Predicting Sentence-Level Factuality of News and Bias of Media Outlets

Francielle Vargas[1,2]**, Kokil Jaidka**[3]**, Thiago A. S. Pardo**[1]**, Fabrício Benevenuto**[2]
[1]Institute of Mathematical and Computer Sciences, University of São Paulo, Brazil
[2]Computer Science Department, Federal University of Minas Gerais, Brazil
[3]Centre for Trusted Internet and Community, National University of Singapore, Singapore
`francielleavargas@usp.br, jaidka@nus.edu.sg`
`taspardo@icmc.usp.br, fabricio@dcc.ufmg.br`

Abstract

Automated news credibility and fact-checking at scale require accurate prediction of news factuality and media bias. This paper introduces a large sentence-level dataset, titled *FactNews*[1], composed of 6,191 sentences expertly annotated according to factuality and media bias definitions proposed by AllSides[2]. We use *FactNews* to assess the overall reliability of news sources by formulating two text classification problems for predicting sentence-level factuality of news reporting and bias of media outlets. Our experiments demonstrate that biased sentences present a higher number of words compared to factual sentences, besides having a predominance of emotions. Hence, the fine-grained analysis of subjectivity and impartiality of news articles showed promising results for predicting the reliability of the entire media outlet. Finally, due to the severity of fake news and political polarization in Brazil, and the lack of research for Portuguese, both dataset and baseline were proposed for Brazilian Portuguese.

1 Introduction

Automated fact-checking and news credibility have become undoubtedly an important research issue mainly due to the potential for misinformation to spread in the modern media ecosystem (Guo et al., 2022). Furthermore, although fake news is spreading on social media, it is necessary a source media where they would have been posted originally. Since websites have published low-credible news in the past, it is likely to happen again (Baly et al., 2018). Nevertheless, automated news credibility to assist human efforts and increase the understanding of the news ecosystem as a whole still requires urgent improvements (Horne et al., 2018).

Nowadays, fact-checking organizations have provided lists of unreliable news articles and media

N.	Sentence-level news article	Label
Title	President **lowers** Brazil's image with repeated misinformation and does not receive attention from global leaders.	Biased
S1	President Jair Bolsonaro **touch a sore point of** Europeans when he pointed out that the increased use of fossil fuels is a **serious** environmental setback, in his opening speech at the UN General Assembly, Tuesday (20).	Biased
S2	Germany received criticism at the UN for the investment agreement with Senegal for the production of gas in the African country.	Factual
S3	"This constitutes a serious setback for the environment", he said, referring to the Europeans	Quotes
S4	However, Bolsonaro signed measures contrary to environmental protection during the four years of the Brazilian government.	Factual
S5	There is **a huge difference** between speaking at the UN and being heard at the UN.	Biased

Table 1: Sentence-level factuality and bias prediction.

sources (Baly et al., 2018). Notwithstanding, these are inefficient once they need to be updated faster, besides being a very time-consuming task and requiring domain expertise.

A strategy to measure the credibility of news sources had already been done using the distribution of biased news in media outlets. While journalism is tied to a set of ethical standards and values, including truth and fairness, it often strays from impartial facts (Mastrine, 2022). As a result, biased news are produced, which may be correlated with the increasing polarization of media (Hamborg, 2020; Prior, 2013; Gentzkow and Shapiro, 2010). Moreover, media outlets play an important role in democratic societies (Baly et al., 2020) against sophisticated strategies of misinformation.

The state-of-the-art media bias detection has centered around predicting political-ideological bias (left, center, right) of news media. Most of the proposals use *lexical bias* that is linked to lexical

---
[1]https://zenodo.org/record/7868597
[2]https://www.allsides.com/

and grammatical cues and typically does not depend on context outside of the sentence. Also, it can be alleviated while maintaining its semantics: polarized words can be removed or replaced, and clauses written in active voice can be rewritten in passive voice (Fan et al., 2019). In the same settings, the definition of *frame bias* (Recasens et al., 2013) is also used to identify media bias, which occurs when subjective or opinion-based words are applied. In a study proposed by Fan et al. (2019), a frame-based analysis was performed for sentence-level media bias detection. The authors suggest that *informational bias* can be considered a specific form of framing in which there is an intention of influencing the reader's opinion of an entity (Fan et al., 2019). In this paper, we identify sentence-level media bias according to a guideline proposed by AllSides (Mastrine, 2022), which describes 16 different types of media bias.

Most researchers address media bias and factuality either at the level of media outlet (Baly et al., 2018) or at the level of individual article (Roy and Goldwasser, 2020; Baly et al., 2020). Nevertheless, each article comprises multiple sentences, which vary in their embedded bias (Lim et al., 2020), as well as factuality and quotes, as shown in Table 1.

Observe that factual sentences are a type of information presented with impartiality, focused on objective facts (e.g. S2, S4). In contrast, biased sentences stray from impartial facts and present the point of view of the journalist (e.g. Title, S1, S5), which may influence readers' perceptions. There are also direct quotes, which are neither biased sentences nor factual sentences (e.g. S3). Therefore, the news media sources may affect the power of swaying public opinion through the practical limitation to impartiality or using deliberate attempts to go against or in favor of something or someone.

Taking advantage of the fact that textual analysis of news articles published by a media outlet is critical for assessing the factuality of its reporting, and its potential bias (Baly et al., 2018), we tackle both biased and factual sentence prediction by using a strategy that has proved to be effective. In accordance with the literature, we created a new dataset titled *FactNews* composed of 6,191 sentences from 100 news stories totaling 300 documents. The same news story was extracted from three different media outlets. Furthermore, each sentence of the dataset was annotated with three different classes according to factuality and media

bias definitions proposed by AllSides (Mastrine, 2022): **(i) factual spans**, which consists of a type of information presented with impartiality focused on the objective fact or, in other words, they are sentences that describe a fact and are committed to objectivity; **(ii) biased spans**, specifically biased spans were classified according to 12 types of media bias proposed by AllSides (Mastrine, 2022), which we describe in detail in Section 3.2.2; additionally, **(iii) quotes** consist of direct statements often followed by quotation marks that journalists in general use to report the speech of someone involved in the reported event. In this paper, we argue that quotes should be defined differently than biased and factual spans.

Furthermore, we trained two different models using fine-tuned BERT. The first model predicts whether the sentence of a given news article is factual or not. The second model predicts whether the sentence of a news article from a given news media outlet is biased or not. As a result, baseline models for sentence-level factuality and sentence-level media bias prediction by BERT fine-tuning were presented in order to provide a more accurate score of the reliability of the entire media source. Our contributions may be summarized as follows:

- We focus on an under-explored and surely relevant problem: predicting the factuality of news reporting and bias of media outlets.

- We create the first large-scale and manually annotated dataset at the sentence-level for both tasks in Portuguese. The dataset, agreements/disagreements, and code are available, which may facilitate future research.

- We present a new annotation schema to identify media bias and factuality, as well as a baseline for the factual sentence prediction task.

- We provide data analysis on factual and biased sentences demonstrating the reliability of the proposed annotation schema and models.

In what follows, in Section 2, related work is presented. Section 3 describes the proposed FactNews dataset, and Section 4 our experimental settings. In Section 5, baseline results for sentence-level factuality and media bias prediction are shown. In Section 6, conclusions are presented.

2 Related Work

2.1 News Credibility

While the assessing of news has been made mainly by journalists, information analysts, and news consumers, this task has become complex due to the ever-growing number of news sources and the mixed tactics of maliciously false sources and misinformation strategies (Horne et al., 2018). News credibility state-of-the-art has been mostly focused on measuring the reliability of news reporting (Pérez-Rosas et al., 2018; Hardalov et al., 2016) or the entire media outlets (Baly et al., 2018; Horne et al., 2018; Baly et al., 2019), as well as social media platforms (Castillo et al., 2011; Mukherjee and Weikum, 2015) in order to mitigate fake news harmful spreading. Furthermore, as stated by Baly et al. (2018), estimating the reliability of a news source is relevant not only when fact-checking a claim (Popat et al., 2016; Nguyen et al., 2018), nevertheless, it provides a surely contribution in order to tackle article-level tasks such as "fake news" detection (De Sarkar et al., 2018; Yuan et al., 2020; Reis et al., 2019; Pan et al., 2018; Vargas et al., 2022; Dong et al., 2015). News credibility information has been studied at different levels (Baly et al., 2018): (i) claim-level (e.g., fact-checking), (ii) article-level (e.g., "fake news" detection), (iii) user-level (e.g., hunting for trolls), and (iv) medium-level (e.g., source reliability estimation). In this paper, we focus on predicting the factuality of reporting and bias of media outlets at medium-level towards source reliability estimation.

2.2 Fact-Checking

According to Guo et al. (2022), fake news detection and fact-checking are different tasks once that fact-checkers focus on assessing news articles and include labeling items based on aspects not related to veracity, besides other factors—such as the audience reached by the claim, and the intentions and forms of the claim—are often considered, as well as the context of propaganda detection (Martino et al., 2020). Fact-checking state-of-the-art at the claim-level, as claimed by (Baly et al., 2018) mostly uses information extracted from social media, i.e., based on how users comment on the target claim (Ribeiro et al., 2022; Baly et al., 2019), so as to the use of the Web data as information source (Mihaylova et al., 2018; Reis et al., 2020; Mihaylova et al., 2018).

2.3 Media Bias Detection

Article-level media bias consists of predicting whether a news article is biased. This task was studied in (Sapiro-Gheiler, 2019). They predicted political ideology using recursive neural networks (Iyyer et al., 2014). Baly et al. (2019) proposed a multi-task regression framework aiming to predict the trustworthiness and ideology of news media. Liu et al. (2022) applied the pre-trained language model for the political domain to characterize political stance. Baly et al. (2020) created a model from media sources, such as a shortcut, for predicting ideology using adversarial networks.

Sentence-level media bias consists of a task aiming to predict whether each sentence of a news report is biased or not. Fan et al. (2019) provided the first sentence-level annotated dataset titled *BASIL*, composed of 300 news articles annotated with 1,727 biased spans and 6,257 non-biased sentences, as well as fine-tuning BERT baseline experiments reaching an F1-Score of 47,27%. Lim et al. (2020) created a new dataset titled *biased-sents*, which is composed of 966 sentences from 46 English-language news articles covering four different events. Färber et al. (2020) proposed a dataset of 2,057 sentences annotated with four labels: hidden assumptions, subjectivity, framing, and bias. Spinde et al. (2021) provided an annotation-expert project through a new dataset titled *BABE*. This dataset consists of 3,700 sentences balanced among topics and outlets, and a fine-tuned BERT baseline reaching an F1-Score of 80,04%. Lastly, Lei et al.(2022) showed that embedded discourse structure for sentence-level media bias effectively increases the recall by 8.27% - 8.62%, and precision by 2.82% - 3.48%.

2.4 Factuality of News Reporting

Predicting the factuality of news reporting is definitely an under-explored research topic. This task consists of predicting whether a news report on news media is factual or not. Baly et al. (2018) studied article-level factuality of news reporting. They proposed a baseline by analyzing textual content (syntactic and semantic) of news reporting given a news media source with features based on sentiment, morality, part-of-speech, etc. The best model obtained 58.02% at F1-Score. Bozhanova et al. (2021) studied the factuality of reporting of news media outlets by studying the user attention cycles in their YouTube channels.

3 FactNews Dataset

We collected, annotated, and released a new dataset titled *FactNews*, which consists of a sentence-level annotated dataset in Brazilian Portuguese that contains 6,191 annotated sentences, as follows: 4,302 sentences annotated as factual spans; 1,389 sentences annotated as quotes, and 558 sentences annotated as biased spans. The entire dataset-building process lasted an average of six months. A dataset overview is shown in Table 4. We first selected three different well-known and relevant media outlets in Brazil, and extracted the same news story from each one of them, as shown in Table 2.

Media	News Reporting
Folha	O presidente Jair Bolsonaro colocou o dedo na ferida dos europeus ao apontar que o aumento do uso de combustíveis fósseis é um grave retrocesso ambiental, em seu discurso de abertura da Assembleia-Geral da ONU na manhã desta terça-feira (20). *President Jair Bolsonaro touch a sore point of Europeans when he pointed out that the increased use of fossil fuels is a serious environmental setback, in his opening speech at the UN General Assembly this Tuesday morning (20) (...)*
Estadão	O presidente Jair Bolsonaro encerrou seu discurso na Assembleia-Geral da ONU, nesta terça-feira, 20, afirmando que o povo brasileiro acredita em "Deus, Pátria, família e liberdade", que tem inspiração no fascismo de Benito Mussolini (1883-1945). *President Jair Bolsonaro ended his speech at the UN General Assembly, this Tuesday, 20, stating that the Brazilian people believe in "God, Fatherland, family and freedom", which has by the fascism of Benito Mussolini (1883-1945) (...)*
O Globo	O presidente Jair Bolsonaro seguiu o roteiro de campanha em seu discurso na Assembleia Geral da Organização das Nações Unidas (ONU), em Nova York (EUA), e aproveitou para atacar o ex-presidente Luiz Inácio Lula da Silva nesta terça-feira (20)(...) *President Jair Bolsonaro followed the campaign script in his speech at the General Assembly of the United Nations (UN) in New York (USA), and took the opportunity to attack former president Luiz Inácio Lula da Silva this Tuesday (20)(...)*

Table 2: The same news story was collected from three different Brazilian media outlets, which reports the Jair Bolsonaro (former President) speech at the UN in 2022.

3.1 Data Collection

As shown in Table 4, the proposed *FactNews* was collected from 100 news articles in triples - the same news story from three different Brazilian media news outlets: Folha de São Paulo[3], O Globo[4], and Estadão[5], resulting in 300 documents.

Furthermore, we used a statistical approach and a search algorithm, in order to collect news related to six different domains (e.g. politics, world, daily, sports, science, and culture) from periods 2006-2007 and 2021-2022. Therefore, in accordance with relevant literature of the area, we selected three news articles from different news outlets related to the same topic or story (Spinde et al., 2021; Baly et al., 2020; Fan et al., 2019).

3.2 Data Annotation

3.2.1 Annotators Profile

In order to ensure the reliability of data annotation, two different annotators, a linguist and a computer scientist from different regions (southeast and northeast) performed the task, both with at least a Ph.D. degree or Ph.D. candidate status. Furthermore, the annotation task was led by an NLP researcher, and the annotators were supported by our annotation schema (see Figure 1), and a guideline with rich examples proposed by AllSides.

3.2.2 Annotation Schema

Corroborating our objective of classifying factuality and bias at the sentence level, we segmented each one of the 300 news articles in sentences and annotated them according to three different classes: (i) factual spans, (ii) biased spans, and (iii) quotes, as shown in Figure 1.

We proposed an expert annotation schema for sentence-level factuality and media bias classification. We first evaluated whether the sentence was committed with impartiality. In other words, whether it presented a type of information focused on objective facts. Whether "yes", it should be classified as factual span. Otherwise, it should be classified as a biased span taking into account 12 types of media bias defined by AllSides (Mastrine, 2022), described as follow. We did not consider 4 types (slant, bias by omission, bias by story choice, and photo bias), from the AllSides guidelines[6], once they did not match our sentence-level proposal.
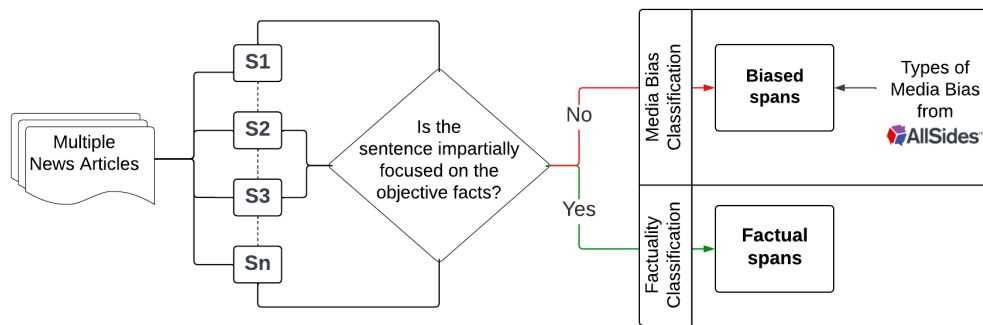

Figure 1: FactNews annotation schema.

1. **Spin**: This type of bias consists of vague, dramatic, or sensational language. For example "President Donald Trump gloated over mass layoffs at multiple news outlets on Saturday". Note that "gloated" is evidence of subjective interpretation from the journalist meaning that Trump's tweet shows he is smug or taking pleasure in the layoffs.

2. **Unsubstantiated Claims**: This bias occurs when journalists provide claims in their reporting without including any evidence. For example, "Sen. Kamala Harris condemned the violent attack on actor Jussie Smollett, calling it an attempted modern-day lynching".

3. **Opinion Statements Presented as Facts**: In this bias, journalists use subjective language or statements under the guise of reporting objectively, which is based on personal opinions, assumptions, beliefs, tastes, preferences, or interpretations. For example, "The EPA is lifting greenhouse gas limits on coal power plants: The latest proposal won't stop the steady decline of the coal industry". Note that the underline statement shows the point of view of the journalist.

4. **Sensationalism/Emotionalism**: Here, the information is presented in a way that provides a shock or triggers a deep impression. For example, "If seats that look like this one in Rio de Janeiro are toss-ups in November, it's going to be a bloodbath".

5. **Mudslinging/Ad Hominem**: This type of media bias occurs when unfair or insulting things are said about someone in order to damage their reputation. For example, "Bret Stephens is not a bedbug. He is a delicate snowflake".

6. **Mind Reading**: This bias occurs when journalists assume they know what another person thinks, or thinks that the way they see the world reflects the way the world really is. For example, "Bolsonaro's hatred of looking foolish and left party' conviction that they have a winning hand is leaving the President with no way out of the stalemate over his gun port legalization.".

7. **Flowed Logic**: This bias consists of a type of faulty reasoning resulting in misrepresenting people's opinions or arriving at conclusions that are not justified by the given evidence (e.g. arriving at a conclusion that doesn't follow from the premise). For example, "Two-time failed Democratic presidential candidate Hillary Clinton snubbed Melania Trump during George H.W. Bush's funeral, refusing to shake her hand (...), and an awkward and bitter nod back from Hillary".

8. **Omission of Source Attribution**: This bias occurs when a journalist does not back up their claims by linking to the source of that information. For example, when journalists claim "critics say" without specific attribution.

9. **Subjective Qualifying Adjectives**: Journalists can reveal this bias when they include subjective, qualifying adjectives in front of specific words or phrases. For example, "Rep. Madison Cawthorn issues sinister warning to anyone opposing Him. The extremist republican ranted about liberals trying to make people "sexless"". Note that subjective qualifiers are closely related to *spin words* and phrases once they obscure the objective truth and insert subjectivity.

10. **Word Choice**: This bias occurs when words and phrases are loaded with political implications. Therefore, the words or phrases a media outlet uses can reveal its perspective or ideology. Examples of Polarizing Word Choices: "pro-choice — anti-choice", "gun rights — gun control", "riot — protest", "illegal immigrants — migrants".

11. **Negativity Bias**: Journalists can emphasize bad or negative news, or frame events in a negative light. For example, news articles related to death, violence, turmoil, and struggle, tend to obtain more attention and elicit more shock, and fear. As a result, we keep reading the news, in order to know more on this issue.

12. **Elite v. Populist Bias**: Journalists can defer to the beliefs, viewpoints, and perspectives of people who are part of society's most prestigious or not prestigious. Furthermore, Elite/populist bias has a geographic component. For example, "The FDA turned a blind eye or colluded with unbelievable harms revealed in the Pfizer documents, so the FDA can't be trusted. The CDC can't be trusted". Here, the journalist pushes back against the elite government, saying they can't be trusted.

3.2.3 Annotation Evaluation

We computed the inter-annotator agreement score using Cohen's kappa (Sim and Wright, 2005). We obtained a kappa score of 82%. We also analyzed the matrix of agreements and disagreements among annotators for each class (e.g. factual, biased, and quotes). Results are shown in Table 3.

FactNews Dataset		Annotator 1			Total
		Factual	*Biased*	*Quotes*	
	Factual	4,211	27	7	4,245
Annotator 2	*Biased*	284	261	1	546
	Quotes	138	6	1,256	1,400
Total		4,633	294	1,264	**6,191**
Kappa		0.82			

Table 3: Inter-annotator agreement by Kappa.

Observe that two annotators, a linguist (expert in media bias) and a computer scientist (non-expert) labeled the FactNews dataset. Moreover, disagreement cases[7] were also judged by two judges, and three meetings were carried out, in which annotators could discuss and re-evaluate the given labels.

[7] https://zenodo.org/record/7868597/

Furthermore, as shown in Table 3, the high values obtained by diagonal lines (e.g. 4,211, 261, 1,256) are indicative of high-human agreement. We also observed that annotator 2, which is a specialist, provides better media bias classification compared with annotator 1, which is a non-specialist. For example, while both annotators agreed on the bias labels with 261 matches between them, 284 labels were classified by annotator 1 as "factual" and by annotator 2 (specialist) as "biased". These cases were mostly decided by judges as being "biased".

3.3 Data Analysis

Table 4 shows the dataset statistics. The FactNews is composed of 6,191 sentences annotated according to three classes: factual spans (4,242), quotes (1,391), and biased spans (558). Most of the sentences (68.51%) are factual spans, in contrast to quotes (22.52%) and biased (8.81%) categories, respectively. Each news article consists of an average of 24.27 sentences of which 14.14 are classified as factual sentences, 7.06 as quotes, and 3.27 as biased sentences. Furthermore, factual sentences contain an average of 20.36 words, biased sentences 22.14 words, and quotes 17.38 words.

Furthermore, biased spans present more words than factual spans in all grammar categories (e.g. nouns, verbs, adjectives), as well as predominance in terms of emotion lexicon. Lastly, the titles of news articles hold 8.36% bias, 5.33% quotes, and 86% of factual sentences. On the other hand, the body of news articles holds 13.35% bias, 20.38% quotes, and 66.27% of factual sentences.

In Figure 2, we also show the distribution of factual and biased sentences across domains according to each media news outlet. Notably, the distribution of factuality is equivalent across different domains. Differently, the distribution of bias varies in accordance with the domain and media outlet. Considering the labels across domains, 62.55% are related to politics; 14.21% world; 7.14 sport; 6.67 daily; 6.65 culture; and 1.98% science.
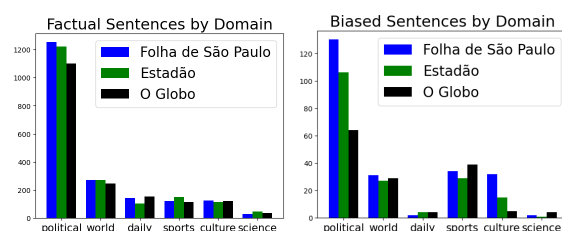

Figure 2: The cross-domain distribution of factual and biased sentences from different media outlets.

4 Baseline Experiments

4.1 Motivations and Goals

As mentioned before, news credibility analysis and fact-checking are both time-consuming tasks. Furthermore, with the amount of new information that appears and the speed with which it spreads, manual validation is insufficient (Guo et al., 2022). Nevertheless, automated approaches present several challenges, since automated trustworthiness analysis is a technically complex issue, besides involving a wide variety of ethical dilemmas.

Instead of analyzing the veracity of news articles, in this paper, we are interested in the fine-grained characterization of the entire media outlet by predicting the factuality of news reporting and bias of media outlets for source reliability estimation.

We aim to predict sentence-level media bias and factuality by analyzing different types of media bias and journalist factuality definitions, both proposed by AllSides (Mastrine, 2022). Specifically, we first built the state-of-art media bias detection models. Secondly, a baseline sentence-level factuality detection model was proposed by analyzing the subjectivity and impartiality of text content. As a result, we hope to explain more accurately the overall reliability of the entire news media source.

4.2 Model Architecture

First of all, we argue that factual spans contain a type of information that deals with facts, hence it is impartially focused on objective facts. In contrast, non-factual information contains a type of information presented subjectively (with partiality) that often strays from objective facts. Taking into account this premise, we describe both model's sentence-level media bias and factuality, as follows:

Sentence-Level Media Bias Model: We implemented the state-of-the-art sentence-level media bias models (Fan et al., 2019) on the FactNews dataset. Our model for media bias uses a binary class variable composed of biased spans (558 labels) versus unbiased spans (558 labels).

Sentence-Level Factuality Model: We hypothesize that the factuality of news reporting may be predicted by analyzing the subjectivity and impartiality of text content, which is inspired by Baly et al., (2018). Since factual sentences are impartially focused on objective facts, in contrast to the biased ones that are partially presented and focused on subjective interpretations, we built a model to predict sentence-level factuality based on aspects

of subjectivity and impartiality. Finally, once both biased spans and quotes present evidence of subjective interpretation of facts (Hu et al., 2023), our sentence-level factuality model is composed of a binary class variable from biased spans and quotes (1,949 labels) versus factual spans (1,949 labels).

4.3 Learning Methods and Features Set

In data preparation, we segmented sentences using the spaCy library and only special characters were removed. As learning method, we used the SVM with linear kernel. We split our data into train (90%), and test (10%), and applied the 10-fold cross-validation. We also used the undersampling (Witten et al., 2016) to balance the classes. Finally, a robust set of experiments was performed using four model architectures inspired by Baly et al. (2018), which we describe in detail as follows:

BERT fine-tuning: We used the best BERT fine-tuned model by Keras, held batch size at 64, maximum of 500 features, learning rate at 2e-05 and number of epochs at 4.

Subjective-lexicons: We evaluated a BoW using features extracted from sentiment and emotion lexicons (Pasqualotti, 2008), which present semantic polarity and emotion types.

Part-of-speech (POS): We evaluated a BoW using features based on POS, more precisely, noun, verb, adjective, adverb, pronoun, and conjunctions, which was supported by the spaCy tagging.

TF-IDF: Baseline vector space model.

5 Results and Discussion

Table 5 summarizes the performance of the models. We further provide a comparison of results in Table 6. The best model for sentence-level factuality prediction obtained 88% of F1-Score. For sentence-level media bias prediction, the best model obtained 67% of F1-score. Notably, the part-of-speech model presented competitive results for both tasks in contrast to the subjective lexicons, which obtained poor results for both tasks.

5.1 Comparing Results

While a direct comparison is unfair (as the authors use different datasets), it offers an idea of the general performance, as shown in Table 6. Note that although it only offers an idea of the general performance, our sentence-level factuality prediction model (88%) significantly outperforms the article-level factuality prediction baseline (58%).

Description		Folha de São Paulo			Estadão			O Globo			All
		factual	quotes	biased	factual	quotes	biased	factual	quotes	biased	
#Articles			100			100			100		300
#Sentences		1,494	450	231	1,428	483	182	1,320	458	145	6191
#Words		30,374	7,946	5,177	30,589	8,504	4,002	25,505	7,740	3,195	123,032
Avg Sentences/Article		14.94	7.03	3.78	14.28	7.00	3.19	13.20	7.15	2.84	8.15
Avg Words/Sentences		20.33	17.65	**22,41**	21,45	17,60	**21,98**	19,32	16,89	**22,03**	19,96
Body/Title	*Body*	1,337	440	207	1,218	473	162	1,089	441	131	5,498
	Title	157	10	24	210	10	20	231	17	14	693
Domains	*Political*	912	340	130	870	352	106	748	351	64	3,873
	World	224	48	31	224	49	27	216	32	29	880
	Sports	100	23	34	124	25	29	98	18	39	490
	Daily	132	11	2	98	7	4	148	7	4	413
	Culture	98	26	32	72	42	15	77	45	5	412
	Science	28	2	2	40	8	1	33	5	4	123
Part-of-speech (Avg)	*Noun*	4.85	4.09	**5.72**	5.21	4.12	**5.60**	4.59	3.82	**5.19**	4.79
	Verb	2.20	2.55	**2.60**	2.28	2.51	**2.53**	2.00	2.44	**2.57**	4.18
	Adjective	1.03	1.03	**1.32**	1.11	1.08	**1.32**	0.94	0.97	**1.48**	1.14
	Adverb	0.67	0.82	**0.93**	0.67	0.94	**0.90**	0.59	0.90	**0.94**	0.81
	Pronoun	0.52	1.02	**0.73**	0.51	0.97	**0.56**	0.47	0.90	**0.59**	0.69
	Conjunction	0.51	0.55	**0.61**	0.54	0.57	**0.73**	0.51	0.88	**0.70**	0.62
Emotion (Avg)	*Happiness*	0.12	0.22	**0.20**	0.16	0.28	**0.26**	0.13	0.28	**0.22**	0.20
	Disgust	0.03	0.06	**0.05**	0.04	0.06	**0.03**	0.04	0.04	**0.04**	0.04
	Fear	**4.18**	3.80	4.63	4.41	3.77	4.56	4.05	3.60	4.50	4.16
	Anger	0.05	0.06	**0.13**	0.07	0.07	**0.12**	0.06	0.08	**0.20**	0.09
	Surprise	0.01	0.03	**0.03**	0.01	0.03	**0.05**	0.01	0.02	**0.01**	0.02
	Sadness	**5.86**	5.71	6.52	6.17	5.55	6.48	5.56	5.40	6.19	5.93
Polarity (Avg)	*Positive*	2.41	3.25	2.93	2.55	3.22	2.95	2.26	3.26	2.96	2.86
	Negative	0.05	0.06	0.05	0.07	0.10	0.09	0.06	0.07	0.06	0.06
	Neutral	9.55	9.77	10.93	9.92	9.52	11.03	8.91	9.28	10.56	9.94

Table 4: FactNews dataset statistics.

Sentence-Level Factuality	Precision	Recall	F1-Score
BERT fine-tuning	0.89	0.89	**0.88**
Part-of-speech	0.77	0.77	0.76
TF-IDF	0.81	0.69	0.66
Polarity-lexicon	0.63	0.62	0.62
Emotion-lexicon	0.61	0.61	0.61
Sentence-Level Media Bias	**Precision**	**Recall**	**F1-Score**
BERT fine-tuning	0.70	0.68	**0.67**
Part-of-speech	0.67	0.66	0.66
Polarity-lexicon	0.50	0.50	0.50
Emotion-lexicon	0.53	0.52	0.50
TF-IDF	0.78	0.58	0.48

Table 5: Sentence-level factuality and bias prediction.

Sentence-Level Media Bias Prediction			
Datasets	**Lang**	**Docum.**	**Sent.**	**F1-Score**
BASIL (baseline)	En	300 news	7,984	**0.47**
Biased-sents	En	46 news	966	-
BABE	En	100 news	3,700	0.80
FactNews	Pt	300 news	6,191	0.67
Sentence-Level Factuality Prediction				
FactNews (baseline)	Pt	300 news	6,191	**0.88**
Article-Level Factuality Prediction				
MBFC (baseline)	En	1,066 medias	-	**0.58**
MBFC corpus	En	489 medias	-	0.76*

Table 6: Result analysis in comparison with literature.

6 Conclusions

Since low-credibility media outlets may potentially be targeted for the spreading of misinformation, we study the factuality of news reporting and bias of media outlets at the sentence-level for fine-grained source reliability estimation. We further provide a new data resource and baselines for Brazilian Portuguese low-resourced language. We first created a large and manually-annotated dataset for sentence-level factuality and media bias prediction. Then, we provided a detailed data analysis, demonstrating the reliability of the annotation schema and models. Finally, baseline models for sentence-level factuality and media bias prediction by BERT were presented in order to provide an accurate score of the reliability of the entire news media. Results also showed that biased spans are more numerous in words and emotions compared to factual spans. Moreover, media outlets presented different proportions of bias, and its distribution in news articles may vary according to the domain, in contrast to factual spans. We also concluded that expert annotators are more successful to identify media bias.

Acknowledgements

References

Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. 2020. We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, page 4982–4991, Held Online.

Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. Predicting factuality of reporting and bias of news media sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3528–3539, Brussels, Belgium.

Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James Glass, and Preslav Nakov. 2019. Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2109–2116, Minneapolis, Minnesota.

Krasimira Bozhanova, Yoan Dinkov, Ivan Koychev, Maria Castaldo, Tommaso Venturini, and Preslav Nakov. 2021. Predicting the factuality of reporting of news media using observations about user attention in their YouTube channels. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 182–189, Held Online.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, page 675–684, New York, United States.

Sohan De Sarkar, Fan Yang, and Arjun Mukherjee. 2018. Attending sentences to detect satirical fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3371–3380, New Mexico, United States.

Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. 2015. Knowledge-based trust: Estimating the trustworthiness of web sources. *Proc. VLDB Endow.*, 8(9):938–949.

Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. 2019. In plain sight: Media bias through the lens of factual reporting. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 6343–6349, Hong Kong, China.

Michael Färber, Victoria Burkard, Adam Jatowt, and Sora Lim. 2020. A multidimensional dataset based on crowdsourcing for analyzing and detecting news bias. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, page 3007–3014, New York, United States.

Matthew Gentzkow and Jesse M Shapiro. 2010. What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. 2022. A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, 10:178–206.

Felix Hamborg. 2020. Media bias, the social sciences, and NLP: Automating frame analyses to identify bias by word choice and labeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 79–87, Held Online.

Momchil Hardalov, Ivan Koychev, and Preslav Nakov. 2016. In search of credible news. In *17th International Conference on Artificial Intelligence: Methodology, Systems, and Application*, pages 172–180, Varna, Bulgaria.

Benjamin D. Horne, William Dron, Sara Khedr, and Sibel Adali. 2018. Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news. In *Proceedings of the The Web Conference 2018*, page 235–238, Geneva, Switzerland.

Tiancheng Hu, Manoel Horta Ribeiro, Robert West, and Andreas Spitz. 2023. Quotatives indicate decline in objectivity in u.s. political news. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 363–374, Limassol, Cyprus.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1113–1122, Baltimore, Maryland.

Yuanyuan Lei, Ruihong Huang, Lu Wang, and Nick Beauchamp. 2022. Sentence-level media bias analysis informed by discourse structures. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10040 – 10050, Abu Dhabi, United Arab Emirates.

Sora Lim, Adam Jatowt, Michael Färber, and Masatoshi Yoshikawa. 2020. Annotating and analyzing biased sentences in news articles using crowdsourcing. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 1478–1484, Marseille, France.

Yujian Liu, Xinliang Frederick Zhang, David Wegsman, Nicholas Beauchamp, and Lu Wang. 2022. POLITICS: Pretraining with same-story article comparison for ideology prediction and stance detection. In *Findings of the Association for Computational Linguistics: 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1354–1374, Seattle, United States.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020. A survey on computational propaganda detection. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence*, pages 4826–4832, Yokohama, Japan.

Julie Mastrine. 2022. *How to Spot 16 Types of Media Bias*. AllSides: Don't be fooled by media bias and misinformation, California, United States.

Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Mitra Mohtarami, Georgi Karadzhov, and James Glass. 2018. Fact checking in community forums. In *The 32th AAAI Conference on Artificial Intelligence*, pages 5309–5316, New Orleans, United States.

Subhabrata Mukherjee and Gerhard Weikum. 2015. Leveraging joint interactions for credibility analysis in news communities. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, page 353–362, New York, United States.

An Nguyen, Aditya Kharosekar, Matthew Lease, and Byron Wallace. 2018. An interpretable joint graphical model for fact-checking from crowds. In *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, 1, pages 1511–1518, Louisiana, United States.

Jeff Z. Pan, Siyana Pavlova, Chenxi Li, Ningxi Li, Yangmei Li, and Jinshuo Liu. 2018. Content based fake news detection using knowledge graphs. In *Proceedings of the 17th International Semantic Web Conference*, pages 669–683, California, Unites States.

P. R Pasqualotti. 2008. Reconhecimento de expressões de emoções na interação mediada por computador. Master's thesis, Dissertação de Mestrado em Ciência da Computação. Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS, Porto Alegre, Brasil.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3391–3401, New Mexico, United States.

Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2016. Credibility assessment of textual claims on the web. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, page 2173–2178, New York, United States.

Markus Prior. 2013. Media and political polarization. *Annual Review of Political Science*, 16(1):101–127.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1650–1659, Sofia, Bulgaria.

Julio C. S. Reis, Andre Correia, Fabricio Murai, Adriano Veloso, and Fabricio Benevenuto. 2019. Explainable machine learning for fake news detection. In *Proceedings of the 11th ACM Conference on Web Science*, pages 17–26, Massachusetts, United States.

Julio C. S. Reis, Philipe Melo, Kiran Garimella, Jussara M. Almeida, Dean Eckles, and Fabrício Benevenuto. 2020. A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In *Proceedings of the 14th International AAAI Conference on Web and Social Media*, pages 903–908, Held Online.

Manoel Horta Ribeiro, Savvas Zannettou, Oana Goga, Fabrício Benevenuto, and Robert West. 2022. Can online attention signals help fact-checkers to fact-check? In *Workshop Proceedings of the 17th International AAAI Conference On Web and Social Media*, pages 1–10, Atlanta, United States.

Shamik Roy and Dan Goldwasser. 2020. Weakly supervised learning of nuanced frames for analyzing polarization in news media. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 7698–7716, Held Online.

Eitan Sapiro-Gheiler. 2019. Examining political trustworthiness through text-based measures of ideology. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 10029–10030, Hawaii, United States.

Julius Sim and Chris C Wright. 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical therapy*, 85(3):257–268.

Timo Spinde, Manuel Plank, Jan-David Krieger, Terry Ruas, Bela Gipp, and Akiko Aizawa. 2021. Neural media bias detection using distant supervision with BABE - bias annotations by experts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1166–1177, Punta Cana, Dominican Republic.

Francielle Vargas, Jonas D'Alessandro, Zohar Rabinovich, Fabrício Benevenuto, and Thiago Pardo. 2022. Rhetorical structure approach for online deception detection: A survey. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5906–5915, Marseille, France.

Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. 2016. *Data Mining: Practical machine learning tools and techniques*, 3ed edition. Morgan Kaufmann Publishers.

Chunyuan Yuan, Qianwen Ma, Wei Zhou, Jizhong Han, and Songlin Hu. 2020. Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5444–5454, Held Online.