VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i4.6645

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

SENTIMENT ANALYSIS OF GOVERNMENT ON TIKTOK AND X PLATFORMS WITH SVM AND SMOTE APPROACH

Dimar Pateman[1*]; Tri Ferga Prasetyo[2]; Harun Sujadi[3]

Informatics[1,2,3]
Universitas Majalengka, Majalengka, Indonesia[1,2,3]
https://unma.ac.id/ [1,2,3]
dimarpateman251@gmail.com[1*], triferga.prasetyo@gmail.com[2], harunsujadi@unma.ac.id[3]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *This study aims to analyze public sentiment toward the government on TikTok and X (formerly Twitter) using the Support Vector Machine (SVM) algorithm optimized with the Synthetic Minority Over-sampling Technique (SMOTE). Data were collected through keyword-based scraping of posts containing the word "pemerintah" (government) and processed using standard NLP pre-processing techniques. Results show that SVM combined with SMOTE significantly improves classification accuracy from 61% to 76% on TikTok, and from 74% to 86% on X. Word cloud analysis confirms these findings: TikTok content tends to reflect neutral and positive sentiments, while X contains predominantly negative expressions. These differences highlight platform-specific public discourse characteristics. The findings suggest that public communication strategies should be tailored accordingly: TikTok for positive narrative and outreach, X for monitoring feedback and criticism. This approach demonstrates the effectiveness of machine learning-based sentiment analysis in supporting data-driven public policy communication.*

Keywords: *sentiment analysis, SMOTE, support vector machine, TikTok, X.*

Intisari— *Penelitian ini bertujuan menganalisis sentimen publik terhadap pemerintah di TikTok dan X (Twitter) menggunakan algoritma Support Vector Machine (SVM) dengan teknik balancing SMOTE. Data dikumpulkan melalui scraping berbasis kata kunci "pemerintah" dan diproses menggunakan tahapan NLP standar. Hasil menunjukkan SVM+SMOTE meningkatkan akurasi klasifikasi secara signifikan: dari 61% ke 76% pada TikTok, dan dari 74% ke 86% pada X. Analisis word cloud mendukung temuan ini; opini di TikTok cenderung netral dan positif, sedangkan X lebih negatif dan kritis. Perbedaan ini mencerminkan karakteristik masing-masing platform. Temuan ini menyarankan strategi komunikasi publik yang disesuaikan: TikTok untuk edukasi dan promosi positif, X untuk pemantauan opini dan kritik kebijakan. Pendekatan ini menunjukkan efektivitas machine learning dalam menganalisis dinamika opini digital berbasis data.*

Kata Kunci: *analisis sentimen, SMOTE, support vector machine, TikTok, X.*

INTRODUCTION

Social media has emerged as a dominant arena for the public to express their opinions on government policies and public affairs. Platforms such as TikTok and X (formerly Twitter) have become critical venues for discourse, where users actively share their sentiments both positive and negative toward government actions and decisions [1]. With the exponential growth in social media usage, researchers and policymakers have shown increasing interest in leveraging artificial intelligence-based sentiment analysis to understand public opinion at scale [2]. To date, research comparing sentiment analysis between TikTok and X remains relatively limited. Therefore, research on sentiment analysis on social media platforms can serve as a trigger for further scientific inquiry, serving as a reference for sentiment classification using the Support Vector Machine

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6645

(SVM) algorithm with the implementation of SMOTE, which is still rarely used today. The results of this study will enhance the understanding of sentiment across social media platforms, particularly TikTok and X (Twitter), with implications ranging from technological to socio-political aspects. A report from[3] recorded that Indonesia has over 125 million TikTok users and 24 million X users, making both platforms the main venues for public policy discussions. Hence, this study sees potential problems that may arise due to differing sentiments among users on the two platforms. Accordingly, this study is a response to the social issues resulting from these differing sentiments on social media. Previous studies have shown that public opinion on social media can influence policies and public perception of the government[4].

Research by [5] revealed that social media is frequently used as a platform to express dissatisfaction with the government, which can affect public perception and political stability. For instance, a study by [6] indicated that government policies related to national debt received mixed reactions on X, with the majority of comments showing negative sentiment. However, most sentiment analysis studies in Indonesia still use conventional methods such as Naïve Bayes and lexicon-based analysis, which are less optimal for handling unstructured data and class imbalance [7]. Therefore, a more sophisticated approach is needed, such as Support Vector Machine (SVM) with the Synthetic Minority Over-sampling Technique (SMOTE), to improve sentiment prediction accuracy. Recent research shows that SMOTE improves SVM performance significantly in imbalanced sentiment datasets [8]. Another study also demonstrated that SVM combined with SMOTE yields better results than Decision Tree or Naïve Bayes in handling social sentiment data [9]. Research by [10] also demonstrated that implementing SMOTE in the SVM algorithm for analyzing public opinion on specific brands improved accuracy in managing imbalanced datasets.

Sentiment analysis is a natural language processing (NLP) technique used to identify and classify opinions in a text as positive, negative, or neutral. This technique's effectiveness depends on the algorithm used and the linguistic complexity of the platform being analyzed [11]. In the context of public policy, this analysis allows the government to understand public responses more objectively and based on data [12]. However, sentiment analysis in the Indonesian language still faces various challenges, such as the use of informal language, slang, and a mix of regional dialects [6]. These previous studies serve as a reference for the author to further explore sentiments on TikTok and X platforms. Thus, this study will be a unique and rarely found investigation, as it delves deeper into social media sentiment, which can be used as a reference for both the general public and policymakers.

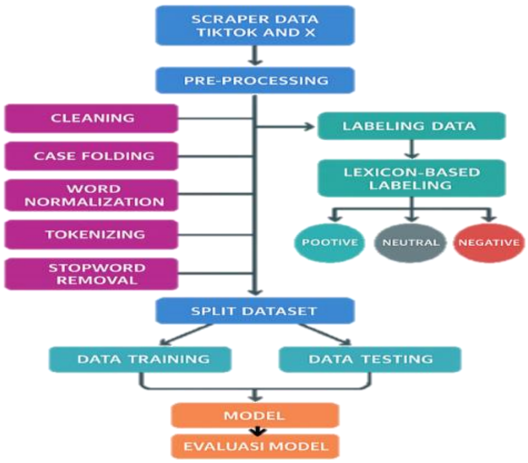
Previous studies have demonstrated that the Support Vector Machine (SVM) method yields high accuracy in classifying public sentiments on government policies shared via Twitter. For example, a study analyzing responses to the Ministry of Education's regulation on the prevention of sexual violence in higher education environments found that SVM achieved an accuracy of 80.3% using optimized parameters (C = 10; gamma = 0.1; kernel = RBF), confirming the algorithm's robustness in handling opinionated and imbalanced tweet data [13]. Nevertheless, this study uses a different locus, where the sentiment analysis focuses on the government using TikTok and X as platforms. The algorithm used also focuses on Support Vector Machine (SVM) with the implementation of SMOTE. Additionally, recent experiments with imbalanced datasets in Indonesia show that the combination of SVM and SMOTE can achieve classification accuracy above 90% for social media data [9]. These findings indicate that the combination of SVM and SMOTE is more effective in handling data imbalance in social media-based sentiment analysis. Therefore, this previous research reaffirms that sentiment comparisons across social media platforms, particularly TikTok and X, have not been extensively studied and/or no studies have yet been conducted focusing on the comparative sentiment between these two platforms.

Although research on sentiment analysis in social media has progressed, several gaps remain. Studies on sentiment analysis toward the government on TikTok are still limited, even though the platform has a very large user base and increasing trends in political discussion [14]. Moreover, exploration of the SMOTE method in Indonesian-language sentiment analysis is also minimal, despite its potential to help balance positive and negative sentiment data distribution [14]. Furthermore, comparisons between TikTok and X in sentiment analysis are still rarely conducted, although both platforms have different content characteristics and user demographics [15]. Additionally, research by [16] shows that the SVM method yields higher accuracy than Naïve Bayes in classifying Indonesian-language sentiment texts, strengthening the argument that this approach is more effective for social media sentiment analysis.

This study aims to analyze public sentiment toward the government based on data from TikTok and X using the Support Vector Machine model. Furthermore, this study explores the effectiveness of SMOTE in improving the accuracy of Indonesian-language sentiment classification models and identifies common patterns of criticism and support toward the government frequently appearing on social media. Theoretically, this study contributes to the development of more accurate sentiment classification approaches for the Indonesian language, particularly in the context of public opinion toward the government. Practically, the results of this study are useful for policymakers to understand public perception based on data, as well as for academics and practitioners in designing public opinion monitoring systems that are more adaptive to the characteristics of each social media platform.

MATERIALS AND METHODS

This research was conducted quantitatively with an exploratory approach to analyze public sentiment toward the government on two social media platforms: TikTok and X (formerly Twitter). The methods used include data collection, pre-processing, sentiment labeling, classification model development, and model performance evaluation.



Source: (Research Results, 2025)

Figure 1. Research Procedure

A. Data Collection

Data in this study were collected from two social media platforms, namely TikTok and X (formerly known as Twitter), during the period of January to March 2025. For the TikTok platform, comments were collected from videos containing the hashtag #pemerintah. This process was carried out automatically using Apify Scraper running on the apify.com platform. The collected comments were generally public and accessible to anyone.

Meanwhile, for the X platform, data were collected through a web crawling process using the Python programming language in the Google Colaboratory environment. Tweets were retrieved based on the keyword "pemerintah" using the auth_token from user accounts to access the Twitter API. Data collection was performed using the tweepy library or similar tools such as tweet-harvest, which allow researchers to filter tweets based on time, language, and keyword parameters. As previously stated in [17] Twitter, now known as X, is used because it is open and public, making data retrieval easier. Tweets can be accessed through the Twitter API by filtering based on keywords, dates, and other parameters relevant to the research topic.

The data obtained from both platforms was then saved in CSV format, with each entry including the text of the comment or tweet, the platform name, and the timestamp. The data was then used in the pre-processing and sentiment analysis stages. Data was collected from two social media platforms, namely TikTok and X (formerly Twitter).

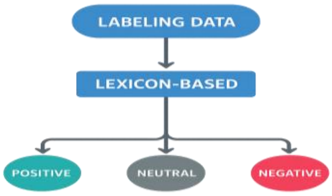
B. Data Pre-Processing

Pre-processing is done to improve the quality of the data and prepare it for sentiment analysis. This process includes text normalisation, punctuation removal, tokenisation and stemming, as described by [18], who showed that these steps significantly improved classification accuracy on Indonesian text:

1. Cleaning
   Remove URLs, mentions (@), hashtags (#), numbers, symbols, and punctuation.
   Example: @user https://link becomes user.
2. Case Folding
   Convert all characters in the text to lowercase to standardize and reduce redundancy due to capitalization.
   Example: Pemerintah Hebat becomes pemerintah hebat.
3. Word Normalization
   Convert informal words (slang, abbreviations) to standard forms using an Indonesian normalization dictionary.
   Example: yg, pdhl, anjg becomes yang, padahal, anjing.
4. Tokenizing
   Split text into tokens (words) using nltk.tokenize.
   Example: pemerintah hedon becomes [pemerintah, hedon].

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6645

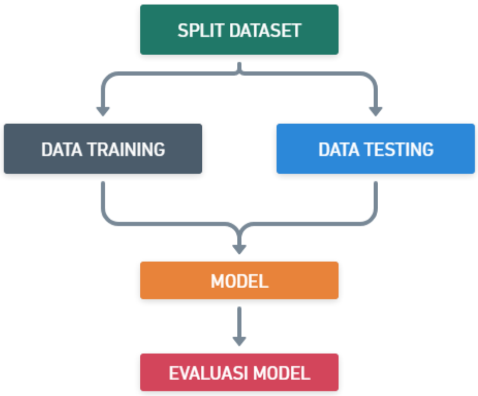
5. Stopword Removal
   Remove common words that have low sentiment significance, such as “yang”, “di”, “dan”.
   Example: pemerintah yang hedon becomes pemerintah hedon.
6. Stemming
   Convert words to their root form using the Sastrawi stemmer.
   Example: menyuguhkan becomes suguh
7. Pemberian Label (Labeling)
   Cleaned data were labeled using a lexicon-based approach with reference to the Indonesian Sentiment Lexicon [19]. Affective word scores in each sentence were used to determine polarity:
   Score > 0 = Positive
   Score < 0 = Negative
   Score = 0 = Neutral
   As a quality control step, manual validation using independent annotators is highly recommended, and to measure the level of agreement Cohen's Kappa coefficient is used. According to [20], a value ≥ 0.80 indicates excellent inter-rater reliability in text-based opinion coding.



Source: (Research Results, 2025)
Figure 2. Data Labeling Process

C. Modeling

After the labeling process, data were split into two subsets using stratified sampling to maintain the proportional distribution of sentiment labels: 80% for the training set and 20% for the testing set. For sentiment classification, the Support Vector Machine (SVM) algorithm remains the top choice. [21] showed that SVM with linear kernel and optimal parameter tuning outperformed other models in social media text classification in Indonesia. The processed data were then converted into numerical representation using the TF-IDF vectorization method to be accepted by machine learning algorithms. To address class imbalance issues, particularly the dominance of negative sentiment data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data. The SMOTE technique is used to balance the distribution of minority classes. [22] proved that the use of SMOTE in combination with SVM improved the accuracy of the model in sentiment analysis of Indonesian online transportation applications.



Source: (Research Results, 2025)
Figure 3. Split Dataset Process For Modeling And Evaluation

D. Model Evaluation

The sentiment classification model was evaluated using metrics of accuracy, precision, recall, and F1-score. Accuracy reflects the proportion of correct predictions, while precision and recall measure the correctness and completeness of predictions per class. F1-score balances both, especially under imbalanced class conditions. Evaluation was conducted under two scenarios: a standard SVM model and an SVM model with SMOTE applied to the training data to balance class distribution. Comparative evaluations were performed separately for the TikTok and X platforms to identify public sentiment differences toward the government on each medium. All implementation and evaluation processes were carried out using Python with the scikit-learn and imbalanced-learn libraries.

RESULTS AND DISCUSSION

A. Data Collection

This study utilized data from two popular social media platforms in Indonesia, namely TikTok and X (formerly Twitter), to capture public opinion regarding the government through user comments and public tweets. TikTok data were obtained using automated scraping via Apify, specifically from videos with the hashtag *#Pemerintah*, resulting in 5,684 relevant comments. Meanwhile, X data were collected through public API and Python-based crawling using the keyword “pemerintah”, producing 5,520 tweets. All data were stored in CSV format to facilitate further analysis. Table 1 and Table 2 present examples of raw data collected from both platforms before further processing.

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i4.6645

Table 1. Example of TikTok Comments Before Pre-Processing and Labeling

No	Comment
1	Gubernur Jawa Barat Dedi Mulyadi melarang orga...
2	Jabarnews.com - Menjelang Hari Raya Idulfitri ...
3	Pemerintah kerja keras untuk rakyat ? #fyp #fy...
4	Pembahasan RUU TNI dalam rapat Panja oleh Komi...
5	Mulai April 2025, kendaraan dengan STNK yang m...

Source: (Research Results, 2025)

Table 2. Example of Tweets from X Before Pre-Processing and Labeling

No	Tweets
1	Beda banget sama pemerintah Indonesia yg suguh...
2	Pemerintah Indonesia melalui lima kementerian ...
3	@weiihyuk Ini yg kurasakan ketika jd minoritas...
4	@irwndfrry Gak ad demo gak ada aksi masa tapi ...
5	Hidup kok disetir pemerintah hidup disetir cho...

Source: (Research Results, 2025)

B. Data Processing

Data from both platforms were processed through several pre-processing steps, including case folding, cleaning, tokenizing, normalizing, stopword removal, stemming, and sentiment labeling. Examples of post-processed data and sentiment labels are shown in Table 3 and Table 4, where each text entry is assigned a sentiment score and final label (positive, neutral, or negative)

Table 3. TikTok Comments After Pre-Processing and Labeling

No	Comment	Score	Sentiment
1	gubernur jawa barat dedi mulyadi larang organi...	2	Positive
2	jabarnewscom jelang raya idulfitri h gubernur ...	1	Positive
3	perintah kerja keras rakyat fyp fyou fyoupage ...	1	Positive
4	bahas ruu tni rapat panja komisi i dpr perinta...	-4	Negative
5	april kendara stnk mati sita polisi wewenang h...	1	Positive

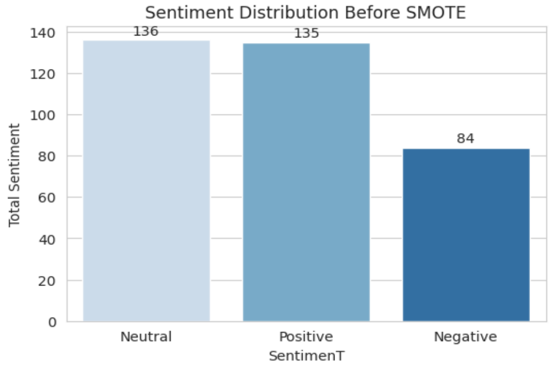
Source: (Research Results, 2025)

Table 4. Tweets from X After Pre-Processing and Labeling

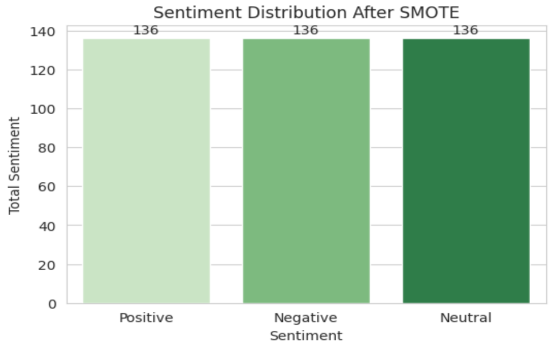
No	Tweets	Score	Sentiment
1	beda banget perintah indonesia suguh hedon ban...	0	Neutral
2	perintah indonesia menteri lembaga sepakat kua...	-1	Negative
3	rasa minoritas keluarga wkwkwk beda pro jokowi...	-1	Negative
4	demo aksi tindak bijak sembrono perintah dpr p...	-3	Negative
5	hidup setir perintah hidup setir choi arin	2	Positive

Source: (Research Results, 2025)

Sentiment distribution across both platforms before and after SMOTE implementation is shown in Figures 4, 5, 6, and 7.

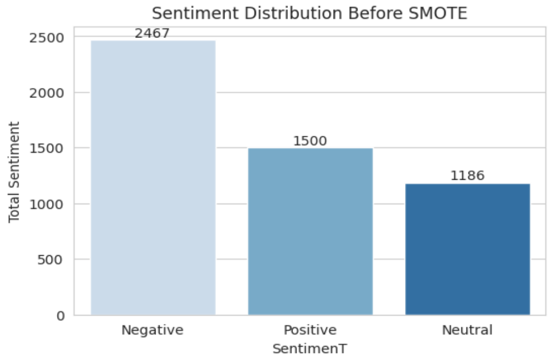


Source: (Research Results, 2025)
Figure 4. Sentiment Distribution on TikTok Before SMOTE



Source: (Research Results, 2025)
Figure 5. Sentiment Distribution on TikTok After SMOTE

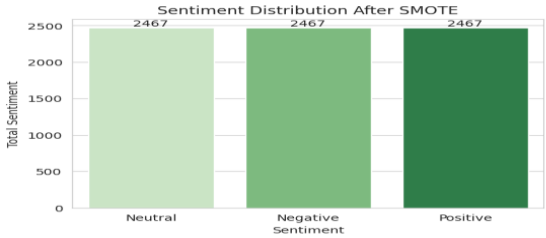
Figures 4 and 5 show TikTok comment distribution: initially dominated by positive and neutral sentiments, but balanced post-SMOTE (136 per class).



Source: (Research Results, 2025)
Figure 6. Sentiment Distribution on X Before SMOTE

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6645



Source: (Research Results, 2025)
Figure 7. Sentiment Distribution on X After SMOTE

Figures 6 and 7 illustrate tweet distribution on X: originally dominated by negative sentiment, balanced to 2,467 for each class after SMOTE.

C. Visualisasi Distribusi Sentimen

To complement the quantitative sentiment analysis, word cloud visualization was used to illustrate the most dominant terms found in public discourse related to government performance. This visual representation, based on cleaned and pre-processed data, effectively highlights the central themes of public attention through word frequency distribution. As demonstrated by [23], word clouds are widely used to present textual insights in a compact and interpretable form, particularly in the context of sentiment toward government policy on social platforms like Twitter.

Source: (Research Results, 2025)
Figure 8. Word Cloud of TikTok Comments About the Government

Figure 8 shows dominant words in TikTok comments such as “Indonesia”, “pemerintah”, “presiden”, “fyp”, and “prabowo”. This pattern indicates informative discourse focused on public figures and institutions, consistent with the positive and neutral sentiment distribution.

Source: (Research Results, 2025)
Figure 9. Word Cloud of Tweets from X About the Government

Conversely, Figure 9 highlights words like “perintah”, “negara”, “rakyat”, “hidup”, “kerja”, along with harsh terms such as “anjing”, “kontol”, and “tolol”. This reflects a more critical and emotional discourse on X, reinforcing the quantitative finding of predominant negative sentiment. Qualitatively, it shows that X users often use the platform to express dissatisfaction with the government.

D. Modeling

The cleaned and labeled dataset was split using stratified sampling to maintain balanced sentiment distribution between training data (80%) and test data (20%). The classification model was built using the Support Vector Machine (SVM) algorithm with a linear kernel, and SMOTE was applied to the training data to balance minority classes.

E. Model Evaluation

The following presents the evaluation results of the SVM model enhanced with SMOTE, including accuracy, precision, recall, and F1-score.

1. SVM Model Performance on the TikTok Platform

Table 5. Model Accuracy Before SMOTE

	Precision	Recall	F1-Score	Support
Negative	0.70	0.30	0.42	23
Neutral	0.63	0.74	0.68	23
Positive	0.57	0.75	0.65	28
Accuracy			0.61	74
Macro Avg	0.63	0.60	0.58	74
Weighted Avg	0.63	0.61	0.59	74

Source: (Research Results, 2025)

Table 6. Accuracy Model After SMOTE

	Precision	Recall	F1-Score	Support
Negative	0.79	0.84	0.81	31
Neutral	0.72	0.72	0.72	25
Positive	0.75	0.69	0.72	26
Accuracy			0.76	82
Macro Avg	0.75	0.75	0.75	82
Weighted Avg	0.76	0.76	0.75	82

Source: (Research Results, 2025)

Table 5 shows performance before SMOTE, while Table 6 shows the result after SMOTE. Accuracy improved from 61% to 76%, especially in the negative class, with recall increasing from 0.30 to 0.84.

2. SVM Model Performance on the X (Twitter) Platform

Table 7. Model Accuracy Before SMOTE

	Precision	Recall	F1-Score	Support
Negative	0.79	0.88	0.83	519
Neutral	0.55	0.40	0.47	252

Table 8. Model Accuracy Before SMOTE (Continue.)

	Precision	Recall	F1-Score	Support
Positive	0.77	0.79	0.78	290
Accuracy			0.74	1061
Macro Avg	0.70	0.69	0.69	1061
Weighted Avg	0.73	0.74	0.73	1061

Source: (Research Results, 2025)

Table 9. Model Accuracy After SMOTE

	Precision	Recall	F1-Score	Support
Negative	0.86	0.84	0.85	511
Neutral	0.80	0.85	0.82	483
Positive	0.91	0.88	0.89	483
Accuracy			0.86	1477
Macro Avg	0.86	0.86	0.86	1477
Weighted Avg	0.86	0.86	0.86	1477

Source: (Research Results, 2025)

Tables 7 and 8 show similar improvement on X(Twitter): accuracy rose from 74% to 86%, with significant gains in the neutral and positive classes.

3. Comparison of Public Opinion Patterns on TikTok and X (Twitter)

This study successfully identified and compared public sentiment patterns toward the government on two major Indonesian social media platforms: TikTok and X (formerly Twitter), using the Support Vector Machine (SVM) algorithm optimized with the Synthetic Minority Over-sampling Technique (SMOTE). In terms of model performance, the evaluation results show a significant improvement in classification accuracy after applying SMOTE. On TikTok, accuracy increased from 61% to 76%, and on X from 74% to 86%. This enhancement was also reflected in the F1-score average across both platforms, indicating improved balance and classification ability, particularly for minority classes such as negative and neutral sentiments that were previously underrepresented.

Substantive findings revealed platform-specific sentiment trends:

a. TikTok is dominated by positive (38%) and neutral (41%) sentiments. Its casual and visual nature leads to more appreciative expressions.
b. X(Twitter) is dominated by negative sentiment (49%), reflecting its role as a channel for social critique and political expression.

These differences underscore the influence of platform characteristics on public expression and have direct implications for the development of targeted public communication strategies.

F. Policy Implications

The findings of this study offer actionable insights for policymakers, especially in designing adaptive, data-driven public communication strategies that reflect the unique characteristics of each digital platform.

1. X (Twitter): Real-Time Monitoring of Public Criticism

Given its dominance in negative sentiment, X can serve as a real-time feedback and monitoring channel. Government agencies and public relations departments can utilize this platform to detect early signs of controversy and respond promptly. Examples of practical strategies include:

a. Hosting Twitter Space discussions moderated by ministry officials to address public criticism around controversial policies (e.g., energy subsidies).
b. Publishing official infographics under viral tweets to counter misinformation.
c. Implementing a machine learning–based digital listening system for continuous sentiment monitoring.

2. TikTok: A Strategic Platform for Positive Narratives and Policy Education

With its informal and engaging nature, TikTok is ideal for building positive sentiment and increasing policy literacy, especially among youth. Its visual format enables creative storytelling that makes government programs more relatable. Examples of narrative strategies:

a. “60 Seconds with the Minister” short videos explaining key government programs.
b. User testimonials – e.g., a citizen sharing how a government initiative changed their life.
c. Interactive campaigns like #SmartChoosePolicy to promote civic awareness.

Such initiatives can shift perception and build public trust through channels that resonate with the digital-native population.

3. Toward a Cross-Platform, Data-Driven Public Opinion Monitoring System

Beyond communication strategies, this research demonstrates the value of sentiment analysis as a foundation for a public opinion monitoring system. The government could institutionalize this by:

a. Establishing a centralized sentiment monitoring unit.
b. Facilitating inter-ministerial collaboration on real-time issue tracking.
c. Formulating responsive and empathetic policies based on digital evidence.

By embracing data-driven communication governance, public institutions can become more transparent, participatory, and adaptive in responding to evolving public discourse across social platforms.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6645

CONCLUSION

This study demonstrates that public perception of the government varies significantly between TikTok and X (formerly Twitter). TikTok predominantly features positive (38%) and neutral (41%) sentiment, aligning with its nature as an entertainment-based platform that encourages appreciation and lighter content. In contrast, X is dominated by negative sentiment (49%), functioning as a space for public criticism and policy-related discourse. The use of Support Vector Machine (SVM) optimized with the Synthetic Minority Over-sampling Technique (SMOTE) significantly improved classification performance. Accuracy increased from 61% to 76% on TikTok and from 74% to 86% on X. This indicates that SMOTE effectively addressed the sentiment class imbalance and enhanced the model's ability to detect underrepresented categories such as negative and neutral sentiments.

Word cloud visualizations further supported the quantitative results, revealing that expressions on X were more emotionally charged and critical, whereas those on TikTok were more informational and neutral. These findings have strategic implications for public sector communication. Policymakers may consider using TikTok as a platform for positive policy narratives and civic education, while employing X as a tool for real-time monitoring of public opinion and early detection of potential public trust issues. This study affirms that machine learning–based sentiment analysis can inform data-driven communication strategies that are responsive and adaptive to the digital public sphere.

REFERENCE

[1] T. Dewi Salma, M. Ferdi Kurniawan, R. Darmawan, and A. Basri, "Analisis Sentimen Berbasis Transformer: Persepsi Publik terhadap Nusantara pada Perayaan Kemerdekaan Indonesia yang Pertama," vol 9, no 2, pp. 757-764, Jan. 2025, doi: https://doi.org/10.35870/jtik.v9i2.3535

[2] Y. Rum Zattayu Mustopo, "Analisis Sentimen Proyek Strategis Nasional Food Estate Menggunakan Algoritma Naïve Bayes, Logistic Regression dan Support Vector Machine," vol 9 no 2, pp. 485-494, 2025, doi: https://doi.org/10.35870/jtik.v9i2.3312

[3] We Are Social Indonesia, "DIGITAL 2023 INDONESIA," 2023. Accessed: Mar. 24, 2025. [Online]. Available: https://wearesocial.com/id/blog/2023/01/digital-2023/

[4] S. Widiasih, F. Julina, and I. Septiani Salsabila, "Analisis Sosial Media Pemerintah Daerah Di Indonesia Berdasarkan Respons Warganet," *Jurnal Ilmiah Riset dan Pengembangan |*, vol. 9, Dec. 2024.

[5] A. Sadat, H. Lawelai, and) Ansar Suherman, "Analisis Sentimen Media Sosial: Hate Speech Kepada Pemerintah Di Twitter," *PRAJA: Jurnal Ilmiah Pemerintahan*, vol. 10, p. 2022, Feb. 2022.

[6] F. Rachmawati, U. Azmi, and R. Azwarini, "Comparison of Lexicon-based Methods and Bidirectional Encoder Representations for Transformers Models in Sentiment Analysis of Government Debt Market Movements," *International Journal of Engi-neering and Computer Science Applications (IJECSA)*, vol. 4, no. 1, pp. 13–28, 2025, doi: 10.30812/ijecsa.v4i1.4832.

[7] A. A. Al Kaafi, S. Suparni, H. Rachmi, A. Maulana, and R. Nurtriani, "Optimizing Twitter Sentiment Analysis on Tapera Policy Using SVM and PSO," *sinkron*, vol. 9, no. 1, pp. 167–176, Jan. 2025, doi: 10.33395/sinkron.v9i1.14227.

[8] Y. A. Singgalen, S. Y. Wahyuningtyas, Y. E. Widodo, M. N. A. Dasra, and R. W. Setiawan, "Travel Vlog Reviews: Support Vector Machine Performance in Sentiment Classification," *Ingenierie des Systemes d'Information*, vol. 30, no. 1, pp. 101–110, Jan. 2025, doi: 10.18280/isi.300109.

[9] P. P. Putra, M. K. Anam, A. S. Chan, A. Hadi, N. Hendri, and A. Masnur, "Optimizing Sentiment Analysis on Imbalanced Hotel Review Data Using SMOTE and Ensemble Machine Learning Techniques," *Journal of Applied Data Sciences*, vol. 6, no. 2, pp. 936–951, May 2025, doi: 10.47738/jads.v6i2.618.

[10] Dewi ArdikaTeresia and Mailoa Evangs, "Perbandingan Implementasi Metode SMOTE pada Algoritma Support Vector Machine (SVM) dalam Analisis Sentimen Opini Masyarakat Tentang Mixue Artikel Ilmiah," Jakarta, Apr. 2023.

[11] J. O. Leandro and M. I. Fianty, "Evaluation of Sentiment Analysis Methods for Social Media Applications: A Comparison of Support Vector Machines and Naïve Bayes," Tanggerang, Apr. 2025. doi: http://dx.doi.org/10.62527/joiv.9.2.2905.

[12] Faisal Muhammad *et al.*, "A Hybrid MOO, MCGDM, and Sentiment Analysis Methodologies for Enhancing Regional Expansion Planning: A Case Study Luwu -

Indonesia.," *International Journal of Mathematical Engineering and Management Sciences*, vol. 10, pp. 163–188, Jan. 2025, doi: 10.33889/IJMEMS.2025.10.1.010.

[13] R. A. Saemani and N. Setiyawati, "Analisis Sentimen Permendikbud Pencegahan Dan Penanganan Kekerasan Seksual Di Lingkungan Perguruan Tinggi Menggunakan Support Vector Machine (SVM)," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 8, no. 1, pp. 65–71, Aug. 2022, doi: 10.33480/jitk.v8i1.2807.

[14] M. I. Prayugah and N. Ariyanti, "Analisis Sentimen Publik Pada Pemerintah Dalam Serangan Ransomware Dengan Pendekatan Smote," *Journal Of Information Systems And Informatics Engineering*, vol. 8, no. 2, pp. 333–343, Nov. 2024, doi: 10.35145/joisie.v8i2.4764.

[15] Y. O. Sihombing, "Optimasi Model IndoRoBERTa-Base Untuk Klasifikasi Sentimen dan Emosi Pada Komentar Publik Twitter BKN," Institut Teknologi Sepuluh Nopember, 2025.

[16] A. Muzaki and A. Witanti, "Sentiment Analysis Of The Community In The Twitter To The 2020 Election In Pandemic Covid-19 By Method Naive Bayes Classifier," *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 2, pp. 101–107, Mar. 2021, doi: 10.20884/1.jutif.2021.2.2.51.

[17] Alisya Mutia Mantika, Agung Triayudi, and Rima Tamara Aldisa, "Sentiment Analysis on Twitter Using Naïve Bayes and Logistic Regression for the 2024 Presidential Election," *SaNa: Journal of Blockchain, NFTs and Metaverse Technology*, vol. 2, no. 1, pp. 44–55, Feb. 2024, doi: 10.58905/sana.v2i1.267.

[18] W. Maharani, H. Daud, N. Muhammad, and E. A. Kadir, "Leveraging Social Media Data for Forest Fires Sentiment Classification: A Data-Driven Method," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 3, pp. 392–407, Oct. 2024, doi: 10.20473/jisebi.10.3.392-407.

[19] D. P. Alamsyah, T. Arifin, Y. Ramdhani, F. A. Hidayat, and L. Susanti, "Classification of Customer Complaints: TF-IDF Approaches," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, 2022, pp. 1–5. doi: 10.1109/CONIT55038.2022.9848056.

[20] A. Gandhi *et al.*, "Hate speech detection: A comprehensive review of recent works," *Expert Syst*, vol. 41, no. 8, p. e13562, 2024, doi: https://doi.org/10.1111/exsy.13562.

[21] Y. A. Singgalen, "Comparative Analysis of DT and SVM Model Performance with SMOTE in Sentiment Classification," 2024, doi: 10.30865/klik.v4i5.1828.

[22] V. H. Pranatawijaya, N. N. K. Sari, R. A. Rahman, E. Christian, and S. Geges, "Unveiling User Sentiment: Aspect-Based Analysis and Topic Modeling of Ride-Hailing and Google Play App Reviews," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 3, pp. 328–339, Oct. 2024, doi: 10.20473/jisebi.10.3.328-339.

[23] M. T. Uliniansyah *et al.*, "Twitter dataset on public sentiments towards biodiversity policy in Indonesia," *Data Brief*, vol. 52, Feb. 2024, doi: 10.1016/j.dib.2023.109890.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri