9-2023

Frank at CheckThat! 2023: Detecting the Political Bias of News Articles and News Media Notebook for the CheckThat! Lab at CLEF 2023

Dilshod Azizov
Mohamed Bin Zayed University of Artificial Intelligence

Preslav Nakov
Mohamed Bin Zayed University of Artificial Intelligence

Shangsong Liang
Mohamed Bin Zayed University of Artificial Intelligence

Frank at CheckThat! 2023: Detecting the Political Bias of News Articles and News Media

Notebook for the CheckThat! Lab at CLEF 2023

Dilshod Azizov[1,*], Preslav Nakov[1] and Shangsong Liang[1]

[1]Mohamed bin Zayed University of Artificial Intelligence, UAE

Abstract

This paper addresses the challenge of detecting political bias in news articles and media outlets from `CheckThat!` lab Task 3 [1, 2] by proposing an automated method for classifying these as left, center, or right-leaning. As mass media consumption continues to grow, the capability to identify bias in news reporting is crucial due to the potential societal impact of unaddressed political bias. To tackle this issue, we present a comprehensive approach employing machine learning techniques to detect political leaning in news media and articles. Our model, CatBoost, is evaluated on a diverse dataset comprising over 55,000 news articles sourced from AllSides[1] at the article-level. For each model, we aggregate predictions made across news items by a single medium using a majority voting system at medium-level. Our dataset gathered and annotated from over 1,000 popular online platforms as rated by Media Bias/Fact Check[2], categorizes political bias into the left, center, or right-wing. We have approximately ten articles from each of these platforms, yielding over 8,000 articles in total. We employ both CatBoost and CatBoost OF[3] for media-level classification. These effectively detect political ideology across various media sources, with our CatBoost model demonstrating robustness and effectiveness in handling diverse data. Our findings suggest that utilizing the majority voting technique at the medium level improves model performance. We also highlight the importance of addressing class imbalance and implementing balanced data splits to enhance model performance. Regarding article-level classification using CatBoost, we achieve a Mean Absolute Error (MAE) of 0.270, an F1 score of 0.690, and an accuracy of 0.694. For media-level classification, we achieve a competitive MAE of 0.320, and with the use of the majority voting classifier, our model attains an F1 score of 0.727 and an accuracy of 0.725.

Keywords

political bias, news articles, news media

1. Introduction

The political leaning of news articles and news media has become an increasingly important topic in today's world of information overload. How news is presented and reported can have a significant impact on people's perceptions, beliefs, and even voting behaviors [3]. A recent study has shown that political bias may influence citizens' voting decisions and can change

[1]www.allsides.com

[2]www.mediabiasfactcheck.com

[3]OF - operating only on the first 300 features from TF-IDF

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece

*Corresponding author.

dilshod.azizov@mbzuai.ac.ae (D. Azizov); preslav.nakov@mbzuai.ac.ae (P. Nakov); shangsong.liang@mbzuai.ac.ae (S. Liang)

the voting preferences of undecided individuals at least by 20% [4]. Therefore, detecting the political leaning of news articles and news media has become crucial for researchers, journalists, and policymakers [5].

One of the biggest challenges in detecting the political leaning of news articles and news media is the lack of a standardized approach [6]. Researchers and journalists use various methods to determine the political leaning of news and news outlets, such as manual coding, content analysis, and sentiment analysis[7, 8]. However, these methods have limitations, including subjectivity, biases, and the inability to detect nuanced political perspectives. Moreover, the increasing use of social media and online news platforms has made it even more challenging to detect the political leaning of news. With the rise of user-generated content, identifying the political orientation of a particular news article or media outlet has become more complex [9, 10, 11].

The `CheckThat!` lab CLEF 2023 [12, 1, 13] has initiated several tasks aimed at contributing to the scientific community. In `CheckThat!` lab CLEF 2023, task 3 [1] seeks to solve the problem of detecting political bias at both the article and medium levels. To address this issue, we propose a CatBoost framework to predict political bias, and we present the results of our model. Our research contributes to the scientific community by proposing a new system for predicting political bias in news articles and news media. We believe that our study will contribute to advancing the field of political bias detection.

2. Related Work

Prior research on detecting ideological biases has been emphasized in several studies [14, 15, 16, 17, 18, 19, 6, 20, 21, 22, 23, 24, 25, 26, 27].

Gentzkow and Shapiro in 2010 [17] slant index was the initial attempt to rate the ideological stance of news providers based on the frequency of partisan phrases or co-allocations used in news content. Lin et al. [18] proposed a statistical framework to identify the perspective from which a document is written with high accuracy. However, their insufficient dataset restricted the use of contemporary deep learning techniques. [14] determined the political stance displayed by a text by using a recursive neural network (RNN) framework. Similarly, [16] examined the selection and framing of political problems in fifteen significant US news organizations using machine learning and crowd-sourcing.

Recent advances in predicting the political ideology of news media and news articles have leveraged various aspects such as media stance, factuality, and media profile [28, 8, 7, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38].

In recent years, Kulkarni et al. [31] used an attention-based multi-view model, leveraging cues from neural inference and natural language processing, to identify news article ideology, achieving higher performance than existing models. However, their method has some potential issues. Horne et al. [32] introduced the News Landscape (NELA) Toolbox. This open-source toolkit allows for investigating news veracity using content-based indicators as a step toward automated news credibility research. Kiesel et al. [33] created a large-scale dataset for hyper-partisan news detection and organized a successful SemEval shared task, with the top team achieving a high accuracy rate. Potthast et al. [34] used a meta-learning approach called unmasking to evaluate style similarity between text categories, revealing significant commonalities and differences between news types, but found it inadequate for fake news identification. Rashkin et al. [35] conducted a study on the language of news media, contrasting the language of legitimate news with satire, hoaxes, and propaganda, and highlighted the potential of stylistic clues in assessing text veracity. Jiang et al. [36] developed a system using averaged word embeddings from a pre-trained ELMo model, achieving first place in a hyper-partisan news detection challenge. Barron et al. developed a model to identify propagandistic material in articles, highlighting the effectiveness of character n-grams and other style criteria over word n-grams while discussing the drawbacks of distant supervision. Martino et al. [37] proposed identifying all fragments containing propaganda techniques in a text and developed a corpus of manually annotated news articles for this purpose, demonstrating the effectiveness of a novel multi-granularity neural network.

Dinkov et al. [39] developed a multimodal deep-learning architecture to predict the political ideology of news media by studying YouTube channels, which resulted in an extensive multi-modal dataset. Another approach by [8] involves assessing each document's stance towards a claim and predicting the claim's factuality while also taking into account the source's credibility. The same corpus was annotated to reflect the interdependencies between these tasks, yielding an advanced Arabic fact-checking corpus. The political bias and factual accuracy of news media have also been studied by [29], where posture detection has been introduced as a crucial part of fact-checking systems. This study by [28] employed information from various sources, such as media-produced pieces, Wikipedia pages, Twitter profile metadata, and online features. Lastly, the issue of assessing the bias and factuality of online news sources has been studied by [38]. The researchers modeled the similarity between media outlets based on audience overlap, contrasting with text-based approaches. The resulting inter-media connectivity graph, processed through graph neural networks, led to better predictions of the factuality and bias of news media sources when supplemented with pre-computed representations from various platforms.

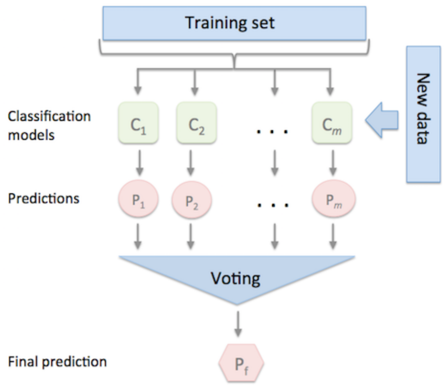
In conclusion, prior and recent research have explored various approaches for predicting the political ideology of news media, examining the general stance. Studies found that joint prediction of factuality and political bias proved more advantageous than predicting each separately [8, 28, 38]. However, more than traditional bias detection methods are required to achieve highly accurate results, and a more nuanced and interdisciplinary approach is necessary for future research. As such, there is a need to undertake further research to develop and quantify more recent studies on detecting ideological biases to enhance the fundamental background of bias detection methods.

3. Method

In the following section, we outline our methodological approach.

TF–IDF. We used the Term Frequency-Inverse Document Frequency (TF-IDF) technique to vectorize our text data [40]. This method calculates the frequency of each term in each article and assigns a weight to each term based on its importance in the article and its frequency in the dataset. This allowed us to feed our textual data into our models and obtain a numerical

Figure 1: Majority voting architecture. Source: www.researchgate.net



representation of the articles.

K–means. We used the K-means algorithm to cluster similar articles based on their political bias [41]. This helped us better understand the distribution of political ideologies in our dataset and enabled us to identify any outliers or anomalies in the data.

CatBoost. CatBoost [42] is an open-source machine learning library developed by Yandex, designed explicitly for gradient boosting on decision trees. It offers a highly efficient and accurate approach to handling categorical features, leveraging a special algorithm to avoid target leakage, and provides numerous options for model interpretation. Its advantages include fast prediction times, robust handling of categorical variables, and a feature that allows it to handle missing data.

Majority voting. We use the majority voting technique as another approach, which is a method employed in ensemble learning [43, 44] by aggregating the predictions made over the news by a single medium at medium level. Figure 1 demonstrates the architecture of our proposed majority voting approach. It determines the final prediction for a given data point by selecting the class or outcome that receives the majority of votes from the ensemble models.

Given a set of m classifiers, $C_1, C_2, \ldots, C_m$, and an input data point x, each classifier makes a prediction for the class label: $P_1, P_2, \ldots, P_n$. The majority voting classifier determines the final output class label P_f by selecting the class with the most votes from the individual classifier predictions.

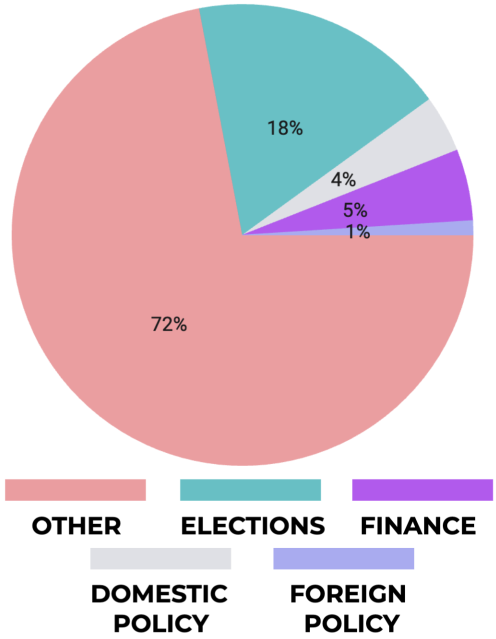
Mathematically, the majority voting classifier can be represented as:

$$P_f = \text{mode}(P_1, P_2, \ldots, P_m) \tag{1}$$

In this Equation 1, the mode function returns the class label that appears most frequently among the individual classifier predictions, where P_f is the final prediction.

For our task, we used hard voting, as we were afraid that the classifiers were not well calibrated, e.g., they can be over-confident in their decisions even when they are wrong; put another way, the classifier might not know when it does not know, and thus their output probability might not be usable directly. Thus, we opted for majority voting. We leave calibration and subsequent soft-voting for future work.

Figure 2: Distribution of topics in news articles (Subtask 3A).



4. Datasets

4.1. Political Bias of News Articles (Subtask 3A)

The dataset for this study includes a comprehensive and diverse collection of news articles from different news agencies collected from Allsides[1].

Data Attributes

For each news article, the following information is available:

- ID: A unique identifier for the article.
- Title: The headline of the article.
- Content: The full text of the article.
- Label: The political leaning of the news article as left, center, or right.

Data Size: The dataset contains a substantial number of articles from each political leaning. In total, we have over 55K articles in a dataset.

In Figure 2, we observe the notable wide range of topic distribution in the dataset. Our dataset has a broad range of topics, including elections, domestic and foreign policy, finance, and others, with "elections" and "other" being the most common topics. Also, Figure 3 shows the class label distribution and the number of articles in each subset. We can see that there are more articles with a right-leaning stance in almost all subsets, with fewer articles in the left-leaning category. Meanwhile, the center class remains consistently in the middle in terms of size.

[1]www.allsides.com

Figure 3: Statistics about the label distribution in the dataset's training, development, and test parts Subtask 3A.

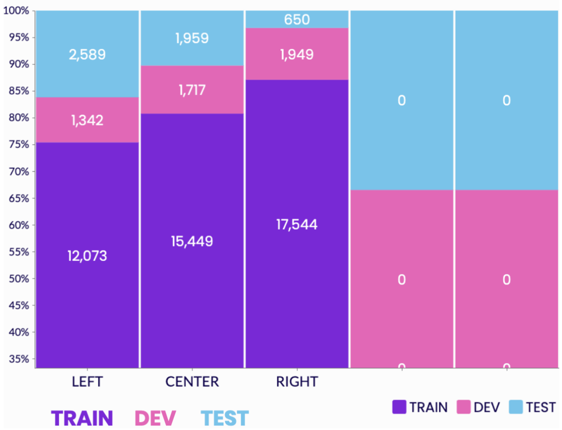


Table 1
Examples of news outlets and their biases.

Left	Center	Right
The Guardian	BBC News	Fox News
The New York Times	Reuters	The Daily Caller
The Washington Post	The Associated Press	The National Review

4.2. Political Bias of News Media (Subtask 3B)

We use a dataset for assessing the political bias of English-language media, sourced from `CheckThat!` lab CLEF 2023 which has been crawled from Media Bias/Fact Check[2]. Table 1 presents examples of news outlets and their corresponding biases.

Data Attributes

This dataset have similar attributes to subtask 3 A, plus the source (name of the medium) as an additional attribute.

Dataset size: We have over 1'000 news media outlets and around 8'000 articles approximately 10 articles per each source.

Figure 4 shows the label distribution for the articles from these media across the three subsets. We can see that the distribution is once again relatively balanced, with similar numbers of instances for each label in each subset. This is important as it ensures that the model trained on the train set will have no biases towards any particular label and will be able to generalize well to the development and the test sets.

Figure 5 shows that over 800 media are used for training set, about 100 for dev, and slightly over 100 for testing. As illustrated in Figure 3, we have a modestly imbalanced data towards the left-leaning news outlets, while the distribution of the other two classes are almost equal.

[2]www.mediabiasfactcheck.com

Figure 4: Statistics about the number of articles of each bias across all news outlets (Subtask 3B).

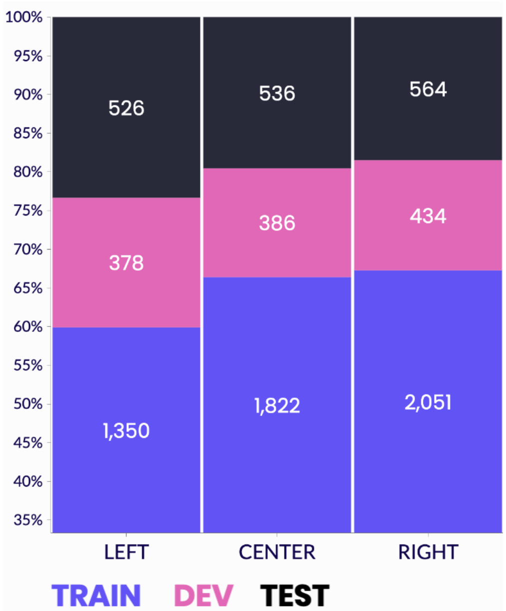


Figure 5: Statistics about the label distribution for news outlets across the dataset splits (Subtask 3B).

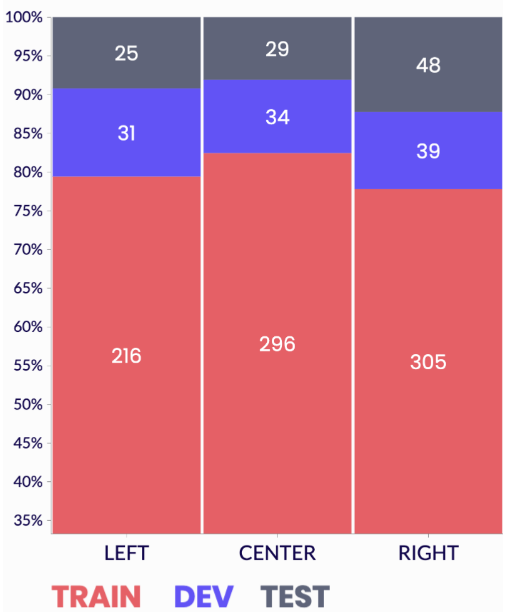


Table 2
Performance of our models at article-level and medium-level dataset.

Task	Model	MAE	F1 Score	Accuracy
Political Bias of News Articles	Baseline	0.877		
Political Bias of News Articles	CatBoost	**0.270**	**0.690**	**0.694**
Political Bias of News Media	Baseline	0.902		
Political Bias of News Media	CatBoost	**0.320**	**0.620**	**0.621**
Political Bias of News Media	CatBoost OF	0.375	0.537	0.538
Political Bias of News Media	Majority voting (CatBoost)		**0.727**	**0.725**
Political Bias of News Media	Majority voting (CatBoost OF)		0.621	0.625

5. Experiments and Results

5.1. Experimental Setup

In this subsection, we provide a detailed description of the experimental setup for our models.

Baseline. The default experimental setup has been used for SVM model with a linear kernel.

CatBoost. For the CatBoost model, we considered the following hyperparameters: learning rate 0.1, depth 6, the maximum number of boosting iterations rate set to 10000, the best model from all iterations is selected as the final model, and the frequency of logging information during a training set to 500, which means that the training progress will be printed every 500 iterations. At medium-level, we modified the number of iterations to 2000, the learning rate to 0.05, and set the progress to log iterations to 100.

CatBoost OF. In the case of the CatBoost model using only the first 300 most important features from the TF-IDF, the experimental setup is slightly modified. The model is trained using only the top 300 most important features derived from the TF-IDF representation of the dataset, focusing on the most relevant information for the classification task. The number of iterations is set to 1000, which means the model trained for 1000 boosting rounds. This parameter determines the maximum number of trees that can be built by the model. The learning rate is set to 0.05, controlling the contribution of each tree to the ensemble model. A lower learning rate generally results in a more robust model, albeit at the expense of longer training times. The frequency of logging the training progress is set to 500 iterations, meaning that the training progress will be logged every 500 iterations, providing less frequent updates compared to the previous setup.

5.2. Results

In this subsection, we present the results of our experiments on predicting the political bias of news articles and news outlets. We applied a majority voting ensemble approach to each model by aggregating the predictions made over the news by a single medium. The performance of the individual models and the majority voting ensemble is summarized in Table 2.

Baseline. The task completion's foundational system was provided by the organizers. A traditional machine learning approach Support Vector Machine (SVM) was selected for this purpose.

Figure 6: Histogram of the article length in terms of symbols for our dataset.

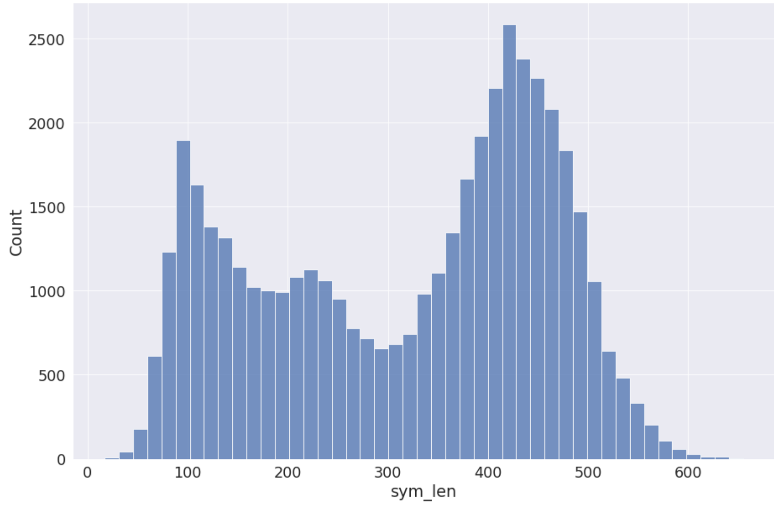
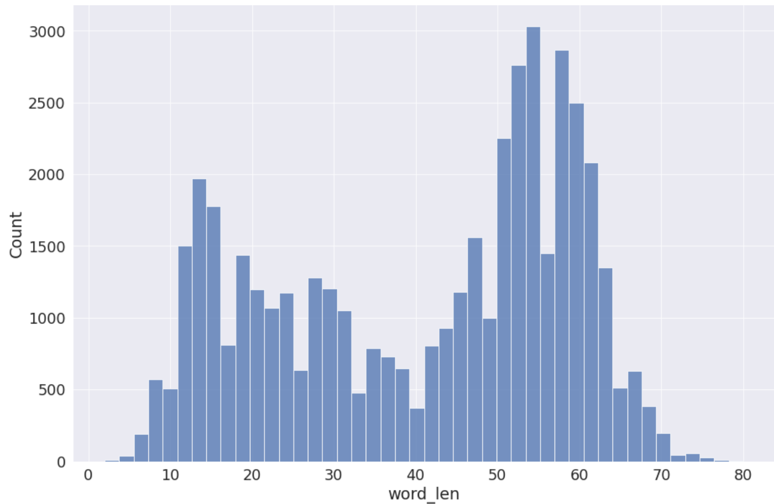


Figure 7: Histogram of the article length in terms of words for our dataset.

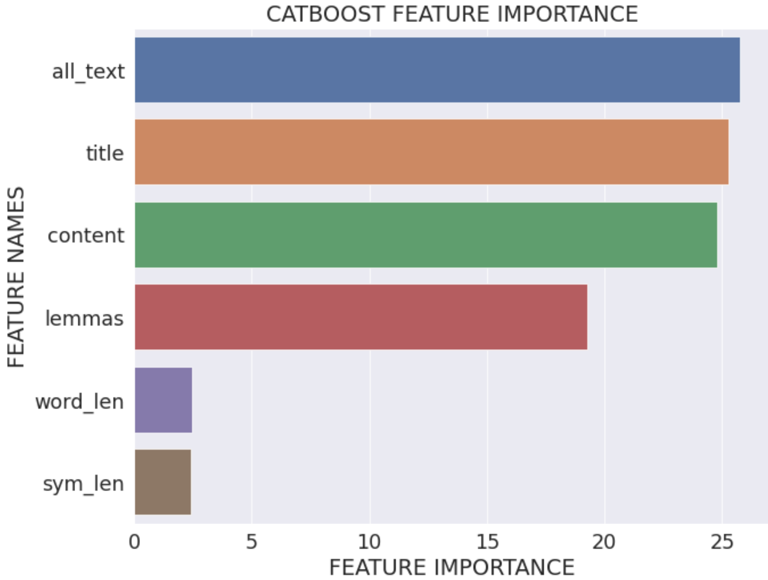


CatBoost (Subtask A). This model consistently yielded the best results in our experiments. In terms of MAE model reached 0.270 an F1 score 0.620 and an accuracy 0.621. This can be attributed to its robust gradient-boosting algorithm and optimized implementation for decision tree learning. Moreover, the algorithm incorporates built-in regularization techniques that help prevent overfitting and improve generalization, which is crucial for maintaining high performance on unseen data.

Figure 6 shows a plot illustrating the distribution of article lengths in terms of dataset characters. Our dataset exhibits a long-tailed distribution, with most articles having symbolic lengths peaking at 100–150 and 350–500 characters.

Similarly, Figure 7 shows the distribution of article lengths in terms of words in our dataset,

Figure 8: CatBoost feature importance on our dataset.



albeit with slightly lower numbers. In our dataset, the majority of the articles have word counts within the ranges of 10–30 and 50–65 words.

The relationship between symbolic count and word count is evident; a higher symbolic length generally corresponds to a higher word count. This observation underscores the interdependence between these two factors in the structure and the content of the articles in our dataset.

Next, we analyzed the feature importance of CatBoost for classifying the political bias in news articles using our dataset. We considered five features: symbolic length, word length, lemmas, content, title, and all text. The results are shown in Figure 8. Our observations indicate that word and symbolic length were not significant features of the model. In our machine learning model, we computed feature importance using the 'SelectPercentile' method and Chi^2 criterion. We ranked features by Chi^2 scores, indicating their relevance to the target variable.

For our dataset, all text emerged as the most important feature, followed by title, content, and lemmas. Our analysis demonstrates that feature importance varies between them. It is important to note that using all text as a feature when training the model on our data accelerates the learning process.

Next, we computed a confusion matrix, which provides a convenient way to assess our model's classification performance by presenting the number of correct and incorrect predictions for each class in a tabular format, allowing us to identify patterns and areas where the model may struggle. As shown in Figure 9, the confusion matrix for our model predictions on our dataset reveals distinct differences in performance across the political bias categories.

In our dataset, the model exhibits a substantial performance in predicting center- and right-leaning articles, whereas it struggles to classify left-leaning ones accurately.

We further investigated the performance of CatBoost a per-class level on our dataset, and the results are shown in Table 3.

Figure 9: Confusion matrix for CatBoost on the test set.

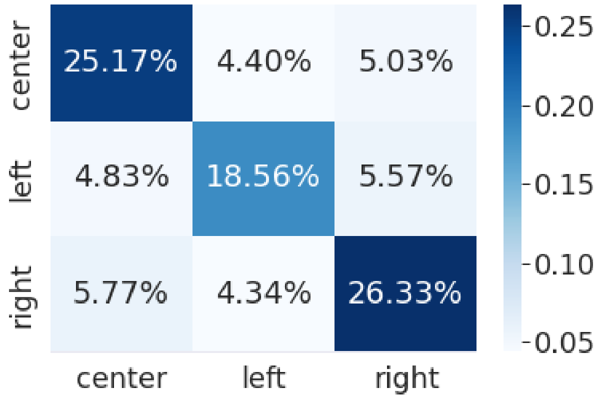


Table 3
Per-class results for CatBoost.

Class	**Precision**	**Recall**	**F1-score**	**Support**
center	0.61	**0.83**	**0.70**	2,959
left	**0.78**	0.36	0.49	2,589
right	0.37	0.75	0.49	650
Accuracy			0.59	5,198
Macro Avg	0.58	0.65	0.56	5,198
Weighted Avg	0.66	0.59	0.57	5,198

Table 3 shows the classification report for CatBoost. The model achieves an overall accuracy of 0.59, with a macro-average F1-score of 0.56 and a weighted F1-score of 0.57. It performs best in classifying center-leaning articles with an F1-score of 0.70, while it has lower F1-scores of 0.49 for both left- and right-leaning articles. Support refers to the number of actual occurrences of the class in the specified dataset. In our case, it means the number of instances for each class ("Center", "Left", and "Right") present in the dataset. For example, there are 2959 instances of "Center", 2589 instances of "Left", and 650 instances of "Right". For "Accuracy", "Macro Avg", and "Weighted Avg", the support is the total number of instances, which is 5198 in this case.

CatBoost (Subtask B). At medium-level CatBoost also consistently achieved superior results throughout our experiments. Moreover, it reached MAE with a score of 0.320 and after applying majority voting we have solid increase in terms of an F1 score and an accuracy which are 0.727 and 0.725, respectively. The confusion matrix in Figure 10 for the test set of the CatBoost model reveals that the model demonstrates strong predictive performance for center-leaning and right-leaning categories. However, its accuracy in predicting left-leaning instances is comparatively lower than that of the other classes.

In Table 4 CatBoost model demonstrates the highest precision for the left class at 0.64, followed closely by the center class at 0.63. The right class exhibits a slightly lower precision of 0.60. In terms of recall, the right class outperforms the others with a value of 0.68. The left and

Figure 10: Confusion matrix of CatBoost on test set.

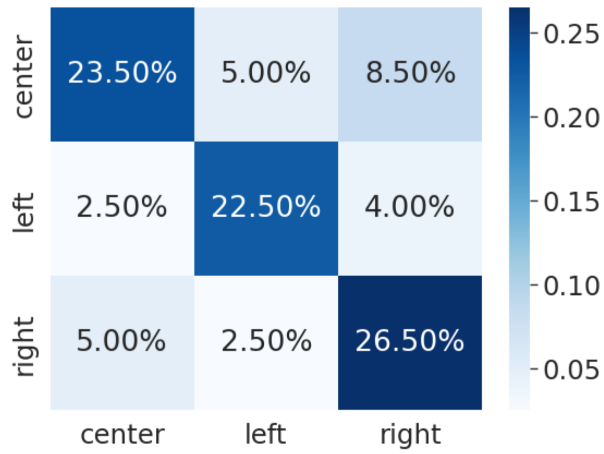
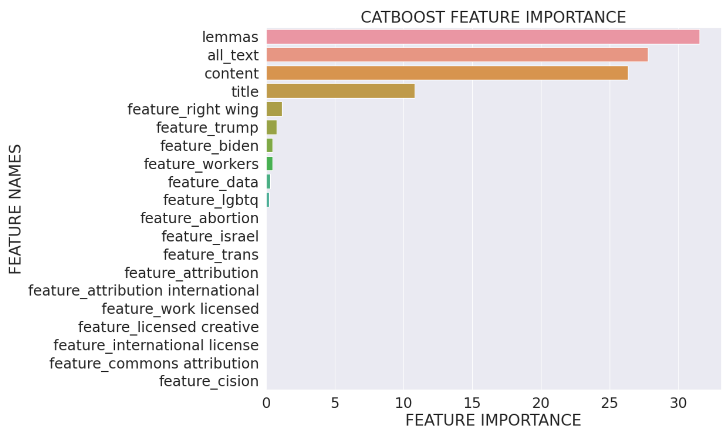


Figure 11: CatBoost feature importance.



center classes show recall values of 0.60 and 0.58, respectively. The F1-score, which balances precision and recall, indicates that the left class achieves a slightly better score of 0.62 compared to the center and right classes, which have F1-scores of 0.60 and 0.63, respectively. The overall accuracy of the CatBoost model stands at 0.62, while the macro and weighted averages for precision, recall, and F1-score are also equal to 0.62.

CatBoost OF. The CatBoos OF model incorporates the top 300 most important features derived from the Term Frequency-Inverse Document Frequency (TF-IDF) method. Despite a relatively modest Mean Absolute Error (MAE) of 0.375, an F1 score of 0.537, and an accuracy of 0.538, the model exhibited less favorable results compared to other CatBoost models, securing the bottom position. This held true even when a majority voting approach was applied to enhance the performance of our model.

Table 4
Per-class results for the Catboost.

Class	Precision	Recall	F1-score	Support
Center	**0.63**	0.58	0.60	526
Left	**0.64**	0.60	0.62	536
Right	0.60	**0.68**	0.63	564
Accuracy			0.62	1626
Macro Avg	0.62	0.62	0.62	1626
Weighted Avg	0.62	0.62	0.62	1626

Table 5
Performance of CatBoost vs. BERT with random vs. balanced splits on our article-level dataset.

Model	Split Type	MAE	F1 Score	Accuracy
CatBoost	Random	0.322	0.585	0.586
CatBoost	Balanced	**0.265**	**0.703**	**0.705**
BERT	Random	0.385	0.491	0.499
BERT	Balanced	0.351	0.565	0.569

6. Discussion

During the initial stage of our analysis, we observed that our dataset exhibited an unbalanced distribution of classes. To investigate the impact of balanced splits on model performance, we created train, test, and development sets with balanced class distributions. Our findings revealed that using balanced splits led to improved results, as shown in Table 5. This suggests that balancing the dataset can contribute to more accurate and reliable predictions. Upon conducting a comparative analysis of the BERT and CatBoost models, we have chosen to implement the CatBoost model in our experiments. This decision was guided by its demonstrably superior performance compared to BERT.

7. Conclusion and Future Work

We explored predicting the political bias of news articles and news media using a dataset comprising 55,000 news articles and over 1,000 news media sources with over 8,000 articles. We proposed a new model, CatBoost, which was used to determine the political leaning. Furthermore, we employed the majority voting technique to enhance our model's performance at the media level.

Our approach effectively classified the political bias, yielding consistent results. The CatBoost model trained on our article-level dataset achieved a classification accuracy of 0.690, an F1 score of 0.694, and a MAE of 0.270. When the model was applied to the medium-level, it reached an MAE of 0.320.

By implementing the majority voting classifier, which aggregates the predictions made over the news by a single medium, we achieved an enhanced F1 score of 0.727 and an accuracy of 0.725. Our comprehensive experiments showed that CatBoost consistently performed effectively. However, the CatBoost OF model delivered the least effective results on the media-level dataset. We noticed that applying majority voting to the news from a single medium improved each model's performance.

In future work, we aim to explore topic-level bias prediction and move beyond the left-center-right political bias classification. This may require collecting additional labels and breaking away from the current 3-way classification. Furthermore, we intend to carry out cross-language experiments to adapt these methods to languages other than English. Another potential area of investigation is predicting veracity and political leaning simultaneously, necessitating the development of models incorporating both textual and non-textual features, such as source credibility or author reputation.

References

[1] G. Da San Martino, F. Alam, M. Hasanain, R. N. Nandi, D. Azizov, P. Nakov, Overview of the CLEF-2023 CheckThat! lab task 3 on political bias of news articles and news media, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[2] A. Barrón-Cedeño, F. Alam, T. Caselli, G. Da San Martino, T. Elsayed, A. Galassi, F. Haouari, F. Ruggeri, J. M. Struss, R. N. Nandi, G. S. Cheema, D. Azizov, P. Nakov, The clef-2023 checkthat! lab: Checkworthiness, subjectivity, political bias, factuality, and authority, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2023, pp. 506–517.

[3] G. Pennycook, D. G. Rand, The psychology of fake news, Trends in cognitive sciences 25 (2021) 388–402.

[4] G. Gezici, Quantifying political bias in news articles, arXiv preprint arXiv:2210.03404 (2022).

[5] R. R. R. Gangula, S. R. Duggenpudi, R. Mamidi, Detecting political bias in news articles using headline attention, in: Proceedings of the 2019 ACL workshop BlackboxNLP: analyzing and interpreting neural networks for NLP, 2019, pp. 77–84.

[6] P. Nakov, H. T. Sencar, J. An, H. Kwak, A survey on predicting the factuality and the bias of news media, arXiv preprint arXiv:2103.12506 (2021).

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

[8] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, P. Nakov, Predicting factuality of reporting and bias of news media sources, arXiv preprint arXiv:1810.01765 (2018).

[9] D. Saez-Trumper, C. Castillo, M. Lalmas, Social media news communities: gatekeeping, coverage, and statement bias, in: Proceedings of the 22nd ACM international conference on Information & Knowledge Management, 2013, pp. 1679–1684.

[10] K. Darwish, P. Stefanov, M. Aupetit, P. Nakov, Unsupervised user stance detection on

Twitter, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 14, 2020, pp. 141–152.
[11] R. Baly, G. Karadzhov, J. An, H. Kwak, Y. Dinkov, A. Ali, J. Glass, P. Nakov, What was written vs. who read it: News media profiling using text analysis and social media context, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL '20, Association for Computational Linguistics, Online, 2020, pp. 3364–3374. URL: https://aclanthology.org/2020.acl-main.308. doi:10.18653/v1/2020.acl-main.308.
[12] F. Alam, A. Barrón-Cedeño, G. S. Cheema, S. Hakimov, M. Hasanain, C. Li, R. Míguez, H. Mubarak, G. K. Shahi, W. Zaghouani, P. Nakov, Overview of the CLEF-2023 CheckThat! lab task 1 on check-worthiness in multimodal and multigenre content, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.
[13] P. Nakov, F. Alam, G. Da San Martino, M. Hasanain, R. N. Nandi, D. Azizov, P. Panayotov, Overview of the CLEF-2023 CheckThat! lab task 4 on factuality of reporting of news media, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.
[14] M. Iyyer, P. Enns, J. Boyd-Graber, P. Resnik, Political ideology detection using recursive neural networks, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014, pp. 1113–1122.
[15] J. H. Gross, B. Acree, Y. Sim, N. A. Smith, Testing the etch-a-sketch hypothesis: a computational analysis of mitt romney's ideological makeover during the 2012 primary vs. general elections, in: APSA 2013 Annual Meeting Paper, American Political Science Association 2013 Annual Meeting, 2013.
[16] C. Budak, S. Goel, J. M. Rao, Fair and balanced? quantifying media bias through crowd-sourced content analysis, Public Opinion Quarterly 80 (2016) 250–271.
[17] M. Gentzkow, J. M. Shapiro, What drives media slant? evidence from us daily newspapers, Econometrica 78 (2010) 35–71.
[18] W.-H. Lin, T. Wilson, J. Wiebe, A. G. Hauptmann, Which side are you on? identifying perspectives at the document and sentence levels, in: Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X), 2006, pp. 109–116.
[19] W.-H. Lin, E. Xing, A. Hauptmann, A joint topic and perspective model for ideological discourse, in: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part II 19, Springer, 2008, pp. 17–32.
[20] K. R. Canini, B. Suh, P. L. Pirolli, Finding credible information sources in social networks based on content and social structure, in: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, IEEE, 2011, pp. 1–8.
[21] C. Castillo, M. Mendoza, B. Poblete, Information credibility on Twitter, in: S. Srinivasan, K. Ramamritham, A. Kumar, M. P. Ravindra, E. Bertino, R. Kumar (Eds.), Proceedings of the 20th International Conference on World Wide Web, WWW, 2011, pp. 675–684.
[22] J. Ma, W. Gao, Z. Wei, Y. Lu, K.-F. Wong, Detect rumors using time series of social context information on microblogging websites, in: Proceedings of the 24th ACM international on conference on information and knowledge management, 2015, pp. 1751–1754.

[23] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K. Wong, M. Cha, Detecting rumors from microblogs with recurrent neural networks, in: S. Kambhampati (Ed.), Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016, 2016, pp. 3818–3824.
[24] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, P. Tolmie, Analysing how people orient to and spread rumours in social media by looking at conversational threads, PloS one 11 (2016) e0150989.
[25] S. Iyengar, K. S. Hahn, Red media, blue media: Evidence of ideological selectivity in media use, Journal of communication 59 (2009) 19–39.
[26] S. DellaVigna, E. Kaplan, The fox news effect: Media bias and voting, The Quarterly Journal of Economics 122 (2007) 1187–1234.
[27] A. Simonov, S. K. Sacher, J.-P. H. Dubé, S. Biswas, The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic, Technical Report, National Bureau of Economic Research, 2020.
[28] R. Baly, G. Karadzhov, A. Saleh, J. Glass, P. Nakov, Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media, arXiv preprint arXiv:1904.00542 (2019).
[29] R. Baly, M. Mohtarami, J. Glass, L. Màrquez, A. Moschitti, P. Nakov, Integrating stance detection and fact checking in a unified corpus, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), New Orleans, Louisiana, 2018, pp. 21–27.
[30] K. Popat, S. Mukherjee, J. Strötgen, G. Weikum, Credeye: A credibility lens for analyzing and explaining misinformation, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 155–158.
[31] V. Kulkarni, J. Ye, S. Skiena, W. Y. Wang, Multi-view models for political ideology detection of news articles, arXiv preprint arXiv:1809.03485 (2018).
[32] B. D. Horne, W. Dron, S. Khedr, S. Adali, Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 235–238.
[33] J. Kiesel, M. Mestre, R. Shukla, E. Vincent, P. Adineh, D. Corney, B. Stein, M. Potthast, Semeval-2019 task 4: Hyperpartisan news detection, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 829–839.
[34] M. Potthast, M. Hagen, T. Gollub, M. Tippmann, J. Kiesel, P. Rosso, E. Stamatatos, B. Stein, Overview of the 5th international competition on plagiarism detection, in: CLEF Conference on Multilingual and Multimodal Information Access Evaluation, CELCT, 2013, pp. 301–331.
[35] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, Y. Choi, Truth of varying shades: Analyzing language in fake news and political fact-checking, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2931–2937.
[36] Y. Jiang, J. Petrak, X. Song, K. Bontcheva, D. Maynard, Team bertha von suttner at SemEval-2019 task 4: Hyperpartisan news detection using ELMo sentence representation convolutional network, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 840–844. URL: https://aclanthology.org/S19-2146. doi:10.18653/v1/S19-2146.

[37] G. D. S. Martino, S. Yu, A. Barrón-Cedeño, R. Petrov, P. Nakov, Fine-grained analysis of propaganda in news articles, arXiv preprint arXiv:1910.02517 (2019).
[38] P. Panayotov, U. Shukla, H. T. Sencar, M. Nabeel, P. Nakov, GREENER: Graph neural networks for news media profiling, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 7470–7480. URL: https://aclanthology.org/2022.emnlp-main.506.
[39] Y. Dinkov, A. Ali, I. Koychev, P. Nakov, Predicting the leading political ideology of youtube channels using acoustic, textual, and metadata information, arXiv preprint arXiv:1910.08948 (2019).
[40] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, Information processing & management 24 (1988) 513–523.
[41] M. E. Celebi, H. A. Kingravi, P. A. Vela, A comparative study of efficient initialization methods for the k-means clustering algorithm, Expert systems with applications 40 (2013) 200–210.
[42] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin, Catboost: unbiased boosting with categorical features, Advances in neural information processing systems 31 (2018).
[43] T. G. Dietterich, et al., Ensemble learning, The handbook of brain theory and neural networks 2 (2002) 110–125.
[44] O. Sagi, L. Rokach, Ensemble learning: A survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8 (2018) e1249.