BAREKENG: Journal of Mathematics and Its Applications
June 2026 Volume 20 Issue 2 Page 1709–1718
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss2pp1709-1718

ANALYSIS OF MULTILINGUAL OPINION POLARIZATION WITH CROSS-LINGUAL LANGUAGE MODEL-ROBUSTLY OPTIMIZED BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS APPROACH (XLM-ROBERTA)

Ghaitsa Shafa Cinta Kananta [1*], **Dewi Retno Sari Saputro** [2], **Sutanto** [3]

[1,2,3] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret*
Jln. Ir.Sutami No. 36, Surakarta, 57126, Indonesia

*Corresponding author's e-mail: * **ghaitsashafa@student.uns.ac.id***

Article Info

Article History:

Received: 14th August 2025
Revised: 5th September 2025
Accepted: 3rd October 2025
Available Online: 26th January 2026

Keywords:

Multilingual;
Opinion Polarization;
Sentiment analysis;
XLM-RoBERTa.

ABSTRACT

The rapid growth of digital communication has intensified opinion exchanges across languages and cultures on social media, enriching public discourse while also increasing the risk of polarization that deepens social divisions. Conventional sentiment analysis methods that rely on translation often distort meaning, overlook emotional nuances, and fail to capture rhetorical devices such as irony and sarcasm, thereby limiting their reliability in multilingual contexts. This study examines the capability of XLM-RoBERTa, a multilingual transformer model pretrained on more than 100 languages, to address these challenges by generating consistent semantic representations and accommodating linguistic and cultural diversity without translation. The research employs bibliometric analysis using VOSviewer on 357 Scopus-indexed publications from 2020 to 2025 to map research trends, combined with a literature review that evaluates XLM-RoBERTa in sentiment and opinion analysis. The findings reveal that although XLM-RoBERTa has been widely employed for sentiment classification, text categorization, and offensive language detection, research explicitly focused on multilingual opinion polarization remains limited. Benchmark evaluations further indicate that XLM-RoBERTa surpasses earlier multilingual models, achieving 79.6% accuracy on XNLI and an 81.2% F1-score on MLQA, confirming its robustness in capturing semantic nuances, cultural variations, and rhetorical complexity without translation. The novelty of this research lies in integrating trend-mapping with methodological evaluation, thereby establishing XLM-RoBERTa as a reliable framework for real-time monitoring of global public opinion, supporting evidence-based policymaking, and advancing scholarly understanding of multilingual communication dynamics in the digital era.

How to cite this article:

G. S. C. Kananta, D. R. S. Saputro and Sutanto., "ANALYSIS OF MULTILINGUAL OPINION POLARIZATION WITH CROSS-LINGUAL LANGUAGE MODEL-ROBUSTLY OPTIMIZED BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS APPROACH (XLM-ROBERTA)", *BAREKENG: J. Math. & App.*, vol. 20, no. 2, pp. 1709-1718, Jun, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The digital industrial revolution has fostered a vibrant environment for exchanging dialogue and opinions on social media platforms [1]. On a global scale, opinions are expressed across diverse languages and cultures, creating challenges for opinion analysis when polarization occurs [2]. Such polarization is intensified by echo chambers, filter bubbles, and algorithmic content curation that limit exposure to contrasting ideas [3][4]. Consequently, polarization impacts not only individual perceptions of specific issues but also has the capacity to foster social divisions across various segments of society [5]. The growing concern over opinion polarization lies in its broader social and political consequences. Rather than fostering dialogue, exposure to diverse viewpoints on social media often heightens conflict, even in countries with strong democratic traditions [6]. This polarization exacerbates divergences in issue perception, undermines consensus, and erodes trust in political institutions [7].

In the era of globalization, issues emerging in one country can rapidly cross borders and be reframed with distinct cultural meanings [8][9]. This presents unique challenges for opinion analysis, as emotional nuances and semantic content are shaped by linguistic structures and cultural conventions [10]. Languages differ in rhetorical techniques, idiomatic expressions, and the use of sarcasm [11][12]. Such variances can lead to confusion. Variations often create ambiguity in interpretation, particularly when texts are examined through traditional translation-based methods [13]. Therefore, analytical approaches must be more sensitive and designed for linguistic and cultural diversity.

An initial approach to understanding public opinion is sentiment analysis, which classifies textual data into positive, negative, or neutral categories [14]. However, traditional sentiment analysis merely assigns categorical labels without revealing the opinion distributions, even though balanced opposing viewpoints may indicate polarization [15][16]. These limitations underscore the need for advanced natural language processing (NLP) methods that can analyze texts in their original languages while producing consistent, reliable semantic representations. Contemporary NLP technologies allow opinion analysis to be conducted directly, eliminating the need for manual translation, which often risks distorting meaning [17]. Moreover, advances in deep learning enhance the capacity to capture complex linguistic patterns, making NLP increasingly suitable for multilingual opinion analysis [18]. NLP also enables real-time monitoring of social media discourse, which is essential for identifying and tracking emerging issues [19].

Among recent developments, XLM-RoBERTa stands out as a cross-lingual transformer that learns linguistic structures across more than 100 languages without the need for translation [20]. Its capacity to capture cultural and linguistic nuances directly addresses the shortcomings of traditional sentiment analysis, making it particularly effective for analyzing opinion polarization in multilingual contexts [21]. Accordingly, this article explores XLM-RoBERTa's ability to analyze multilingual opinion polarization on social media and assesses its effectiveness in addressing linguistic and cultural challenges. The aim is to address gaps in traditional approaches, enhance detection accuracy, and provide methodological insights for real-time monitoring of global public opinion.

2. RESEARCH METHODS

This research employs two main methods: a bibliometric analysis using the Visualization of Similarities Viewer (VOSviewer) software and a literature review of the Cross-Lingual Language Model–Robustly Optimized Bidirectional Encoder Representations from Transformers (XLM-RoBERTa) approach. The objective of this approach is to map research trends on multilingual opinion polarization and to evaluate and examine opinion-polarization strategies in the XLM-RoBERTa model.

2.1 Bibliometric Analysis Using the Visualization of Similarities Viewer (VOSViewer)

The use of VOSviewer enables the illustration of relevant current research topics. It is a bibliometric software capable of creating visual maps based on keyword relationships (co-word), author collaborations (co-author), and citation patterns (citation network), thereby enabling systematic knowledge mapping [22]. This analysis aims to identify research trends, collaboration networks, and commonly employed research topics [23].

Bibliometric data were obtained from Scopus-indexed journal publications within the publication period 2020–2025. A total of 357 documents were retrieved using the keywords “XLM-RoBERTa”, "sentiment analysis", "multilingual NLP", and "opinion polarization". The inclusion criteria were limited to journal articles, conference papers, and book chapters. The bibliometric procedure follows the bibliometric steps established by Indriati et al. [24], including keyword determination, literature source search, document restriction, metadata collection, network analysis, and visualization interpretation. The bibliometric mapping of XLM-RoBERTa, created with VOSviewer using Scopus data, is depicted Fig. 1.

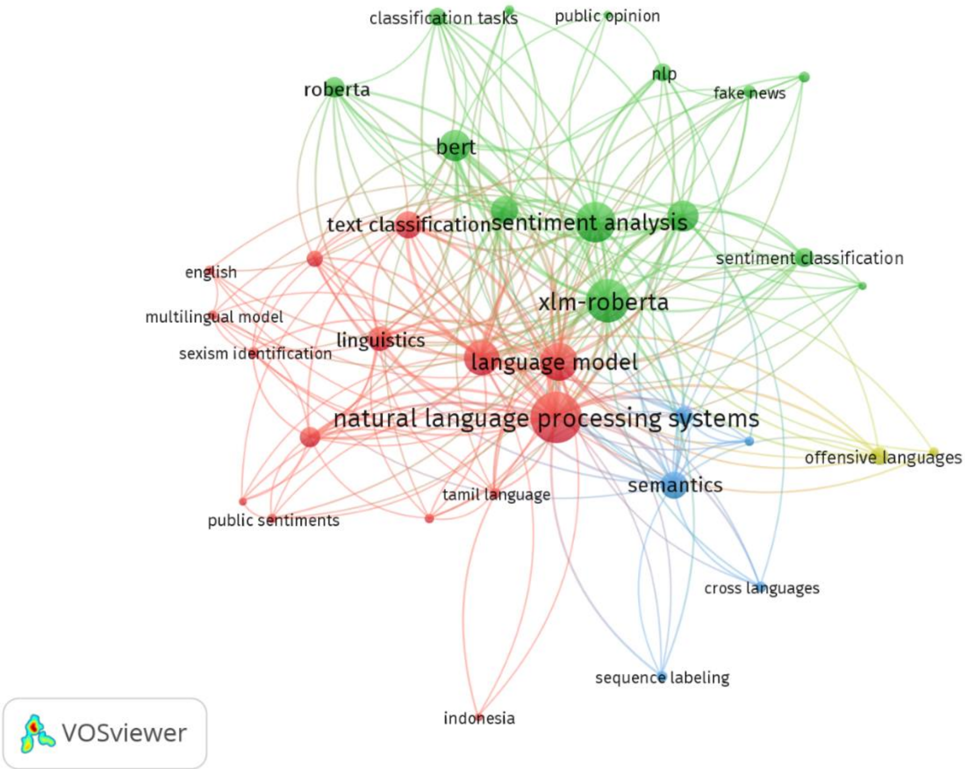

Figure 1. Visualization Based on Titles and Abstracts Generated using VOSviewer

Fig. 1 presents the bibliometric network related to XLM-RoBERTa. Nodes represent frequently occurring keywords, with node size reflecting their frequency. Connecting lines indicate relationships between terms, and node colors denote thematic clusters based on contextual similarity [25]. The visualization indicates that "model", "sentiment", and "language" dominate the network, signifying the prevalence of sentiment analysis models and natural language processing in multilingual contexts.

Furthermore, the visualization highlights four thematic clusters, each reflecting key areas where XLM-RoBERTa is applied. The green cluster focuses on sentiment analysis, text classification, and fake news detection, demonstrating its significant role in analyzing public opinion and categorizing sentiment. The red cluster is related to linguistics, language models, and NLP systems, highlighting XLM-RoBERTa's contribution to the development of sophisticated language models and NLP frameworks. The blue cluster emphasizes semantics, cross-lingual tasks, and sequence labeling, showcasing XLM-RoBERTa's capabilities in multilingual processing and sequence-based tasks. Lastly, the orange cluster focuses on offensive language detection, showing its application in identifying harmful or toxic language.

Despite these varied clusters, the visualization shows that no cluster is dedicated solely to multilingual opinion polarization, suggesting that this area remains underexplored. Accordingly, the bibliometric mapping underscores XLM-RoBERTa's versatility across diverse NLP domains while simultaneously revealing the research gap in multilingual opinion polarization, thereby offering opportunities for further in-depth investigations.

2.2 Cross-Lingual Language Model-Robustly Optimized Bidirectional Encoder Representations from Transformers Approach (XLM-RoBERTa)

XLM-RoBERTa represents a multilingual advancement of XLM and RoBERTa, created by Facebook AI to encompass a wider range of languages. This model aims to enhance language modeling across multiple languages by being pretrained on an extensive dataset spanning over 100 languages, with the goal of addressing the challenges of comprehending multiple languages simultaneously, particularly those that are less resourced [20].

In contrast to XLM, which employs sentence-parallel Translation Language Modeling (TLM) to establish cross-lingual connections between words and phrases, XLM-RoBERTa adopts a purely Masked Language Modeling (MLM) approach. It is trained without supervision on the CommonCrawl corpus (CC 100) spanning over 100 languages. This method enhances XLM-RoBERTa's ability to accommodate languages with fewer resources, unlike XLM, which relies heavily on parallel sentence pairs [20][26].

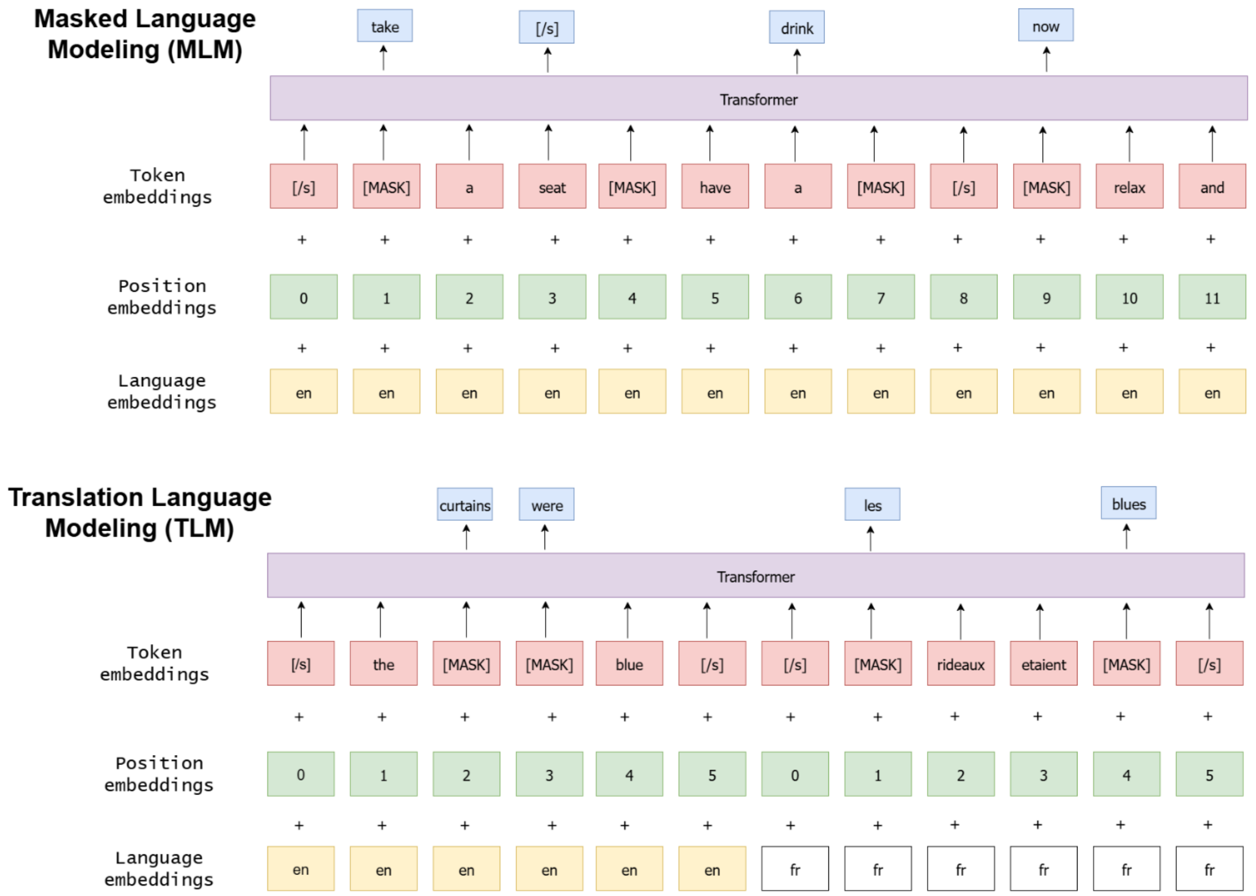

Figure 2. Architectural Differences Between MLM and TLM

Fig. 2 illustrates the architectural differences between MLM and TLM, which represent two fundamental training approaches in multilingual language modeling. To provide a clearer comparative perspective, Table 1 summarizes widely used multilingual models, outlining their architectures, training objectives, pretraining datasets, language coverage, and benchmark performance.

Table 1. Comparative Overview of Multilingual Language Models

Model	Architecture	Training Objective	Pretraining Dataset	Language Covered	Benchmark Performance
mBERT [27]	Transformer (BERT)	MLM	Wikipedia	104	Baseline, moderate on XNLI & MLQA
XLM [26]	Transformer	MLM + TLM	Wikipedia + Parallel corpora	15	Outperforms mBERT on XNLI
RoBERTa [28]	Transformer (optimized BERT)	MLM (optimized)	BooksCorpus, CC-News, OpenWebText, Stories	1	SOTA on GLUE, RACE (monolingual)
XLM-RoBERTa [20]	Transformer (RoBERTa-based)	MLM (large-scale)	CommonCrawl (CC100)	100+	SOTA on XNLI, MLQA, XTREME

Beyond its pretraining method, the XLM-RoBERTa framework incorporates several architectural enhancements compared to XLM. XLM-RoBERTa employs a transformer encoder based on RoBERTa, featuring a self-attention mechanism that enables the model to grasp global relationships among tokens. This mechanism relies on three primary vectors: (Q), key (K), and value (V). The scaled dot-product attention is formulated as demonstrated in Eq. (1).

$$Attention = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{1}$$

where Q represents the embedding vector of a word within a sentence, K represents the embeddings of all words in that sentence, V is the vector obtained from the dot product of a word embedding and specific parameters, and d_k is the dimension of the $Q, K,$dan V vectors, used to dynamically evaluate the relevance of a word within its context [29]. The resulting score is normalized using the softmax function in Eq. (2).

$$softmax(z_i) = \frac{e^{zi}}{\sum_j^K e^{zj}}, \tag{2}$$

where e^{zi} represents the exponent of the input vector, K denotes the total number of classes, ande^{zj} corresponds to the exponent of the output vector. This probability distribution assigns dynamic contextual weights to tokens, ensuring the sum of all weights equals one [29].

This mechanism is further strengthened by multi-head attention, which allows numerous attention heads to concurrently identify distinct patterns of relationships between tokens, yielding more comprehensive and nuanced representations [29]. The model is structured with 12 transformer layers, 768 hidden units, and a self-attention mechanism that exhibits reliable and effective performance in cross-lingual classification, sequence labeling, and question answering [20]. The fundamental architectural design is depicted in Fig. 3, which includes an encoder transformer and a fully connected layer, aimed at generating an output that captures both the representation of each token and the overall meaning of the sentence.

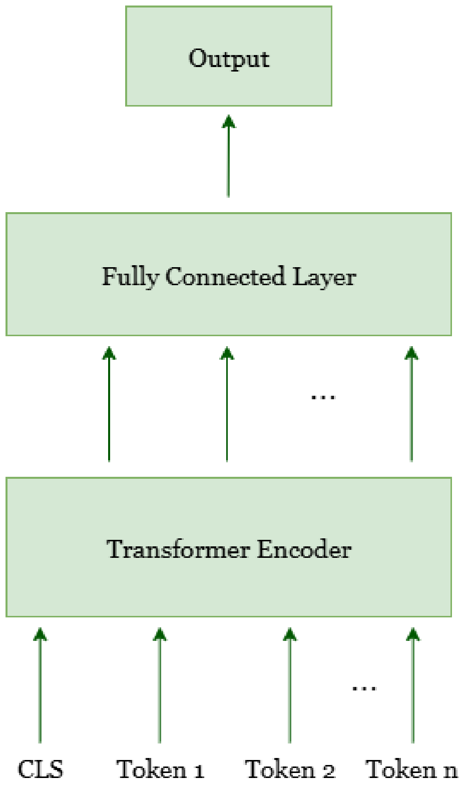

Figure 3. Fundamental Architecture of XLM-RoBERTa

To strengthen its multilingual capabilities, XLM-RoBERTa uses a shared vocabulary and subword tokenization via SentencePiece. This approach enables effective processing across diverse languages, including low-resource ones, thereby reinforcing its robustness as a multilingual model.

3. RESULTS AND DISCUSSION

Multilingual polarization of opinion reflects the fragmentation of public opinion caused by differences in language, culture, and emotional expression. Opinions and emotions vary not only in the words used but also in the way they are culturally understood and expressed. When opinions are expressed in different languages, differences in meaning can arise from mismatches between language use and the social norms underlying each language [30]. Balahur and Turchi [13] observed that differences in how emotions are expressed in Spanish and English significantly affect sentiment classification. Traditional approaches using monolingual models or automatic translation-based models have limitations in distinguishing between pro and con opinions.

Conventional methods often struggle with this differentiation, as machine translation frequently introduces distortions that undermine the accuracy of cross-lingual sentiment classification [31]. These challenges underscore the need for multilingual models that comprehend text in its native language, reducing reliance on translation and thereby preserving emotional significance and cultural nuance more effectively.

The strength of XLM-RoBERTa lies in its technical mechanisms that directly address these challenges. First, the model uses the SentencePiece (SPM) with Byte-Pair Encoding (BPE) [20]. SentencePiece performs subword tokenization, transforming text into a sequence of IDs without the need for a language-specific dictionary [32]. This approach ensures rare or morphologically complex words from multiple languages are represented consistently, a critical feature for multilingual opinion analysis. Second, the model utilizes a self-attention mechanism to dynamically capture semantic relations among tokens.

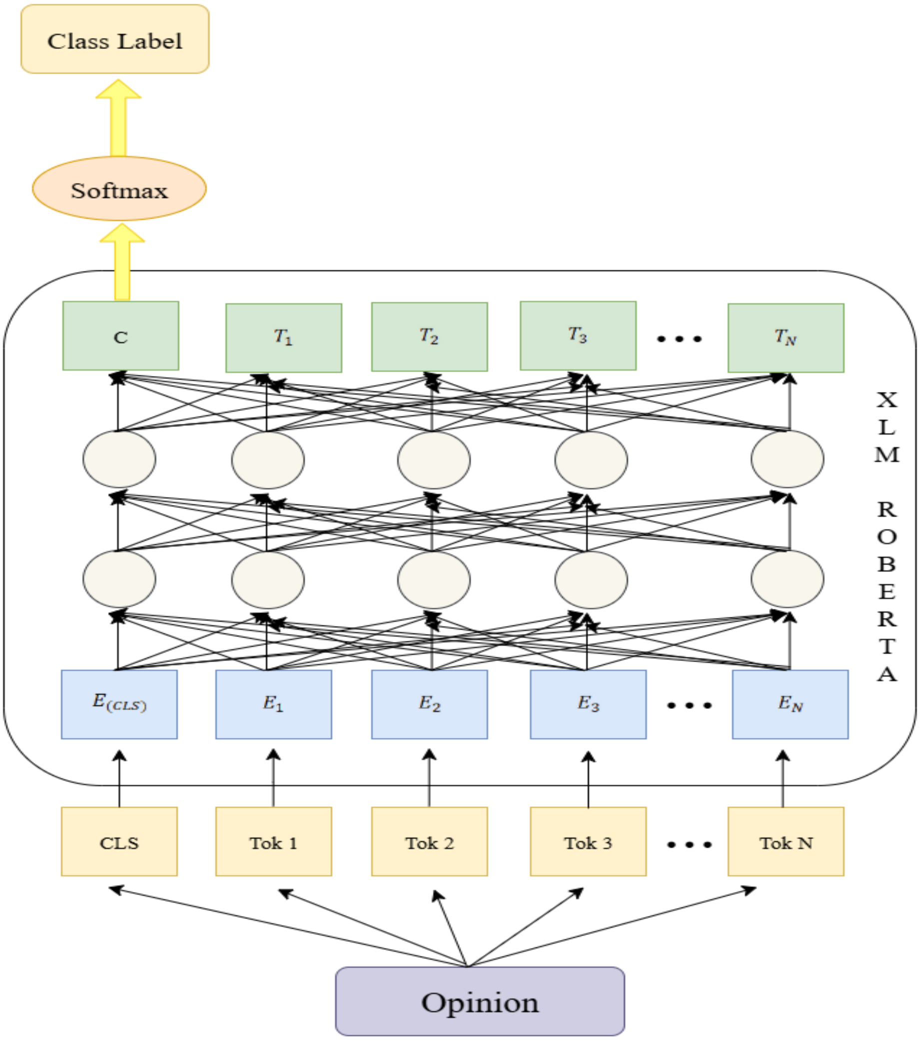

Figure 4. XLM-RoBERTa Architecture

Fig. 4 illustrates the classification process, which begins with tokenization and proceeds to embedding representations that pass through the transformer encoder layers. The special [CLS] token aggregates the contextual meaning of the entire sentence, which is then processed by dropout and dense layers with a softmax activation to generate sentiment probabilities [33]. This architecture is not merely a technical detail; it is essential for detecting polarized opinions across diverse languages. By ensuring consistent semantic representation across languages, XLM-RoBERTa enables more reliable cross-lingual comparisons of public sentiment. To demonstrate the effectiveness of XLM-RoBERTa's design, Table 2 compares its performance with earlier multilingual models across widely used benchmark datasets.

Table 2. Performance Comparison between Multilingual Models

Model	Task (Dataset)	Metric	Result
mBERT	XNLI (Cross-lingual NLI)	Accuracy	65.0%
XLM	MLQA (Multilingual QA)	F1 Score	68.2%
XLM-RoBERTa	XNLI (Cross-lingual NLI)	Accuracy	79.6%
XLM-RoBERTa	MLQA (Multilingual QA)	F1 Score	81.2%

Data source: Adapted from [20]

The results in Table 2 clearly highlight the advantage of XLM-RoBERTa over earlier multilingual models, confirming that the architectural improvements in Fig. 4, particularly the use of SentencePiece tokenization and transformer-based self-attention, translate into tangible performance gains. Furthermore, case studies reinforce this evidence. Beyond the empirical improvements shown in Table 2, XLM-RoBERTa's architecture has specific implications for opinion polarization analysis. The use of SentencePiece tokenization ensures a consistent representation of linguistically diverse and informal expressions, including code-switching and colloquial terms commonly found on social media. This feature is crucial for analyzing polarized debates, where users often mix languages or employ non-standard vocabulary. Moreover, the self-attention mechanism enables the model to capture subtle rhetorical devices such as sarcasm, irony, and emphasis, which are common in polarized discussions. By assigning higher weights to emotionally charged or contextually critical words, the model can more accurately distinguish between pro and con opinions. Finally, the softmax output provides probabilistic scores that reflect the strength of these opposing stances, thereby enabling quantifiable measurements of polarization.

The model shows significant improvements across various benchmarks, achieving a +14.6% accuracy gain on XNLI compared to mBERT and a +13% F1 score on MLQA [20]. This capability is especially crucial for analyzing opinion polarization, as consistent representations enable cross-language comparisons of opinions without losing their intended meaning [34]. In practice, Wiciaputra et al. [35] showed that applying transfer learning with XLM-RoBERTa for bilingual classification (English–Indonesian) achieved 90.2% accuracy, outperforming traditional baselines. Similarly, Kumar and Albuquerque [36] demonstrated the model's flexibility through zero-shot transfer learning, achieving 70.5% accuracy in Hindi sentiment analysis despite limited training resources. These results highlight the model's ability to incorporate minority or low-resource languages into global opinion analysis [30].

Additional validation comes from the TweetEval benchmark conducted by Barbieri et al. [37], where XLM-RoBERTa consistently outperformed alternatives in emotion and hate-speech detection across more than 10 languages. Likewise, Ranasinghe and Zampieri [21] highlight the model's strength in cross-lingual offensive language detection. Gaurav et al. [38] applied XLM-RoBERTa to categorize public sentiment toward the Metaverse and 6G technology, revealing that sentiment was predominantly neutral and positive, demonstrating the model's versatility in addressing modern technological topics. These findings underline the model's versatility across domains, while maintaining strong performance in multilingual contexts.

Beyond performance metrics, XLM-RoBERTa offers a methodological advantage for studies of opinion polarization. Its ability to provide direct multilingual representations minimizes reliance on machine translation, which often distorts meaning [39]. This capability is crucial for multilingual social media platforms where users frequently mix languages or employ culture-specific expressions [40]. It helps preserve subtleties such as sarcasm or irony that machine translation often fails to convey accurately [41]. Consequently, the model offers a robust foundation for detecting polarization dynamics and enhances the reliability of real-time monitoring of global discourse.

4. CONCLUSION

This research underscores XLM-RoBERTa's ability as a multilingual transformer model to effectively capture opinion polarization across diverse languages and cultural contexts. Through its multilingual pretraining, the model provides a consistent representation of diverse languages and cultural contexts, preserving semantic, contextual, and emotional nuances while mitigating distortions that commonly arise in translation-based or monolingual methods. Its technical strengths, including the self-attention mechanism and SentencePiece tokenization, enable reliable detection of nuanced opinion polarity, rhetorical devices such as sarcasm and irony, and cross-lingual sentiment variations. Bibliometric analysis further reveals that research on multilingual semantic representation for opinion polarization remains limited, underscoring the importance of advancing models such as XLM-RoBERTa. Accordingly, the model not only improves the accuracy and robustness of sentiment analysis but also offers a methodological framework for real-time monitoring of global public opinion, contributing to a deeper understanding of complex social dynamics and communication patterns across diverse societies.

Despite these advantages, this research recognizes several limitations. The current investigation relies primarily on theoretical discussion and existing benchmark evaluations, with limited large-scale empirical validation across multiple social media platforms, which may constrain the generalizability of findings. Future research should therefore emphasize comparative analyses with other multilingual transformer-based models, exploration of domain-specific applications, and examination of ethical considerations in cross-cultural opinion mining. Expanding these research directions will enhance the practical applicability of XLM-RoBERTa, enabling its use in policy analysis, public sentiment monitoring, and decision-making processes. Furthermore, subsequent studies could evaluate its performance in low-resource languages, cross-domain adaptation, and real-time detection of emerging trends, thereby consolidating XLM-RoBERTa as a strategic tool for capturing and interpreting complex multilingual opinion dynamics in an increasingly interconnected digital environment.

Author Contributions

Ghaitsa Shafa Cinta Kananta: Conceptualization, Methodology, Writing-Original Draft, Visualization. Dewi Retno Sari Saputro: Conceptualization, Writing-Review and Editing, Supervision, Validation. Sutanto: Supervision, Writing-Review, and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research was financially supported by the Institute for Research and Community Service (LPPM) Universitas Sebelas Maret (UNS), through a research group grant as stated in the letter of approval for the implementation of non-APBN-funded research No. 370/UN27.22/PT.01.03/2025.

Acknowledgment

Sincere appreciation is extended to the Softcomputing Mathematics Research Group for their continuous sustained support, insightful discussions, and constructive feedback, all of which have been invaluable in refining both the conceptual framework and the overall presentation of this paper. Appreciation is also conveyed to colleagues and peers whose constructive contributions have significantly enhanced the quality of this manuscript.

Declarations

The authors declare that there are no conflicts of interest to report study.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

[1] N. A. P. Masaling and D. Suhartono, "UTILIZING ROBERTA AND XLM-ROBERTA PRE-TRAINED MODEL FOR STRUCTURED SENTIMENT ANALYSIS," *International Journal of Informatics and Communication Technology*, vol. 13, no. 3, pp. 410–421, Dec. 2024, doi: https://doi.org/10.11591/ijict.v13i3.pp410-421.

[2] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, "POLITICAL DISCOURSE ON SOCIAL MEDIA: ECHO CHAMBERS, GATEKEEPERS, AND THE PRICE OF BIPARTISANSHIP," in *The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018*, Association for Computing Machinery, Inc, Apr. 2018, pp. 913–922. doi: https://doi.org/10.1145/3178876.3186139.

[3] E. Pariser, *THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU*. Penguin Group, The, 2011.

[4] E. Bakshy, S. Messing, and L. A. Adamic, "EXPOSURE TO IDEOLOGICALLY DIVERSE NEWS AND OPINION ON FACEBOOK", doi: https://doi.org/10.7910/DVN/LDJ7MS.

[5] S. Iyengar, G. Sood, and Y. Lelkes, "AFFECT, NOT IDEOLOGY: A SOCIAL IDENTITY PERSPECTIVE ON POLARIZATION," Sep. 2012. doi: https://doi.org/10.1093/poq/nfs038.

[6] C. A. Bail *et al.*, "EXPOSURE TO OPPOSING VIEWS ON SOCIAL MEDIA CAN INCREASE POLITICAL POLARIZATION," *Proc Natl Acad Sci U S A*, vol. 115, no. 37, pp. 9216–9221, Sep. 2018, doi: https://doi.org/10.1073/pnas.1804840115.

[7] S. Iyengar and S. J. Westwood, "FEAR AND LOATHING ACROSS PARTY LINES: NEW EVIDENCE ON GROUP POLARIZATION," *Am J Pol Sci*, vol. 59, no. 3, pp. 690–707, Jul. 2015, doi: https://doi.org/10.1111/ajps.12152.

[8] D. Yuna, L. Xiaokun, L. Jianing, and H. Lu, "CROSS-CULTURAL COMMUNICATION ON SOCIAL MEDIA: REVIEW FROM THE PERSPECTIVE OF CULTURAL PSYCHOLOGY AND NEUROSCIENCE," Mar. 08, 2022, *Frontiers Media S.A.* doi: https://doi.org/10.3389/fpsyg.2022.858900.

[9] M. Falkenberg, F. Zollo, W. Quattrociocchi, J. Pfeffer, and A. Baronchelli, "PATTERNS OF PARTISAN TOXICITY AND ENGAGEMENT REVEAL THE COMMON STRUCTURE OF ONLINE POLITICAL COMMUNICATION ACROSS COUNTRIES," *Nature Communications* , vol. 15, no. 1, Dec. 2024, doi: https://doi.org/10.1038/s41467-024-53868-0.

[10] K. Dashtipour *et al.*, "MULTILINGUAL SENTIMENT ANALYSIS: STATE OF THE ART AND INDEPENDENT COMPARISON OF TECHNIQUES," *Cognit Comput*, vol. 8, no. 4, pp. 757–771, Aug. 2016, doi: https://doi.org/10.1007/s12559-016-9415-7.

[11] B. Ghanem, J. Karoui, F. Benamara, P. Rosso, and V. Moriceau, "IRONY DETECTION IN A MULTILINGUAL CONTEXT," Feb. 2020.

[12] D. Jain, A. Kumar, and G. Garg, "SARCASM DETECTION IN MASH-UP LANGUAGE USING SOFT-ATTENTION BASED BI-DIRECTIONAL LSTM AND FEATURE-RICH CNN," *Applied Soft Computing Journal*, vol. 91, Jun. 2020, doi: https://doi.org/10.1016/j.asoc.2020.106198.

[13] A. Balahur and M. Turchi, "COMPARATIVE EXPERIMENTS USING SUPERVISED LEARNING AND MACHINE TRANSLATION FOR MULTILINGUAL SENTIMENT ANALYSIS," *Comput Speech Lang*, vol. 28, no. 1, pp. 56–75, 2014, doi: https://doi.org/10.1016/j.csl.2013.03.004.

[14] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "SENTIMENT ANALYSIS: A REVIEW AND COMPARATIVE ANALYSIS OF WEB SERVICES," *Inf Sci (N Y)*, vol. 311, pp. 18–38, Aug. 2015, doi: https://doi.org/10.1016/j.ins.2015.03.040.

[15] S. Poria, D. Hazarika, N. Majumder, and R. Mihalcea, "BENEATH THE TIP OF THE ICEBERG: CURRENT CHALLENGES AND NEW DIRECTIONS IN SENTIMENT ANALYSIS RESEARCH," *IEEE Trans Affect Comput*, vol. 14, no. 1, pp. 108–132, Jan. 2023, doi: https://doi.org/10.1109/TAFFC.2020.3038167.

[16] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito, "MEASURING POLITICAL POLARIZATION: TWITTER SHOWS THE TWO SIDES OF VENEZUELA," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 3, Mar. 2015, doi: https://doi.org/10.1063/1.4913758.

[17] M. Abdalla and G. Hirst, "CROSS-LINGUAL SENTIMENT ANALYSIS WITHOUT (GOOD) TRANSLATION," Oct. 2017.

[18] T. Young, D. Hazarika, S. Poria, and E. Cambria, "RECENT TRENDS IN DEEP LEARNING BASED NATURAL LANGUAGE PROCESSING," Nov. 2018.

[19] E. Cambria, Y. Li, F. Z. Xing, S. Poria, and K. Kwok, "SENTICNET 6: ENSEMBLE APPLICATION OF SYMBOLIC AND SUBSYMBOLIC AI FOR SENTIMENT ANALYSIS," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, New York, NY, USA: ACM, Oct. 2020, pp. 105–114. doi: https://doi.org/10.1145/3340531.3412003.

[20] A. Conneau *et al.*, "UNSUPERVISED CROSS-LINGUAL REPRESENTATION LEARNING AT SCALE," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 8440–8451. doi: https://doi.org/10.18653/v1/2020.acl-main.747.

[21] T. Ranasinghe and M. Zampieri, "MULTILINGUAL OFFENSIVE LANGUAGE IDENTIFICATION WITH CROSS-LINGUAL EMBEDDINGS."

[22] D. R. S. Saputro, H. Prasetyo, A. Wibowo, F. Khairina, K. Sidiq, and G. N. A. Wibowo, "BIBLIOMETRIC ANALYSIS OF NEURAL BASIS EXPANSION ANALYSIS FOR INTERPRETABLE TIME SERIES (N-BEATS) FOR RESEARCH TREND MAPPING," *Barekeng*, vol. 17, no. 2, pp. 1103–1112, Jun. 2023, doi: https://doi.org/10.30598/barekengvol17iss2pp1103-1112.

[23] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "HOW TO CONDUCT A BIBLIOMETRIC ANALYSIS: AN OVERVIEW AND GUIDELINES," *J Bus Res*, vol. 133, pp. 285–296, Sep. 2021, doi: https://doi.org/10.1016/j.jbusres.2021.04.070.

[24] S. P. Indriati, D. R. S. Saputro, and P. Widyaningsih, "THE BIBLIOMETRIC NETWORK TO IDENTIFY RESEARCH TRENDS IN MULTI-INPUT TRANSFER FUNCTION," *Barekeng*, vol. 18, no. 3, pp. 1931–1938, Aug. 2024, doi: https://doi.org/10.30598/barekengvol18iss3pp1931-1938.

[25] N. J. van Eck and L. Waltman, "SOFTWARE SURVEY: VOSVIEWER, A COMPUTER PROGRAM FOR BIBLIOMETRIC MAPPING," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010, doi: https://doi.org/10.1007/s11192-009-0146-3.

[26] G. Lample and A. Conneau, "CROSS-LINGUAL LANGUAGE MODEL PRETRAINING," Jan. 2019.

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: PRE-TRAINING OF DEEP BIDIRECTIONAL TRANSFORMERS FOR LANGUAGE UNDERSTANDING," May 2019.

[28] Y. Liu *et al.*, "ROBERTA: A ROBUSTLY OPTIMIZED BERT PRETRAINING APPROACH," Jul. 2019, [Online]. Available: https://arxiv.org/abs/1907.11692.

[29] A. Vaswani *et al.*, "ATTENTION IS ALL YOU NEED," Aug. 2023.

[30] S. Ruder, I. Vulić, and A. Søgaard, "A SURVEY OF CROSS-LINGUAL WORD EMBEDDING MODELS," Oct. 2019, doi: https://doi.org/10.1613/jair.1.11640.

[31] J. Barnes, R. Klinger, and S. Schulte Im Walde, "PROJECTING EMBEDDINGS FOR DOMAIN ADAPTATION: JOINT MODELING OF SENTIMENT ANALYSIS IN DIVERSE DOMAINS TITLE AND ABSTRACT IN BASQUE."

[32] T. Kudo and J. Richardson, "SENTENCEPIECE: A SIMPLE AND LANGUAGE INDEPENDENT SUBWORD TOKENIZER AND DETOKENIZER FOR NEURAL TEXT PROCESSING," Aug. 2018, [Online]. Available: https://arxiv.org/abs/1808.06226.

[33] M. Mosbach, M. Andriushchenko, and D. Klakow, "ON THE STABILITY OF FINE-TUNING BERT: MISCONCEPTIONS, EXPLANATIONS, AND STRONG BASELINES," Jun. 2020, [Online]. Available: https://arxiv.org/abs/2006.04884.

[34] J. Hu, S. Ruder, A. Siddhant, G. Neubig, O. Firat, and M. Johnson, "XTREME: A MASSIVELY MULTILINGUAL MULTI-TASK BENCHMARK FOR EVALUATING CROSS-LINGUAL GENERALIZATION," Sep. 2020.

[35] Y. K. Wiciaputra, J. C. Young, and A. Rusli, "BILINGUAL TEXT CLASSIFICATION IN ENGLISH AND INDONESIAN VIA TRANSFER LEARNING USING XLM-ROBERTA," *International Journal of Advances in Soft Computing and its Applications*, vol. 13, no. 3, pp. 72–87, 2021, doi: https://doi.org/10.15849/ijasca.211128.06.

[36] A. Kumar and V. H. C. Albuquerque, "SENTIMENT ANALYSIS USING XLM-R TRANSFORMER AND ZERO-SHOT TRANSFER LEARNING ON RESOURCE-POOR INDIAN LANGUAGE," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, Sep. 2021, doi: https://doi.org/10.1145/3461764.

[37] F. Barbieri, L. E. Anke, and J. Camacho-Collados, "XLM-T: MULTILINGUAL LANGUAGE MODELS IN TWITTER FOR SENTIMENT ANALYSIS AND BEYOND," May 2022.

[38] A. Gaurav, B. B. Gupta, S. Sharma, R. Bansal, and K. T. Chui, "XLM-ROBERTA BASED SENTIMENT ANALYSIS OF TWEETS ON METAVERSE AND 6G," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 902–907. doi: https://doi.org/10.1016/j.procs.2024.06.110.

[39] E. Briakou and M. Carpuat, "DETECTING FINE-GRAINED CROSS-LINGUAL SEMANTIC DIVERGENCES WITHOUT SUPERVISION BY LEARNING TO RANK," Oct. 2020, [Online]. Available: https://arxiv.org/abs/2010.03662.

[40] K. Bali, J. Sharma, M. Choudhury, and Y. Vyas, "'I AM BORROWING YA MIXING ?' AN ANALYSIS OF ENGLISH-HINDI CODE MIXING IN FACEBOOK," in *Proceedings of the First Workshop on Computational Approaches to Code Switching*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 116–126. doi: https://doi.org/10.3115/v1/W14-3914.

[41] A. Joshi, P. Bhattacharyya, and M. J. Carman, "AUTOMATIC SARCASM DETECTION," *ACM Comput Surv*, vol. 50, no. 5, pp. 1–22, Sep. 2018, doi: https://doi.org/10.1145/3124420.