International Journal Software Engineering and Computer Science (IJSECS)
6 (3), 2026, 976-986
Published Online December 2026 in IJSECS (http://www.journal.lembagakita.com/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v6i3.8086.

RESEARCH ARTICLE Open Access

Public Sentiment Analysis of the #KaburAjaDulu Hashtag Using a Combination of Support Vector Machine (SVM) and Random Forest Algorithms

Muhammad Derry Oktaviandi [1*], **Yuma Akbar** [2], **Mesra Betty Yel** [3]

[1*,2,3] Department of Informatics, Faculty of Computer Technology, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.

Article History:

Received: July 18, 2026.
Revised: July 20, 2026.
Accepted: August 27, 2026.
Available online: September 20, 2026.
Published: December 1, 2026.

***Corresponding Author:**
derrymuhammad0@gmail.com

Abstract: This study aims to analyze public sentiment toward the #KaburAjaDulu hashtag on platform X and compare the classification performance of Support Vector Machine (SVM), Random Forest, and a Voting Classifier ensemble. A total of 1,502 posts were collected via web scraping. The text preprocessing pipeline comprised case folding, text cleaning, tokenization, stopword removal, and stemming using the Sastrawi library. Processed texts were transformed into numerical feature vectors using Term Frequency-Inverse Document Frequency (TF-IDF), followed by an 80:20 train-test split. The empirical distribution revealed an extreme class imbalance: negative sentiment dominated at 96.54% (1,450 posts), followed by neutral at 3.33% (50 posts) and positive at 0.13% (2 posts), generating an imbalance ratio of 725:1. SVM and Random Forest achieved identical aggregate scores with 95.35% accuracy, 90.91% precision, 95.35% recall, and a 93.08% F1-score; however, both completely failed to detect neutral and positive classes. The Voting Classifier achieved the highest aggregate performance, reaching 96.01% accuracy, 95.84% precision, 96.01% recall, and a 94.55% F1-score by successfully identifying two neutral instances. Nevertheless, the ensemble model was unable to recognize positive sentiment, yielding zero sensitivity for the extreme minority class. These findings demonstrate that combining SVM and Random Forest offers marginal improvements in aggregate metrics but remains constrained by data distribution. Consequently, relying solely on aggregate accuracy produces misleading evaluations in severely skewed datasets, emphasizing the necessity of per-class metrics, confusion matrices, and data-balancing strategies for social media sentiment classification.

Keywords: Sentiment Analysis; Random Forest; Support Vector Machine; TF-IDF; Voting Classifier.

1. Introduction

The advancement of digital technology has fundamentally transformed how public discourse is produced, disseminated, and negotiated. Social media platforms serve not only as instruments for interpersonal networking but also as critical arenas where individuals articulate grievances, support, socio-political critique, and perspectives on emerging issues. Among these platforms, X (formerly Twitter) remains a prominent medium for text-based public opinion. Its brief, open, rapid, and easily shareable post formats generate massive volumes of textual data suitable for examining public sentiment toward economic conditions, public policy, and governance. Discourse on X is typically organized through hashtags, which enable users to aggregate discussions, monitor trending topics, and participate in shared digital conversations. When a hashtag gains viral momentum, the resulting corpus reflects shifting public attitudes. Nevertheless, the scale and unstructured nature of social media text impede manual interpretation. Posts frequently contain colloquialisms, abbreviations, non-standard spelling, irony, slang, punctuation noise, hyperlinks, and account mentions, necessitating systematic computational text-processing pipelines to clean, represent, and classify data effectively. A notable socio-digital phenomenon on platform X is the emergence of the #KaburAjaDulu hashtag. This hashtag has been widely used to discuss desires to leave Indonesia temporarily, seek educational and employment opportunities abroad, or articulate general dissatisfaction with domestic living conditions. Furthermore, users employ the tag to share satirical commentary, jokes, emotional distress, critique, and expressions of mutual encouragement. Consequently, the sentiment of these posts cannot be inferred merely from the presence of the word "kabur" (escape). Each post must be analyzed within its broader linguistic and contextual framing to capture the underlying emotional valence. Such analysis is essential to discern whether the prevailing discourse leans toward negative grievances, positive encouragement, or neutral informational statements.

Sentiment analysis, as a branch of text mining and natural language processing, extracts subjective opinion from unstructured text through data collection, preprocessing, feature extraction, model training, and performance evaluation. In social media text processing, preprocessing is vital for reducing linguistic noise prior to feature conversion using Term Frequency-Inverse Document Frequency (TF-IDF), which weights terms according to their local frequency and global rarity across the corpus. Prior studies have established the efficacy of Support Vector Machine (SVM) models in classifying social media text across diverse domains, including e-commerce, streaming services, movie reviews, and public policy [add reference: cite relevant prior studies]. However, existing research has predominantly focused on single-algorithm implementations evaluating structured product reviews or institutional policies. Empirical studies examining spontaneous public discourse driven by viral socio-political hashtags remain scarce, particularly regarding the combined implementation of SVM and Random Forest for three-class sentiment classification on Indonesian texts. Furthermore, a methodological shortcoming in prior literature is the disproportionate reliance on overall accuracy as the primary benchmark. High aggregate accuracy often conceals severe misclassification of minority classes in skewed datasets; therefore, robust performance assessment demands detailed evaluation through precision, recall, F1-score, and confusion matrices. This study implements an ensemble approach combining Support Vector Machine and Random Forest via a Voting Classifier to address multi-class sentiment classification. SVM is well suited for text classification due to its proficiency in handling high-dimensional, sparse feature spaces through maximum-margin hyperplane separation. In contrast, Random Forest operates as an ensemble of randomized decision trees that captures complex feature interactions while mitigating model variance. Given their distinct mathematical formulations and decision boundaries, integrating both classifiers into a voting framework is hypothesized to enhance prediction stability and classification robustness beyond individual baseline models.

The novelty of this study encompasses three primary dimensions: object of study, computational methodology, and evaluation framework. First, it investigates the #KaburAjaDulu hashtag on platform X as an empirical Indonesian socio-digital phenomenon. Second, it systematically compares SVM, Random Forest, and their combined Voting Classifier on an identical TF-IDF feature space. Third, rather than relying solely on aggregate accuracy, it examines per-class performance to determine whether ensemble gains translate to improved minority-class recognition. The findings provide both empirical insight into public sentiment regarding the hashtag and methodological evidence concerning the capabilities and limitations of ensemble machine learning under extreme class imbalance in Indonesian social media corpora. Based on these considerations, this study aims to examine the empirical distribution of positive, negative, and neutral sentiment in posts associated with the #KaburAjaDulu hashtag on platform X. Additionally, this study aims to construct and evaluate the classification performance of Support Vector Machine, Random Forest, and a Voting Classifier using accuracy, precision, recall, F1-score, and confusion matrices, specifically assessing whether the ensemble approach improves multi-class sentiment discrimination compared to individual classifiers.

2. Related Work

2.1 Related Works

The rapid growth of social media has generated unprecedented volumes of textual data, serving as an empirical source for capturing public opinion on socio-economic and political developments. Platform X (formerly Twitter) is frequently utilized in sentiment analysis research due to its rapid dissemination of concise public posts. The unstructured nature of this data, characterized by non-standard terminology, colloquial abbreviations, emojis, and emerging slang, positions sentiment analysis as a critical methodology within text mining and natural language processing (NLP) to classify public perceptions into positive, negative, or neutral orientations. Numerous empirical studies have affirmed the effectiveness of Support Vector Machine (SVM) models in textual classification, largely attributed to their capacity to manage high-dimensional and sparse feature representations. Ramlan *et al*. (2023) analyzed public sentiment regarding fuel price adjustments, demonstrating that SVM attained 82.69% accuracy in binary classification. Similarly, Khairudin *et al*. (2023) applied SVM coupled with Term Frequency-Inverse Document Frequency (TF-IDF) feature weighting to Indonesian movie reviews, achieving 87.5% accuracy and confirming the efficacy of structured preprocessing on short texts. Furthermore, Alhaq *et al*. (2021) utilized crawled Twitter data to evaluate consumer perceptions of the Bukalapak marketplace, recording a 93% accuracy rate. These studies substantiate the resilience of SVM across diverse linguistic styles in digital corpora.

Beyond individual linear and kernelized classifiers, the Random Forest algorithm is widely implemented in text categorization due to its capacity to capture non-linear feature relationships and resist overfitting. Operating as an ensemble of randomized decision trees aggregated via majority voting, Random Forest provides robust generalization performance across complex datasets; however, its efficacy in high-dimensional text spaces remains heavily contingent upon feature sparsity and class balance. Recent developments have consequently transitioned toward ensemble learning frameworks. Notably, the Voting Classifier combines predictions from diverse model architectures to maximize decision stability and classification accuracy over standalone baselines. Concurrently, while existing sentiment literature predominantly addresses consumer goods, digital applications, commercial services, and formal institutional policies, empirical investigations into spontaneous public discourse generated through viral socio-political hashtags remain scarce.

2.2 Comparative Analysis of Prior Studies

Extensive research has contributed to the evolution of machine-learning-based sentiment analysis, with SVM frequently serving as the benchmark due to its strong discriminatory performance on text corpora. Nevertheless, variations in dataset characteristics, feature representations, and classification tasks yield divergent performance benchmarks across domains. Husada and Paramita (2021) implemented multi-class SVM for airline service feedback, obtaining 84.37% accuracy, while Giovani *et al*. (2020) demonstrated that combining SVM with feature optimization yielded superior classification over alternative classical baselines. Despite these favorable outcomes, existing studies exhibit three notable limitations. Most implementations rely on single-model architectures or evaluate multiple classifiers purely in isolation rather than exploring collaborative ensemble architectures. Additionally, the dominant empirical focus on e-commerce, commercial products, and institutional services overlooks the dynamic, highly reactive nature of organic public discourse on social networks. A further critical methodological limitation concerns performance evaluation: prior investigations frequently rely on overall accuracy as the primary benchmark. On imbalanced datasets, aggregate accuracy disproportionately reflects majority-class dominance, thereby obscuring severe classification failures on minority categories. Comprehensive evaluation thus necessitates per-class precision, recall, F1-score, and confusion matrix assessments in evaluating classification performance across individual sentiment classes.

2.3 Research Positioning

Based on the synthesis of existing literature, several critical research gaps remain unaddressed. First, the socio-digital phenomenon surrounding the #KaburAjaDulu hashtag on platform X has received minimal empirical investigation using machine learning pipelines, despite reflecting public responsiveness to socio-economic conditions in Indonesia. Second, empirical implementations combining Support Vector Machine and Random Forest within an ensemble Voting Classifier remain sparse in Indonesian natural language processing. Integrating these two complementary algorithms addresses this gap by combining maximum-margin hyperplane optimization with randomized decision ensembles to enhance classification stability. Finally, this study addresses the methodological shortcoming of aggregate-metric reliance by evaluating model performance through detailed per-class precision, recall, F1-scores, and confusion matrices. The novelty of this work thus centers on the application of an ensemble Voting Classifier to organic socio-political Indonesian hashtag discourse under an evaluation framework capable of diagnosing minority-class classification behavior.

2.4 Review of Algorithms and Computational Pipeline

This study implements a text mining pipeline to extract sentiment patterns from unstructured social media corpora through systematic text preprocessing, feature weighting, model training, and multi-metric evaluation. The preprocessing pipeline applies case folding, text cleaning, tokenization, stopword removal, and morphological stemming to eliminate non-informative noise from the raw corpus. The cleaned texts are subsequently transformed into sparse numeric matrices using Term Frequency-Inverse Document Frequency (TF-IDF), which scales word importance based on local document frequency relative to corpus-wide occurrence. Classification is executed using Support Vector Machine (SVM), Random Forest, and a Voting Classifier ensemble. SVM maps input feature vectors into a high-dimensional space to establish an optimal separating hyperplane that maximizes the geometric margin between sentiment classes, ensuring robust generalization on sparse feature spaces. Conversely, Random Forest constructs multiple randomized decision trees and aggregates individual outputs through majority voting, mitigating variance and reducing overfitting risks. The Voting Classifier aggregates the predictive outputs of both SVM and Random Forest to determine final class labels. Model performance is comprehensively evaluated using confusion matrices alongside accuracy, precision, recall, and F1-score across all sentiment categories.

3. Methodology

This study employed a quantitative approach with a computational experimental design. The empirical corpus comprised public posts on platform X containing the #KaburAjaDulu hashtag, collected via web scraping using a GitHub-based crawler. Sampling was conducted through purposive sampling by filtering posts strictly matching the specified research hashtag. The data collection process yielded 1,502 posts archived in CSV format and processed through the full_text attribute. Each post was categorized into one of three sentiment classes: positive, neutral, or negative. Positive sentiment denoted expressions of support, appreciation, or favorable assessments; negative sentiment encompassed socio-political critique, rejection, dissatisfaction, or frustration; and neutral sentiment comprised purely informational statements devoid of clear emotional polarity. The assigned sentiment labels served as the target variable, while the processed text represented the input features. Data labeling was conducted manually by a human annotator possessing native fluency in Indonesian and contextual familiarity with the socio-digital dynamics of the #KaburAjaDulu hashtag. To ensure annotation consistency, a test-retest procedure was implemented by re-annotating the entire corpus after a two-week interval. Intra-coder reliability between the two annotation iterations was evaluated using Cohen's Kappa metric to mitigate classification ambiguities arising from figurative language, irony, and sarcasm in online discourse. The textual preprocessing pipeline involved five sequential operations: case folding, text cleaning, tokenization, stopword removal, and morphological stemming. The text cleaning stage systematically stripped URLs, user mentions, hashtag symbols, numerical digits, punctuation marks, emojis, and special characters. Tokenization segmented sentences into constituent word tokens, whereas stemming utilized the Sastrawi library to reduce inflected Indonesian words to their canonical root forms. Following preprocessing, the dataset retained all 1,502 complete textual records. The cleaned texts were subsequently transformed into numerical feature representations via Term Frequency-Inverse Document Frequency (TF-IDF), weighting terms by their local document frequency against their corpus-wide rarity based on unigram tokenization. The dataset was then partitioned into an 80% training set (1,201 posts) and a 20% testing set (301 posts) using an identical partition split across all experimental runs to guarantee objective baseline comparisons. To prevent data leakage and ensure rigorous generalization, the TF-IDF feature vocabulary was fitted strictly on the training partition and subsequently projected onto the testing partition.

Three classification architectures were developed and tested: Support Vector Machine (SVM), Random Forest, and an ensemble Voting Classifier. SVM was selected for its proficiency in optimizing decision hyperplanes within high-dimensional sparse spaces, while Random Forest served as a decision-tree ensemble baseline designed to capture non-linear feature interactions and minimize variance. The Voting Classifier integrated the class predictions of both base classifiers using a majority hard-voting mechanism, evaluating whether ensemble decision fusion improved multi-class sentiment identification over standalone algorithms. Computational workflows were implemented in Python utilizing the pandas, NLTK, Sastrawi, and scikit-learn libraries within the Visual Studio Code integrated development environment, with preliminary data management conducted via Google Sheets. Model efficacy was evaluated using multi-class confusion matrices alongside aggregate and per-class metrics of accuracy, precision, recall, and F1-score. Classification metrics were mathematically derived from true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes across each sentiment category. Furthermore, Balanced Accuracy—calculated as the macro-averaged arithmetic mean of recall scores across all three individual classes—was incorporated to provide an unweighted assessment of model performance. This metric provides a rigorous diagnostic framework to

evaluate whether aggregate accuracy gains genuinely reflect balanced predictive capabilities or merely mirror the majority-class bias inherent in severely skewed social media corpora.

4. Results and Discussion

4.1 Results

The scraping procedure on platform X yielded 1,502 public posts containing the #KaburAjaDulu hashtag. All records contained complete textual entries within the full_text attribute with no missing values. The sequential execution of text preprocessing—comprising case folding, text cleaning, tokenization, stopword removal, and morphological stemming—reduced orthographic variations while preserving the complete volume of 1,502 posts. The manual annotation process revealed a severe class imbalance across the dataset, as detailed in Table 1 and illustrated in Figure 1.

Table 1. Empirical Sentiment Distribution of the #KaburAjaDulu

Sentimen	Jumlah unggahan	Persentase
Negatif	1.450	96,54%
Netral	50	3,33%
Positif	2	0,13%
Total	1.502	100,00%

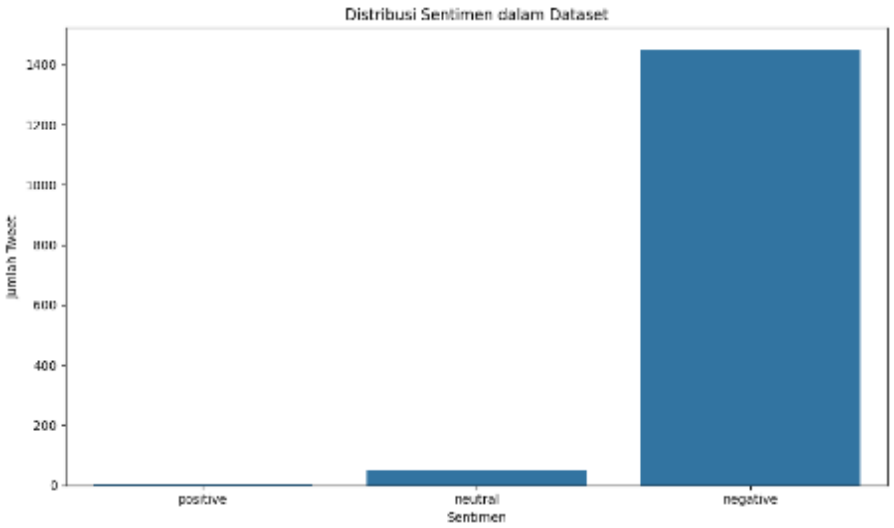

Figure 1. Empirical Sentiment Distribution of the Corpus

As shown in Table 1 and Figure 1, the corpus is overwhelmingly dominated by negative sentiment (96.54%), whereas neutral posts represent 3.33%, and positive posts constitute an extreme minority at 0.13%, establishing a negative-to-positive class imbalance ratio of 725:1. Following the 80:20 train-test split, the test partition comprised 301 posts consisting of 287 negative, 13 neutral, and 1 positive instance. This stratified partitioning preserved the empirical distribution of the overall corpus, ensuring that the evaluation partition realistically simulated the severe skewness encountered during real-world social media monitoring. When trained on the TF-IDF feature space, the Support Vector Machine (SVM) model demonstrated rapid convergence and yielded seemingly superior initial evaluation scores, achieving an aggregate accuracy of 95.35%, a precision of 90.91%, a recall of 95.35%, and an aggregate F1-score of 93.08%. However, in the context of extreme class disproportion where negative cases account for over 95% of the total instances, reliance on aggregate performance indicators can obscure critical algorithmic deficiencies. High global accuracy may merely reflect the classifier's proficiency in optimizing loss functions by predicting the overwhelming majority class while entirely ignoring sparse minority patterns. Consequently, to uncover the granular classification behavior across each individual polarity and determine whether the high aggregate accuracy represents genuine discriminative ability, an empirical cross-tabulation of ground-truth annotations against model outputs was constructed. The detailed distribution of these predictions is illustrated in the confusion matrix for SVM presented in Figure 2.

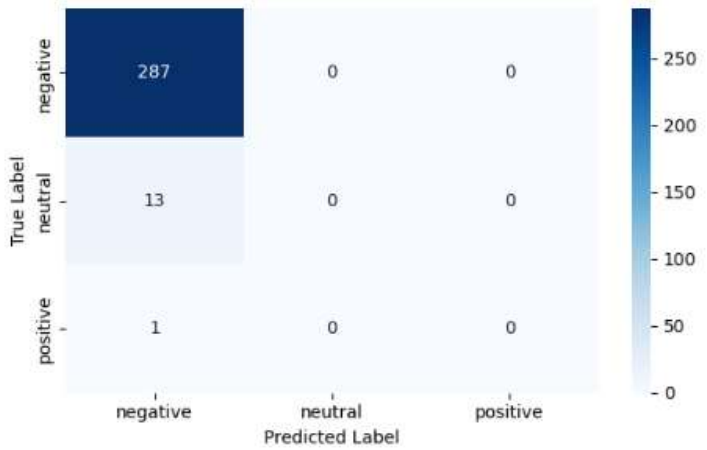

Figure 2. Confusion Matrix for the Support Vector Machine Classifier

As illustrated in Figure 2, the SVM classifier correctly predicted all 287 negative instances, yielding a per-class recall of 1.00 for the negative class. However, it misclassified all 13 neutral posts and the single positive post as negative, producing zero predictions for both minority classes. A baseline model assigning the majority negative label to all test instances attains an identical accuracy, as defined in Equation (1):

$$\frac{287}{301} \times 100\% = 95{,}35\% \quad (1)$$

The Random Forest classifier yielded performance metrics identical to those of the SVM model, achieving an aggregate accuracy of 95.35%, precision of 90.91%, recall of 95.35%, and an F1-score of 93.08%. The corresponding confusion matrix is displayed in Figure 3.

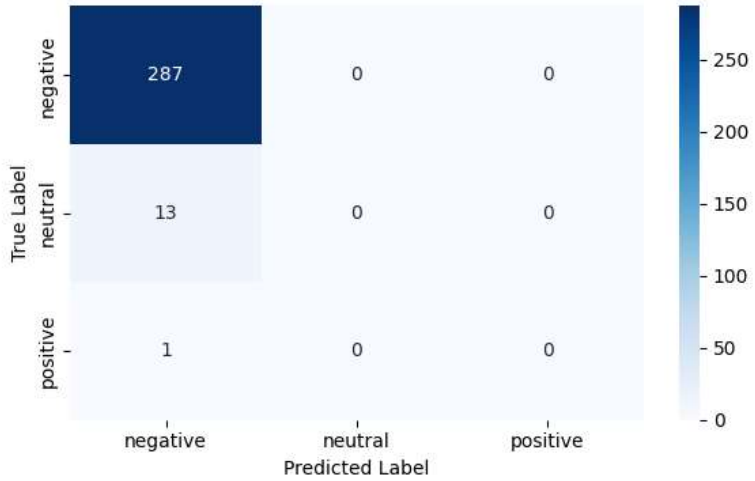

Figure 3. Confusion Matrix for the Random Forest Classifier

As indicated in Figure 3, Random Forest also classified all 287 negative posts correctly while misclassifying all 13 neutral instances and the 1 positive instance as negative, generating zero true positive predictions for the minority classes. This outcome demonstrates that despite its ensemble architecture of bagging decision trees and random feature sub-spacing, Random Forest suffered from the identical majority-class collapse observed in SVM. During recursive node partitioning, the impurity reduction criteria (such as the Gini impurity) consistently favored splits that isolated the pervasive negative instances, leaving the sparse neutral and positive samples submerged in majority-dominated terminal leaves. Consequently, both standalone models functioned as trivial zero-rule classifiers for the minority categories. To address the predictive rigidity of these individual learners, decision-level ensemble fusion was introduced via a hard-voting mechanism. By synthesizing the distinct boundary estimations of SVM's maximal-margin hyperplane and Random Forest's stochastic recursive partitions, the Voting Classifier aimed to resolve ambiguous decision spaces where solitary algorithms defaulted to the majority class. The objective of this integration was to determine whether cross-algorithmic consensus could reactivate decision boundaries for sparse textual instances without degrading

dominant-class accuracy. The resulting classification distribution across all sentiment classes is depicted in the confusion matrix for the ensemble architecture presented in Figure 4.

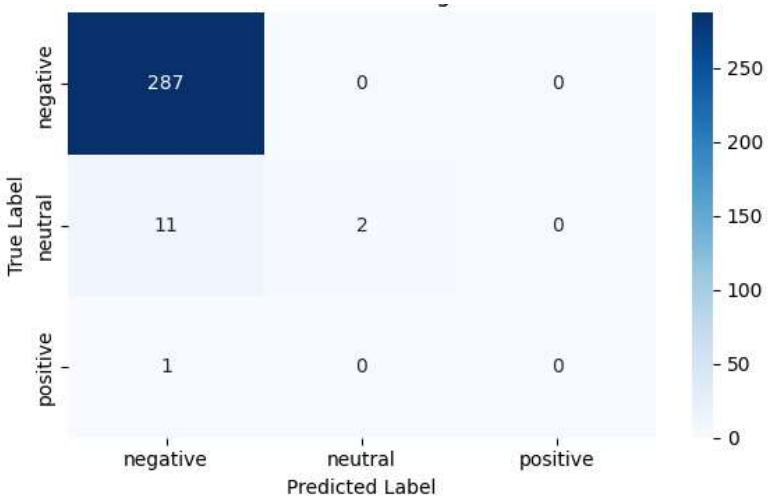

Figure 4. Confusion Matrix for the Voting Classifier Ensemble

As shown in Figure 4, the Voting Classifier correctly identified all 287 negative instances and successfully classified 2 out of the 13 neutral instances, while the remaining 11 neutral posts and the single positive instance were predicted as negative. This yielded a total of 289 correct classifications out of 301 test samples. The aggregate accuracy improvement of 0.66 percentage points over the baseline models is calculated via Equation (2):

$$\frac{2}{301} \times 100\% = 0{,}66\% \tag{2}$$

The per-class recall for the neutral category under the Voting Classifier is calculated as shown in Equation (3):

$$\frac{2}{13} = 0{,}1538 \tag{3}$$

Equation (3) indicates that the ensemble model captured 15.38% of the neutral test samples. Because both positive neutral predictions were accurate, the per-class precision for the neutral category reached 1.00; however, the model remained unable to detect the single positive instance. A consolidated comparison of aggregate metrics and balanced accuracy across all three models is presented in Table 2 and visualized in Figure 5.

Table 2. Classification Performance Comparison Across Evaluated Models

Model	Akurasi	Presisi	Recall	F1-score
SVM	0,9535	0,9091	0,9535	0,9308
Random Forest	0,9535	0,9091	0,9535	0,9308
Voting Classifier	0,9601	0,9584	0,9601	0,9455

As summarized in Table 2, the empirical metrics reveal a clear divergence between aggregate scoring and minority-class responsiveness. While the baseline Support Vector Machine and Random Forest architectures yielded identical aggregate performance across all standard metrics, the ensemble Voting Classifier achieved consistent, albeit modest, numeric improvements across accuracy, precision, recall, and F1-score. The identical precision of 0.9091 and recall of 0.9535 recorded by both standalone classifiers directly stem from their systematic defaulting to majority-class predictions, demonstrating that different algorithmic paradigms converged onto the same failure mode when unassisted by resampling or class-weighting mechanisms. In contrast, the integration of SVM and Random Forest within the hard-voting architecture enabled the model to break this empirical deadlock, yielding an aggregate accuracy gain of 0.66 percentage points and elevating precision to 0.9584. To facilitate a clearer visual inspection of these inter-model performance variations and juxtapose the uniform baseline scores against the ensemble gains, the evaluation metrics are graphically contrasted in Figure 5.

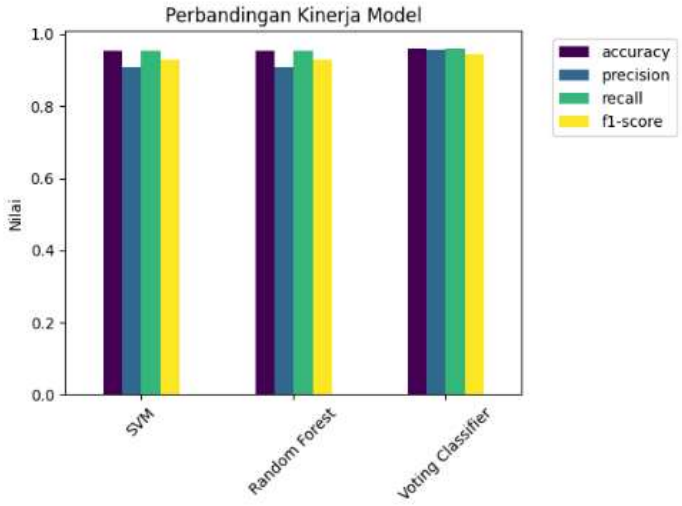

Figure 5 Comparative Performance Metrics Across Evaluated Models

The comparative metrics visualized in Figure 5 illustrate that while the ensemble Voting Classifier achieved the highest aggregate performance across all standard indicators, the nominal margin of improvement over the standalone Support Vector Machine and Random Forest baselines remained modest. The identical aggregate scores of SVM and Random Forest further underscore that algorithm selection alone does not alleviate the learning bias imposed by a heavily skewed class distribution. When evaluated through balanced accuracy to account for class representation, SVM and Random Forest achieved a balanced accuracy of only 33.33%, whereas the Voting Classifier reached 38.46%. These empirical outcomes indicate that aggregate accuracy metrics fail to reflect the true discriminatory capability of the models across all sentiment categories. A deeper contextual interpretation of these empirical findings, along with an analysis of lexical patterns and comparative benchmarks from prior studies, is detailed in the subsequent discussion section.

4.2 Discussion

The pronounced dominance of negative sentiment (96.54%) reflects the specific socio-digital context through which the #KaburAjaDulu hashtag proliferated on platform X. Rather than serving as an open, balanced forum for general migration deliberations, the hashtag functioned predominantly as a vehicle for public grievance, socio-economic frustration, and political cynicism. Users frequently engaged with the tag to express desires to escape domestic difficulties, economic constraints, or governance dissatisfaction. Consequently, the empirical distribution is an artifact of the hashtag's organic conversational framing. Nevertheless, this high proportion must not be interpreted as representing the collective sentiment of the wider Indonesian populace, as it characterizes only the active user cohort contributing to this specific hashtag during the scraping window. A critical linguistic factor underlying classification performance is the semantic overlap of dominant lexical features across distinct sentiment categories, as summarized in Table 3.

Table 3. Dominant Lexical Features Categorized by Sentiment Polarity

Sentimen	Kata dominan
Positif	sukses, selamat, gerakan, teman, negara
Netral	kabur, bagus, keren, mantap, dukung
Negatif	kabur, gak, orang, negara, gue

As detailed in Table 3, specific vocabulary items span opposing polarities; for instance, "kabur" occurs in both negative and neutral expressions, while "negara" appears across negative and positive posts. Under Term Frequency-Inverse Document Frequency (TF-IDF) representation, word vectors capture term frequency and corpus rarity but fail to encode contextual semantics, pragmatic nuance, and syntactical intent. A statement such as "Gak perlu kabur aja dulu soalnya udah diusir" uses "kabur" within a sarcastic, negative expression, whereas identical terms can occur in neutral news sharing or positive well-wishing. Similarly, positive lexical tokens such as "bagus" and "keren" emerge within neutral posts as sarcastic remarks or rhetorical questions. Because TF-IDF generates adjacent vector representations for texts sharing lexical tokens regardless of context, the decision boundaries separating neutral and positive instances become diffuse. This semantic ambiguity, compounded by the extreme scarcity of positive instances, severely restricts the classifiers from learning discriminative minority patterns. The experimental outcomes provide vital methodological insights into evaluation practices for highly skewed social media corpora. The identical performance of SVM and Random

Forest demonstrates that structural model differences—margin-maximizing hyperplanes versus randomized recursive partitioning—are secondary to training distribution characteristics. With negative cases exceeding 96%, both algorithms minimize empirical loss by defaulting to majority-class assignment. While the Voting Classifier achieved modest gains by correctly resolving two neutral instances, its aggregate accuracy of 96.01% is profoundly deceptive. When evaluated via balanced accuracy, SVM and Random Forest achieve only 33.33% (reflecting 100% negative recall and 0% recall for both minority classes), while the Voting Classifier reaches only 38.46%. This discrepancy highlights that aggregate accuracy metrics alone are insufficient and potentially misleading for imbalanced datasets.

When benchmarked against prior literature, the aggregate accuracy recorded in this study exceeds several published baselines, such as Ramlan *et al*. (2023) at 82.69% for fuel subsidy reactions, Khairudin *et al*. (2023) at 87.5% for film reviews, Tineges *et al*. (2020) at 87% for telecommunications feedback, and Darwis *et al*. (2020) at 82% for anti-corruption discourse. Conversely, it falls below the 98% accuracy reported by Idris *et al*. (2023) for commercial mobile application reviews. However, direct nominal comparisons across these benchmarks are methodologically invalid due to distinct class proportions, corpus genres, and task definitions. In commercial reviews, linguistic patterns are direct and structured around product utility, whereas viral hashtag discourse is characterized by colloquial sarcasm, contextual irony, and extreme class imbalance. The findings align with the observations of Husada and Paramita (2021) and Giovani *et al*. (2020), affirming that data distribution, preprocessing, and feature representations exert a greater influence on multi-class boundary formation than the classifier architecture itself. These findings carry important practical and methodological implications. For public policy analysts, digital hashtag tracking provides real-time access to public grievances; however, analytical pipelines must account for sampling bias and hashtag-specific emotional polarization before drawing broader societal inferences. For natural language processing practitioners, this study underscores that ensemble methods offer marginal stability gains but cannot resolve foundational data deficiencies. Evaluating models on skewed datasets demands the standard reporting of per-class precision, recall, confusion matrices, and balanced accuracy. To overcome the semantic limitations of bag-of-words representations and extreme skewness, subsequent research should integrate contextual transformer embeddings (such as IndoBERT) alongside targeted minority oversampling techniques (such as SMOTE) or synthetic text augmentation.

5. Conclusion and Recommendations

This study demonstrates that public discourse surrounding the #KaburAjaDulu hashtag on platform X was predominantly characterized by negative sentiment, which accounted for 96.54% of the corpus, while neutral and positive expressions constituted 3.33% and 0.13%, respectively. In model evaluations, the Voting Classifier ensemble attained the highest aggregate performance, achieving 96.01% accuracy, 95.84% precision, 96.01% recall, and a 94.55% F1-score, marginally surpassing the identical aggregate accuracy of 95.35% recorded by both the Support Vector Machine and Random Forest classifiers. Nevertheless, the performance gain achieved by the ensemble model was strictly limited. While the Voting Classifier resolved a small fraction of neutral instances (yielding a 15.38% per-class recall), it completely failed to detect the positive sentiment class. These findings indicate that integrating Support Vector Machine and Random Forest provides marginal gains in aggregate stability but cannot overcome severe class imbalance. Furthermore, the results emphasize that high aggregate accuracy can be profoundly misleading in skewed datasets, reinforcing the critical need to evaluate classification systems through per-class metrics, balanced accuracy, and confusion matrices. Future research should prioritize balancing strategies, such as the Synthetic Minority Over-sampling Technique (SMOTE) or targeted text augmentation, and investigate contextual language representations using transformer-based architectures, such as IndoBERT, to capture semantic nuance and enhance minority-class discrimination in Indonesian social media text.

References

Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. (2021). Penerapan metode support vector machine untuk analisis sentimen pengguna Twitter. *Journal of Information System Management (JOISM)*, *3*(2), 44–49. https://doi.org/10.24076/joism.2021v3i2.558

Bourequat, W., & Mourad, H. (2021). Sentiment analysis approach for analyzing iPhone release using support vector machine. *International Journal of Advances in Data and Information Systems*, *2*(2), 75–84. https://doi.org/10.25008/ijadis.v2i2.1228

Cahyani, R. D., & Prasetyaningrum, P. T. (2026). Sentiment analysis of user reviews for AI applications: Evaluating SVM, logistic regression, and random forest. *Journal of Information Systems and Informatics*, *8*(1), 1–12.

Darwis, D., Pratiwi, E. S., & Pasaribu, A. F. O. (2020). Penerapan algoritma SVM untuk analisis sentimen pada data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Edutic: Scientific Journal of Informatics Education*, *7*(1), 1–11. https://doi.org/10.21107/edutic.v7i1.8779

Ditami, G. R., Ripanti, E. F., & Sujaini, H. (2022). Implementasi support vector machine untuk analisis sentimen terhadap pengaruh program promosi event belanja pada marketplace. *Jurnal Edukasi dan Penelitian Informatika*, *8*(3), 508–517. https://doi.org/10.26418/jp.v8i3.56478

Fikri, M., & Sarno, R. (2020). A comparative study of sentiment analysis using support vector machine and SentiWordNet. *International Journal of Electrical and Computer Engineering*, *10*(6), 6340–6348. https://doi.org/10.11591/ijece.v10i6.pp6340-6348

Fitriyah, N., Warsito, B., & Maruddani, D. A. I. (2020). Analisis sentimen Gojek pada media sosial Twitter dengan klasifikasi support vector machine (SVM). *Jurnal Gaussian*, *9*(3), 376–390. https://doi.org/10.14710/j.gauss.v9i3.28913

Giovani, A. P., Ardiansyah, Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis sentimen aplikasi Ruangguru di Twitter menggunakan algoritma klasifikasi. *Jurnal Teknoinfo*, *14*(2), 115–123. https://doi.org/10.33365/jti.v14i2.679

Hagi, A., & Rarasati, D. B. (2024). Sentiment analysis of Sirekap application review using logistic regression algorithm. *Jurnal Informatika*, *11*(2), 55–64. https://doi.org/10.31294/inf.v11i2.22066

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and. *Techniques, Waltham: Morgan Kaufmann Publishers*, *13*.

Handayani, R. N. (2021). Optimasi algoritma support vector machine untuk analisis sentimen pada ulasan produk Tokopedia menggunakan PSO. *Media Informatika*, *20*(2), 97–108. https://doi.org/10.37595/mediainfo.v20i2.59

Husada, H. C., & Paramita, A. S. (2021). Analisis sentimen pada maskapai penerbangan di platform Twitter menggunakan algoritma Support Vector Machine. *Teknika*, *10*(1), 18–26. https://doi.org/10.34148/teknika.v10i1.311

Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis sentimen terhadap penggunaan aplikasi Shopee menggunakan algoritma Support Vector Machine. *Jambura Journal of Electrical and Electronics Engineering*, *5*(1), 32–35. https://doi.org/10.37905/jjeee.v5i1.16830

Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Sentimen analisis publik terhadap kebijakan lockdown Pemerintah Jakarta menggunakan algoritma SVM. *Jurnal Data Mining dan Sistem Informasi*, *2*(1), 31–37. https://doi.org/10.33365/jdmsi.v2i1.1021

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed., draft). Stanford University. https://web.stanford.edu/~jurafsky/slp3/

Khairudin, M., Sukendar, A., & Somantri, A. (2023). Analisis sentimen film di Twitter menggunakan metode support vector machine. *Jurnal Sains dan Sistem Teknologi Informasi (SANDI)*, *5*(1), 97–102. https://doi.org/10.59811/sandi.v5i1.47

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Mariska, I. V., Meiriza, A., & Lestarini, D. (2025). Comparison of support vector machine and random forest algorithms in sentiment analysis of the JMO Mobile application. *Journal of Applied Informatics and Computing*, *9*(5), 789–798. https://doi.org/10.30871/jaic.v9i5.10764

Nurmadewi, D., Jailani, Z. F., Rafi, H., & Anggoro, D. A. (2026). Perbandingan support vector machine dan Naïve Bayes untuk klasifikasi sentimen ulasan e-commerce. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *6*(2), 145–154.

Pranata, A., Budianita, E., Yusra, Y., & Cynthia, E. P. (2022). Klasifikasi sentimen terhadap Maxim menggunakan algoritma support vector machine. *Jurnal Nasional Komputasi dan Teknologi Informasi*, *5*(3), 334–342. https://doi.org/10.32672/jnkti.v5i3.4412

Putra, A. R. P., Wibowo, J. S., & Juang, J. T. L. (2024). Analisa sentimen Twitter terhadap Capres Indonesia 2024 menggunakan metode K-Nearest Neighbor. *Jurnal Ilmiah Elektronika dan Komputer*, *17*(1), 111–119. https://doi.org/10.51903/elkom.v17i1.1685

Ramlan, R., Satyahadewi, N., & Andani, W. (2023). Analisis sentimen pengguna Twitter menggunakan support vector machine pada kasus kenaikan harga BBM. *Jambura Journal of Mathematics*, *5*(2), 431–445. https://doi.org/10.34312/jjom.v5i2.20860

Safitri, T., Umaidah, Y., & Maulana, I. (2023). Analisis sentimen pengguna Twitter terhadap grup musik BTS menggunakan algoritma Support Vector Machine. *Journal of Applied Informatics and Computing*, *7*(1), 28–35. https://doi.org/10.30871/jaic.v7i1.5039

Suryana, A., Purnamasari, A. I., & Ali, I. (2024). Mengoptimalkan kepuasan pengguna: Analisis sentimen review aplikasi Grab di Indonesia. *JATI: Jurnal Mahasiswa Teknik Informatika*, *8*(3), 3396–3404. https://doi.org/10.36040/jati.v8i3.9688

Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis sentimen terhadap layanan IndiHome berdasarkan Twitter dengan metode klasifikasi Support Vector Machine. *Jurnal Media Informatika Budidarma*, *4*(3), 650–658. https://doi.org/10.30865/mib.v4i3.2181

Widiarta, I. P. A. P., Dwiyansaputra, R., & Aranta, A. (2023). Analisis sentimen masyarakat terhadap kebijakan penerapan PPKM di media sosial Twitter menggunakan metode XGBoost. *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya*, *5*(2), 154–163. https://doi.org/10.29303/jtika.v5i2.342