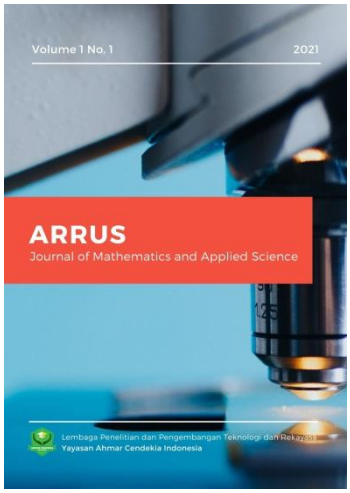
https://doi.org/10.35877/mathscience5586

OPEN ACCESS

ISSN : 2776-7922 (Print) / 2807-3037 (Online)



*Corresponding author: Agung Tri Utomo, Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Makassar, Indonesia

E-mail: agungtriutomo@unm.ac.id

RESEARCH ARTICLE

Sentiment Analysis of TikTok Comments on Law Enforcement Misconduct Cases Using BiLSTM with RoBERTa Labeling

Ajeng Windhyantika Dwi Ramadhani, Agung Tri Utomo*, Patonangi M., Emanuel Gepetrix Excelcis Badin,& Alfia Fharanita[1]

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Makassar, South Sulawesi, Indonesia.

Abstract: This study examines the sentiment distribution of TikTok comments on four law enforcement misconduct cases, explores how the comments relate to officer accountability and expectations of government responsiveness, and evaluates sentiment classification with a bidirectional long short-term memory (BiLSTM) network. The study uses a quantitative approach with comments from four purposively selected TikTok posts. Web scraping collects 3,685 comments, which decrease to 3,446 after the removal of empty and duplicate texts. RoBERTa assigns negative, neutral, and positive labels, and the normalized texts with these labels train the BiLSTM. A stratified split reserves 20% of the data for testing, and random oversampling balances the classes in the training data. The labeling results show negative sentiment at 72.37%, followed by positive at 16.37% and neutral at 11.26%. On 690 test comments, the BiLSTM obtains an accuracy of 71.01% and a macro-F1 of 0.580. The accuracy is lower than the 72.32% of a majority-class baseline, but the macro-F1 is higher than the baseline of 0.280. The findings suggest that the comments express criticism, demands for sanctions, and sympathy toward victims, so positive sentiment does not always signal support for officers. Recognition of the neutral and positive classes remains limited, and the automatic labels lack human validation.

Keywords: Sentiment analysis, Officer accountability, TikTok, RoBERTa, BiLSTM

1. INTRODUCTION

Public trust in law enforcement institutions is an important foundation for the rule of law. Indonesia, as a state based on law under Article 1 paragraph (3) of the 1945 Constitution, places fair treatment of citizens within the responsibility of the state. In this context, the actions of officers relate not only to the exercise of authority but also to the legitimacy of the institutions they represent. A survey by Lembaga Survei Indonesia conducted on 31 March–4 April 2023 shows that trust in the police stands at 60%, lower than trust in the Attorney General's Office at 72% and the Corruption Eradication Commission at 66% (Lembaga Survei Indonesia, 2023). This finding provides context for understanding public responses to officer conduct, although it does not directly describe the level of trust during the study period.

In early 2026, four cases involving officers attract wide attention on TikTok: an assault that causes the death of a student in Tual, a rape case involving two members of the Indonesian National Police (Polri), the alleged assault of an ice cake vendor in Kemayoran, and the

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5586

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

shooting of a teenager in Makassar (Table 1). These events raise questions about the use of authority, the protection of citizens, and institutional accountability. Public responses to these cases are relevant to study because they show how social media users evaluate officer conduct and what they expect from government handling.

Based on the authors' review of the literature, studies on the use of RoBERTa labels to train a BiLSTM on Indonesian-language TikTok comments about officer misconduct remain limited. To address this gap, the study uses RoBERTa to label comments from the four cases and trains a BiLSTM on the labeled data. The study aims to analyze the sentiment distribution across the four cases, to examine expressions related to officer accountability and expectations of government responsiveness, and to evaluate the ability of the BiLSTM to classify comments based on RoBERTa labels. The results are expected to describe how TikTok users respond to the cases studied and to inform public communication and more responsive and accountable complaint handling. The interpretation is limited to the comments analyzed, and the sentiment does not serve as a direct measure of national public trust or of the effectiveness of the government response.

2. Literature Review

Social media allows events involving officers to develop into open public conversations. Mohler et al. (2021) show that exposure to videos of police–citizen interactions can affect perceptions of police legitimacy, especially when the interactions are portrayed negatively. The comment section becomes a space where users express criticism, anger, support, and empathy for victims. These expressions help in understanding sentiment tendencies and demands for accountability, without treating them as representative of the whole society. Sentiment analysis helps group responses in large numbers systematically so that emerging opinion patterns can be examined further (Wankhade et al., 2022; Zhang et al., 2018).

TikTok comments are short, informal, and often contain expressions whose meaning depends on context. For this reason, the study uses RoBERTa to label comments as negative, neutral, or positive. As a transformer-based model, RoBERTa learns word representations from sentence context, so labeling does not rely only on the presence of words with a positive or negative tone (Liu et al., 2019). The development of Indonesian language resources and models through IndoLEM and IndoNLU also shows the relevance of contextual approaches for Indonesian sentiment analysis tasks (Koto et al., 2020; Wilie et al., 2020).

The labels produced by RoBERTa then serve as training targets for the BiLSTM. The BiLSTM architecture learns relationships between words from two directions, so the context before and after a word is considered in classification (Zhang et al., 2023). Thus, RoBERTa serves for labeling, whereas the BiLSTM learns sentiment patterns in the labeled comments. However, automatic labels can still contain errors, so classification results need to be interpreted with the quality of those labels in mind (Northcutt et al., 2021; Song et al., 2023). Evaluation also needs to consider class imbalance so that performance assessment is not determined only by the ability to recognize the dominant sentiment (Buda et al., 2018; Johnson & Khoshgoftaar, 2019).

3. Research Method and Materials

3.1. Research Design and Data Source

This study uses a quantitative approach to analyze public comment sentiment on officer misconduct cases and its relation to officer accountability and government responsiveness. The research stages include data collection, preprocessing, sentiment labeling with RoBERTa, BiLSTM training based on those labels, and model evaluation. Sentiment falls into three classes: negative, neutral, and positive.

The data come from comments on four TikTok posts that are purposively selected based on their relevance to cases involving officers (Table 1). Collection through web scraping with Apify Client yields 3,685 comments stored in CSV format. The analysis focuses on the

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5586

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

comment text, as illustrated in Section 4.1. Usernames are not used, and the identities of individuals in the case descriptions are abbreviated to protect privacy.

Table 1. Summary of the Four Officer Misconduct Cases Analyzed

Case	Parties involved	Event summary	Institutional follow-up	Source
Tual, Maluku	Brimob (Mobile Brigade) member, Maluku Regional Police (MS); victim is a 14-year-old Islamic junior high school (MTs) student (AT)	Assault with a helmet that ends in the victim's death	Dishonorable discharge (PTDH)	Liputan6.com (2026a)
Rape by Polri members	Two Polri members; victim is an 18-year-old female teenager (C)	Declared guilty of a serious violation and disgraceful conduct in a Polri Code of Ethics Commission hearing	PTDH on 6 February 2026	Deadline.co.id (2026)
Kemayoran, Central Jakarta	50-year-old ice cake vendor; TNI and Polri members	The vendor is accused of selling ice made of sponge and reports being surrounded, hit, and kicked	Dokpol of Metro Jaya Regional Police: examination finds the product contains no hazardous substances	Liputan6.com (2026b)
Makassar, South Sulawesi	Iptu N (Head of the Criminal Investigation Unit, Panakukang Sector Police); 18-year-old teenage victim	Arrest of a jelly-gun player accompanied by a warning shot; the firearm discharges and the victim dies	Awaiting autopsy results (according to news reports)	Tvonenews.com (2026)

3.2. Data Preprocessing

Preprocessing prepares the comments for labeling and model training. The original comment text is retained for sentiment labeling with RoBERTa, whereas the normalized version serves as the BiLSTM input, with the RoBERTa labels as training targets. Normalization includes case unification, cleaning of punctuation and repeated characters, and adjustment of non-standard words. Stopword removal and stemming are not applied so that the elements that form sentiment meaning remain intact. Empty and duplicate comments are then removed.

3.3. Sentiment Labeling with RoBERTa

Labeling uses a pre-trained Indonesian RoBERTa-based sentiment model accessed through Hugging Face Transformers, without further training. The model classifies comments as negative, neutral, or positive based on sentence context. This approach allows the meaning of comments to be considered contextually, so sentiment determination does not depend only on particular words (Liu et al., 2019). Each prediction comes with a confidence value as an indicator of model certainty. The automatic labels are treated as a silver standard that can still contain errors.

3.4. Data Splitting and Class Balancing

The data are split in a stratified manner so that class proportions remain represented. A total of 20%, or 690 comments, is used for testing, and the remainder is divided into training and validation data. Class imbalance is handled through random oversampling on the training data to broaden the representation of minority classes (Lemaître et al., 2017). The validation and test data keep the original distribution so that the evaluation reflects the condition of the research data.

3.5. Model Architecture and Training

The sentiment classifier is a BiLSTM built with TensorFlow and Keras. The model consists of an embedding layer, a BiLSTM layer, dropout, and a softmax output layer for the three sentiment classes. The BiLSTM learns relationships between words from two directions so

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5586

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

that the comment context is considered in classification. Training uses the Adam optimizer, while dropout, early stopping, and learning rate adjustment control overfitting. Evaluation on the test data uses accuracy, precision, recall, and F1-score. Table 2 presents the model configuration.

Table 2. Architecture and Training Configuration of the BiLSTM Model

Parameter	Setting
Optimizer	Adam
Loss function	Sparse categorical crossentropy
Batch size	32
Dropout rate	Spatial dropout 0.3; LSTM dropout 0.3; dropout before the dense layer 0.5
Validation data proportion	15% of the data after the test split, equal to about 12% of the total data
Early stopping	Monitors validation loss, patience 5, restores the best weights
Learning rate adjustment	ReduceLROnPlateau: factor 0.5, patience 2, minimum learning rate 0.00001
Random seed	42

3.6. Model Evaluation

Model performance is evaluated using the confusion matrix, accuracy, precision, recall, and F1-score. Accuracy shows the overall correctness of predictions, whereas precision and recall describe the ability of the model to recognize each sentiment class. The F1-score summarizes the balance between the two, while the macro-F1 gives equal weight to the three classes (Grandini et al., 2020).

$$Accuracy = \frac{\sum_{k=1}^{K} n_{kk}}{N} \quad (1)$$

$$P_k = \frac{n_{kk}}{\sum_{i=1}^{K} n_{ik}} \quad (2)$$

$$R_k = \frac{n_{kk}}{\sum_{j=1}^{K} n_{kj}} \quad (3)$$

$$F1_k = 2 \times \frac{P_k \times R_k}{P_k + R_k} \quad (4)$$

where n_{ij} is the number of test comments with actual class i and predicted class j, K is the number of classes, and N is the total number of test comments. The analysis is complemented by Wilson 95% confidence intervals, Cohen's kappa coefficient to measure the agreement of predictions with RoBERTa labels, and a chi-square test to assess the sentiment distribution. The performance of the BiLSTM is also compared with a baseline that always predicts the majority class so that the evaluation accounts for the imbalance in the number of comments across classes. Thus, the evaluation results describe the ability of the model to learn sentiment patterns based on RoBERTa labels on the test data.

4. Results and Discussion

4.1. Preprocessing Results

Preprocessing yields 3,446 comments from the initial 3,685 comments after removing 52 empty comments and 187 duplicate comments. Most comments are short, with a median of 7 words, so the input limit of 32 tokens accommodates most texts in full.

Table 3 presents examples of original text, normalization results, and RoBERTa labels. Some non-standard forms remain, and some labels do not match the meaning of the comments, such as the expression "DEMI ALLAH GA TERIMA" ("I swear to God, I do not accept this"), which receives a positive label. The table also shows that positive sentiment can take the form of sympathy for victims, so it cannot be interpreted directly as support for officers. Therefore, the interpretation of sentiment needs to consider the context and the target addressed in the comment.

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5586

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Table 3. Examples of Original Text, Normalized Text, and RoBERTa Labels

No.	Original text	Normalized text	English translation of the original text	Label	Confidence
1	PECAT!!!!!!!!!!	pecat	DISMISS HIM!!!!!!!!!!	Negative	0.999
2	DEMI ALLAH GA TERIMA	demi allah tidak terima	I SWEAR TO GOD, I DO NOT ACCEPT THIS	Positive	0.707
3	semog bapak penjual es, rejekinya berlimpah aminnn	semog bapak penjual es rejekinya berlimpah amin	May the ice vendor's fortune be abundant, amen	Positive	0.960
4	maaf tapi aku setuju smaa mreka karna es nya bkin batuk	maaf tapi aku setuju smaa mreka karena es nya bkin batuk	Sorry, but I agree with them because the ice makes people cough	Negative	0.412
5	logikanya aja masa spons bisa dikunyah	logikanya saja masa spons bisa dikunyah	Just use logic, how can a sponge be chewed	Positive	0.925
6	Aparat apa seperti ini?	aparat apa seperti ini	Is this what officers are like?	Negative	0.999
7	sumpah itu abis di peras baru dari tangan ny lgi lansung di mkn	sumpah itu habis di peras baru dari tangannya lagi langsung di makan	I swear, it was just squeezed and then eaten straight from his hand	Negative	0.981

4.2. Sentiment Distribution

Table 4 shows that the RoBERTa labeling is dominated by negative sentiment at 72.37%, followed by positive sentiment at 16.37% and neutral sentiment at 11.26%. The negative class also has the highest mean confidence, although this value does not guarantee label correctness.

Table 4. Distribution of RoBERTa Labels

Class	Count	Proportion (%)	Wilson 95% CI (%)	Mean confidence	Confidence < 0.7 (%)
Negative	2,494	72.37	70.86–73.84	0.956	5.0
Neutral	388	11.26	10.25–12.36	0.855	21.9
Positive	564	16.37	15.17–17.64	0.872	16.0
Total	3,446	100.00	–	–	–

The dominance of negative sentiment is consistent with the experimental finding that negative videos reduce the perceived legitimacy of the police the most (Mohler et al., 2021) and with the lower position of the police compared with two other institutions in the LSI survey (Lembaga Survei Indonesia, 2023). The three sources measure different things. The survey assesses the trust of respondents selected through sampling, whereas comments are reactions of users who choose to react after watching an emotionally charged event, and user characteristics and platform algorithms shape who comments (Olteanu et al., 2019). The figure of 72.37% therefore describes the tone of conversation in these four cases, not the level of national public trust. The small neutral class (11.26%) also deserves a careful reading. In cases that provoke anger, truly impartial comments may indeed be rare, but it is also possible that ambiguous comments are pushed into the extreme classes. The study has no human annotation to distinguish between the two explanations.

4.3. Labeler Confidence

Table 4 shows that the highest mean confidence is in the negative class (0.956), followed by the positive class (0.872) and the neutral class (0.855). The proportion of comments with confidence below 0.7 is also smaller in the negative class, at 5.0%, compared with 21.9% in the neutral class and 16.0% in the positive class. These results indicate that RoBERTa more often gives low-confidence predictions for the two minority classes. However, high confidence does not guarantee label correctness because neural networks can be overconfident in their predictions (Guo et al., 2017).

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5586

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

4.4. Evaluation on the Test Data

The evaluation results on the 690 test comments appear in Figure 1 and Table 5. The BiLSTM classifies 490 comments in agreement with the RoBERTa labels, with an accuracy of 71.01%, a macro-F1 of 0.580, and a Cohen's kappa of 0.384. As a comparison, a prediction that always selects the negative class produces an accuracy of 72.32% and a macro-F1 of 0.280. The accuracy of the BiLSTM is 1.31 percentage points lower, but its macro-F1 is higher because the model also recognizes some neutral and positive comments.

Table 5 shows that the recognition of negative sentiment is better than that of the other two classes. The F1-score of the negative class reaches 0.814, while the neutral and positive classes obtain 0.410 and 0.516, respectively. This difference shows that the ability of the model is not yet even across the three classes. The smaller number of neutral and positive comments also produces wider confidence intervals for recall, so the precision of their estimates is more limited.

The confusion matrix in Figure 1 shows 200 predictions that disagree with the RoBERTa labels. A total of 102 errors occur between the negative and positive classes, namely 64 negative comments predicted as positive and 38 positive comments predicted as negative. These errors matter for the discussion because they can change the reading of criticism, rejection, or sympathy in the comments. Therefore, sentiment predictions need to be examined together with the context and the target of the comment.

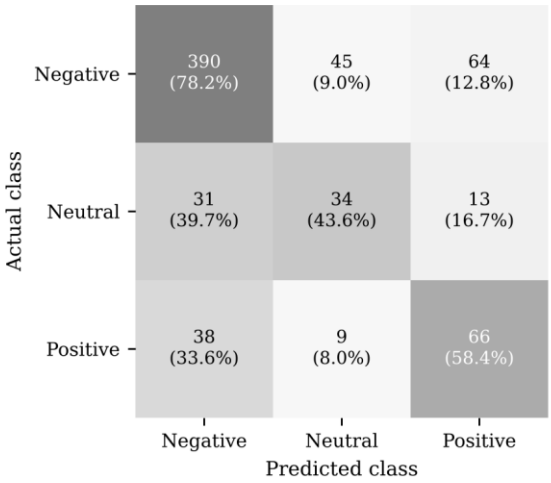


Figure 1. Confusion Matrix on the Test Data

Table 5. Evaluation Metrics of the BiLSTM per Class on the Test Data

Class	Precision	Recall	F1-score	Wilson 95% CI of recall (%)	Support (n)
Negative	0.850	0.782	0.814	74.3–81.6	499
Neutral	0.386	0.436	0.410	33.1–54.6	78
Positive	0.462	0.584	0.516	49.2–67.1	113
Macro average	0.566	0.601	0.580	–	690
Weighted average	0.734	0.710	0.720	–	690
Accuracy			0.710	67.52–74.28	690
Baseline: always negative (macro-F1)			0.280	–	690

Errors from the negative class also reduce the precision of the minority classes. Of the 143 comments predicted as positive, only 66 agree with the RoBERTa labels, whereas 64 come

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5586

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

from the negative class. For neutral predictions, only 34 of 88 comments agree, while 45 come from the negative class. Given the much larger number of negative comments, a portion of misclassified comments already exerts a large influence on the precision of the neutral and positive classes. This pattern needs to be read together with the labeling uncertainty in Section 4.3, but it does not prove that label errors cause the low performance. Validation through human annotation is needed to assess the correctness of sentiment based on the meaning of the comments.

4.5. Research Implications

The dominance of negative sentiment at 72.37% indicates strong criticism of officers in the comments on the four cases studied. When parties who hold the authority to protect citizens are themselves involved in violence, the problem does not end with the actions of the perpetrators. These events also raise questions about oversight, the use of authority, and institutional accountability.

TikTok becomes a space where users express rejection and demands, as seen in the comments "PECAT!!!!!!!!!!" ("DISMISS HIM!!!!!!!!!!") and "Aparat apa seperti ini?" ("Is this what officers are like?"). These expressions show that the attention of users is directed not only at the events but also at the actions expected from institutions. Digital conversation allows criticism to be voiced openly and answered by other users, in line with the changes in public communication discussed by Chadwick and Vaccari (2019). However, the data do not show whether users comment because formal complaint channels are perceived as unresponsive.

Positive sentiment also needs to be read with caution because the presence of positive comments cannot serve as a reason to assume that criticism of the institution has subsided. Positive sentiment can take the form of sympathy for victims, so the interpretation must consider the target of the comment.

Therefore, this study does not yet assess whether the official response meets these expectations. However, the criticism that emerges gives reason to place citizen safety and justice for victims as the primary concern in handling cases. Efforts to restore the image of an institution are unlikely to answer the criticism if the public has not obtained clarity about the resolution of cases and the prevention of similar violations.

4.6. Study Limitations

This study has several limitations. First, the data come from four posts on a single platform, so the results cannot be generalized to Indonesian public opinion and differences between cases cannot be analyzed (Olteanu et al., 2019). Second, the sentiment labels come entirely from RoBERTa without checking by human annotators. Label accuracy and inter-annotator agreement are therefore unknown (Bobicev & Sokolova, 2017). Third, sarcasm, irony, and the target of comments are not specifically distinguished, although they can affect the interpretation of attitudes toward officers (Nip & Berthon, 2024). Fourth, the results come from a single training run without cross-validation, without repetition with different seeds, and without other comparison models such as SVM with TF-IDF or a tuned transformer model, so the variability and relative position of the BiLSTM are unknown. Finally, the small neutral and positive test data make the estimates for both classes imprecise.

5. Conclusion

The study of 3,446 TikTok comments on four officer misconduct cases shows the dominance of negative sentiment at 72.37%, followed by positive at 16.37% and neutral at 11.26%. The criticism and demands for sanctions in the comments show the attention of users to officer accountability. Meanwhile, positive comments can take the form of sympathy for victims, so they cannot be interpreted directly as support for institutions. These findings describe responses to the cases studied, not the level of public trust in Indonesia or the effectiveness of the government response.

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5586

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

The BiLSTM trained with RoBERTa labels obtains a test accuracy of 71.01% and a macro-F1 of 0.580. Its accuracy does not exceed the majority-class baseline of 72.32%, but the higher macro-F1 shows an ability to recognize some neutral and positive comments. The performance on these two classes remains limited.

Future research is suggested to check labels through at least two annotators and to distinguish the target of comments among officers, institutions, and victims. Model development can consider Indonesian language resources through IndoLEM and IndoNLU (Koto et al., 2020; Wilie et al., 2020). Evaluation needs to be complemented by cross-validation, multiple seeds, and comparison models, with class balancing applied only to the training data. For the related institutions, the criticism in the comments can serve as material for evaluating the openness of case handling, victim protection, and oversight of the use of authority.

References

Bobicev, V., & Sokolova, M. (2017). Inter-annotator agreement in sentiment analysis: Machine learning perspective. *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP 2017)*, 97–102. https://aclanthology.org/R17-1015/

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Chadwick, A., & Vaccari, C. (2019). Digital media, power, and democracy: The impact of social platforms on political communication. *Journal of Communication*, *69*(2), 212–238.

Deadline.co.id. (2026). *Calon polwan diperkosa di Jambi, 2 polisi dipecat* [Female police candidate raped in Jambi, 2 police officers dismissed]. Retrieved April 21, 2026, from https://deadline.co.id/calon-polwan-diperkosa-di-jambi-2-polisi-dipecat/

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for multi-class classification: An overview*. arXiv. https://arxiv.org/abs/2008.05756

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning, PMLR 70*, 1321–1330. https://proceedings.mlr.press/v70/guo17a.html

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*, Article 27. https://doi.org/10.1186/s40537-019-0192-5

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, 757–770. https://doi.org/10.18653/v1/2020.coling-main.66

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, *18*(17), 1–5. https://jmlr.org/papers/v18/16-365.html

Lembaga Survei Indonesia. (2023). *Rilis survei: Kepercayaan publik terhadap lembaga penegak hukum, isu Piala Dunia U-20, aliran dana tak wajar di Kemenkeu, dugaan korupsi BTS, dan peta politik terkini* [Survey release: Public trust in law enforcement institutions, the U-20 World Cup issue, unusual fund flows at the Ministry of Finance, BTS (base transceiver station) corruption allegations, and the latest political map]. Retrieved April 21, 2026, from https://www.lsi.or.id/post/rilis-survei-lsi-09-april-2023

Liputan6.com. (2026a). *Anggota Brimob aniaya pelajar hingga tewas di Tual Maluku dipecat dari Polri* [Brimob member who beat a student to death in Tual, Maluku, dismissed from Polri]. Retrieved April 21, 2026, from https://www.liputan6.com/news/read/6284699/anggota-brimob-aniaya-pelajar-hingga-tewas-di-tual-maluku-dipecat-dari-polri

Liputan6.com. (2026b). *Pengakuan pedagang es kue jadul yang dituduh berbahan spons: Saya dikepung, ditonjok* [Account of the traditional ice cake vendor accused of using sponge: I was surrounded and punched]. Retrieved April 21, 2026, from

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

https://www.liputan6.com/news/read/6266169/pengakuan-pedagang-es-kue-jadul-yang-dituduh-berbahan-spons-saya-dikepung-ditonjok

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv. https://arxiv.org/abs/1907.11692

Mohler, M., Campbell, C. M., Henderson, K. S., & Renauer, B. C. (2021). Policing in an era of sousveillance: A randomised controlled trial examining the influence of video footage on perceptions of legitimacy. *Policing and Society*. Preprint available at https://pdxscholar.library.pdx.edu/ccj_fac/85

Nip, J. Y. M., & Berthon, B. (2024). Social media sentiment analysis. *Encyclopedia*, *4*(4), 1591–1603. https://doi.org/10.3390/encyclopedia4040104

Northcutt, C. G., Jiang, L., & Chuang, I. L. (2021). Confident learning: Estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research*, *70*, 1373–1411. https://doi.org/10.1613/jair.1.12125

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, *2*, Article 13. https://doi.org/10.3389/fdata.2019.00013

Song, H., Kim, M., Park, D., Shin, Y., & Lee, J.-G. (2023). Learning from noisy labels with deep neural networks: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(11), 8135–8153. https://arxiv.org/abs/2007.08199

Tvonenews.com. (2026). *Remaja tewas ditembak Iptu N usai tawuran pakai senjata water jelly di Makassar, polisi masih tunggu hasil autopsi* [Teenager shot dead by Iptu N after a brawl using a water jelly weapon in Makassar, police still await autopsy results]. Retrieved April 21, 2026, from https://www.tvonenews.com/berita/nasional/421230-remaja-tewas-ditembak-iptu-n-usai-tawuran-pakai-senjata-water-jelly-di-makassar-polisi-masih-tunggu-hasil-autopsi

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 843–857. https://doi.org/10.18653/v1/2020.aacl-main.85

Zhang, H., Xu, J., Lei, L., Qiu, J., & Alshalabi, R. (2023). A sentiment analysis method based on bidirectional long short-term memory networks. *Applied Mathematics and Nonlinear Sciences*, *8*(1), 55–68. https://doi.org/10.2478/amns.2022.1.00015

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, *8*(4), e1253. https://arxiv.org/abs/1801.07883

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.