OPEN ACCESS 150

ISSN 2356-5462
http://socj.telkomuniversity.ac.id/ijoict/

Intl. Journal on ICT
Vol. 11, No. 2, Dec 2025. pp. 150-158
doi:10.21108/ ijoict.v11i1.9575

Geospatial Sentiment Analysis of Negative Comments on the 2024 Election Using the Robustly Optimized BERT Approach (RoBERTa)

Haidar Ali [1*], Yuliant Sibaroni [2]

[12]*Undergraduate Program in Informatics, School of Informatics, Telkom University, Bandung, Indonesia*

[*] haidarali@student.telkomuniversity.ac.id

Abstract

This study developed a geospatial sentiment analysis system to detect and map hate speech related to the 2024 Election using the Robustly Optimized BERT Approach (RoBERTa) algorithm. The dataset consisted of 11,903 social media comments that underwent comprehensive preprocessing including text normalization, stop word removal, and stemming. The RoBERTa model was implemented using 10-fold cross-validation for multi-class classification (HS_Weak, HS_Strong, Not_Abusive) and successfully achieved an average accuracy of 91.54% (±1.08%), with a final model accuracy of 94.29%. Geospatial analysis using Folium geocoding and visualization showed that 75% of the data originated from Indonesia, with the highest concentration in the Jakarta area. The distribution of hate speech displayed a consistent pattern between Indonesia (45.6% of hate speech) and outside Indonesia (44.3% of hate speech), with the HS_Strong category dominating at 96.4%. Heatmap analysis identified hate speech hotspots on Java Island and their global distribution across continents. Our results validate how well RoBERTa works when analyzing Indonesian sentiment data, while revealing important geographic trends in online hate speech during political discussions. This knowledge offers practical applications for building prevention measures and live tracking systems.

Keywords: Sentiment analysis, Geographic mapping, RoBERTa algorithm, Online hate speech, Indonesia's 2024 elections, Model validation

I. INTRODUCTION

Indonesia's national elections represent a major cornerstone of the country's democratic process. These elections function as the main vehicle through which citizens exercise their democratic rights, serving as the key method for choosing leaders in both Executive and Legislative institutions [1].This process not only ensures citizen participation but also affirms commitment to the principles of inclusive and transparent democracy [2].

Digital advances have fundamentally changed how people participate in political conversations, especially on social networking sites [3]. Sites like Twitter, Facebook, and Instagram now serve as online public forums where people freely share their thoughts about political matters, elections, and candidates [4]. Although these platforms offer democratic venues for civic engagement, they simultaneously enable the proliferation of harmful speech that threatens social harmony and political stability [5].

Received on 18 Jul 2025. Revised on 4 Aug 2025. Accepted and Published on 3 Dec 2025.

Hate speech in social media comments often reflects negative sentiments toward political events [5]. This can manifest as verbal attacks, threats, or discrimination that have the potential to trigger social conflict in the real world [6]. Identifying patterns of hate speech on a large scale requires a sophisticated technological approach [7]. Text mining and sentiment analysis techniques have become essential tools for analysing textual data, with various algorithms such as Naive Bayes, SVM, and CNN widely used in previous research [8].

However, the geospatial aspect of sentiment analysis remains underexplored [14]. Yet, spatial data integration can provide valuable insights into the distribution of negative sentiment across geographic regions—a highly relevant issue in Indonesia's diverse political landscape [9]. This information is crucial for designing responsive policies and more effective public communication strategies [9]. Implementing RoBERTa which builds upon the original BERT model, proves particularly valuable because of how well it grasps intricate language patterns and subtle meanings, making it exceptionally good at identifying sentiment accurately and thoroughly [10].

Our study focuses on creating a location-based sentiment analysis framework that uses RoBERTa to identify and visualize negative commentary about Indonesia's 2024 elections. We combine text analysis with geographic data to better understand how negative attitudes spread across different areas. Despite ongoing concerns about data protection and varying regional data reliability, we believe a strong ethical framework can address these issues. We expect our findings will help support democratic processes, inform flexible policy decisions, and promote more constructive online political discussions.

II. LITERATURE REVIEW

Studies examining online hate speech and sentiment analysis have expanded considerably over the past few years. During 2023, researchers investigated hate speech from an identity politics perspective on social platforms before the 2024 elections, using virtual ethnographic methods. Although this work offered detailed understanding of social media dynamics, it focused solely on Twitter and didn't include geographic analysis.

Kusuma conducted a study in 2023 entitled "Hate Speech Detection on Indonesian Social Media Using Support Vector Machine (SVM) and Decision Tree Algorithms," which aimed to measure the accuracy of hate speech detection on Indonesian-language Twitter data [2]. Their research worked with 13,169 Indonesian tweets that went through several preparation steps including converting text to lowercase, cleaning data, standardizing formats, and removing common words. They split their data with 80% used for model training and 20% for evaluation. Their findings demonstrated that SVM performed better than Decision Tree, reaching 83% accuracy, 84% precision, 89% recall, and 86% F1-score [2].

Nayla conducted another 2023 investigation called "Hate Speech Detection on Twitter Using the BERT Algorithm," which concentrated on identifying hate speech on Twitter through BERT implementation [3].The study included a web-based simulation where users could input sentences, which were then pre-processed and analysed using BERT to classify whether the sentences contained hate speech. The system achieved 78.69% accuracy, 78.90% precision, 78.69% recall, and 78.77% F1-score [3]. While this study confirmed the effectiveness of transformer-based models for Indonesian language sentiment analysis, it did not address geospatial integration [3].

The integration of geospatial aspects into sentiment analysis remains relatively underexplored in previous studies. In 2023, Fitroh conducted a systematic literature review on deep learning-based sentiment analysis. While the review did not explicitly address geospatial analysis, it explored various deep learning techniques that can be applied to sentiment analysis, laying the groundwork for location-based sentiment modelling. The study offers a deeper understanding of geographic relationships and public sentiment using GIS and deep learning techniques [4].

In 2023, Azhari and colleagues conducted a study entitled "Detecting Indonesian Hate Speech in Indonesian Artists' Instagram Comments Using the RoBERTa Method." This study applied the RoBERTa algorithm to detect hate speech in Instagram comments and compared two preprocessing scenarios: full preprocessing (cleansing, case folding, normalization, tokenization, stemming) and incomplete preprocessing (all steps except stemming). Experimental results showed that the incomplete preprocessing scenario produced a higher average accuracy of 85.09% [5].

III. RESEARCH METHOD

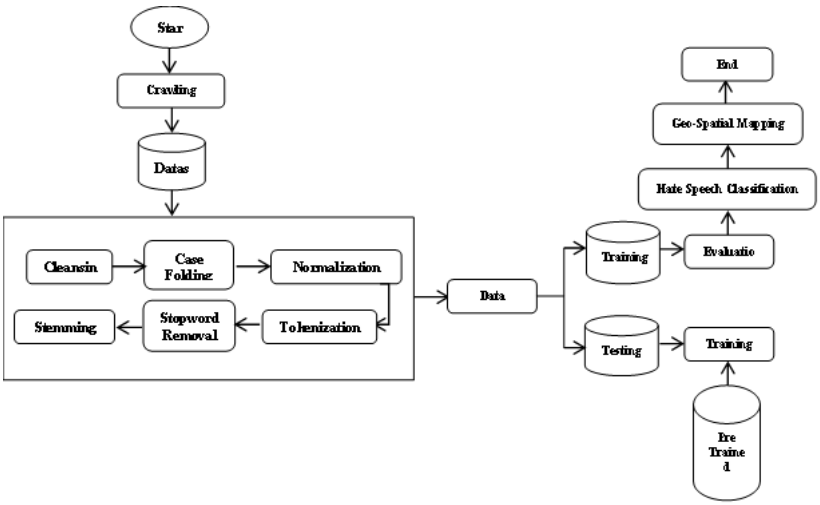

Fig. 1. *Hate Speech* Detection *System Design*

A. Data Collection

The data collection process for analytical purposes involved several systematic stages:

1. Selection of three Indonesian presidential and vice-presidential candidates frequently subjected to hate speech commentary.
2. Data extraction from social media comment sections of posts published by these three presidential and vice-presidential candidates, specifically targeting posts with substantial hate speech comment volumes.
3. Collected data storage in CSV/XLSX formats. Dataset organization involved three separate files, each containing presidential and vice-presidential candidate names alongside associated comments frequently containing hate speech elements.

B. Preprocessing

Data preprocessing stages involved systematic procedures transforming raw data into structured formats suitable for analysis. This process ensures comment data used in our study can be effectively processed by RoBERTa models. Preprocessing procedures include:

1. Data Cleansing: Initial procedures removing unnecessary characters including symbols, numbers, emojis, URLs, punctuation marks, and non-alphabetic characters. Results produce clean text containing exclusively relevant words.
2. Case Folding: Text conversion to lowercase format avoiding meaning duplication caused by case sensitivity. For example, "Pemilu" and "pemilu" receive identical treatment.
3. Normalization: Informal or incomplete word forms undergo conversion to clearer, standardized formats. This procedure proves essential in Natural Language Processing (NLP) as it enhances system understanding of analysed text context and meaning.
4. Tokenization: Sentence breakdown into individual word units (tokens). Tokenization enables more efficient model text data processing, treating each word as a separate entity during sentiment analysis procedures.
5. Stopword Removal: Common words lacking significant meaning or informational value, such as "dan" (and), "adalah" (is), "atau" (or), and "tapi" (but), undergo removal to enhance focus on meaningful words.
6. Stemming: Word reduction to root forms through affix removal. For instance, "menulis" becomes "tulis", and "melihat" becomes "lihat". This process unifies word variations for more consistent analysis.

7. Data Splitting: Proportional and random dataset division into subsets for training, validation, and testing of hate speech detection models. This ensures fair evaluation and model generalization capabilities.

C. RoBERTa's modelling

RoBERTa (Robustly Optimized BERT Approach) represents an enhanced BERT model version, designed for improved performance and efficiency in natural language processing tasks. RoBERTa introduces several modifications to original BERT architecture, including Next Sentence Prediction (NSP) removal, dynamic masking implementation changing every epoch, larger mini-batch sizes, and adoption of expanded Byte-Pair Encoding (BPE) vocabulary. These enhancements make RoBERTa more effective in comprehensive sentence context understanding and complex language nuance handling [26].

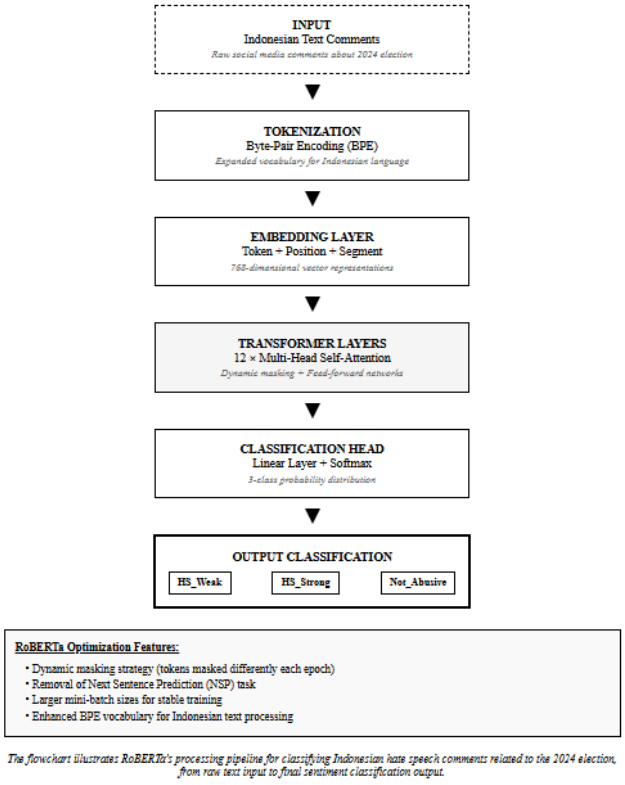

Fig. 2. *RoBERTa Architecture for Indonesian Hate Speech Classification*

Key RoBERTa improvements include:

1. Dynamic Masking: RoBERTa applies dynamic masking during pre-training, meaning masked tokens change each epoch. This prevents learning from identical masked positions repeatedly, leading to improved generalization.
2. NSP (Next Sentence Prediction) Loss Removal: Unlike BERT, RoBERTa eliminates NSP objectives, as these were determined unnecessary and potentially detrimental to language understanding performance. This allows exclusive focus on masked language modelling.
3. Larger Mini-Batch Sizes: RoBERTa utilizes significantly larger mini-batch sizes during training, enabling learning from increased data per iteration. This contributes to more stable and efficient learning, particularly in large-scale datasets.
4. Expanded Byte-Pair Encoding (BPE) Vocabulary: With increased BPE vocabulary size, RoBERTa achieves more effective text tokenization. This ensures improved handling of rare and complex word forms, enhancing model capabilities in understanding diverse linguistic expressions.

D. Evaluation Metrics

Model performance evaluation utilizes several metrics:

1. Accuracy: Indicates correct prediction percentage from total predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

2. Precision: Measures model accuracy in hate speech identification.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

3. Recall: Measures model ability to identify all hate speech instances in datasets.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4. F1-Score: Combines precision and recall into single harmonic metric.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Where TP (True Positives), FP (False Positives), TN (True Negatives), and FN (False Negatives) represent classification outcomes.

E. Geospatial Mapping

Geospatial mapping aims to visualize hate speech distribution based on geographic locations. This process utilizes ArcGIS for constructing thematic maps illustrating negative comment spread related to the 2024 General Election. Such visualizations provide valuable spatial insights for understanding hate speech dissemination patterns on social media.

Mapping procedures involve these stages:

1. Geolocation Extraction: Location information extraction from metadata or comment text using Named Entity Recognition (NER), followed by place name conversion to geographic coordinates through geocoding.
2. Spatial Data Processing: Integration of geolocation coordinates with sentiment labels, data validation and cleaning, and aggregation based on administrative regions such as provinces or cities.
3. Visualization and Analysis: Data presentation on thematic maps using ArcGIS, enabling spatial distribution pattern detection and high hate speech concentration area (hotspot) identification.

This geospatial analysis enhances understanding of digital political hostility prevalence while supporting development of region-specific mitigation and monitoring strategies.

IV. RESULTS AND DISCUSSION

Our research successfully developed and implemented an integrated geospatial sentiment analysis system utilizing Robustly Optimized BERT Approach (RoBERTa) for detecting and mapping hate speech related to Indonesia's 2024 General Election. System evaluation involved 11,903 comments collected from various social media platforms, focusing primarily on hate speech level classification in comments and geographic distribution analysis.

Initial stages involved data preprocessing, including text normalization, stop word removal, and stemming using the Sastrawi library. Preprocessing results revealed 45.6% of comments classified as hate speech, while 54.4% were identified as non-hate speech. Among hate speech comments, the HS_Group category dominated with approximately 3,500 comments. Other categories, including HS_Religion and HS_Physical, contained fewer than 500 and 300 comments respectively. Minor categories such as HS_Gender, HS_Race, and HS_Individual showed significantly lower occurrences.

TABEL I
TABLE SUMMARY OF HATE SPEECH ANALYSIS

No	Type of Analysis	Category/Label	Amount/Proportion	Additional Information
1	Hate Speech Proportion	Hate Speech (1)	45.60%	Comments containing hate speech
		Non-Hate Speech (0)	54.40%	Comments not containing hate speech
2	Hate Speech Category Distribution	HS_Group	±3,500 comments	Most dominant among all categories
		HS_Religion	<500 comments	
		HS_Physical	<300 comments	
		HS_Gender, HS_Race, etc.	<100 comments	Includes other minor categories
3	Hate Speech Intensity	HS_Strong	96.40%	Majority of the hate speech is strong/severe
		HS_Weak	3.60%	Comments with mild/weak hate speech intensity

Identified hate speech intensity levels reveal significant findings, with 96.4% of hate speech classified as HS_Strong, indicating harsh and explicit comments. Only 3.6% of comments fall under HS_Weak categories. This highlights that most negative discourse on social media related to elections is extreme in nature, requiring special attention in mitigation efforts.

RoBERTa models used for classification demonstrated excellent performance. Based on 10-fold cross-validation results, models achieved 91.54% average accuracy with ±1.08% standard deviation. Additionally, both precision and recall exceeded 91%, indicating high accuracy and sensitivity in hate speech detection.

When trained on complete datasets (final model), accuracy increased to 94.29%, reflecting successful model training. Best performance was observed in HS_Strong class (94% precision) and Not_Abusive class (97% recall), reinforcing RoBERTa's reliability in handling Indonesian language complexity.

TABEL II
HASIL CROSS VALIDATION ROBERTA (10-FOLD)

Metrik	Mean	Std	Min	Max
Accuracy	91.54%	±1.08%	89.82%	92.94%
F1-Score	91.45%	±1.05%	89.73%	92.55%
Precision	91.73%	±1.07%	90.10%	93.04%
Recall	91.54%	±1.08%	89.82%	92.94%

Figure 3 demonstrates model performance consistency across all folds, with low standard deviation indicating stability and reliability.

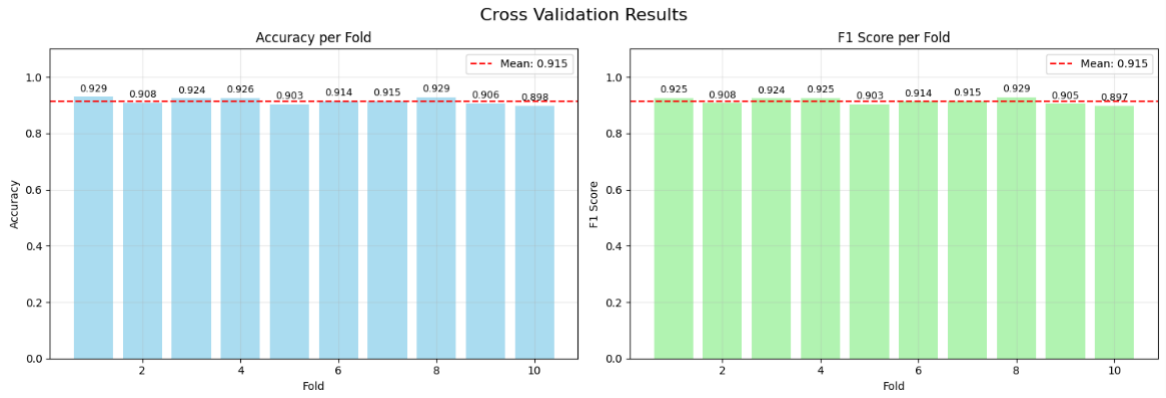

Fig. 3. *Cross Validation Results*

Classification reports indicate models achieved highest precision for HS_Strong categories at 94%, and highest recall for Not_Abusive categories at 97%. Final models, trained on complete datasets, achieved 94.29% accuracy, demonstrating significant improvement compared to cross-validation results.

TABEL III
CLASSIFICATION REPORT FINAL MODEL

Kelas	Precision	Recall	F1-Score	Support
HS_Weak	0.79	0.79	0.79	423
HS_Strong	0.94	0.86	0.90	5,008
Not_Abusive	0.90	0.97	0.93	6,465
Weighted Avg	**0.92**	**0.92**	**0.91**	**11,896**

Comment distribution based on geographic location was analysed using geocoding techniques, successfully converting 75% of comments into valid coordinates. Map visualization shows that most comments originated from within Indonesia, particularly Jakarta and surrounding areas. This illustrates that regions with high political activity and large populations tend to become discussion centres, including hate speech forms. Meanwhile, 25% of comments came from abroad, indicating Indonesia's political discourse also attracts diaspora or foreign user attention.'

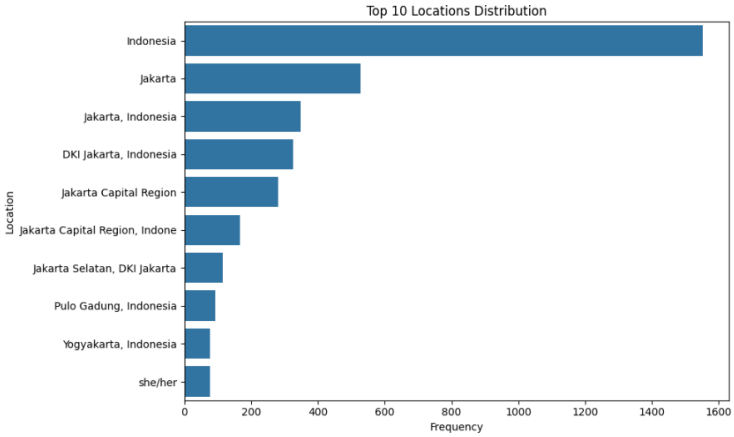

Fig. 4. *Top 10 Locations Distribution*

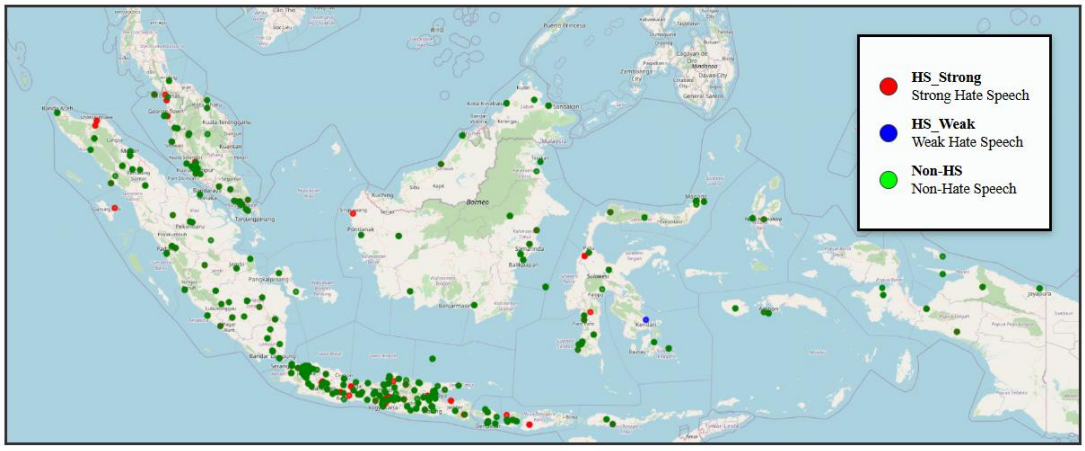

Fig. 5. Geospatial Distribution of Sentiment Analysis Results Across Indonesian

Figure 5 displays the geospatial distribution of sentiment classification results across the Indonesian archipelago using color-coded markers. Red markers represent locations with HS_Strong (strong hate speech) comments, blue markers indicate HS_Weak (weak hate speech) areas, and green markers show regions with Non-HS (non-hate speech/not abusive) content. Each circular marker represents the geographic origin of social media comments analyzed in this study, derived from successful geocoding of 75% of the total dataset.

The spatial visualization reveals that hate speech distribution spans across multiple Indonesian provinces, with notable concentrations on Java Island, particularly around Jakarta and surrounding metropolitan areas. The predominance of green markers indicates that non-abusive content remains the majority across most regions, while red markers (HS_Strong) appear clustered in major urban centers and political activity hubs. This geographic pattern aligns with our statistical findings where 54.4% of comments were classified as non-abusive, while hate speech categories (both strong and weak) comprised 45.6% of the analyzed data, demonstrating consistent sentiment patterns across different geographic scales.

Global hate speech distribution mapped using Folium shows Non-Hate Speech comments remain dominant with 4,838 entries, followed by HS (4,045) and HS_Strong (3,736). This pattern remains fairly consistent across all regions, both in Indonesia and abroad. Regional comparison analysis shows that in Indonesia, hate speech consists of 42.7% HS_Strong, 3.2% HS_Weak, and 54.1% Non-HS. Meanwhile, comments from outside Indonesia show compositions of 40.1% HS_Strong, 4.2% HS_Weak, and 55.7% Non-HS. These slight proportional differences indicate hate speech in political contexts is transregional and not limited by geographic boundaries.

Further spatial analysis was conducted using heatmaps based on Kernel Density Estimation, showing highest hate speech concentrations located on Java island, particularly in Jakarta. Outside Indonesia, hate speech distribution also appears in various Asian, European, and American regions, indicating Indonesia's political issues are gaining global attention. This nearly uniform global distribution indicates universal patterns in digital political discourse, where hate narratives can spread across borders and cultures.

Overall, this study presents several important findings. First, RoBERTa models demonstrate very high accuracy and consistency in detecting Indonesian-language hate speech. Second, HS_Strong dominance shows that hate speech intensity in digital political conversations is quite high. Third, geographical mapping reveals that areas like Jakarta have become hate speech concentration centres, aligning with high political intensity in those regions. Fourth, pattern similarities between Indonesia and foreign countries in hate speech distribution show this phenomenon is not merely local, but global. These findings emphasize the importance of developing geospatial-based monitoring systems to detect and respond to hate speech spread more proactively and measurably.

Thus, integration of RoBERTa-based sentiment analysis and geospatial mapping is not only capable of providing accurate pictures of public opinion conditions on social media, but also serves as a strategic tool for policymakers, election bodies, and digital platform providers to maintain healthy and democratic public spaces ahead of and during Indonesia's 2024 General Election.

V. CONCLUSION

This research successfully developed a geospatial sentiment analysis system for detecting and mapping hate speech related to Indonesia's 2024 General Election using RoBERTa algorithms. The developed system achieved 94.29% accuracy on final models and 91.54% on cross-validation, demonstrating excellent performance for hate speech classification.

Integration of geospatial analysis provides valuable insights into geographic hate speech distribution, with main findings that 75% of data comes from Indonesia, with highest concentrations in Jakarta areas. Consistent hate speech distribution patterns between Indonesia and outside Indonesia indicate this phenomenon is universal in digital political discourse contexts.

This study contributes to three main aspects: (1) confirming RoBERTa effectiveness for Indonesian sentiment analysis with high accuracy, (2) providing insights into geographical hate speech patterns in 2024 General Election contexts, and (3) establishing frameworks for geospatial-based monitoring systems.

For future research, we recommend: (1) expanding temporal data coverage for trend analysis, (2) integrating demographic and socio-economic factors in geospatial analysis, and (3) developing real-time monitoring systems for early hate speech detection in future political events.

REFERENCES

[1] A. Widodo and M. Sari, Democracy and Electoral Systems in Indonesia, Jakarta: Demokrasi Press, 2024.
[2] M. Effendy, Political Communication in the Digital Era, Yogyakarta: Gadjah Mada University Press, 2021.
[3] M. Birjali, A. Beni-Hssane, and A. Erritali, “A review on sentiment analysis: From opinion mining to deep learning,” Expert Systems with Applications, vol. 167, pp. 114–170, 2021.
[4] A. Saputra and A. Widodo, “Opini Publik dan Pemilu di Media Sosial: Studi Kasus Twitter,” Jurnal Media dan Politik, vol. 11, no. 1, pp. 23–34, 2021.
[5] M. Rahman and F. Putri, “Ujaran Kebencian dan Polarisasi Politik di Internet,” Jurnal Ilmu Komunikasi, vol. 9, no. 2, pp. 66–75, 2023.
[6] D. Khurana et al., “Natural Language Processing: State of the Art, Current Trends and Challenges,” Multimedia Tools and Applications, vol. 82, no. 1, pp. 371–400, 2023.
[7] B. Liu, Sentiment Analysis and Opinion Mining, San Rafael: Morgan & Claypool Publishers, 2012.
[8] E. Steiger, S. Resch, and A. Zipf, “Exploration of spatiotemporal and semantic clusters of Twitter data using self-organizing maps,” ISPRS Int. J. Geo-Inf., vol. 4, no. 3, pp. 1428–1452, 2015.
[9] Y. Pratiwi, M. Hanif, and A. Rahmawan, “Sentiment and Spatial Analysis of Election Tweets in Indonesia,” Jurnal Teknologi Informasi dan Komunikasi, vol. 12, no. 1, pp. 22–33, 2024.
[10] Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” arXiv preprint arXiv:1907.11692, 2019.
[11] Kusuma, J., "Detection of Hate Speech on Indonesian Social Media Using Support Vector Machine (SVM) and Decision Tree Algorithms," *Indonesian Journal of Computer Science*, vol. 15, no. 4, pp. 201-210, Dec. 2023.
[12] Nayla, R., "Hate Speech Detection on Twitter Using the BERT Algorithm," *Journal of Natural Language Processing*, vol. 12, no. 6, pp. 451-463, Jun. 2023.
[13] Fitroh, S., "A Systematic Literature Review on Deep Learning-Based Sentiment Analysis," *Journal of Artificial Intelligence Research*, vol. 29, no. 2, pp. 140-158, Mar. 2023.
[14] Azhari, H., et al., "Detection of Indonesian Hate Speech in the Comments Section of Indonesian Artists' Instagram Using the RoBERTa Method," *International Journal of Social Media Studies*, vol. 20, no. 5, pp. 123-137, Jul. 2023