Identifying Political Bias in Arabic News Articles

Saoussen Ben Chaabane[1,3], Omar Trigui[2,3], Maher Jaoua[1,3]

[1] Faculty of Economic Sciences and Management of Sfax (FSEGS), University of Sfax, Tunisia,
[2] The Higher Institute of Management of Sousse (ISGS), University of Sousse, Tunisia,
[3]ANLP Research Group, MIRACL laboratory, Sfax, Tunisia
saoussenbenchabane@isgs.u-sousse.tn, omar.trigui@isgs.u-sousse.tn, maher.jaoua@fsegs.usf.tn

Abstract

A comprehensive framework was developed to detect political bias in Arabic news articles, with a case study focusing on media reporting of the Palestinian issue. The methodology integrates MARBERT contextual embeddings with classical and deep learning classifiers, including SVM, Logistic Regression, Random Forest, and LSTM. The scalability of data processing was ensured through Apache Spark for potential real-time deployment. Experimental results showed that fine-tuned MARBERT embeddings combined with LSTM achieved the highest classification accuracy of 0.87, along with notable improvements in F1-scores across the pro, against, and neutral categories. These findings highlight the effectiveness of domain-specific fine-tuning of transformer models for political bias classification. The study also addressed class imbalance using SMOTE and class weighting strategies, and assessed feature robustness using multiple vectorization techniques.

Keywords: Political bias detection in Arabic news, Word embeddings

1. Introduction

In an age where information is abundant and narratives shape perceptions, political bias in media has become a pressing issue, particularly in regions marked by deep-rooted geopolitical conflict. Bias whether explicit or implicit frames ideologies, distorts public opinion, and deepens divides. This challenge is especially salient in the Arabic media landscape, where coverage of sensitive topics such as the Palestinian cause is often influenced by partisan rhetoric, sociocultural factors, and algorithmic amplification. Detecting bias in Arabic-language media presents unique difficulties due to linguistic complexity, dialectal diversity, and the scarcity of labeled data. Despite global advances in natural language processing (NLP), political bias detection in Arabic remains underexplored. Addressing this gap requires approaches that are both linguistically sensitive and computationally scalable. To this end, the present study proposes a domain-aware framework for detecting political bias in Arabic news articles. The framework leverages MARBERT embeddings combined with deep learning classifiers, ensuring robust performance in recognizing nuanced biases. Through rigorous evaluation, the study demonstrates the superior performance of fine-tuned MARBERT models in capturing subtle rhetorical and emotional cues. The Palestinian case serves as a focal point, illustrating how political leanings manifest across social and institutional dimensions. This contribution not only advances Arabic NLP research but also provides practical insights for automated bias detection in politically sensitive contexts.

2. Background

Political bias is when someone consistently leans toward or against a particular ideology, party, or policy, instead of staying neutral. It shows up in the way news is reported, how institutions make decisions, and even in personal judgments. This bias can shape how the public sees issues and weaken fair, balanced discussion. As D'Alessio and Allen (Dave D'Alessio, Mike Allen, 2000) put it, political bias is "the systematic deviation from objective neutrality in the treatment of political information, where coverage or analysis disproportionately favors one party, ideology, or policy outcome."

2.1 Origins of Political Bias

Political bias has many roots, shaped by our biology, psychology, social environment, and the systems we live in. At its core, it comes from survival instincts and tribal loyalty, which make us favor our own group. Our minds also rely on shortcuts like confirmation bias, where we seek out information that supports what we already believe, or authority bias, where we trust those in power. These tendencies are reinforced by family, education, religion, and peers, and amplified by media algorithms that thrive on polarizing content. On a larger scale, the way political elites frame issues, the concentration of media ownership, and partisan systems push society further into ideological camps. Over time, this makes political identity so deeply tied to personal identity that the two become almost inseparable.

2.2 Types of Political Bias

Political bias has many roots, shaped by our biology, psychology, social environment, and the systems we live in. At its core, it comes from survival instincts and tribal loyalty, which make us favor our own group. Our minds also rely on shortcuts like confirmation bias, where we seek out information that supports what we already believe, or authority bias, where we trust those in power. These tendencies are reinforced by family, education, religion, and peers, and amplified by

Proceedings of The 3rd Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2026) @ LREC 2026, pages 228–233
16 May 2026, Palma, Mallorca (Spain).

media algorithms that thrive on polarizing content. On a larger scale, the way political elites frame issues, the concentration of media ownership, and partisan systems push society further into ideological camps. Over time, this makes political identity so deeply tied to personal identity that the two become almost inseparable.

2.3 Political Bias in the Palestinian Case

The Palestinian issue shows how political bias plays out across many levels biological, psychological, social, media, and political. Different narratives whether pro-Palestinian, anti-Palestinian, or neutral reveal the full range of bias, from emotionally charged reporting to selective framing or even leaving things out. This makes the case a powerful example for studying how bias appears and how it can be detected, especially in Arabic media.

3. Related work

Research on political bias detection has evolved across multiple domains, ranging from traditional news analysis to multimodal approaches. Early work by Azizov et al (Dilshod Azizov, Preslav Nakov and Shangsong Liang, 2023). applied CatBoost classifiers with TF-IDF and stylistic features to categorize news articles into left, center, and right orientations, achieving moderate accuracy but facing challenges related to class imbalance and limited representation of smaller media outlets. Tran et al (Sieu Tran, Paul Rodrigues, Benjamin Strauss and Evan M. Williams, 2023) extended this line of research by leveraging RoBERTa with data augmentation techniques such as back-translation, reporting improved robustness but sensitivity to annotation biases.

Transformer-based and zero-shot approaches have also gained traction. Türkman et al (M.D. Coşgun, G. Kutlu, Mücahid Türkmen, 2023). explored the use of ChatGPT for zero-shot and few-shot classification, achieving promising results across German and English datasets, though performance varied significantly by language. Jiang et al (Ye Jiang, Johann Petrak, Xingyi Song, Kalina Bontcheva, Diana Maynard, 2019). combined ELMo embeddings with convolutional neural networks to detect hyperpartisan news, reaching over 82% accuracy but remaining sensitive to training data biases.

Beyond textual analysis, multimodal approaches have broadened the scope of bias detection. Dinkov et al (Yoan Dinkov, Ahmed Ali, Ivan Koychev, Preslav Nakov, 2019). investigated political ideology in YouTube channels using acoustic, textual, and metadata features, achieving a 6% improvement over text-only models but relying heavily on external annotations. Lin et al (Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, Alexander Hauptmann, 2006). applied statistical models to articles on the Israeli Palestinian conflict, successfully identifying political perspectives but limited to a single geopolitical context.

Ben Amor et al (Ghada Ben Amor, Nawres Medimagh, Saoussen Ben Chaabene, and Omar Trigui, 2025) addressed bias detection in social media by focusing on sexism in Spanish tweets. Their system combined contextual embeddings from the multilingual T5 (mT5) model with both traditional machine learning classifiers (Logistic Regression, SVM) and deep learning architectures (RNN, GRU, hybrid models). They emphasized preprocessing, dimensionality reduction, and class balancing techniques to improve robustness, highlighting methodological parallels with political bias detection in Arabic news articles.

In the Arabic media landscape, research remains relatively scarce. Darwish et al (Saad M. Darwish, Abdul Rahman M. Sabri, Dhafar Hamed Abd, Adel A. Elzoghabi, 2024). proposed a CatBoost-based system using TF-IDF, lexicons, and linguistic patterns, achieving high accuracy (98.14%) but showing sensitivity to linguistic variation and source biases. Baly et al (Ramy Baly, Mitra Mohtarami, James Glass, Lluís Màrquez, Alessandro Moschitti, Preslav Nakov, 2018). introduced a unified corpus combining stance detection and fact-checking, limited to Arabic sources but valuable for advancing regional NLP research.

El Abed et al (Yasmine El Abed, Mariem Ben Arbia, Saoussen Ben Chaabene, Omar Trigui, 2025) tackled the detection of hate and hope speech in Arabic social media as part of the MAHED shared task. Their system combined XLM-RoBERTa embeddings with Logistic Regression and LSTM classifiers, supported by Arabic-specific preprocessing, SMOTE for class balancing, and back-translation for data augmentation. Their results highlight the challenges of handling morphologically rich languages like Arabic, particularly in imbalanced datasets, and demonstrate the effectiveness of multilingual transformers for nuanced text classification. These methodological insights are directly relevant to political bias detection in Arabic news, where similar linguistic and contextual complexities arise.

These works underscore the importance of domain-specific models tailored to Arabic linguistic and sociopolitical contexts.

4. Proposed Method

The initial stage of this study involved collecting a large corpus of Arabic news articles related to the Palestinian issue. Given the impracticality of manually annotating the entire dataset, a carefully curated subset of 218 articles was manually labeled with the support of an expert in Arabic political discourse, balanced across three categories: Pro (73), Against (73), and Neutral (72).

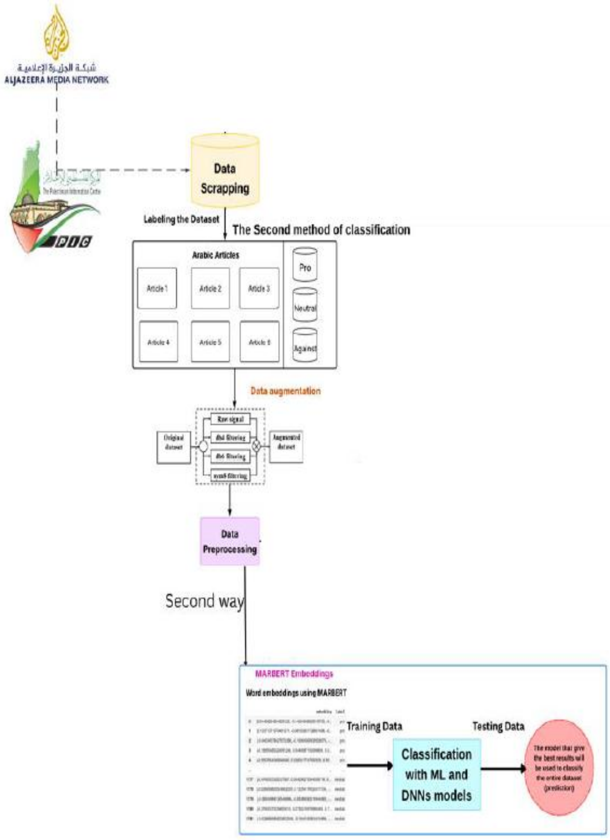

Figure 1: Classification Using MARBERT embeddings

To expand and diversify this limited dataset, several augmentation techniques were applied, including multi-step back-translation using the argostranslate library, GPT-2-based paraphrasing, summarization with the multilingual mBART model, and synonym replacement via AraBERT. These processes increased linguistic variation and lexical diversity, resulting in an augmented dataset of 1,782 articles (670 Pro, 584 Neutral, 528 Against). To capture nuanced semantic features, MARBERT a transformer pretrained on Arabic text was fine-tuned and employed to extract contextual embeddings using the [CLS] token. The embeddings were organized into a structured DataFrame with numerically encoded class labels, and the dataset was split into stratified training and testing subsets to ensure balanced representation. Multiple machine learning and deep learning models were then trained, with the best-performing classifier selected to categorize the remaining unlabeled articles into the three political stance categories: Pro-Palestine, Against, and Neutral.

Finally, to ensure scalability when processing large volumes of text data, we implement the solution using Apache Spark's distributed. computing framework.

5. Experimental Study

5.1 Data Preparation

To build a representative dataset, Arabic news articles were collected from two major outlets: Al

	embedding	label
0	[0.04466264694929123, -0.14264848828315735, -0...	pro
1	[0.12071371078491211, -0.08193931728601456, -0...	pro
2	[-0.04034578427672386, -0.10004305839538574, -...	pro
3	[-0.18855485320091248, 0.5446987152099609, 0.0...	pro
4	[-0.5857664346694946, 0.2380317747592926, 0.39...	pro
...	...	...
1777	[-0.4746353328227997, 0.044636279344558716, 0....	neutral
1778	[-0.028930820524692535, 0.12254178524017334, -...	neutral
1779	[-0.08504968136548996, -0.05386582016944885, -...	neutral
1780	[-0.37900373339653015, 0.2730216979980469, 0.7...	neutral

Figure 2: Processed data (MARBERT)

Jazeera Arabic[1] and The Palestinian Information Centre[2]. These sources were selected based on three criteria:

- The feasibility of automated extraction through web scraping,
- The availability of a large volume of Arabic content, and the relevance of the material for evaluating the generalization capacity of classification models.

The dataset covers events occurring both before and after October 7, 2023, ensuring temporal diversity and contextual richness.

Table 1 presents a structured description of the datasets collected from the Arabic news sources Al Jazeera and Palinfo. It details the key elements of the data collection process, including the source URLs, the tools used for scraping, the type of pagination implemented by each site, the number of articles extracted, the browser mode utilized, the format in which the data was stored, and the content structure of the collected articles.

Table 1: Description of the datasets

Aspect	Al Jazeera Dataset	Palinfo Dataset
Source URL	https://www.aljazeera.net/opinion/	https://palinfo.com/category/opinion
Max Articles	2,000 (script-defined limit)	All available articles

[1] https://www.aljazeera.net/opinion/

[2] https://palinfo.com/category/opinion

		scraped until exhaustion
Article Structure	• Title. • Content.	• Title. • Author. • Date. • Link. • Content.

إتمت ما فرحة ..المصالحة

Author: رزقة يوسف .د

Date: Unknown Date

Link: https://palinfo.com/news/2020/10/22/218857/

Content:

اجتماعات أرسلت لقد !المصالحة عن أمس أخبار حملته ما على اللحظة هذه في المناسب التعليق هو ربما هذا تمت، ما فرحة يا
أننا حتى تحققها، قرب عن عريضة آمالا الصحفية المؤتمرات ونشرت المصالحة، عن إيجابية رسائل وحماس فتح بين إسطنبول
الملف هذا حمل في الأحمد عزام مباحثات من أفضل كانت الرجوب مباحثات أن ظننا.
سنشهد وإننا بجدية، مسكونة المرة هذه إن قلنا حتى الرجوب، العاروري اجتماعات من تولدت جيدة جديدة روحا أن شك من ما
الوطني والمجلس والرئاسية، والتشريعية، المحلية، :الانتخابات ومرسوم المصالحة، حفل قريبا.
المصالحة، نحو التقدم على مصرون نحن قال الأول الأحمد، لعزام والآخر للرجوب، أحدهما تصريحين، حملت أمس أخبار ولكن
وأن والانتخابات، المصالحة حول ردها ترسل لم وأنها بالتراجع، حماس فيه اتهم للثاني والثاني محالة، لا العقبات على وسنتغلب
نواب مستحقات وصرف التشريعي، المجلس بشرعية الإقرار ومنها الأجواء، تهيئة ضرورة عن حدثه مرزوق أبو موسى
!عزام تعبير حدّ على الشعب،
إطار في زال ما تقدم من كان ما أن ندرك إسطنبول، لقاءات على انقضى الذي الزمن وقاربنا معا، التصريحين جمعنا إذا إننا
!تمت ما فرحة يا :آسفين قلنا لذا النظر، وجهات واختلاف المعوقات، من تعاني تزال لا العملية الأمور وأن الأمل، وبث الشكل،
،(أبيب تل) في سفارتها إقامة نحو وتقدمت ،(إسرائيل) مع صلحا عقدت المتحدة العربية الإمارات إن :للطرفين قوله أود ما
اللاءات شرع الذي السودان وهذا العربية، القدس في الجوز وادي هايتك مشروع في كبيرة أموالا ستستثمر أنها الأخبار وتحمل
:المشهورة الثلاث
فهل ،(إسرائيل) مع وتطبيع دبلوماسي وتبادل مصالحة اتفاقية توقيع نحو يسير للمصالحة، لا للاعتراف، لا الأرض، لاحتلال لا
الشعوب غضب ظل في هؤلاء ينجح لماذا !وفتح؟ حماس بين المشتركة القواسم على يتفوق مشتركة، قواسم من الدول هذه بين ما
!والشراكة؟ المصالحة على لهما الشعب تشجيع مع وحماس فتح تفشل حين في منهم،
وكلاهما واحد، ودينهم واحدة، ولغتهم الشعب، قيادة في وشركاء تاريخ، وشركاء نضال، وشركاء قضية، شركاء وفتح حماس
ولا اللغة، في ولا القضية، في لا (إسرائيل) وبين بينهم شراكة لا من وينجح يفشلون، فلماذا العدو، أتباع ومن العدو، من مستهدف
!التاريخ؟ في ولا الدين، في
أم ذاتية، لأسباب هل التراجع؟ لماذا !إسطنبول في أبدته ما رغم عن تتراجع قل أو إتفشل الفلسطينية والمصالحة تنجح، (إسرائيل)
لتدخلات إخارجية؟ يا فرحة إن شاء الله تتم.

Figure 3: Example of an article scraped

Ensuring data quality is a crucial step in Natural Language Processing (NLP) and Machine Learning (ML), particularly when dealing with Arabic text, which is characterized by morphological richness and orthographic variation. To improve model performance and reduce noise, several preprocessing operations were applied. General steps included the removal of stop words, digits, and special characters that carry little semantic weight or hinder readability. In addition, Arabic-specific techniques were implemented to address linguistic complexity: diacritics such as fatHa, damma, kasra, and sukun were stripped to reduce variability, while different forms of Alif and Ya were normalized to ensure consistency in tokenization and embedding.

5.2 Experimental Results

The class distribution across training and testing datasets for a three-class sentiment analysis task (Pro, Against, Neutral). The data shows a relatively balanced distribution, with the Pro class being slightly more prevalent than the other categories in both sets. In the training data, Pro appears 536 times compared to 467 Against and 422 Neutral instances, while the testing set follows a similar pattern with 134 Pro, 117 Against, and 106 Neutral examples.

To perform this classification, we have employed several models, including Logistic Regression, Random Forest, SVM, and LSTM, ensuring a comparative evaluation of different approaches.

5.2.1 Logistic Regression

Table 2 shows the results of the classification obtained by the Logistic Regression model.

Pro class		
Precesion	Recall	F1-score
0.81	0.83	0.82
Against class		
Precesion	Recall	F1-score
0.76	0.75	0.75
Neutral class		
Precesion	Recall	F1-score
0.81	0.80	0.81
Accuracy: 0.80		

Table 2: Results of logistic regression

5.2.2 Random Forest

Table 3 shows the results of the classification obtained by the Random Forest model.

Pro class		
Precesion	Recall	F1-score
0.78	0.95	0.86
Against class		
Precesion	Recall	F1-score
0.88	0.72	0.79
Neutral class		
Precesion	Recall	F1-score
0.83	0.77	0.80
Accuracy: 0.82		

Table 3: Results of Random Forest

5.2.3 SVM

Table 4 shows the results of the classification obtained by the SVM model.

Pro class		
Precesion	Recall	F1-score
0.84	0.80	0.82
Against class		
Precesion	Recall	F1-score
0.72	0.79	0.76
Neutral class		
Precesion	Recall	F1-score
0.82	0.79	0.81
Accuracy: 0.80		

Table 4: Results of SVM

5.2.4 LSTM

Table 5 shows the results of the classification obtained by the LSTM model.

Pro class		
Precesion	Recall	F1-score
0.86	0.93	0.89
Against class		
Precesion	Recall	F1-score
0.86	0.83	0.85
Neutral class		
Precesion	Recall	F1-score
0.89	0.84	0.86
Accuracy: 0.87		

Table 5: Results of LSTM

The tables show a clear comparison of how models perform using fine-tuned MARBERT embeddings, and some interesting patterns stand out.The LSTM model reaches its best accuracy right at it scored 0.87, just a bit better than the other methods. Among classic machine learning models, both SVM and Logistic Regression tend to give solid results.The accuracy is about 0.80, and Random Forest gives a similar result.It has about 0.82 accuracy, but the results can vary quite a bit.Overall, the LSTM turned out to be the best model, so it was used to automatically sort the rest of the unlabeled articles into three political views: Pro-Palestine, Anti-Palestine, and Neutral.

The histogram in Figure 4 illustrates a well-balanced dataset across the three sentiment categories: "Against", "Neutral" and "Pro" with each class comprising roughly 6000 instances.

6. Discussion

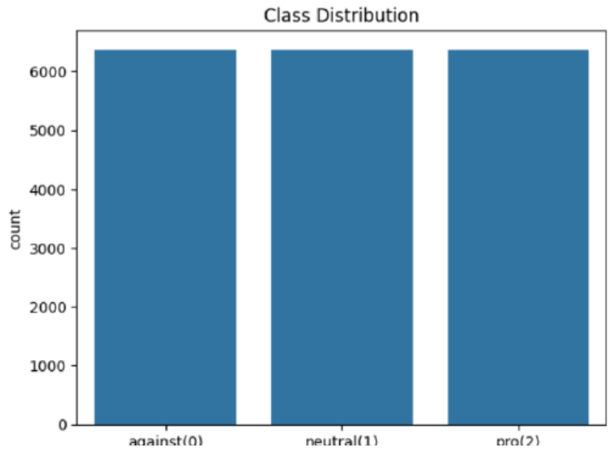

Figure 4: Balanced Sentiment Class Distribution After the classification

In our case, the neutral class represented the biggest challenge, as it was very difficult to achieve balance between the classes. It consistently appeared as the minority class throughout the entire solution. This leads to two possible explanations: either it is genuinely difficult to find neutral articles in the context of political bias, or the models we used were unable to identify neutrality.

The method proved to be the most effective. It relies on labeled data and word embeddings generated with the powerful transformer model MARBERT, combined with fine-tuning. As the results indicate, this approach achieved the best performance, with a balanced class distribution and high classification accuracy. To classify the entire dataset, we employed an LSTM model, which reached the highest accuracy of 0.87. The combination of MARBERT embeddings and LSTM classification thus proved highly effective.

7. Conclusion

This study looked into spotting political bias in Arabic news articles, especially when it came to how they covered the Palestinian issue. A framework was put together that mixes deep learning and transformer-based models, running on scalable processing with Apache Spark. By adjusting MARBERT and using careful preprocessing along with data augmentation and class balancing, the system did a good job sorting articles into Pro-Palestine, Neutral, and Anti-Palestine categories. These results show how well domain-specific contextual embeddings can pick up on subtle rhetorical and emotional cues in

Arabic texts. There are still some challenges left. The Neutral category was hard to classify because it wasn't very clear-cut, and Spark was run in a local setup instead of a full distributed cluster, which meant it couldn't scale as well. Future work will look at expanding the framework to cover more Arabic dialects and include different types of sources like text, images, audio, and video. We can improve real-time scalability by combining Apache Kafka with distributed Spark clusters, which allows streaming pipelines to run constantly and catch bias as it happens. Aside from political bias, this method can also be used in areas like spotting fake news and analyzing subjectivity. Furthermore, this approach opens a significant perspective in the field of Artificial Intelligence by positioning political bias detection as a specialized sub-domain of anomaly behavior detection. Combining contextual embeddings with other linguistic features, along with using large language models (LLMs), can help improve accuracy and adaptability when analyzing media on a large scale.

8. References

Dave D'Alessio, Mike Allen. (2000). Media Bias in Presidential Elections: A Meta-Analysis. *Journal of Communication,*, 133–156.

Dilshod Azizov, Preslav Nakov and Shangsong Liang. (2023). Frank at CheckThat! 2023: Detecting the Political Bias of News Articles and News Media. *Conference and Labs of the Evaluation Forum.* Thessaloniki, Greece.

Ghada Ben Amor, Nawres Medimagh, Saoussen Ben Chaabene, and Omar Trigui. (2025). ANLP-Uniso at EXIST 2025: Sexism Identification and Characterization in Tweets. *Conference and Labs of the Evaluation Forum.* Madrid, Spain.

M.D. Coşgun, G. Kutlu, Mücahid Türkmen. (2023). TOBB ETU at Checkthat! 2023: Utilizing Chatgpt To Detect Subjective Statements and Political Bias. *The Conference and Labs of the Evaluation Forum.* Thessaloniki, Greece.

Ramy Baly, Mitra Mohtarami, James Glass, Lluís Màrquez, Alessandro Moschitti, Preslav Nakov. (2018). Integrating Stance Detection and Fact Checking in a Unified Corpus. *The 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* New Orleans, Louisiana.

Saad M. Darwish, Abdul Rahman M. Sabri, Dhafar Hamed Abd, Adel A. Elzoghabi. (2024). An Expert System to Detect Political Arabic Articles Orientation Using CatBoost Classifier Boosted by Multi-Level Features. *Computer Systems Science and Engineering*, 1595-1624.

Sieu Tran, Paul Rodrigues, Benjamin Strauss and Evan M. Williams. (2023). Accenture at CheckThat! 2023: Learning to Detect Political Bias of News Articles and Sources. *Conference and Labs of the Evaluation Forum.* Thessaloniki, Greece.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, Alexander Hauptmann. (2006). Which Side are You on? Identifying Perspectives at the Document and Sentence Levels. *Conference on Computational Natural Language Learning.* New York City,.

Yasmine El Abed, Mariem Ben Arbia, Saoussen Ben Chaabene, Omar Trigui. (2025). ANLP-UniSo at MAHED Shared Task: Detection of Hate and Hope Speech in Arabic Social Media based on XLM-RoBERTa and Logistic Regression. *Proceedings of the Third Arabic Natural Language Processing Conference (ArabicNLP 2025)* (pp. 639–644). China: Association for Computational Linguistics.

Ye Jiang, Johann Petrak, Xingyi Song, Kalina Bontcheva, Diana Maynard. (2019). Team Bertha von Suttner at SemEval-2019 Task 4: Hyperpartisan News Detection using ELMo Sentence Representation Convolutional Network. *The 13th International Workshop on Semantic Evaluation.* Minnesota, USA.

Yoan Dinkov, Ahmed Ali, Ivan Koychev, Preslav Nakov. (2019). Predicting the Leading Political Ideology of YouTube Channels Using Acoustic, Textual, and Metadata Information. *INTERSPEECH.* Graz, Austria.