Comparing Naïve Bayes and SVM for Public Sentiment on Online Gambling Policies

Siti Qomariah[1], Widya Noviana Noor [2], Nazwa Risanti[3]
[1,3]Faculty of Ushuluddin, Adab and Da'wah, Universitasn Islam Negeri Sultan Aji Muhammad Idris Samarinda, Indonesia
[2]Faculty of Tarbiyah and Teacher Training, Universitas Islam Negeri Sultan Aji Muhammad Idris Samarinda, Indonesia

ABSTRACT

Online gambling remains a persistent digital-governance challenge in Indonesia, as gambling links, promotional accounts, and user complaints continue to surface on social media despite government blocking and reporting mechanisms. This study examines public sentiment regarding the implementation of anti-online gambling policies in Indonesia and compares the performance of Naïve Bayes and the Linear Support Vector Machine (SVM) for classifying Indonesian social media comments. Data were collected from Instagram and TikTok between January 2024 and November 2025. The initial dataset comprised 746 user comments (153 from Instagram and 593 from TikTok). For supervised learning, only manually labelled comments were included in the analysis. After excluding unlabeled records, irrelevant entries, and texts rendered empty by preprocessing, 338 valid labelled comments remained. The corpus was divided using an 80:20 stratified split. To prevent data leakage, SMOTE was applied exclusively to the training set, while the testing set retained its original distribution. A target-balanced SMOTE strategy generated 1,110 training instances, with 370 instances per sentiment class. Text features were represented using unigram TF-IDF. SVM achieved the highest accuracy (73.53%) and weighted recall (0.735), with a weighted F1-score of 0.698. Naïve Bayes attained 67.65% accuracy, higher weighted precision (0.745), and a weighted F1-score of 0.693. Therefore, SVM demonstrated superior accuracy and recall-oriented weighted evaluation, while Naïve Bayes remained competitive in precision. Public comments generally supported online-gambling eradication but also criticised the persistence of gambling sites, advertisements, and perceived enforcement delays.

Keywords: sentiment analysis, Naïve Bayes, SVM, online gambling, Instagram, TikTok, TF-IDF, SMOTE

Corresponding Author:
Siti Qomariah,
Faculty of Ushuluddin, Adab, and Da'wah
Universitas Islam Negeri Sultan Aji Muhammad Idris Samarinda
Jl. H.A.M. Rifaddin, Loa Janan Ilir, Samarinda 75251, Kalimantan Timur, Indonesia
Email: siti.qomariah@uinsi.ac.id

1. INTRODUCTION

Online gambling in Indonesia is a legal, moral, and digital-governance issue. Gambling-related links, promotional comments, and user reports circulate rapidly on social media, increasing pressure on government agencies to enhance blocking, monitoring, and public reporting mechanisms (Ahyani et al., 2024). In this context, public comments on social media serve as evidence for assessing citizens' responses to anti-online-gambling policies (Liu, 2012; A. D. Putra et al., 2022) . Prior research has addressed online gambling from legal, social, religious, psychological, and policy perspectives. Additional studies have applied sentiment analysis to investigate public opinion on government regulations, public services, and digital issues(Pangestu & Harahap, 2024) . However, computational studies specifically examining public sentiment toward the implementation of anti-online gambling policy in Indonesia remain limited, particularly those using Instagram and TikTok comment data (Antonius et al., 2024; Miranda et al., 2021).

A further gap concerns methodological transparency. Sentiment analysis studies frequently report model accuracy without adequately detailing the original dataset size, labelling procedures, class imbalance, SMOTE implementation, or the distinction between original comments and synthetic training instances (Santos & Henriques, 2023) . Such details are critical, as oversampling prior to train-test splitting can artificially inflate model performance due to data leakage (Yahya et al., 2021).

This study addresses these gaps by analysing manually labelled Instagram and TikTok comments related to online gambling policy enforcement and by comparing Naïve Bayes and SVM within a transparent TF-IDF and SMOTE workflow. The article contributes methodologically by reporting the complete data-screening process, class distribution, SMOTE strategy, and model configuration. In practical terms, it provides evidence of how Indonesian social media users express support, criticism, and neutrality toward efforts to eradicate online gambling (Fahrudin et al., 2024; Muflih et al., 2021; Susilowati et al., 2015)

2. RESEARCH METHOD

This study used a quantitative computational design with Natural Language Processing (NLP) and supervised machine learning (Medhat et al., 2014). The unit of analysis was a social media comment related to online gambling, blocking, reporting, gambling advertisements, policy enforcement, or government digital monitoring. The workflow consisted of scraping, data screening, sentiment labelling, preprocessing, TF-IDF feature extraction, training-set balancing with SMOTE, model training, and testing on the original, unseen comments.(Feldman & Sanger, n.d.; Isnain et al., 2021)

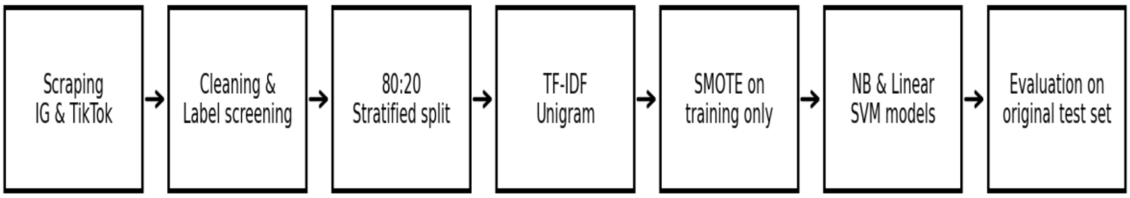

Figure 1. Research workflow used in the revised experiment

A. Data Collection and Dataset Description

Data were collected from Instagram and TikTok from January 2024 to November 2025. These platforms were selected because Indonesian users frequently use both to respond to public issues through captions, short videos, and comment threads . The scraping process used keyword-based filtering related to online gambling and policy enforcement, including terms such as "judi online", "blokir situs", "iklan judi", "rekening dormant", "Kominfo", and related expressions found in the comment threads (Kowalczyk, 2017; Z. D. W. Putra, 2022; Taruk et al., 2023)

The raw scraping file contained 746 user comments, comprising 153 from Instagram and 593 from TikTok. Only comments with manually assigned sentiment labels were used for supervised classification. After excluding unlabeled comments and comments that became empty after preprocessing, 338 valid labelled comments were retained for analysis. The cleaned analytical corpus consisted of 143 Instagram comments and 195 TikTok comments .

Table 1. Dataset source and screening summary

Platform	Raw comments	Labeled comments	Valid labeled comments	Scraping period
Instagram	153	153	143	January 2024-November 2025
TikTok	593	201	195	January 2024-November 2025
Total	746	354	338	January 2024-November 2025

Source: Researcher's Processed Data (2025)

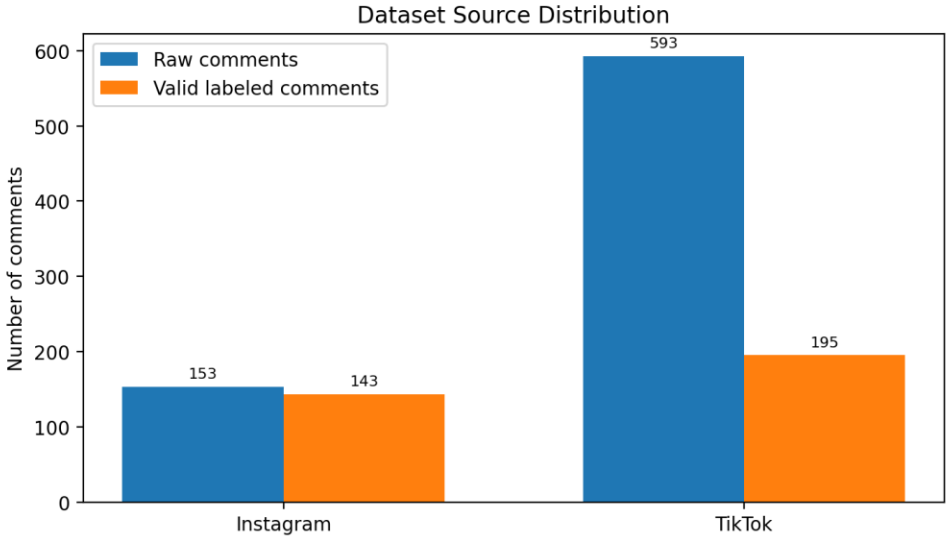

Figure 2. Distribution of raw and valid labeled comments by platform.

B. Sentiment Labelling Procedure

The dataset included a manual sentiment label for supervised learning. The labels were normalised into three categories: Negative, Neutral, and Positive. Comments were labelled negative when they expressed dissatisfaction, criticism, complaints, distrust, or reports about the persistence of online gambling content. Comments were labelled positive when they expressed support for enforcement, appreciation for blocking efforts, or encouragement to eradicate online gambling. Comments were labelled neutral when they were descriptive or informational, or when they did not convey a clear evaluative stance.

Table 2. Sentiment distribution in the valid analytical corpus

Sentiment class	Instagram	TikTok	Total
Negative	125	121	246
Neutral	5	21	26
Positive	13	53	66
Total	143	195	338

Source: Researcher's Processed Data (2025)

C. Text Preprocessing and TF-IDF Representation

The comment texts were preprocessed before modelling. The preprocessing stages included lowercasing, removal of URLs, mentions, excessive symbols, punctuation, emojis, and duplicate spaces, followed by token normalisation, stopword removal, and stemming. The cleaned text was then transformed into numerical vectors using Term Frequency-Inverse Document Frequency (TF-IDF). A unigram representation was used to capture individual terms commonly appearing in Indonesian social media comments.

D. Train-Test Split and SMOTE Implementation

The valid labelled dataset was divided using an 80:20 stratified split. The training set contained 270 comments, and the testing set contained 68 original comments. SMOTE was applied only to the training set to avoid data leakage. The testing set was not oversampled and was preserved in its original distribution so that model performance was evaluated on real, unseen comments. In the final experiment, a target-balanced SMOTE strategy was set to 370 instances per sentiment class, yielding 1,110 balanced training instances.

Table 3. Sentiment class distribution before and after SMOTE

Sentiment class	Training before SMOTE	Training after SMOTE	Original testing set
Negative	196	370	50
Neutral	21	370	5
Positive	53	370	13
Total	270	1110	68

Source: Researcher's Processed Data (2025)

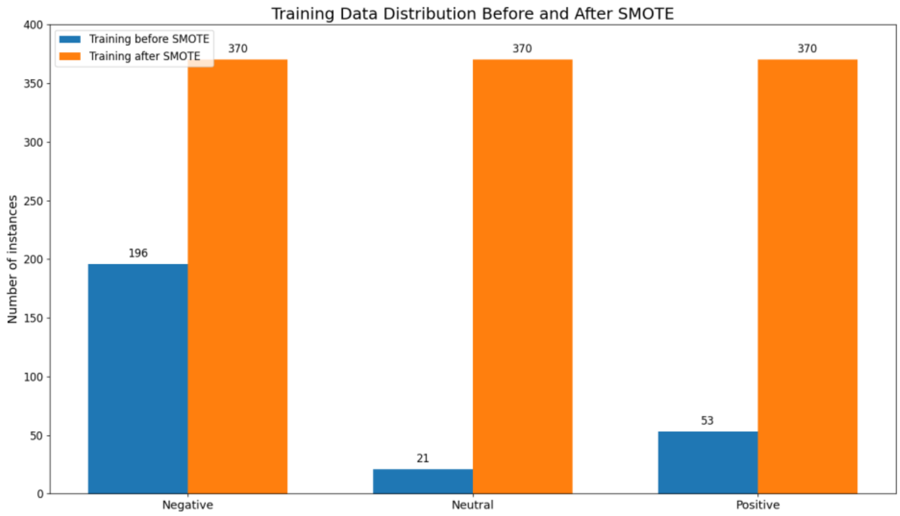

Figure 3. Training-set class distribution before and after target-balanced SMOTE.

E. Model Configuration and Evaluation

Two classifiers were compared using the same TF-IDF features: Multinomial Naïve Bayes and SVM. Naïve Bayes was used as a probabilistic baseline because of its efficiency in text classification, whereas SVM was selected because sparse TF-IDF vectors are commonly separable in high-dimensional feature spaces. Model performance was evaluated using accuracy, weighted precision, weighted recall, weighted F1-score, macro averages, class-level metrics, and confusion matrices.(Susilowati et al., 2015)

Table 4. Software and model configuration

Component	Configuration
Programming language	Python
Libraries	pandas, scikit-learn, imbalanced-learn, matplotlib
Feature extraction	TF-IDF, unigram
Data split	80:20 stratified split; random_state = 42
Balancing method	SMOTE applied to training set only
SMOTE target	370 instances per class; total 1,110 training instances
Naïve Bayes	MultinomialNB, alpha = 1.0
SVM	LinearSVC, C = 1.0
Evaluation	Accuracy, precision, recall, F1-score, confusion matrix

3. RESULTS AND DISCUSSION

This section presents the experimental results obtained from sentiment classification using Naïve Bayes and Support Vector Machine (SVM). The analysis includes quantitative evaluation using classification metrics as well as qualitative interpretation through confusion matrices and WordCloud visualisations (Liu, 2012), (Medhat et al., 2014).

A. Classification Performance

The performance comparison between Naïve Bayes and SVM is presented in Table 1.

Table 5. Performance Comparison of Naïve Bayes and SVM

Evaluation Metrics	Naïve Bayes	Support Vector Machine (SVM)	Performance Difference
Accuracy	67.65%	73.53%	+5.88%
Weighted precision	0.745	0.686	-0.059
Weighted recall	0.676	0.735	+0.059
Weighted F1-score	0.693	0.698	+0.005

Source: Researcher's Processed Data (2025)

The results indicate that SVM achieved the highest accuracy, weighted recall, and weighted F1-score. Specifically, SVM attained 73.53% accuracy, weighted recall of 0.735, and a weighted F1-score of 0.698, whereas Naïve Bayes achieved 67.65% accuracy, weighted recall of 0.676, and a weighted F1-score of 0.693. Although the difference in weighted F1-score is modest, SVM demonstrated stronger classification coverage on the original test set. Naïve Bayes yielded higher weighted precision and macro F1-score, suggesting a more conservative approach and improved balance on certain minority-class indicators. Consequently, SVM should be regarded as superior primarily in terms of accuracy, weighted recall, and weighted F1-score, but not across all evaluation metrics. These findings support the study's contribution by showing that SVM is the stronger option for this sentiment-classification task under the reported workflow.

Table 6. Class-level performance on the original testing set

Model	Class	Precision	Recall	F1-score	Support
Naïve Bayes	Negative	0.818	0.720	0.766	50
Naïve Bayes	Neutral	1.000	0.400	0.571	5
Naïve Bayes	Positive	0.364	0.615	0.457	13
SVM	Negative	0.780	0.920	0.844	50
SVM	Neutral	0.667	0.400	0.500	5
SVM	Positive	0.333	0.154	0.211	13

B. Confusion Matrix Analysis

The confusion matrices show that both models classified the negative class more accurately than the neutral and positive classes. This pattern is expected because negative comments dominated the original corpus and the test set. SVM correctly classified 46 of 50 negative comments but struggled with positive comments, classifying only 2 of 13 correctly. This result suggests that positive comments in the corpus often shared vocabulary with negative complaints, such as "judi", "blokir", "situs", and "vpn", making separation difficult with unigram TF-IDF features.

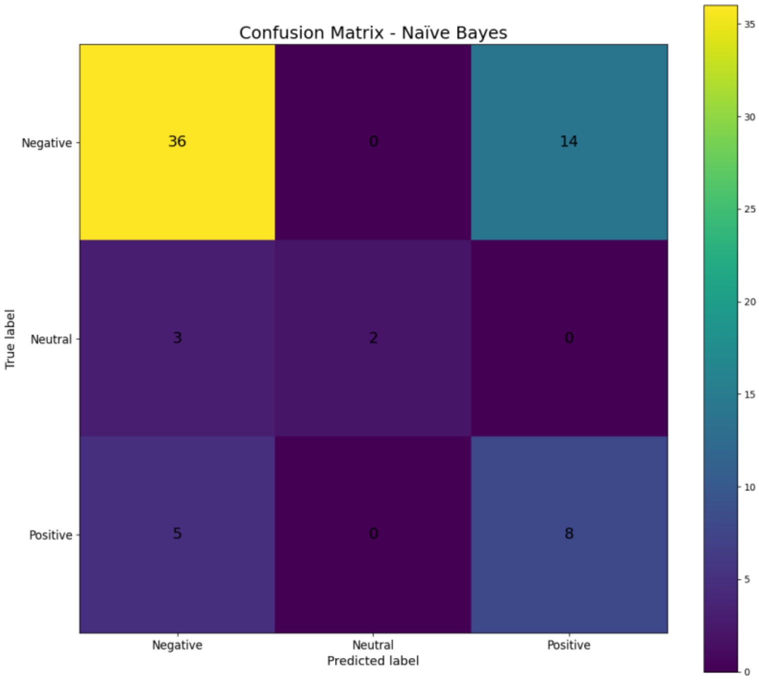

Figure 4. Confusion matrix of Naïve Bayes on the original test set.

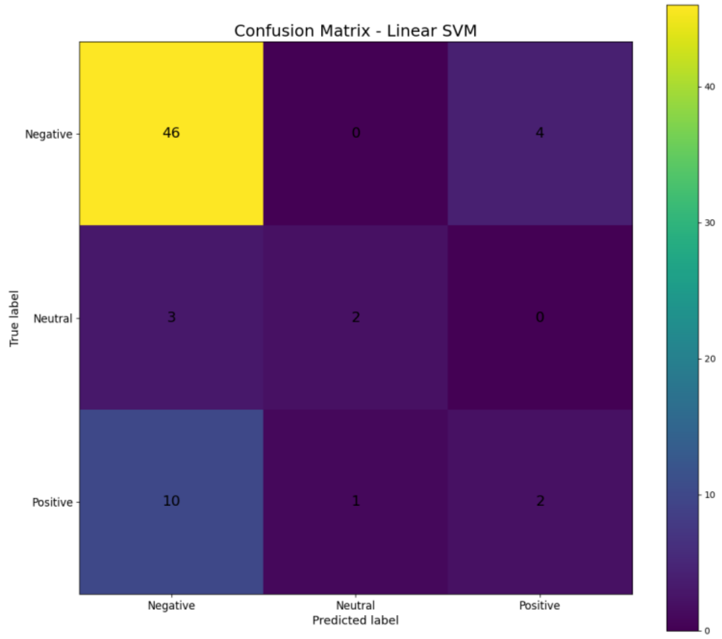

Figure 5. Confusion matrix of SVM on the original test set

C. WordCloud Analysis

WordCloud visualisation was used to identify dominant terms appearing within each sentiment category. The negative sentiment WordCloud reveals that terms such as "iklan," "terlambat," "masih," and "muncul" frequently appeared in user discussions. These terms indicate public dissatisfaction regarding the persistence of illegal gambling advertisements and the perceived slow response of enforcement authorities. Similar concerns regarding online gambling exposure and ineffective monitoring have also been discussed in previous studies.

Figure 3. Negative Sentiment WordCloud

In contrast, the positive sentiment WordCloud highlights terms such as "dukung," "berantas," "blokir," and "polisi," reflecting public support toward government efforts to eradicate online gambling activities and strengthen digital monitoring systems (Miranda et al., 2021). The neutral sentiment category mainly contains descriptive and informational terms with limited emotional expression. This characteristic explains why neutral sentiment was classified more accurately than other categories.

Figure 4. Positive Sentiment WordCloud

The WordCloud analysis complements the quantitative findings by illustrating dominant public concerns and attitudes toward anti–online gambling policies in Indonesia.

D. Discussion

The findings support the expectation that SVM is suitable for sparse TF-IDF representations of Indonesian social media comments, particularly when accuracy and weighted recall are prioritised. SVM constructs a separating hyperplane in a high-dimensional feature space, which helps when comments contain many sparse lexical cues. Naïve Bayes remains a useful baseline and produced higher weighted precision and macro F1-score in this experiment, but its conditional-independence assumption may be too restrictive for informal comments where meaning depends on combinations of words, context, and platform-specific expressions.

The results also demonstrate the importance of transparent reporting of SMOTE implementation. The original corpus contained 338 valid comments, whereas 1,110 refers to balanced training instances generated after SMOTE, not the number of original social media comments. By applying SMOTE only to the training set and preserving the test set's original distribution, the evaluation avoids data leakage and yields a more credible estimate of model performance.

From a policy perspective, the comments suggest that citizens generally support the eradication of online gambling, but they question implementation effectiveness when gambling advertisements, links, or applications remain visible. Therefore, sentiment analysis can help policymakers identify public concerns, recurring complaint themes, and perceived weaknesses in digital enforcement mechanisms.

E. Limitations

This study has several limitations. First, the supervised experiment used 338 valid labelled comments; therefore, the results should be interpreted as an initial computational analysis rather than a definitive representation of all Indonesian social media users. Second, the neutral class was small, reducing model stability in that category. Third, the study used unigram TF-IDF features; future studies may incorporate character n-grams, contextual embeddings, or transformer-based Indonesian language models to better capture sarcasm, informal spelling, and short-comment ambiguity.

4. CONCLUSION

This study analysed public sentiment toward the implementation of anti-online gambling policies in Indonesia using Instagram and TikTok comments collected during the scraping period from January 2024 to November 2025. The raw dataset consisted of 746 comments, but only 338 valid labelled comments were used for supervised classification after screening and preprocessing. The data were split into training and testing sets using an 80:20 stratified split. SMOTE was applied only to the training set, producing 1,110 balanced training instances, while the test set remained in its original distribution. The results show that SVM achieved the highest accuracy of 73.53%, weighted recall of 0.735, and weighted F1-score of 0.698, while Naïve Bayes achieved 67.65% accuracy, weighted recall of 0.676, and weighted F1-score of 0.693. However, Naïve Bayes achieved higher weighted precision and macro F1-score, so SVM should be understood as stronger mainly in terms of accuracy and weighted recall/F1 evaluation. Substantively, negative comments were linked to dissatisfaction with persistent gambling sites, advertisements, reporting mechanisms, and perceived enforcement delays, whereas positive comments reflected support for stricter blocking and eradication efforts. Future research should expand the labelled corpus, include more balanced sentiment categories, and test advanced Indonesian-language models such as IndoBERT or other transformer-based classifiers.

ACKNOWLEDGEMENTS

The authors express their appreciation to Universitas Islam Negeri Sultan Aji Muhammad Idris Samarinda for institutional support during manuscript preparation. This study was conducted independently without external funding. The authors also thank colleagues who provided input during data processing, model evaluation, and manuscript revision.

REFERENCES

Ahyani, H., Harahap, A. M., Solehudin, E., Adnan, N. I. M., Mutmainah, N., Siripipatthanakul, S., Lousada, S. A. N., Putra, H. M., & Novita, D. (2024). Effectiveness of Law Enforcement Against Online Gambling Practices in Indonesia in Supporting the Achievement of SDGs. *Journal of Lifestyle and SDGs Review*, *5*(1), e03686. https://doi.org/10.47172/2965-730X.SDGsReview.v5.n01.pe03686

Antonius, R., Zulkarnain, A. R., & Irsyad, H. (2024). Pendekatan TF-IDF, SMOTE, dan SVM dalam Klasifikasi Sentimen Masyarakat terhadap Pemblokiran Judi Online. *Buletin Ilmiah Informatika Teknologi*, *2*(3), 115–122.

Fahrudin, A., Satispi, E., Subardhini, M., Rinda Andayani, R. H., Jayaputra, A., Yuniarti, L., Wijayanti, F., & Suryani, S. (2024). Online gambling addiction: Problems and solutions for policymakers and stakeholders in Indonesia. *Journal of Infrastructure, Policy and Development*, *8*(11), 9077. https://doi.org/10.24294/jipd.v8i11.9077

Feldman, R., & Sanger, J. (n.d.). *The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data*.

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. https://doi.org/10.33365/jdmsi.v2i1.1021

Kowalczyk, A. (2017). Support vector machines succinctly. *Syncfusion Inc*, *317*.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Miranda, E., Elias, R. A., Kibtiah, T. M., & Permana, A. (2021). Indonesia China Trade Relations, Social Media and Sentiment Analysis: Insight from Text Mining Technique. *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, 334–339. https://doi.org/10.1109/ICCSAI53272.2021.9609735

Muflih, H. N., Putra, W. H. N., & Arwani, I. (2021). Analisis Sentimen pada Media Sosial Twitter terhadap AturanPemberlakuan Pembatasan Kegiatan Masyarakat Kota Malang denganMetode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *5*(10), 4486–4493. http://j-ptiik.ub.ac.id

Pangestu, A. D., & Harahap, L. S. (2024). Analisis Sentimen Terkait Judi Online di Media Sosial Instagram Menggunankan Naïve Bayes. *Indonesian Journal of Education and Development Research*, *3*(1), 556–561.

Putra, A. D., Pudjiantoro, T. H., Renaldi, F., & Hadiana, A. I. (2022). Analyzing sentiments on official online lending platform in Indonesia with a Combination of Naive Bayes and Lexicon Based Method. *2022 International Conference on Science and Technology (ICOSTECH)*, 1–6. https://doi.org/10.1109/ICOSTECH54296.2022.9829141

Putra, Z. D. W. (2022). The Sentiments of Indonesian Urban Citizens Regarding the Lockdown-Like Policy During the COVID-19 Pandemic. *International Journal of E-Planning Research*, *11*(1), 1–20. https://doi.org/10.4018/IJEPR.297515

Santos, R. M., & Henriques, R. (2023). Accurate, timely, and portable: Course-agnostic early prediction of student performance from LMS logs. *Computers and Education: Artificial Intelligence*, *5*(October), 100175. https://doi.org/10.1016/j.caeai.2023.100175

Susilowati, E., Sabariah, M. K., & Gozali, A. A. (2015). Implementasi Metode Support Vector Machine untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter. *E-Proceeding of Engineering*, *2*(1), 1478–1484.

Taruk, M., Septiarini, A., & Al Akbar, F. A. (2023). Analisis Sentimen Terhadap Kebijakan Pemerintah Terkait Pandemi Covid-19 Pada Twitter. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, *7*(1), 104–112.

Yahya, A. A., Sulaiman, A. A., Mashraqi, A. M., Zaidan, Z. M., & Halawani, H. T. (2021). Toward a better understanding of academic programs educational objectives: A data analytics-based approach. *Applied Sciences (Switzerland)*, *11*(20). https://doi.org/10.3390/app11209623