



NLP-Based Analysis of Indonesian Public Opinion on the Nadiem Makarim Chromebook Case Using Instagram and YouTube Comments

Rifky Naufal Athaya^{1*}, Syabihul Khairy², Hana Amirah³,
Muhammad Aris⁴, Abdi Hasanuddin⁵

^{1,2}Department of Information Systems, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

³Department of Business Administration, Faculty of Business,
Karabük Üniversitesi, Türkiye

⁴Department of Usul Fiqh, Faculty of Sharia wal Qonun,
Wasathiyah Sharia University of Hadramaut, Yemen

⁵Department of Islamic Sharia, Faculty of Sharia wal Qonun, Al-Azhar University, Egypt

E-Mail: 112350312904@students.uin-suska.ac.id, 212350312975@students.uin-suska.ac.id,
hanaamirah2704@gmail.com, arismuhammad0104@gmail.com, abdihasanuddin87@gmail.com

Received Jul 14th 2026; Revised Aug 15th 2026; Accepted Sep 18th 2026; Available Online Sep 30th 2026

Corresponding Author: Rifky Naufal Athaya

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

This study examines Indonesian public sentiment toward the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek)'s Chromebook procurement under Nadiem Makarim and evaluates emoji tokenization in text classification. Between 2021 and 2024, the initiative sparked debate over device specifications, internet connectivity in frontier, outermost, and underdeveloped (3T) regions, and budget accountability. A balanced dataset of 4,794 comments (2,397 from Instagram and 2,397 from YouTube) was analyzed after removing duplicates. Sentiment pseudo-labels (negative, positive, and neutral) were generated using Indonesian RoBERTa fine-tuned on the SmSA IndoNLU benchmark. Two preprocessing scenarios were compared: removing emojis and emoticons (E1) and converting them into descriptive tokens (E2). Unigram and bigram TF-IDF features were classified using class-weighted Multinomial Logistic Regression within a leakage-free pipeline. Negative sentiment dominated (76.93%; 3,688 comments), followed by positive (17.86%; 856 comments) and neutral (5.21%; 250 comments). In the 80:20 holdout evaluation, E2 achieved 0.8352 accuracy and 0.6722 macro F1, while E1 achieved 0.8321 accuracy and 0.6725 macro F1. Five-fold cross-validation showed modest improvements with E2 in mean accuracy (0.8204 vs. 0.8169) and macro F1 (0.6433 vs. 0.6333). Performance was higher on Instagram than on YouTube. Overall, findings indicate predominantly negative public reception and modest benefits from emoji tokenization.

Keywords: Chromebook Procurement, Emoji Tokenization, Indonesian NLP, Logistic Regression, Public Opinion

1. INTRODUCTION

Public reactions toward government procurement programs often surface first on social media, offering an unfiltered view of citizen sentiment. In Indonesia, the digital education initiative launched by the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek), led by Nadiem Makarim, allocated over Rp 17.42 trillion between 2021 and 2024 for Information and Communication Technology (ICT) infrastructure, including approximately Rp 3.7 trillion for laptop computers. The selection of Chrome OS-based laptops (Chromebooks) at an estimated unit price of Rp 10 million sparked national debate. Educators, technology analysts, and civil society groups questioned the utility of cloud-dependent hardware in schools across frontier, outermost, and underdeveloped (3T) regions, where electricity grids and internet coverage remain inconsistent. Widespread concerns over budget allocation, hardware specifications, and policy alignment made the Chromebook procurement one of the most visible public controversies in Indonesian education governance [1].

Despite the national prominence of this controversy, policy evaluations have largely relied on formal parliamentary hearings and internal administrative audits, leaving the actual distribution and nuance of citizen opinion largely unquantified. Conventional public opinion surveys are constrained by sampling delays, high



costs, and social desirability bias. Social media data can fill this empirical gap by capturing spontaneous public reactions in real time, yet analyzing this discourse requires multi-platform observation. Instagram and YouTube reflect two distinct communication styles in Indonesia: Instagram comments are typically concise, reactive, and emoji-heavy, whereas YouTube comment sections under investigative news reports feature longer, multi-clause arguments. Evaluating both platforms within a balanced framework helps avoid single-source sampling bias and clarifies how platform interfaces influence public discourse [19].

From an analytical standpoint, user-generated comments in Indonesian pose specific challenges for Natural Language Processing (NLP). Informal text (bahasa gaul) contains frequent spelling variations, slang, word shortenings, and code-mixing between Indonesian, regional languages (such as Javanese and Sundanese), and English [2]. Within this unstructured text, emojis and emoticons serve as visual cues for sentiment, stance, and tone [6]. While text preprocessing pipelines often strip non-alphanumeric characters, removing emojis risks discarding affective signals [11]. Conversely, retaining emojis as text tokens is not straightforward. In Indonesian political discussions, laughing and clown emojis often function as markers of sarcasm or mockery rather than literal amusement [8]. Whether emoji tokenization aids or hinders sentiment classification on informal policy discourse remains an open empirical question [9].

Earlier sentiment studies on Indonesian government policies, including evaluations of the BPJS health system, the Omnibus Law, and curriculum reforms, have largely centered on single-platform Twitter datasets using standard cleaning routines [3], [13]. Although international studies show that emojis can provide useful features for sentiment and sarcasm detection [4], [5], [7], others report that adding emoji tokens can increase feature sparsity without consistent performance gains [9], [14]. Furthermore, while pretrained transformer models such as IndoBERT and IndoRoBERTa produce high-quality representations [2], their computational demands limit practical deployment for real-time monitoring. Using a fine-tuned transformer teacher to generate pseudo-labels for training an interpretable, lightweight linear classifier provides a practical alternative [24], but the cross-platform stability and emoji sensitivity of this approach have not been tested on Indonesian public policy controversies.

To address these gaps, this study investigates three research questions: (RQ1) What is the overall and platform-specific sentiment distribution regarding the Nadiem Makarim Chromebook case on Instagram and YouTube? (RQ2) How does converting emojis and emoticons into textual tokens (E2) compare with removing them (E1) when using TF-IDF features and class-weighted Multinomial Logistic Regression? (RQ3) How stable are the resulting models and cross-platform differences under stratified holdout testing and five-fold cross-validation in a leakage-free pipeline? This work provides three main contributions: an empirical assessment of public opinion on the Chromebook procurement controversy using balanced data from two platforms; a controlled comparison of emoji preprocessing methods in informal Indonesian text; and a reproducible workflow that combines transformer pseudo-labeling with an interpretable linear baseline.

2. MATERIAL AND METHOD

This study adopts a quantitative text-mining and machine learning design structured into six main stages: data acquisition and integration, text cleaning and platform balancing, construction of dual emoji representations, NLP-based teacher pseudo-labeling, TF-IDF feature extraction, and class-weighted Logistic Regression modeling evaluated under holdout and five-fold cross-validation protocols. The end-to-end analytical workflow is illustrated in Figure 1.

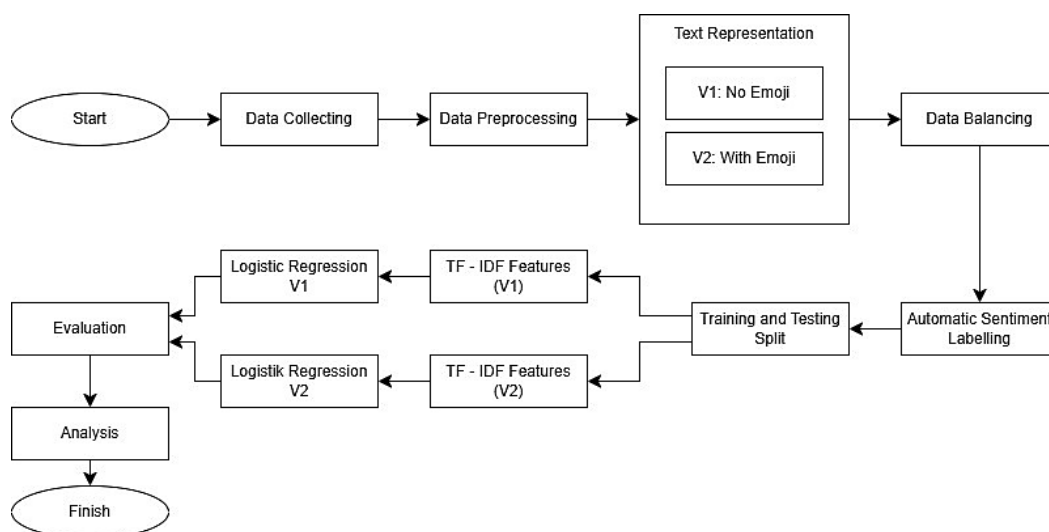


Figure 1. Research Methodology

2.1. Data Sources, Acquisition, and Scope

The research corpus was assembled from publicly accessible user comments on Instagram and YouTube addressing the Kemendikbudristek Chromebook procurement controversy during the peak discussion period between 2021 and 2024. Data acquisition targeted high-engagement posts from verified national news outlets, tech reviewers, and official educational channels using specific Indonesian keywords, including 'Chromebook Nadiem', 'pengadaan laptop Kemendikbud', 'laptop 30 triliun', and 'Chromebook sekolah 3T'. The initial raw collection yielded 5,002 comments, comprising 2,502 comments from Instagram and 2,500 comments from YouTube. Non-textual modalities, such as embedded video content, user avatars, and social network graphs, were excluded to focus strictly on natural language discourse.

Table 1. Data volume across preprocessing stages

Processing Stage	Instagram	Youtube	Total
Raw uploaded data	2,502	2,500	5,002
After removing empty comments	2,502	2,500	5,002
After the within-platform duplicate removal	2,397	2,481	4,878
After text preprocessing	2,397	2,481	4,878
After platform balancing	2,397	2,397	4,794

2.2. Data Cleaning and Platform Balancing

Exported comment files were standardized across platform fields and parsed to ensure UTF-8 compliance. Missing and blank comments were verified; no records were empty. Duplicate comments were detected by lowercasing text and collapsing redundant whitespace. Deduplication was performed strictly within each platform rather than globally, recognizing that identical comments across platforms represent distinct user expressions in separate social ecosystems. This reduced Instagram comments to 2,397 and YouTube comments to 2,481. Basic cleaning converted text to lowercase, removed URLs, stripped user mentions (@username), eliminated system interface artifacts, and normalized repetitive punctuation. Hashtags were stripped of the '#' symbol while retaining their lexical content. Because YouTube contained slightly more comments than Instagram, random undersampling (random_state=42) was applied to YouTube, yielding a balanced corpus of exactly 2,397 comments per platform and 4,794 comments in total [12], [13]. Balancing prevents evaluation metrics from being skewed toward the platform with greater observation volume.

2.3. Emoji and Emoticon Processing Pipeline

Unicode emojis were identified using the Python emoji library, while popular ASCII text emoticons (such as ':)', ':(', ':D', 'XD', 'T_T', and ':v') were extracted via compiled regular expressions. An indicator variable was created to verify the presence of expressive symbols. In the finalized balanced corpus, every retained comment contained at least one Unicode emoji or textual emoticon. Consequently, the analyzed corpus specifically reflects expressive, emoji-bearing public discourse [8].

2.4. Controlled Preprocessing Scenarios

To assess the empirical impact of emoji treatment, two controlled text representations were constructed from the identical balanced corpus: (1) Scenario E1 (No Emoji): All Unicode emojis were stripped using emoji.replace_emoji() and textual emoticons were removed via regex; (2) Scenario E2 (Emoji Tokens): Emojis were converted into standardized descriptive text using emoji.demajize(language='en') and prefixed with 'emoji_', while text emoticons were translated to canonical Indonesian tokens (e.g., 'emot_senyum', 'emot_sedih', 'emot_tertawa'). The underscore connector was preserved to treat multi-word emoji descriptors as atomic unigram features in TF-IDF vectorization [4], [6].

Table 2. Controlled preprocessing scenarios

Experiment	Output field	Emoji treatment	Example
E1: No emoji	text_tanpa_emoji	Unicode emojis and textual emoticons removed	`semangat pak` → `semangat pak` `semangat pak` → `semangat pak`
E2: Emoji tokens	text_dengan_emoji_token	Unicode emojis demojized; emoticons mapped to tokens	→ `semangat pak` emoji_folded_hands`

2.5. NLP-Based Automatic Sentiment Pseudo-Labeling

To establish reproducible sentiment labels across the corpus without prohibitive manual annotation costs, automatic pseudo-labeling was conducted using a state-of-the-art Indonesian deep language model. The teacher model was w11wo/indonesianroberta-base-sentiment-classifier, an Indonesian RoBERTa architecture fine-tuned on the SmSA IndoNLU benchmark [2]. SmSA is the definitive sentence-level Indonesian sentiment

resource, covering positive, negative, and neutral classes. Inference was performed using the Hugging Face Transformers pipeline with a batch size of 32 and a maximum sequence length of 512 tokens. URLs and user mentions were removed prior to labeling, while full lexical and emoji context was preserved.

The teacher model assigned a pseudo-label and associated confidence score to each comment. Across the 4,794 balanced comments, the distribution was 3,688 negative (76.93%), 856 positive (17.86%), and 250 neutral (5.21%), as detailed in Table 3. These labels serve as machine-annotated reference targets. In this teacher-student setup, downstream Logistic Regression models evaluate how effectively a lightweight linear classifier can reproduce deep contextual sentiment representations under different text preprocessing schemes [24].

Table 3. NLP-generated pseudo-label distribution

Pseudo-label	Number of Comments	Percentage
Negative	3,688	76.93%
Positive	856	17.86%
Neutral	250	5.21%
Total	4,794	100.00%

2.6. TF-IDF Feature Extraction

Text representations E1 and E2 were converted into high-dimensional numerical feature spaces using scikit-learn's TfidfVectorizer. The extraction configuration used unigrams and bigrams (`ngram_range=(1,2)`), sublinear term frequency scaling, `min_df=2`, `max_df=0.98`, and a vocabulary limit of 30,000 features. The Term Frequency-Inverse Document Frequency weight for term t in comment d is computed as Equation 1.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log\left(\frac{N}{\text{DF}(t)}\right) \quad (1)$$

where $\text{TF}(t, d)$ is the term frequency in comment d , N is the total number of comments, and $\text{DF}(t)$ is the document frequency. To strictly prevent data leakage, vectorizers were fitted solely on training partitions [15], [16].

2.7. Downstream Classifier: Scientific Justification for Logistic Regression

Multinomial Logistic Regression was chosen as the downstream classifier based on several practical and theoretical considerations [13], [17], [33]: (1) Compatibility with High-Dimensional Sparse Feature Spaces: Extracting unigram and bigram TF-IDF features yields a 30,000-dimensional sparse feature space. In high-dimensional text settings, linear decision boundaries are often sufficient to separate sentiment classes, avoiding the computational cost and potential overfitting of non-linear models; (2) Probabilistic Interpretability: Unlike deep neural networks, Logistic Regression yields explicit weight coefficients (w_k) corresponding to log-odds changes, allowing direct inspection of which terms and emoji tokens drive class predictions; (3) Computational Efficiency: With linear training complexity and $O(d)$ inference time, Logistic Regression operates efficiently on standard hardware without GPU requirements, making it suitable for continuous data processing; (4) Regularization and Class Weighting: An L2 penalty ($C=1.0$) limits extreme weight values caused by collinear informal words, while setting `class_weight='balanced'` adjusts loss penalties inversely proportional to class frequencies, directly addressing the 76.93% negative class majority [12], [23].

$$P(y = k | x) = \frac{\exp(w_k^T x + b_k)}{\sum_j \exp(w_j^T x + b_j)} \quad (2)$$

The classifier predicts the sentiment class k achieving the maximum posterior probability. Hyperparameter tuning validated $C=1.0$ with L-BFGS solver and `max_iter=2000`, ensuring guaranteed numerical convergence. Two separate vectorizer-classifier pipelines were trained independently for E1 and E2.

Table 4. Experimental configuration

Component	Implemented setting
Platform scope	Instagram and YouTube
Final dataset size	4,794 comments; 2,397 per platform
Teacher labeler	Indonesian RoBERTa sentiment classifier fine-tuned on SmSA
Text inputs	No-emoji text and emoji-token text
TF-IDF	Unigrams-bigrams; <code>min_df=2</code> ; <code>max_df=0.98</code> ; sublinear TF; <code>max_features=30,000</code>
Classifier	Logistic Regression; $C=1.0$; <code>class_weight=balanced</code> ; <code>max_iter=2,000</code>
Holdout test	Stratified 80:20 split; <code>random_state=42</code>

Component	Implemented setting
Cross-validation	Stratified 5-fold; shuffled; random_state=42
Primary metrics	Accuracy and macro-F1; macro-precision, macro-recall, classification report, confusion matrix

2.8. Holdout Training and Testing Protocol

The corpus was partitioned using an 80:20 stratified holdout split (random_state=42), yielding 3,835 training instances and 959 testing instances (738 negative, 171 positive, and 50 neutral). Stratification preserves exact class proportions across splits [12]. Identical row indices were used for E1 and E2 to ensure a controlled, paired experimental comparison free of test-set variation bias. Models were evaluated overall and separately across Instagram (482 test comments) and YouTube (477 test comments) to detect platform-specific behaviors [19].

2.9. Stratified Five-Fold Cross-Validation

To mitigate variance from single-split partitioning, stratified five-fold cross-validation was conducted with shuffled folds (random_state=42). TF-IDF vectorization and Logistic Regression were encapsulated inside scikit-learn Pipelines, guaranteeing that vocabulary extraction and IDF weighting were learned solely within the training folds. Mean and standard deviation were computed for accuracy and macro F1-score across all five folds [15], [20].

2.10. Evaluation Metrics and Model Artifact Export

Model evaluation incorporated Accuracy, Macro-Precision, Macro-Recall, and Macro-F1 score, supplemented by class-specific classification reports and confusion matrices. Macro-F1 was designated as the primary benchmark metric because it weights all classes equally and penalizes models that achieve high accuracy merely by overpredicting the majority negative class [21], [22]. Following evaluation, full-dataset models and transformation pipelines were serialized with complete metadata to ensure scientific reproducibility [25].

2.11. Ethical Considerations and Research Scope

All analyzed comments were retrieved exclusively from public social media posts, in compliance with the respective platforms' terms of service and ethical guidelines for internet research [26]. Direct personal identifiers, usernames, handles, and URLs were purged prior to analysis. The scope of this research is intentionally bounded: the findings characterize Indonesian public sentiment in emoji-bearing comments on Instagram and YouTube about the Chromebook case, providing empirical evidence of online public discourse without asserting legal conclusions.

3. RESULTS AND DISCUSSION

3.1. Data Acquisition, Cleaning, and Balancing Results

The raw dataset comprised 5,002 comments (2,502 Instagram and 2,500 YouTube). Deduplication per platform eliminated 105 redundant entries from Instagram (reducing the total to 2,397) and 19 from YouTube (reducing the total to 2,481). Platform balancing via random undersampling on YouTube standardized the corpus to exactly 2,397 comments per platform, yielding 4,794 final comments. Table 5 details the stages of data attrition.

Table 5. Data Cleaning and Balancing Results

Stage	Instagram	Youtube	Total
Initial data	2,502	2,500	5,002
After removing blank comments	2,502	2,500	5,002
After removing duplicates per platform	2,397	2,481	4,878
After text preprocessing	2,397	2,481	4,878
After platform balancing	2,397	2,397	4,794

3.2. Emoji and Emoticon Detection Results

Every comment in the balanced corpus contained at least one detected emoji or textual emoticon. Common emojis included face with tears of joy, clown face, rolling on the floor laughing, and praying/folded hands, alongside text emoticons like ':' and ':v'. This verified that the dataset provided a rich empirical basis for evaluating the comparative utility of emoji tokenization versus emoji removal in Indonesian text [8], [30].

3.3. Sentiment Pseudo-Label Distribution across Platforms

The RoBERTa pseudo-labeling revealed that negative sentiment dominated Indonesian public opinion regarding the Chromebook procurement, accounting for 76.93% (3,688 comments) of the total corpus. Positive

sentiment comprised 17.86% (856 comments), while neutral sentiment accounted for 5.21% (250 comments). As visualized in Figure 2, this distribution was consistent across both platforms: Instagram registered 77.22% negative, 17.40% positive, and 5.38% neutral sentiment, while YouTube registered 76.64% negative, 18.31% positive, and 5.05% neutral sentiment. This cross-platform consistency demonstrates a clear public consensus critical of the Chromebook initiative.

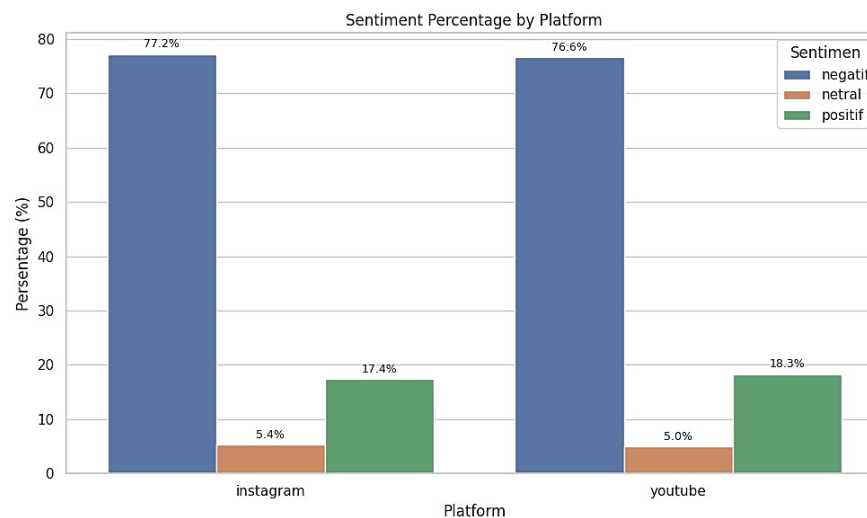


Figure 2. Sentiment Percentage by Platform

3.4. Holdout Evaluation and Confusion Matrix Analysis

In the stratified holdout evaluation (959 test comments: 738 negative, 171 positive, 50 neutral), the no-emoji representation (E1) achieved an accuracy of 0.8321, macro-precision of 0.6738, macro-recall of 0.6711, and macro-F1 of 0.6725. The emoji-token representation (E2) achieved an accuracy of 0.8352, macro-precision of 0.6621, macro-recall of 0.6847, and macro-F1 of 0.6722, as shown in Table 6. The overall performance differences between E1 and E2 were modest, with E2 offering a slight gain in overall accuracy (+0.0031) while maintaining nearly identical macro-F1.

Table 6. Overall Holdout Evaluation Results

Text Representation	Accuracy	Macro Precision	Macro Recall	Macro F1
Without Emoji	0.8321	0.6738	0.6711	0.6725
With Emoji Token	0.8352	0.6621	0.6847	0.6722

Figure 3 illustrates the confusion matrix comparison between E1 and E2. Both models demonstrated high recall on the dominant negative class, correctly predicting 663 negative comments in E1 and 661 in E2. Notably, E2 improved sensitivity toward the minority classes: correctly classified positive comments increased from 112 to 116 (improving positive class recall from 0.655 to 0.678), and correctly classified neutral comments increased from 23 to 24 (raising neutral recall from 0.460 to 0.480). However, this was accompanied by a minor increase in false-positive classifications as negative, explaining why macro-F1 remained steady despite higher accuracy and improved recall on minority classes.

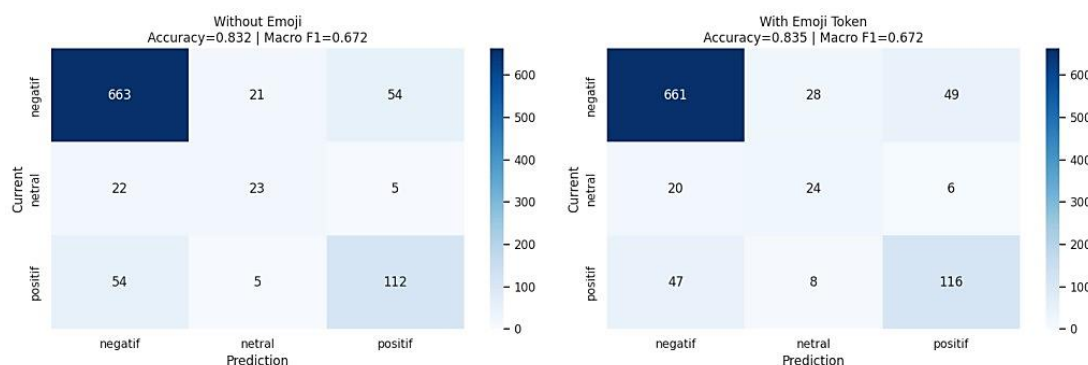


Figure 3. Comparison between no-emoji and with-emoji-token text representations in the holdout evaluation

3.5. Platform-Specific Classification Performance

Evaluating the holdout test set by the platform revealed marked performance differences between Instagram (482 comments) and YouTube (477 comments), as summarized in Table 7. For Instagram, both representations achieved high efficacy: E1 yielded 0.8672 accuracy and 0.7579 macro-F1, while E2 achieved 0.8672 accuracy and 0.7582 macro-F1. In contrast, YouTube performance was lower: E1 reached 0.7966 accuracy and 0.5937 macro-F1, whereas E2 achieved 0.8029 accuracy and 0.5897 macro-F1. The model showed higher classification stability on Instagram, whereas YouTube's macro-F1 was 16.85 percentage points lower, underscoring clear differences in cross-platform text characteristics.

Table 7. Platform-Specific Evaluation Results

Text Representation	Platform	Test Samples	Accuracy	Macro Precision	Macro Recall	Macro F1
Without Emoji	Instagram	482	0.8672	0.7588	0.7592	0.7579
Without Emoji	YouTube	477	0.7966	0.6016	0.5875	0.5937
With Emoji Token	Instagram	482	0.8672	0.7337	0.7883	0.7582
With Emoji Token	YouTube	477	0.8029	0.5952	0.5872	0.5897

3.6. Five-Fold Cross-Validation Performance

Stratified five-fold cross-validation confirmed the stability of the classification architecture across randomized data partitions. The emoji-token representation (E2) achieved higher average performance than E1, increasing mean accuracy from 0.8169 ± 0.009 to 0.8204 ± 0.008 , and macro-F1 from 0.6333 ± 0.015 to 0.6433 ± 0.014 . These cross-validated gains demonstrate that emoji tokenization provides a small but consistent positive inductive bias when evaluated across multiple partitions, helping the model capture subtle nuances in sentiment.

3.7. Discussion and Scholarly Interpretation

The distribution of sentiment reflects clear public skepticism toward the Chromebook procurement program. Negative sentiment accounted for 76.93% of the corpus, maintaining consistent levels on both Instagram (77.22%) and YouTube (76.64%). Reviewing top TF-IDF n-grams and misclassified comments indicates three primary areas of public concern: (1) Price-to-specification ratio, where users questioned spending approximately Rp 10 million per unit on dual-core Chrome OS devices, pointing out that general-purpose Windows laptops with broader offline functionality were available at similar price points; (2) Infrastructure readiness in 3T schools, with commenters noting that cloud-centric laptops are impractical in rural areas with unreliable electricity and limited internet access; and (3) Budget transparency, reflected in frequent comments questioning public procurement procedures. The close alignment in negative proportion across Instagram and YouTube indicates that critical feedback was consistent across both media types.

A notable outcome of the evaluation is the performance gap between Instagram (macro F1: 0.7582) and YouTube (macro F1: 0.5897), representing a 16.85 percentage-point difference. This difference relates to the interaction between platform affordances and language use [19]. Instagram comments in the dataset are largely short and direct ('kemahalan untuk spek segitu', 'gak guna di desa'), producing clear unigram and bigram TF-IDF features that align well with linear boundaries. In contrast, YouTube comments under news and tech review videos tend to be longer and contain multi-sentence commentary. Commenters frequently use rhetorical framing, such as opening with sarcastic praise ('bagus sekali idenya') before criticizing the lack of electricity in village schools. Because TF-IDF treats words independently, positive terms within sarcastic statements shift the decision score, leading to false positives and lower macro F1 on YouTube. The limited gain from emoji tokenization (+0.0031 in holdout accuracy, +0.0035 in cross-validation accuracy) illustrates the pragmatic complexity of emojis in Indonesian political commentary. In discussions of public policy, emojis rarely convey straightforward emotional states. Tokens such as emoji_face_with_tears_of_joy, emoji_clown_face, and emot_tertawa are often used to express ridicule or cynicism toward policy announcements [8], [32]. When mapped to text tokens, their co-occurrence with both positive words (in ironic statements) and negative words (in direct critiques) weakens their discriminative utility for linear models. While emoji tokens helped identify additional positive and neutral cases (raising positive recall from 0.655 to 0.678), their overall effect remains small unless combined with explicit sarcasm detection methods [9], [30].

Comparing these outcomes with prior work supports several established patterns in text mining. The effectiveness of TF-IDF paired with class-weighted Logistic Regression aligns with findings by Fadri et al. [3], Pasha et al. [17], and Riswawan [33], indicating that regularized linear models offer dependable, interpretable baselines for noisy social text. In addition, the teacher-student setup showed that a lightweight classifier can approximate RoBERTa pseudo-labels with over 83% accuracy without requiring specialized hardware for inference, providing a viable workflow for continuous monitoring of public discourse [24].

4. CONCLUSION

This study examined Indonesian public sentiment regarding the Nadiem Makarim Chromebook procurement controversy across Instagram and YouTube comments, addressing all three research questions. Regarding RQ1, public sentiment was predominantly negative across both platforms, accounting for 76.93% of the 4,794 balanced comments (77.22% on Instagram and 76.64% on YouTube), driven by concerns about device pricing, 3T infrastructure constraints, and procurement transparency. Regarding RQ2 and RQ3, converting emojis and emoticons into text tokens (E2) yielded a slight improvement over emoji removal (E1) in five-fold cross-validation (accuracy: 0.8204 vs. 0.8169; macro F1: 0.6433 vs. 0.6333) and improved positive recall in holdout testing, though holdout macro F1 was essentially equivalent (0.6722 vs. 0.6725). Furthermore, classification performance was higher on Instagram (macro F1: 0.7582) than on YouTube (macro F1: 0.5897), indicating that comment length and rhetorical style influence the stability of linear models.

The novelty of this study lies in presenting a platform-balanced empirical analysis of an Indonesian educational procurement controversy, paired with a controlled evaluation of emoji tokenization in informal Indonesian text. The results provide an empirical reference point for policy feedback and demonstrate that a simple, interpretable linear model can effectively replicate deep transformer sentiment labels in resource-constrained environments.

Future research can build on these findings in three ways: First, incorporating human-labeled evaluation sets with inter-annotator agreement metrics (such as Cohen's or Fleiss' Kappa) would provide an external benchmark for transformer pseudo-label accuracy. Second, integrating dedicated sarcasm-detection techniques could help disambiguate the role of humorous emojis in political comments. Third, applying Aspect-Based Sentiment Analysis (ABSA) would enable sentiment evaluation across specific policy components, such as device specifications, procurement pricing, and regional connectivity.

Declaration of Generative AI in Scientific Writing

During the preparation of this manuscript, the authors used AI-assisted language-editing tools solely to refine English grammar, syntax, and phrasing clarity. Following these stylistic enhancements, the authors thoroughly reviewed, edited, and validated the manuscript. The authors assume full responsibility for the originality, scientific validity, methodological execution, analytical interpretations, and conclusions presented in this publication.

ACKNOWLEDGEMENT

The authors express their sincere gratitude to the Department of Information Systems at Universitas Islam Negeri Sultan Syarif Kasim Riau, Karabük Üniversitesi, Wasathiyah Sharia University of Hadramaut, and Al-Azhar University for supporting this collaborative research. The authors also extend their appreciation to the editorial board and anonymous peer reviewers of IJATIS for their constructive critiques and insightful guidance, which greatly enhanced the quality of this paper.

REFERENCES

- [1] J. Y. M. Nip and B. Berthelie, "Social Media Sentiment Analysis," *Encyclopedia*, vol. 4, no. 4, pp. 1590-1598, 2024.
- [2] F. Koto, J. H. Lau, and T. Baldwin, "IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2021, pp. 10660–10668.
- [3] A. I. Fadri, N. Fitri, A. Jelita, D. D. Bagaskara and R. Zahra, "Sentiment Analysis of Public Opinion on the Gaza Conflict Using Machine Learning," vol. 3, no. 2, pp. 138-149, 2026.
- [4] H. Alawneh, A. Hasasneh and M. Maree, "On the Utilization of Emoji Encoding and Data Preprocessing with a Combined CNN-LSTM Framework for Arabic Sentiment Analysis," *Modelling*, vol. 5, no. 4, pp. 1469-1489, 2024.
- [5] A. Khan, D. Majumdar and B. Mondal, "Sentiment analysis of emoji fused reviews using machine learning and Bert," *Sci. Rep.*, vol. 15, no. 1, 2025.
- [6] W. Yin, R. Alkhalifa and A. Zubiaga, "The emojiification of sentiment on social media: Collection and analysis of a longitudinal Twitter sentiment dataset," 2026.
- [7] D. Yang, L. Kejian, Y. Cheng, F. Yuanyuan and L. Weihao, "Emoji-based Fine-grained Attention Network for Sentiment Analysis in the Microblog Comments," Jun. 2022, 2022.
- [8] R. S. Dewi and R. M. M. S. Almadinah, "Emojis Affect The Meaning Of Sarcasm Sentences On Social Media Twitter: Evidence From Indonesia Millennial Generations," *Indonesian Journal of Education*, vol. 4, no. 3, pp. 147-152, 2023.
- [9] G. H. Aleryani, W. Deabes, K. Albishre and A. E. Abdel-Hakim, "Impact Of Emoji Exclusion On The Performance Of Arabic Sarcasm Detection Models A Preprint," 2024., 2024.
- [10] Y. Lou, J. Zhou, J. Zhou, D. Ji and Q. Zhang, "Emoji multimodal microblog sentiment analysis based on mutual attention mechanism," *Sci. Rep.*, vol. 14, no. 1, 2024.

- [11] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences (Switzerland)*, vol. 12, no. 17, 2022.
- [12] Z. A. Sayyed, "Study of sampling methods in sentiment analysis of imbalanced data," Jun. 2021, 2021.
- [13] N. Chamidah, D. Widiyanto, H. B. Seta and A. A. Aziz, "The Impact of Oversampling and Undersampling on Aspect-Based Sentiment Analysis of Indramayu Tourism Using Logistic Regression," *Revue d'Intelligence Artificielle*, vol. 38, no. 3, pp. 795-804, 2024.
- [14] B. S. Wiguna, C. V. Hudiyanti, A. A. Rausanfitra and A. Z. Arifin, "Sarcasm Detection Engine for Twitter Sentiment Analysis using Textual and Emoji Feature," *Jurnal Ilmu Komputer dan Informasi*, vol. 14, no. 1, pp. 1-8, 2021.
- [15] A. Ichwani, R. I. Kesuma, A. Setiawan, E. Wicaksono and R. Hanifah, "Preventing Data Leakage in Classification via Integrated Machine Learning Pipelines: Preprocessing, Feature Transformation, and Hyperparameter Tuning," *Jurnal Teknik Informatika (JUTIF)*, vol. 7, no. 1, 2026.
- [16] A. M. Abdulabbas, R. Alkanany, Y. A. K. Al-Nuaimi and Z. M. A. Al-Hamdawee, "A Sequential Data Preprocessing Pipeline for Diabetes Prediction: A Data Leakage Prevention and Dual-Validation Approach," *Engineering*, vol. 15, no. 6, pp. 30059-30066, 2025.
- [17] A. N. I. Pasha, E. F. Putri, L. Muthoharoh, A. Satria and M. C. T. Manullang, "Hybrid TF--IDF Logistic Regression and MLP Neural Baseline for Indonesian Three-Class Sentiment Analysis on Social Media Text," May 2026, 2026.
- [18] J. Demšar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.
- [19] K. Kapur and R. Harikrishnan, "Comparative Study Of Sentiment Analysis For Multi- Sourced Social Media Platforms," 2026.
- [20] D. Z. Abidin, L. Afuan, A. N. Toscani and N. Nurhadi, "A Comprehensive Benchmarking Pipeline for Transformer-Based Sentiment Analysis using Cross-Validated Metrics," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 1797-1810, 2025.
- [21] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, 2024.
- [22] M. C. H. Lee, J. Braet and J. Springael, "Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores," *Applied Sciences (Switzerland)*, vol. 14, no. 21, 2024.
- [23] D. Harbecke, Y. Chen, L. Hennig and C. Alt, "Why only Micro-F 1 ? Class Weighting of Measures for Relation Classification," 2022.
- [24] M. Jafarlou and M. M. Kubek, "Reducing Labeling Costs in Sentiment Analysis via Semi-Supervised Learning," Oct. 2024, 2024.
- [25] O. E. Gundersen, K. Coakley, C. Kirkpatrick and Y. Gil, "Sources of Irreproducibility in Machine Learning: A Review," Apr. 2023, 2023.
- [26] T. Coffin, "Ethical Considerations for the Use of Social Media in the Human Subjects Research Setting," 2025, 2025.
- [27] Y. Otmakhova, S. Khanezhari and L. Frermann, "Media Framing: A Typology and Survey of Computational Approaches Across Disciplines," 2026.
- [28] Y. Wang, Y. Yin and P. Li, "Towards the Mitigation of Confirmation Bias in Semi-supervised Learning: a Debiased Training Perspective," Sep. 2024, 2024.
- [29] F. B. Kader, N. H. Nujat, T. B. Sogir, M. Kabir, H. Mahmud and K. Hasan, "Computational Sarcasm Analysis on Social Media: A Systematic Review," Sep. 2022, 2022.
- [30] V. Grover and H. Banati, "Task 6: Exploring Emojis and Sentiments for Sarcasm Detection," 2026.
- [31] S. Kusal, S. Patil and K. Kotecha, "Multimodal text-emoji fusion using deep neural networks for text-based emotion detection in online communication," *J. Big Data*, vol. 12, no. 1, 2025.
- [32] V. Grover and H. Banati, "An emoji centric approach to sarcasm detection in online discourse," *Sci. Rep.*, vol. 16, no. 1, 2026.
- [33] A. Riswawan, "Implementing TF-IDF and Logistic Regression for Sentiment Analysis of YouTube Comments on the iPhone 16," *Jurnal Teknologi Dan Open Source*, vol. 8, no. 2, pp. 604-611, 2025.