consists of 1,976 headlines and 2,439 targets. 1,567 headlines have exactly one target, and the remaining have two or more. Among these, 165 headlines feature mixed polarities, mostly in the negative/positive form. Table 2 summarizes the dataset statistics.

Figure 1 illustrates some examples of the dataset. More information about co-ocurrences and mentions of candidates can be found in Appendix A.

Table 2. Dataset statistics

Num. headlines	Different targets	Total Targets	Positive	Negative	Neutral
1976	70	2439	962	710	767

Fig. 1. Examples of the dataset



5 Classification Experiments

To determine whether the classification algorithms can benefit from target information in the dataset, we conducted a classification experiment using both target-aware and target-unaware models. The target-unaware model only considers the headline and predicts an overall polarity, while the target-aware model takes both the headline and the target into account, predicting a polarity for the pair.

In the target-unaware model, the headline is presented as a single sentence, whereas in the target-aware model, the headline and target are presented as a pair of sentences. Figure 2 depicts the classification pipeline for both type of inputs.

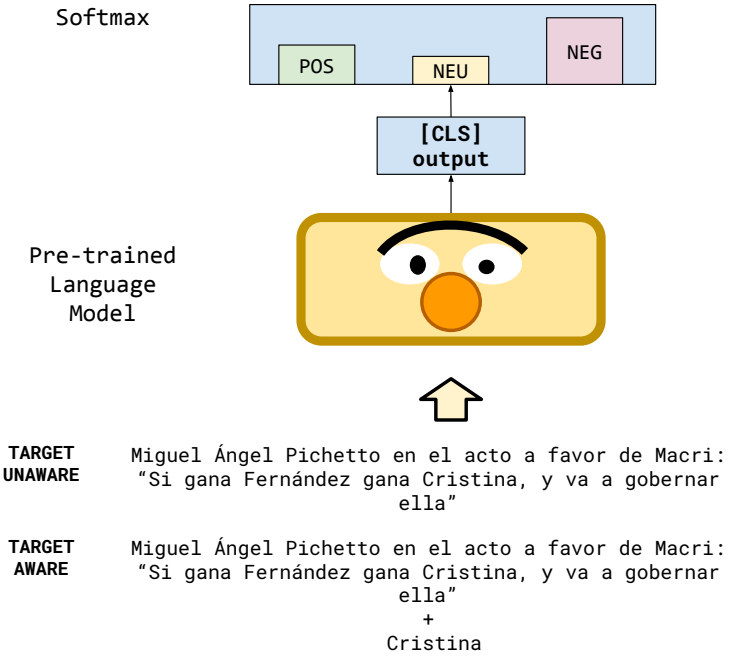
5.1 Training

In our study, we explored a range of different pretrained language models to evaluate their performance on the targeted sentiment analysis task in news headlines. Specifically, we compared the effectiveness of four different models, including BETO, RoBERTa$_es$, BERTin, ELECTRICIDAD, and RoBERTuito.

BETO [4] is a pretrained language model based on BERT pretraining guidelines, but also incorporates some elements of RoBERTa style training. Meanwhile, RoBERTa$_es$ is a RoBERTa model pre-trained on a corpus collected by the National Library of Spain, as part of a national initiative to develop language technologies [11]. BERTin is another RoBERTa model that utilizes a smart sampling strategy to optimize the pretraining process [31]. ELECTRICIDAD, on the other hand, is an ELECTRA model specifically designed for the Spanish language. Finally, we also evaluated the performance of RoBERTuito [28], a RoBERTa model trained on a corpus of 600 million tweets in Spanish. While this model is specially crafted for social media text, it has been shown to be effective in various Spanish NLP tasks, including sentiment analysis.

Given that our dataset was relatively small, we also fine-tuned models trained on the TASS 2020 dataset for polarity detection, which were provided by the *pysentimiento* library. We hypothesized that models trained on a similar task and domain would be useful for our targeted sentiment analysis task in news headlines.

Fig. 2. Classification models for the task. Target-unaware classifier consumes only the headline, while the target-aware version consumes both headline and target



For each pretrained model, we fine-tuned two different classifiers: a target-unaware classifier and a target-aware classifier. The only difference between the two versions is the input: in the target-unaware classifier the input consists of the headline alone, while in the target-aware version it is made of the headline and the target with the special token `[SEP]` in-between them. While there are many options to construct a target-aware input [36], we opted for the simplest one.

The hyperparameters used for training our models were carefully chosen through a hyperparameter search, as recommended by **(author?)** [7]. Specifically, we performed a search for the optimal values of the learning rate, warmup ratio, and number of epochs hyperparameters while keeping the batch size fixed at 32 and the weight decay at 0.1. Additionally, we set the β_1 and β_2 parameters to 0.9 and 0.999, respectively. For optimization, we used the AdamW algorithm [**?**] and a linear scheduler for the learning rate.

Table 3 shows the hyperparameter space used for the classification models. We used a random search to find the best combination of hyperparameters for each model. The search was performed using the *wandb* [3] library.

Our models were trained on a single NVIDIA GTX 1080Ti GPU. We used the *transformers* library [40] library to implement and train the models.

Table 3. Hyperparameter space for the classification models

Hyperparameter	Values
Learning rate	[2e−5, 3e−5, 5e−5, 6e−5, 7e−5, 8e−5]
Warmup ratio	[0.06, 0.08, 0.10]
Number of epochs	[3, 4, 5]

Table 4. Results of the classification experiments, expressed as the mean ± standard deviation. For each base pre-trained model, we analyzed two variants: target unaware and target aware, the latter marked with a + sign.

	POS F1	NEG F1	NEU F1	Macro F1
BERTin	68.6 ± 3.6	65.3 ± 4.5	63.2 ± 4.6	65.7 ± 3.3
BERTin$_{+T}$	67.3 ± 5.6	63.6 ± 5.5	58.0 ± 5.5	63.0 ± 4.7
ELECTRicidad	68.0 ± 2.8	61.4 ± 5.5	59.3 ± 3.5	62.9 ± 3.0
ELECTRicidad$_{+T}$	67.3 ± 1.3	63.1 ± 6.0	55.2 ± 4.9	61.9 ± 2.7
BETO	70.7 ± 3.7	66.9 ± 4.4	64.4 ± 3.8	67.3 ± 3.1
BETO$_{+T}$	73.8 ± 2.1	72.9 ± 3.3	65.9 ± 4.5	70.9 ± 2.3
RoBERTa	72.3 ± 3.5	68.6 ± 4.5	66.8 ± 4.0	69.3 ± 2.9
RoBERTa$_{+T}$	75.3 ± 3.9	74.2 ± 2.9	68.5 ± 4.7	72.7 ± 2.6
RoBERTuito	72.6 ± 2.6	68.9 ± 3.1	67.5 ± 4.0	69.7 ± 2.5
RoBERTuito$_{+T}$	75.0 ± 2.6	73.8 ± 2.3	67.8 ± 3.2	72.2 ± 1.4

5.2 Evaluation

Instead of performing a single train/dev/split, we used a Monte Carlo cross-validation [10] with 30 folds. We used a $N(0.2, 0.02)$ test size proportion, and this was made at the headline level, in order to ensure that no headlines were repeated in the train and test sets.

We used the Macro F1 score as the main evaluation metric, and also report F1 scores for each of the three classes. To keep record of the results, we used the *wandb* library [3].

6 Results

Table 6 displays the results of the classification experiments, expressed as the mean and its standard deviation of the Monte Carlo runs. We can observe that, in general, the target-aware models outperform their target-unaware counterparts, with the only exception of BERTin and ELECTRicidad, which are the worse-performing models.

For the three best models (BETO, RoBERTa and RoBERTuito), the differences in performance between target-aware and unaware models are statistically significant ($p < 0.01$, one-sided Wilcoxon Mann-Whitney U-test, FDR corrected) [39] for all metrics but for the neutral class. This is in line with our data: as neutral targets mostly do not mix with positive or negative polarities (see Figure 3

Table 5. Results of classification experiments for RoBERTa and RoBERTuito, analyzed for two variants: training from scratch and training from a checkpoint in the TASS 2020 dataset. Results are expressed as the mean ± standard deviation.

Model	POS F1	NEG F1	NEU F1	Macro F1
RoBERTa	75.3 ± 3.9	74.2 ± 2.9	68.5 ± 4.7	72.7 ± 2.6
RoBERTa$_{TASS}$	75.9 ± 1.5	74.8 ± 3.3	67.0 ± 3.3	72.6 ± 2.0
RoBERTuito	75.0 ± 2.6	73.8 ± 2.3	67.8 ± 3.2	72.2 ± 1.4
RoBERTuito$_{TASS}$	75.1 ± 3.2	74.4 ± 2.7	68.7 ± 3.0	72.7 ± 2.2

in the Appendix) predicting a general polarity is good enough. The gain in performance is more significant for the other two classes —positive and negative— where the target-aware models are able to better capture the polarity of the target, particularly in a mixed-polarity scenario.

Some of the models (BERTin and ELECTRicidad) are not able to leverage the target information, and their performance is similar to the target-unaware models. This is in line with poor performance for some other opinion-mining tasks in Spanish [28].

Among all the models, RoBERTuito and RoBERTa are the best performing ones. Table 6 shows the results for training the models from a checkpoint in the TASS 2020 dataset, and fine-tuning them on our dataset, both for RoBERTuito and RoBERTa. We can observe that the performance of the models is similar to the one obtained when training from scratch, not obtaining any significant improvement even when our training set is relatively small.

7 Discussion

Our experiments show that incorporating target information is valuable in identifying subjective content in news headlines, particularly for the dataset we constructed. The target-aware versions of RoBERTa$_{es}$ and RoBERTuito achieved the best performance, with Macro F1 scores of 75.3 and 75.0 respectively.

The significance of this task stems from the fact that news headlines can be politically biased, depending on the news outlet, and may influence readers' opinions when evaluating candidates during a presidential election. Our research project aims to evaluate the cognitive processes underlying presidential elections, including the hypothesis that increased exposure to positive news content mentioning certain candidates leads to a more positive perception, increased trust, and higher likelihood of being elected. Sentiment analysis is an important tool in evaluating how news outlets and social media content influence voters' subjective perceptions.

Among the explanations discussed in this field is the Agenda-setting theory [20], which states that the media or social networks, through a process of selection of information content, influence the issues that society considers relevant [8]. Another possible explanation of how these contents can influence voters' decisions is the priming effect, favouring a perception of familiarity and trust [2].

In any case, sentiment analysis algorithms are a useful tool to evaluate these effects and how they can influence an electoral outcome. Different research works, coming from very different theoretical fields and disciplines, have analysed and discussed the role of social media, print media, campaigns or fake news in the formation of judgement criteria and political decisions [35, 5, 6, 33, 29].

In order to build better democratic institutions, it is becoming increasingly important to provide empirical evidence of these influences, and thus alert the population of these implicit manipulations.

8 Conclusion and future work

In this work, we presented a dataset of news headlines from the main Argentinean press media, published during the election period, labeled by 3 annotators as positive, negative or neutral based on a target. By performing experiment with target-aware and target-unaware classifiers with state-of-the-art pre-trained models, we showed that these models can leverage target information, obtaining a small-yet-significant gain over predicting the overall sentiment in our data.

As future work, we plan to enlarge the dataset with new sources and other events, and explore new classification techniques.

We make the dataset available at the huggingface hub [5], and our code at github [6].

Acknowledgements

This work used computational resources from CCAD – Universidad Nacional de Córdoba [7], which are part of SNCAD – MinCyT, República Argentina

[5] https://huggingface.co/datasets/pysentimiento/spanish-targeted-sentiment-headlines
[6] https://github.com/pysentimiento/sentiment-elecciones
[7] https://ccad.unc.edu.ar/

Bibliography

[1] Bago, B., Rand, D.G., Pennycook, G.: Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. Journal of experimental psychology: general **149**(8), 1608 (2020)

[2] Bernal, F.A., Salgueiro, T.A., Brzostowski, A., Zapata, E.R., Carames, A., Pérez, J.M., Furman, D., Graziano, M., Larrosa, P.N.F.: Top-down modulation impairs priming susceptibility in complex decision-making with social implications. Scientific Reports **12**(1), 17867 (2022)

[3] Biewald, L.: Experiment tracking with weights and biases (2020), https://www.wandb.com/, software available from wandb.com

[4] Canete, J., Chaperon, G., Fuentes, R., Pérez, J.: Spanish pre-trained bert model and evaluation data. PML4DC at ICLR **2020** (2020)

[5] Cistulli, M., Snyder, L.B.: Priming, repetition, and the effects of multiple messages on perceptions of a political candidate. Michael G. Elasmar Boston Univ **2**, 44 (2009)

[6] Claibourn, M.P.: Making a connection: Repetition and priming in presidential campaigns. The Journal of Politics **70**(4), 1142–1159 (2008)

[7] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding pp. 4171–4186 (Jun 2019). https://doi.org/10.18653/v1/N19-1423, https://aclanthology.org/N19-1423

[8] García Beaudoux, V., D'Adamo, O.: Campañas electorales y sus efectos sobre el voto. análisis de la campaña electoral presidencial 2003 en argentina (2004)

[9] García-Vegaa, M., Díaz-Galianoa, M.C., García-Cumbrerasa, M.Á., del Arcoa, F.M.P., Montejo-Ráeza, A., Jiménez-Zafraa, S.M., Cámarab, E.M., Aguilarc, C.A., Antonio, M., Cabezudod, S., et al.: Overview of tass 2020: introducing emotion detection (2020)

[10] Gorman, K., Bedrick, S.: We need to talk about standard splits. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 2786–2791. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1267, https://aclanthology.org/P19-1267

[11] Gutiérrez-Fandiño, A., Armengol-Estapé, J., Pàmies, M., Llop-Palao, J., Silveira-Ocampo, J., Carrino, C.P., Armentano-Oller, C., Rodriguez-Penagos, C., Gonzalez-Agirre, A., Villegas, M.: Maria: Spanish language models. Procesamiento del Lenguaje Natural **68**, 39–60 (2022)

[12] Hanson, G., Haridakis, P.M., Cunningham, A.W., Sharma, R., Ponder, J.D.: The 2008 presidential campaign: Political cynicism in the age of facebook, myspace, and youtube. Mass Communication and Society **13**(5), 584–607 (2010)

[13] Krippendorff, K.: Content analysis: An introduction to its methodology. Sage publications (2018)

[14] Kucharski, A.: Study epidemiology of fake news. Nature **540**(7634), 525–525 (2016)
[15] Kushin, M.J., Yamamoto, M.: Did social media really matter? college students' use of online media and political decision making in the 2008 election. Mass communication and society **13**(5), 608–630 (2010)
[16] Kushin, M.J., Yamamoto, M.: Did social media really matter? college students' use of online media and political decision making in the 2008 election. In: New Media, Campaigning and the 2008 Facebook Election, pp. 63–86. Routledge (2013)
[17] Lau, R.R., Sigelman, L., Rovner, I.B.: The effects of negative political campaigns: A meta-analytic reassessment. The Journal of Politics **69**(4), 1176–1209 (2007)
[18] Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., et al.: The science of fake news. Science **359**(6380), 1094–1096 (2018)
[19] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
[20] McCombs, M.E., Shaw, D.L.: The agenda-setting function of mass media. Public opinion quarterly **36**(2), 176–187 (1972)
[21] Miranda, C.H., Guzman, J.: A review of sentiment analysis in spanish. Tecciencia **12**(22), 35–48 (2017)
[22] Mitchell, M., Aguilar, J., Wilson, T., Van Durme, B.: Open domain targeted sentiment. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. pp. 1643–1654 (2013)
[23] Nozza, D., Bianchi, F., Hovy, D.: What the [mask]? making sense of language-specific bert models. arXiv preprint arXiv:2003.02912 (2020)
[24] Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (jan 2008). https://doi.org/10.1561/1500000011, https://doi.org/10.1561/1500000011
[25] Pavlopoulos, I.: Aspect based sentiment analysis. Athens University of Economics and Business (2014)
[26] Pennycook, G., Rand, D.G.: Who falls for fake news? the roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of personality **88**(2), 185–200 (2020)
[27] Pinkleton, B.E., Um, N.H., Austin, E.W.: An exploration of the effects of negative political advertising on political decision making. Journal of Advertising **31**(1), 13–25 (2002)
[28] Pérez, J.M., Furman, D.A., Alonso Alemany, L., Luque, F.M.: Robertuito: a pre-trained language model for social media text in spanish. In: Proceedings of the Language Resources and Evaluation Conference. pp. 7235–7243. European Language Resources Association, Marseille, France (June 2022), https://aclanthology.org/2022.lrec-1.785
[29] Reiter, F., Matthes, J.: "The good, the bad, and the ugly": A panel study on the reciprocal effects of negative, dirty, and positive campaigning on political distrust. Mass Communication and Society **25**(5), 649–672 (2022)

[30] Rosa, J.D.L., Ponferrada, E.G., Romero, M., Villegas, P., de Prado Salas, P.G., Grandury, M.: Bertin: Efficient pre-training of a spanish language model using perplexity sampling. Procesamiento del Lenguaje Natural **68**(0), 13–23 (2022), http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6403

[31] Rosa, J.D.L., Ponferrada, E.G., Romero, M., Villegas, P., de Prado Salas, P.G., Grandury, M.: Bertin: Efficient pre-training of a spanish language model using perplexity sampling. Procesamiento del Lenguaje Natural **68**(0), 13–23 (2022), http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6403

[32] Rosenthal, S., Farra, N., Nakov, P.: SemEval-2017 task 4: Sentiment analysis in Twitter. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). pp. 502–518. Association for Computational Linguistics, Vancouver, Canada (Aug 2017). https://doi.org/10.18653/v1/S17-2088, https://aclanthology.org/S17-2088

[33] Schul, Y., Peri, N.: Influences of distrust (and trust) on decision making. Social Cognition **33**(5), 414–435 (2015)

[34] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. pp. 1631–1642. Association for Computational Linguistics, Seattle, Washington, USA (Oct 2013), https://aclanthology.org/D13-1170

[35] Stern, C.: Priming in political judgment and decision making. In: Oxford Research Encyclopedia of Politics (2019)

[36] Sun, C., Huang, L., Qiu, X.: Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. In: Proceedings of NAACL-HLT. pp. 380–385 (2019)

[37] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)

[38] Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., Bowman, S.: GLUE: A multi-task benchmark and analysis platform for natural language understanding. In: Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. pp. 353–355. Association for Computational Linguistics, Brussels, Belgium (Nov 2018). https://doi.org/10.18653/v1/W18-5446, https://aclanthology.org/W18-5446

[39] Wilcoxon, F.: Individual comparisons by ranking methods. In: Breakthroughs in statistics, pp. 196–202. Springer (1992)

[40] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A.: Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Sys-

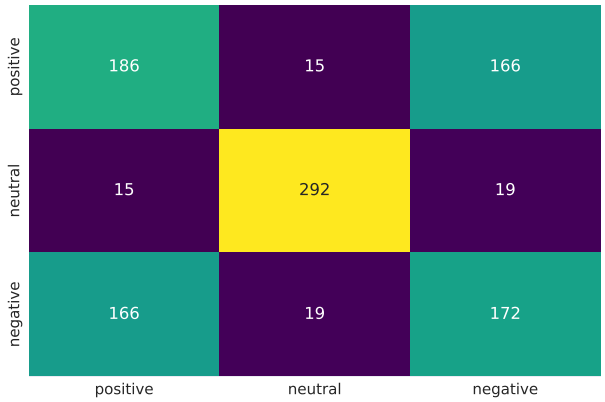


Fig. 3. Distribution of the number of mentions per candidate.

tem Demonstrations. pp. 38–45. Association for Computational Linguistics, Online (Oct 2020). https://doi.org/10.18653/v1/2020.emnlp-demos.6, https://aclanthology.org/2020.emnlp-demos.6

A Additional information about the dataset

Figure 3 shows the co-occurrence of labels within the same headline. The most frequent co-occurrence is neutral-neutral labels, followed by positive-positive and negative-negative labels. It is interesting to note that neutral headlines do not usually mix neither with positive nor negative labels. Also, positive and negative labels are more likely to co-occur. This is an indication that the polarity of the headline is not necessarily related to the polarity of the target.

Figure 4 shows the disaggregated mentions of each candidate by media. We can observe that these mentions are not uniformly distributed, as expected by the editorial line of each outlet [2]. Figure 4A shows this asymmetry in the dissemination of content in favour of certain candidates, and increased regardless of the media outlet. When analysing the frequency of mentions, tagged as positive or negative, there was also an asymmetry with respect to the candidates and the media outlets (Figure 4B,C).

Fig. 4. Frequency of mentions of each candidate in news headline disaggregated by media outlet