International Journal Of Health, Engineering And Technology (IJHET) E-ISSN 2829 - 8683
Volume 5, Number 1, May 2026, *Page. 248 - 257*
Email : editorijhess@gmail.com

Comparative Study Of Public Sentiment Analysis On Ikn (Nusantara Capital City) Across X And Youtube Using Pso-Optimized Svm

Awaliah Fitri Nur Ananda[1)]
[1)] Informatics and Computer Engineering Education, Faculty of Engineering, Universitas Negeri Makassar, Makassar

*Corresponding Author
Email : awaliahftrr@gmail.com

Abstract

This study aims to compare public sentiment regarding the IKN on Platform X and YouTube using a Polynomial Kernel SVM algorithm optimized by PSO. The method used is an experimental study. Data was obtained through web scraping using Google Colab, yielding 1,413 tweets for the X dataset and 814 for the YouTube dataset. The collected data underwent data cleaning, followed by sentiment labeling of each data point into three classes. Following this, TF-IDF vectorization, PSO-based feature selection, and classification using the Polynomial Kernel SVM were performed. The results of the study showed an accuracy of 76% for the PSO-SVM on the X platform and 75% for the YouTube platform. These results indicate that the PSO-SVM algorithm performs better on the X platform compared to the YouTube platform.

Keywords: Comparative Study, IKN, PSO, SVM, Sentiment Analysis, X, Youtube.

INTRODUCTION

In recent years, the public has been captivated by the discourse surrounding the development of the National Capital City (IKN). The relocation of the National Capital City (IKN) from Jakarta to East Kalimantan has become one of the most significant policies in the history of the Indonesian government. The government has offered various reasons for this relocation, such as reducing the burden on overcrowded Jakarta, accelerating economic development outside Java, and addressing worsening environmental problems (Puspitaningrum & Sundoko, 2023). However, this policy has generated both pros and cons on Indonesian social media. Due to the numerous pro and con comments circulating among the public, it is crucial to understand public sentiment more deeply.

In this context, sentiment analysis is crucial in understanding how the public responds to policies and plans for the development of the IKN. Sentiment analysis is a data processing method that analyzes text and identifies the content of opinions within the text (S. Vanaja, 2018) in (Utama & Masruro, 2022). Sentiment analysis is applied to identify positive, negative, or neutral opinions (Villavicencio et al., 2021). Data obtained from sentiment analysis can provide new insights for policymakers in formulating communication and policy approaches that are responsive to public needs, typically derived from data found on social media.

X is one of the most popular social media platforms that helps raise awareness of these crucial issues (Aslan et al., 2023). With its characteristics that enable rapid and widespread information dissemination, X is a rich data source for sentiment analysis. Other research suggests that using X data can provide a comprehensive picture of public opinion on various issues, including government policies and development projects such as the New Capital City (IKN) (Rusydi Umar et al., 2023). This data processing utilizes algorithms. Based on the X platform's requirements, data crawling using the X API key can only retrieve 500 tweets per day, while based on the requirements of Google applications like YouTube, the data can be retrieved up to 5,000 per day.

One algorithm widely used in sentiment analysis is the Support Vector Machine (SVM), known for its ability to handle complex and multidimensional data. The SVM algorithm works by finding the optimal hyperplane to separate classes in the data, often resulting in superior accuracy

compared to the Naïve Bayes Classifier in certain contexts. Research suggests that the SVM algorithm exhibits excellent accuracy and is easy to apply. A study analyzing public sentiment toward the Jakarta government's lockdown policy using the SVM algorithm yielded 68.75% positive sentiment and 31.25% negative sentiment, with an accuracy rate of 74% (Isnain et al., 2021).

A study by Jamil et al. (2024) comparing the SVM and Naïve Bayes Classifier algorithms on disaster management using the YouTube platform showed that, in terms of performance measurement evaluation using a pre-processed dataset, the Naïve Bayes Classifier achieved the highest accuracy rate of 0.804 (80.4%) with an execution time of 0.0097 seconds. Meanwhile, with an execution time of 193.48 seconds, an accuracy rate of 0.723, or 72.3%, was achieved using the SVM approach.

Referring to previous research, this study will analyze public sentiment towards the new capital city using the SVM algorithm and Particle Swarm Optimization (PSO) using data from social media platform X and YouTube. This research is expected to provide insight into the effectiveness of the SVM algorithm, which uses PSO as a feature selection tool, in the context of analyzing public sentiment towards the new capital city.

RESEARCH METHODS

This study is an experimental study using a quantitative approach. This study will conduct an experiment by applying an SVM algorithm optimized by Particle Swarm Optimization (PSO) to the data to assess the algorithm's performance in detecting public sentiment toward the IKN. The study will then generate numerical data as the results of the experiment, such as accuracy, precision, recall, and F1-score. The following is the research design that will be conducted:

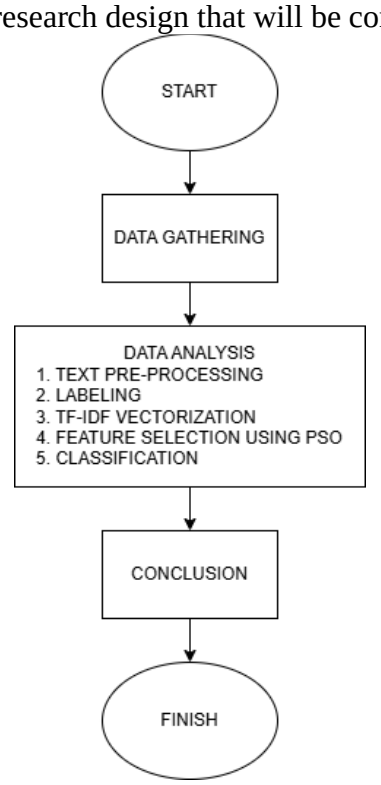

Figure 1 Research design flowchart

A. Data Collection: In this stage, data is collected through the tweet-harvest website and Google Colab.

B. Data Analysis: In this stage, the collected data will be analyzed.

1. Text Pre-Processing: Cleaning the text data from X and YouTube into its root words so that it is ready for use in the next stage using Google Colab.

2. Labeling: Labeling each comment using IndoBERT, which produces positive and negative data. IndoBERT is a pre-trained BERT for Indonesian that uses a large Indonesian corpus of four billion pre-trained words (Putri et al., 2024).
3. TF-IDF Vectorization: Vectorizing using TF-IDF to prepare the data for use in the feature selection process. TF-IDF serves as a basic classification medium, which improves text classification performance by evaluating the weights of terms in the dataset (Jain et al., 2024). The formula for finding the TF-IDF equation is as follows (Sari et al., 2021):

The formula for finding the Term Frequency (TF) value is as in Equation 1

$$TF_{t,d} = 1 + log\, tf \quad (1)$$

where:

$TF_{t,d}$: total occurrences of term t in document d
Tf : total occurrences of term in document

The formula for finding the Inverse Document Frequency (IDF) value is as in Equation 2

$$IDF_t = log \frac{N}{df_t} \quad (2)$$

where:

IDF_t : inverse document frequency or global weighting
N : number of documents
df_t : number of documents containing term t

The formula for calculating the TF-IDF weight is as in Equation 3

$$W_{t,d} = TF_{t,d} * IDF_t \quad (3)$$

where:

$W_{t,d}$: the weight value of document d relative to term t
$TF_{t,d}$: the total occurrence of term t in document d
IDF_t : the inverse document frequency or global weighting

The TF-IDF normalization formula is as in equation 4

$$W_{t,d} = \frac{w_{t,d}}{\sqrt{\sum_{t=1}^{n}(W_{t,d})^2}} \quad (4)$$

where n is the number of terms and t is the iteration of terms starting from the first.

4. Feature Selection using PSO: Feature selection using PSO aims to improve the performance of the SVM algorithm. The Particle Swarm Optimization (PSO) algorithm is a form of evolutionary computing that works by emulating the behavior of a swarm of digital particles passing through a described search area with the primary goal of finding the most satisfactory solution to a given problem (Harron et al., 2024).
5. In high-dimensional spaces, hyperplane identification will be carried out that can maximize the margin between data classes (Darmawan et al., 2022). This is illustrated in Figure 2 below:

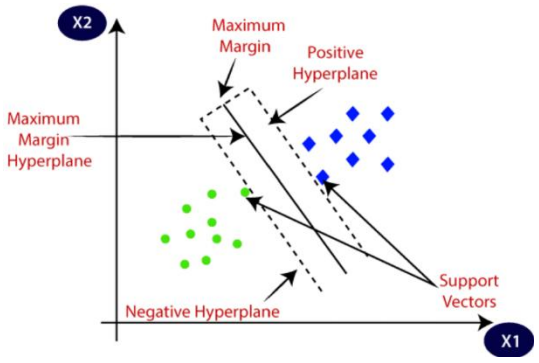

Figure 2 SVM Classification

The linear SVM classification hyperplane equation is as follows (Santosa in Darmawan et al., 2022):

$$f(x) = w^T x + b \quad (5)$$

According to Vapnik and Cortes in (Darmawan et al., 2022), the following equation is obtained:

$$[(w^T . x_i) + b] \geq 1 \text{ for } y_i = +1 \quad (6)$$
$$[(w^T . x_i) + b] \leq 1 \text{ for } y_i = -1$$

Where:
f(x) : hyperplane or class separation function
w : term weight association in the document
b : bias (W0)
x : term in the training data

Linear hyperplanes are unable to handle data-based problems, so the SVM algorithm applies kernel techniques to transform the data. The purpose of this transformation is to address the nonlinearity of the input space, thus obtaining the optimal boundary.
A kernel will be applied in SVM is polynomial kernel:

$$K(x_i, x_j) = (\gamma x_i^T x_j + c)^d, \gamma > 0 \quad (7)$$

Classification process will be using SVM, which produces a Confusion Matrix. This stage also conducts an evaluation to determine the performance of the SVM algorithm optimized using PSO and analyze the results of the Confusion Matrix evaluation between the two platforms, X and YouTube.

RESULTS AND DISCUSSION

Result

1. Data X

Previously, 1,210 data sets were collected, then split into 80% training data (968 data sets) and 20% testing data (242 data sets). The next step was to calculate the performance of the Linear Kernel SVM algorithm using a Confusion Matrix, as shown in Figure 2.

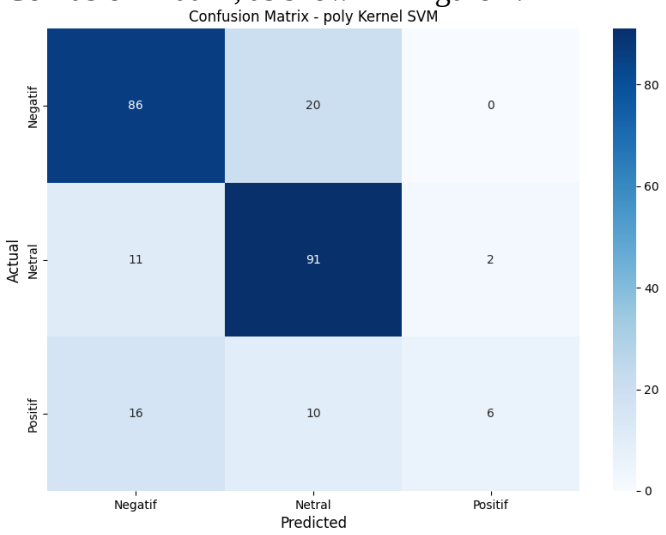

Figure 3. Confusion Matrix for Data X

The following table shows the results of the Accuracy calculation and the Recall, Precision, and F-1 Score values for each class.

TableError! No text of specified style in document.**.1 Confusion Matrix Calculation Results for Data X**

	Positive	Neutral	Negative
Accuracy		76%	
Recall	19%	87%	81%
Precision	75%	75%	76%
F-1 Score	30%	81%	79%

From the Confusion Matrix calculation results above, it can be concluded that the Accuracy of the sentiment analysis using Polynomial Kernel SVM is 76%. The results of the Recall calculation for the negative class were 81%, Precision was 76%, and F-1 Score was 79%. The results of the Recall calculation for the neutral class were 87%, Precision was 75%, and F-1 Score was 81%. In addition, the results of the Recall calculation for the positive class were 19%, Precision was 75%, and F-1 Score was 30%.

2. Data Youtube

The 749 data sets collected were then split into 80% training data (599) and 20% testing data (150). The next step was to calculate the performance of the Polynomial Kernel SVM algorithm using a Confusion Matrix.

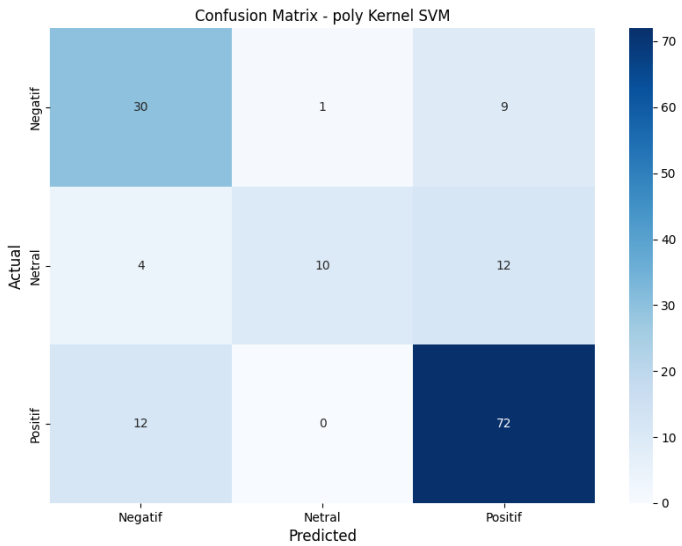

Figure 4. Confusion Matrix for Data Youtube

The following table shows the results of the Accuracy calculation and, for each class, the Recall, Precision, and F-1 Score.

Table.2 YouTube Confusion Matrix Calculation Results

	Positive	Neutral	Negative
Accuracy		75%	
Recall	86%	38%	75%
Precision	77%	91%	65%
F-1 Score	81%	54%	70%

From the Confusion Matrix calculation results above, it can be concluded that the Accuracy of the X sentiment analysis using Polynomial Kernel SVM is 75%. The results of the Recall calculation for the negative class were 75%, Precision was 65%, and F-1 Score was 70%. The results of the Recall calculation for the neutral class were 38%, Precision was 91%, and F-1 Score was 54%. In addition, the results of the Recall calculation for the positive class were 86%, Precision was 77%, and F-1 Score was 81%.

3. Comparison between X and Youtube

After the calculations outlined above, we can compare the kernels on both platforms, X and YouTube. The following is a breakdown of the Confusion Matrix calculations:

a. Accuracy

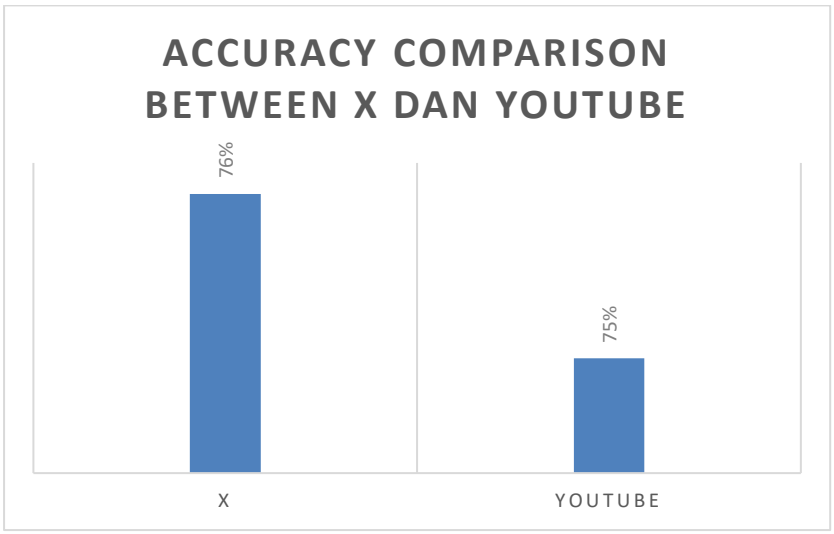

Figure 5. Accuracy comparison between X and Youtube

Based on the image above, it is clear that PSO-SVM performance on platform X is superior to that of the YouTube data. The X data yields 76% accuracy, a 1% difference compared to the YouTube data, which is at 75%.

b. Precision

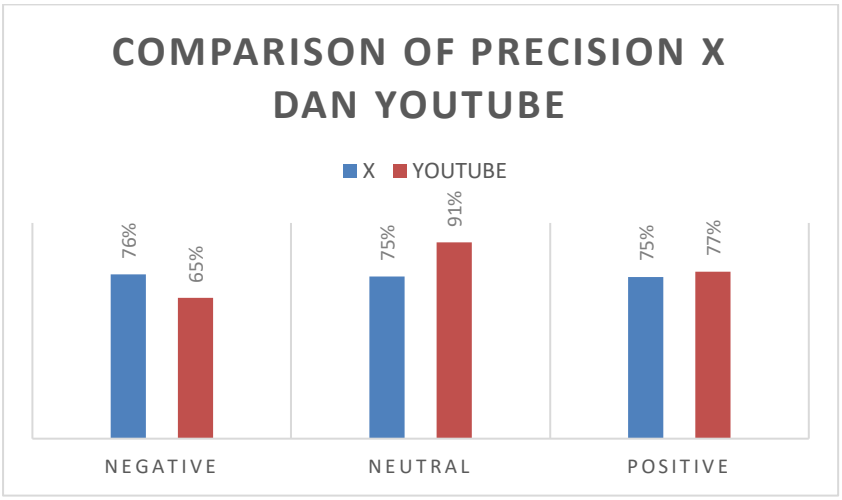

Figure 6. Comparison of precision X and Youtube

Based on Figure 5 which shows the comparison of Precision data X and Youtube, it is known that in the negative class, data X produces a higher Precision, namely 76% compared to the Youtube platform data which only gets a value of 65%. Furthermore, for the neutral class, X gets a lower result, namely 75% compared to 91% of Youtube data. Then, in the positive class X gets a lower Precision value compared to Youtube data, namely 75% compared to 77%.

c. Recall

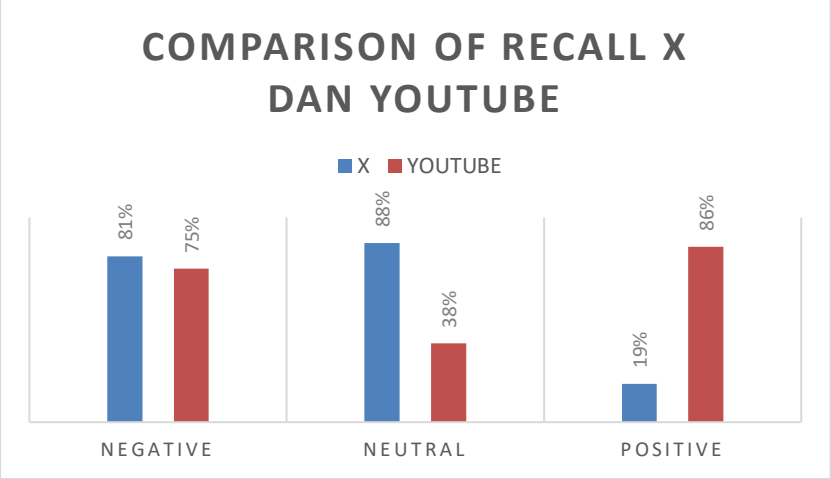

Figure 7. Comparison of Recall X and Youtube

Based on Figure 6 which shows the comparison of Recall data X and Youtube, it is known that in the negative class, data X produces a higher Recall, namely 81% compared to the Youtube platform data which gets a value of 75%. Furthermore, for the neutral class, X gets a higher result, namely 88% compared to 38% of Youtube data. Then, in the positive class Youtube gets a higher Precision value compared to data X, namely 86% compared to 19%.

d. F-1 Score

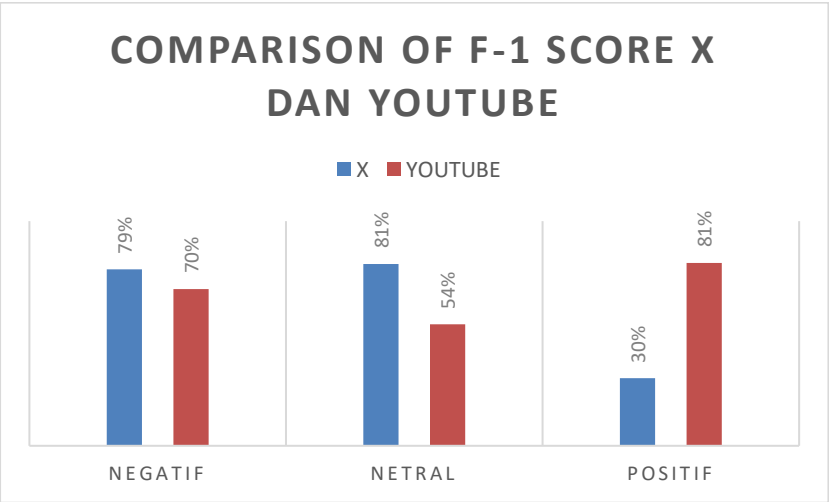

Figure 8. Comparison of F-1 Score X and Youtube

The F-1 Score is the result of the calculation between Precision and Recall, which produces a balance between the two values. Based on Figure 4.10, it is known that in the negative class, data X produces a higher F-1 Score, namely 79% compared to the YouTube platform data which gets a value of 70%. Furthermore, for the neutral class, X gets a higher result, namely 81% compared to 54% of YouTube data. Conversely, in the positive class, YouTube gets a higher Precision value compared to data X, namely 81% compared to 30%.

Discussion

Sentiment analysis of X and YouTube regarding the IKN using the PSO-SVM algorithm was conducted to determine the algorithm's performance in analyzing public sentiment regarding the IKN on platforms X and YouTube. This sentiment analysis used Google Colab as the sentiment analysis platform, while platforms X and YouTube served as data collection platforms.

The research method employed was quantitative experimental. The first step in the research process was to collect X and YouTube data uploaded between February 2024 and April 2024 using

the keyword IKN using Google Colab and tweet-harvest. The data collected from platform X amounted to 1,413 entries, and from YouTube 814 entries. After the data collection, the next stage was Text Preprocessing. In this stage, the text was examined for missing values, followed by case folding, normalization, stopword filtering, and stemming.

The next stage was data labeling using IndoBERT. In this stage, the data was automatically labeled using a pre-trained Indonesian language model called IndoBERT. The results of the IndoBERT labeling were positive, neutral, and negative. Public sentiment toward the new capital city (IKN) on platform X showed negative and neutral reactions, while on YouTube, the majority of people expressed positive reactions to the topic.

The next step after labeling the data was vectorization using TF-IDF, which converts the text into tokens, which serve to assign weights to each word in the data. The weighted data, then optimized using PSO in a Polynomial Kernel SVM (SVM), was classified using the SVM algorithm with an 80:20 ratio of training data to testing data, resulting in a Confusion Matrix. The Confusion Matrix was then evaluated, resulting in an accuracy of 76% for platform X and 75% for YouTube data. Furthermore, the precision for the negative class of X data was 76%, neutral 75%, and positive 75%. Meanwhile, the negative class of YouTube data was 65%, neutral 91%, and positive 77%.

The recall for data X was 81% for the negative class, 88% for the neutral class, and 19% for the positive class. The YouTube data showed 75% for the negative class, 38% for the neutral class, and 86% for the positive class. The F-1 score for data X was 79% for the negative class, 81% for the neutral class, and 30% for the positive class. The YouTube data showed 70% for the negative class, 54% for the neutral class, and 81% for the positive class.

Based on previous research, Darmawan et al. (2022) in their study entitled Optimization of Support Vector Machine (SVM) Based on Particle Swarm Optimization (PSO) in Sentiment Analysis of Ruang Guru's Official Account on Twitter, the results showed that the SVM algorithm optimized with PSO achieved 89.20%. However, unlike the previous study, this study used three classes. Research conducted by Du & Le (2024) suggests that binary classification (2 classes) consistently outperforms ternary classification (3 classes). In other words, the performance of the Polynomial Kernel SVM algorithm using PSO has achieved good results.

CONCLUSION

This study found that the PSO-optimized Polynomial Kernel SVM algorithm achieved higher accuracy on Platform X data (76%) compared to YouTube (75%), with negative and neutral sentiment dominating on X and positive sentiment on YouTube. Performance metrics such as precision, recall, and F1-score indicate X's superiority in the negative and neutral classes, although the positive class is weak on both platforms, reflecting the differences in the characteristics of short text data on X versus long comments on YouTube.

Nevertheless, the study's limitations include the use of a limited dataset (1,413 X tweets and 814 YouTube comments from February–April 2024), reliance on IndoBERT for automatic labeling—which is susceptible to multilingual bias in Indonesian—and accuracy below 80% due to the complexity of the three sentiment classes. Suggestions for further research include testing other SVM kernels, integrating deep learning models such as fine-tuned IndoBERT, and utilizing larger real-time datasets. Practically, these results have implications for IKN policymakers to strengthen communication on YouTube to boost positive sentiment and monitor critical opinions on X for a swift response.

REFERENCES

Ajay Jadhav, Pranjal Jagtap, Suraj Gurav, Shivani Jadhav, Nikita Jadhav, & Afsha Akkalkot. 2023. A survey on text mining - techniques, application. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *3307*, 338–343. https://doi.org/10.32628/cseit2390391

Aslan, S., Kızıloluk, S., & Sert, E. 2023. TSA-CNN-AOA: twitter sentiment analysis using cnn optimized via arithmetic optimization algorithm. *Neural Computing and Applications*, *35*(14), 10311–10328. https://doi.org/10.1007/s00521-023-08236-2

Broklyn, P., Olukemi, A., & Bell, C. 2024. *S*ocial media sentiment analysis for brand reputation management social media sentiment analysis for brand reputation management. https://doi.org/10.20944/preprints202408.0029.v1

Chakraborty, S. 2023. Sentiment analysis in the perspective of natural language processing. *International Journal for Research in Applied Science and Engineering Technology*, *11*(11), 2235–2241. https://doi.org/10.22214/ijraset.2023.56925

Darmawan, R., Indra, I., & Surahmat, A. 2022. Optimalisasi support vector machine (svm) berbasis particle swarm optimization (PSO) pada analisis sentimen terhadap official account ruang guru di twitter. *Jurnal Kajian Ilmiah*, *22*(2), 143–152. https://doi.org/10.31599/jki.v22i2.1130

Fauzi, A., & Yunial, A. H. 2024. Analisis sentimen US airline pada media sosial menggunakan perbandingan algoritma data mining. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *10*(2), 277. https://doi.org/10.26418/jp.v10i2.76024

Frissen, I., & Evans, M. M. 2024. Experimental research in knowledge management. *Knowledge and Process Management*, *31*(1), 60–68. https://doi.org/10.1002/kpm.1764

Gudankwar, A., Mendhe,P. M., Oghare, L. N., & Yemde, A. R. 2024. Sentiments of public opinion. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *10*(2), 459–461. https://doi.org/10.32628/cseit2410239

Harron, S., Saxena, V., & Kumari, N. 2024. Exploring the use of particle swarm optimization algorithms to enhance evolutionary computing. *2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC)*, 1–6. https://doi.org/10.1109/ICOCWC60930.2024.10470721

Hutapea, P. S., & Maharani, W. 2023. Sentiment analysis on twitter social media towards shopee e-commerce through support vector machine (SVM) method. *JINAV: Journal of Information and Visualization*, *4*(1). https://jinav.org/index.php/jinav/article/view/1504

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

Jain, S., Jain, S. K., & Vasal, S. 2024. An effective TF-IDF model to improve the text classification performance. *2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*, 1–4. https://doi.org/10.1109/csnt60213.2024.10545818

Jamil, M., Hadiyanto, H., & Sanjaya, R. 2024. Sentiment analysis: classifying public comments on *Youtube* in disaster management simulation in Indonesia using naïve bayes and support vector machine. *Ingenierie Des Systemes d'Information*, *29*(2), 437–446. https://doi.org/10.18280/isi.290205

Karunanithi, M. 2023. An improved particle swarm optimization algorithm. *2023 IEEE 64th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, 1–6. https://doi.org/10.1109/ITMS59786.2023.10317698

Novianti, D. N., Shiddieq, D. F., Roji, F. F., & Susilawati, W. 2024. Comparison of support vector machine and naïve bayes algorithms for sentiment analysis of the metaverse. *MALCOM:*

Indonesian Journal of Machine Learning and Computer Science, *4*(April), 231–239.

Priyatno, A. M., & Ningsih, L. 2022. TF-IDF weighting to detect spammer accounts on twitter based on tweets and retweet representation of tweets. *Sistemasi*, *11*(3), 614. https://doi.org/10.32520/stmsi.v11i3.1995

Puspitaningrum, S. R., & Sundoko, H. F. 2023. Pemindahan ibu kota negara : pembangunan kota inklusif dan berkelanjutan. *Jurnal Sosial Politik RESOLUSI*, *6*(2), 127–147. https://doi.org/10.32699/resolusi.v6i2.6096

Putri, D. I., Alfian, A. N., Putra, M. Y., & Mulyo, P. D. 2024. IndoBERT model analysis: twitter sentiments on indonesia's 2024 presidential election. *Journal of Applied Informatics and Computing*, *8*(1), 7–12. https://doi.org/10.30871/jaic.v8i1.7440

Rusydi Umar, Sunardi, & Nuriyah, M. N. A. 2023. Comparing the performance of data mining algorithms in predicting sentiments on twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(4), 817–823. https://doi.org/10.29207/resti.v7i4.4931

Sari, H., Leonarde Ginting, G., & Zebua, T. 2021. Penerapan algoritma text mining dan TF-IDF untuk pengelompokan topik skripsi. *Terapan Informatika Nusantara*, *2*(7), 414–432. https://ejurnal.seminar-id.com/index.php/tin

Setiawan, Y., Gunawan, D., & Efendi, R. 2022. Feature extraction TF-IDF to perform cyberbullying text classification: a literature review and future research direction. *2022 International Conference on Information Technology Systems and Innovation, ICITSI 2022 - Proceedings*, 283–288. https://doi.org/10.1109/ICITSI56531.2022.9970942

Silva, J. G. C. 2022. Experimental research. *SpringerBriefs in Applied Sciences and Technology*, *16*(November), 25–46. https://doi.org/10.1007/978-3-030-92130-9_4

Tan, J. J., Firdaus, A., & Aksar, I. A. 2024. Social media for political information: a systematic literature review. *Jurnal Komunikasi: Malaysian Journal of Communication*, *40*(1), 77–98. https://doi.org/10.17576/JKMJC-2024-4001-05

Utama, H., & Masruro, A. 2022. Analisis sentimen pada twitter menggunakan word embedding dengan pendekatan word2vec. *Jurnal Sistem Cerdas*, *5*(2), 128–134. https://doi.org/10.37396/jsc.v5i2.242

Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J. H., & Hsieh, J. G. 2021. Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes. *Information (Switzerland)*, *12*(5). https://doi.org/10.3390/info12050204

Wang, C. K. 2023. *S*entiment analysis using support vector machines, neural networks, and random forests. *Atlantis Press International BV*. https://doi.org/10.2991/978-94-6463-300-9_4

Xiang, L. 2022. Application of an improved tf-idf method in literary text classification. *Advances in Multimedia*, *2022*. https://doi.org/10.1155/2022/9285324

Yulfa, R. I., Setiawan, B. H., Lourensius, G. G., & Purwandari, K. 2023. Enhancing hate speech detection in social media using indoBERT model: a study of sentiment analysis during the 2024 indonesia presidential election. *ICCA 2023 - 2023 5th International Conference on Computer and Applications, Proceedings*, *7*, 1–6. https://doi.org/10.1109/ICCA59364.2023.10401700