Jurnal CoreIT, Vol.11, No.1, June 2025
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Sentiment Analysis on Reviews of the Documentary Film "Dirty Vote" Using Lexicon-Based and Support Vector Machine Approaches

Apri Ramadhan[1], Suhendro Yusuf Irianto[2*]
[1,2]Faculty of Computer Science, Institute of Informatics and Business Darmajaya, Lampung, Indonesia
[1]apriramadh.2121210030@mail.darmajaya.ac.id, [2]suhendro@darmajaya.ac.id

Abstract. The general election is a state agenda in Indonesia held every five years. On February 11, 2024, during the silence period, a video titled "Dirty Vote" was uploaded on YouTube, drawing significant public attention. Its release during the silence period sparked controversy and prompted various opinions in the video's comment section. Sentiment analysis is a suitable method to determine whether public opinions regarding the video are predominantly positive, negative, or neutral. This study utilized the Support Vector Machine (SVM) classification method with different kernels, including linear and non-linear (polynomial, RBF, and sigmoid). Support Vector Machine (SVM) was chosen in this study because it has a high accuracy value. It also requires labels and training data. To accelerate labeling for large datasets for example 10,000 – 60,000 data such as in this study, a Lexicon-Based approach was employed. The SVM approach can contribute to lexicon-based, and lexicon-based can help label datasets on SVM to produce good accuracy. The combination of SVM and Lexicon-Based methods demonstrated that the linear kernel outperformed others, achieving evaluation metrics of 91.1% accuracy, 91.1% recall, 90.9% precision, and 90.8% F1-score. Based on these values, the linear kernel model demonstrates good performance in classifying sentiment in textual data, such as comments or reviews. This model can be used to determine whether a comment or review is positive, negative, or neutral.

Keywords: Sentiment Analysis, Dirty Vote, Lexicon Based, Pemilu, Support Vector Machine, Youtube

Received December 2024 / **Revised** January 2025 / **Accepted** February 2025

INTRODUCTION

The general elections are a key instrument of democracy and a vital pillar of democratic systems, allowing citizens to freely and fairly choose their leaders [1]. Elections in Indonesia follow the principles of direct, general, free, confidential, honest, and fair, as stated in Article 22E(6) of the 1945 Constitution and regulated by Law No. 7 of 2017 on General Elections.

Law No. 7 of 2017 outlines the "election silence period" as part of the election stages, regulated under Article 167(4). For the 2024 election, this period lasts three days, from February 11–13, aiming to give voters time to reflect objectively and calmly, free from external pressure [2]. During the election silence period, Article 287(5) of the Election Law prohibits print media, online media, social media, and broadcasting institutions from publishing news, advertisements, candidate profiles, or any content that benefits or harms election participants.

On February 11, 2024, during the election silence period, a YouTube video titled "Dirty Vote" was released. Nearly two hours long, it features constitutional law experts Feri Amsari (legal activist and academic), Zainal Arifin Mochtar (constitutional law lecturer), and Bivitri Susanti (academic and legal expert). The film exposes power instruments allegedly used to win elections and undermine democracy, showcasing blatant misuse of power to maintain the status quo. Fraud tactics are analyzed through their expertise in constitutional law. The film quickly became a topic of public debate, with supporters of certain presidential candidates accusing it of being a black campaign deliberately released during the silence period [3]. A black campaign refers to a form of campaigning aimed at dividing political parties, inciting hatred, and defaming individuals or groups, either personally or collectively [4]. According to Ninik Rahayu, Chairperson of the Indonesian Press Council, "Dirty Vote" is not fiction or fake news, but a documentation of presidential election events, court facts, and academic analysis [5]. Dr. Suko Widodo, a communication lecturer at Universitas Airlangga and political communication expert, believes that "Dirty Vote" falls into two categories: educational and critical. It educates viewers on how political contests are ultimately about power struggles, while also serving as a critique of the government. However, the truth of the narrative in the film should be confirmed through political dialogue [6].

DOI: 10.24014/coreit.v11i1.34603

In the comment section on YouTube about the film, many people expressed their opinions. These comments reflect individuals' feelings, judgments, and responses to the content of the documentary. The sentiment of these comments will be measured using sentiment analysis strategies. The development of sentiment analysis has been rapid due to its benefits. Around 20 to 30 businesses in the United States focus on providing sentiment analysis services [7]. Sentiment analysis, also known as opinion mining, is a field of study that analyzes reviews or opinions, sentiments, evaluations, judgments, attitudes, and emotions of individuals toward objects such as products, services, organizations, individuals, issues, events, topics, and their attributes [8]. Sentiment analysis categorizes opinions into positive, negative, and neutral, with neutral representing a sentiment that is neither positive nor negative.

Research on sentiment analysis of YouTube comments has been previously conducted by Afdhal et al [9]. The data collection process was conducted in two stages: crawling and labeling the comments on videos containing Islamophobic content to generate the dataset. A similar approach was taken by Soemedhy et al [10] in their research, which involved data collection with a case study on sexual violence. The data was sourced from the YouTube channel Narasi, specifically from a video titled "Bunuh Diri NW: Bripda Randy Tersangka, Penanganan Polisi Dikritik | Narasi Newsroom". Jonathan et al [11] also collected data from YouTube comment sections, focusing on the Flat Earth Theory, using Python as a tool for data scraping. Based on these three research mentioned that successfully conducted collecting the data through crawling on YouTube, this study will also collect data by crawling comments from the documentary film Dirty Vote on YouTube.

Sentiment analysis can leverage machine learning-based classification algorithms such as K-Nearest Neighbor (KNN), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Naive Bayes (NB), among others, for effective results [12]. Machine learning-based approaches require a dataset and the development of a model to train the classifier, referred to as the training data. During the classification process, this training data is tested using testing data to evaluate the model's performance [13].

Research conducted by Ditami et al [14] utilized data from Twitter to analyze public sentiment toward promotional event programs in marketplaces. By employing the SVM classification algorithm combined with grid search for optimal parameter tuning, the study achieved an accuracy improvement of 0.54% to 1.44%. Similarly, Fadila et al [15] applied the SVM algorithm alongside the AdaBoost boosting algorithm to enhance sentiment analysis accuracy, achieving an impressive accuracy rate of 99.31%. Another study by Salsabila et al [16] compared classification algorithms for sentiment analysis of restaurant reviews in Malang, using both SVM and Naive Bayes (NB) algorithms. The results indicated that SVM outperformed NB, with an accuracy of 92.74% compared to 91.67%. Additionally, Oktaviana et al [17] employed Lexicon-Based Features for feature extraction based on word semantics and used the SVM algorithm to classify sentiments (positive or negative) regarding online learning policies during the COVID-19 pandemic. The evaluation results showed a 12% improvement in accuracy after incorporating Lexicon-Based Features. The previous studies mentioned shows that SVM performs better than other method, even with parameter tuning, SVM stills performs better and maintain high accuracy.

Based on these conditions, this study applies Support Vector Machine (SVM) with Lexicon-Based Features to conduct sentiment analysis on reviews of the documentary film "Dirty Vote". The purpose of this research is to classify opinions into categories of positive, negative, or neutral based on their sentiment tendencies. The classification of three labels, positive, negative, and neutral is implemented to gain better understanding. Moreover, the large dataset used in this study give more variation of neutral possibility expressions.

METHODS

In the research methodology section, the structured steps for conducting sentiment analysis on the YouTube comments of the documentary film "Dirty Vote" will be outlined. The steps include labeling using Lexicon-Based Features, word weighting with TF-IDF, and the application of the Support Vector Machine (SVM) classification model. The research framework is illustrated in the research flowchart shown in Figure 1.

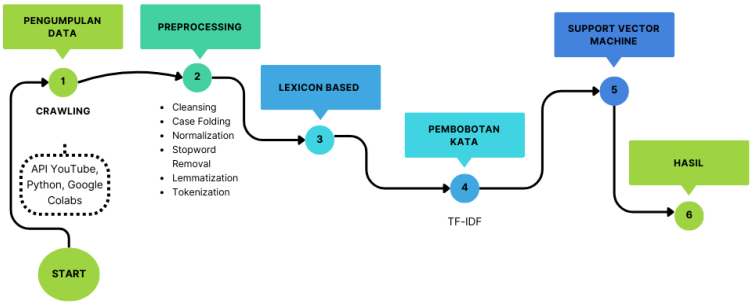

Figure 1. Research Stage

Dataset

The dataset used in this study consists of data crawled from the YouTube comment section, totalling 63,146 entries, collected from the documentary film or video titled "Dirty Vote". This data was then stored in an Excel file with the .xlsx format. Table 1 below presents a sample of the YouTube comments used in this study.

Table 1. Youtube Comment Dataset Film Dirty Vote

Comment
mantap...
Di tunggu part II nya pasca pemilu ,, rekap suara, deklarasi awal, pengangkatan menteri, hak angket, waaaah...banyak deh.. ayo bang observasi dulu.. biar JD dokumen untuk anak2 cucu kita nanti
😊😊😊cemen kritikan gue dihapus wkwkwk cemennnnn

Preprocessing

The collected data will be pre-processed, converting raw, unstructured data with noise into clean, structured data, ready for analysis [18]. After that, the data will be ready for processing. The following are the steps involved in pre-processing.

1) Cleansing: This stage involves removing links, emoticons, hashtags, tagging, punctuation, newlines, and numbers from the data [19] because these components are considered irrelevant to the classification process.
2) Case Folding: It is a pre-processing step that involves cleaning the data by removing or converting uppercase letters to lowercase. This results in the text being uniformly in lowercase, as shown in Table 2.

Table 2. Case Folding Sample

Comment After Case Folding
mantap
i tunggu part ii nya pasca pemilu rekap suara deklarasi awal pengangkatan menteri hak angket waah banyak deh ayo bang observasi dulu biar j dokumen untuk anak cucu kita nanti
cemen kritikan gue dihapus cemenn

3) Normalization: The process of restoring a word to its correct form [20]. Normalizing words to their standard form by converting abbreviations or non-standard expressions into their proper equivalents, following a predefined corpus.
4) Stopword Removal: This step involves removing or minimizing words that do not carry significant meaning in the document, based on a word list from the Sastrawi library. The results of this stage are shown in Table 3.

Table 3. Stopword Removal Sample

Comment After Stopword Removal
mantap
i tunggu part ii nya pasca pemilu rekap suara deklarasi awal pengangkatan menteri hak angket waah banyak deh ayo bang observasi biar j dokumen anak cucu nanti
cemen kritikan dihapus cemenn

5) Lemmatization: This step involves transforming words into their valid forms by reducing variability through the removal of inflectional suffixes, creating words according to the dictionary [21]. In other words, this step converts words into their root forms as per the dictionary without altering their meaning. The results of the lemmatization process are shown in Table 4.

Table 4. Lemmatization Sample

Comment After Lemmatization
mantap
i tunggu part ii nya pasca milu rekap suara deklarasi awal angkat menteri hak angket waah banyak deh ayo bang observasi biar j dokumen anak cucu nanti
cemen kritik hapus cemenn

6) Tokenization: This step processes the word structure by splitting or separating it into individual word tokens [22]. The sentences in the dataset are then converted into word-by-word tokens, as shown in the example in Table 5.

Table 5. Tokenization Sample

Comment After Tokenization
['mantap']
['i', 'tunggu', 'part', 'ii', 'nya', 'pasca', 'milu', 'rekap', 'suara', 'deklarasi', 'awal', 'angkat', 'menteri', 'hak', 'angket', 'waah', 'banyak', 'deh', 'ayo', 'bang', 'observasi', 'biar', 'j', 'dokumen', 'anak', 'cucu', 'nanti']
['cemen', 'kritik', 'hapus', 'cemenn']

Sentiment Labelling

After completing the pre-processing stage, the next step is labelling. Labelling is necessary to transfer the learning process to the SVM. Given the large volume of data, a lexicon-based method is used in the labelling process to save time [18]. The lexicon-based labeling method requires a dictionary as a reference to calculate the polarity of opinions or sentiments. In this study, the Inset (Indonesian Sentiment Lexicon) dictionary is used as the reference. Each word in the review is assigned a weight based on the positive and negative sentiment words found in the dictionary [23].

$$Score = (total\ positive\ words) - (total\ negative\ words) \tag{1}$$

If the total number of positive words in a review exceeds the number of negative words, it is labelled as positive sentiment. If the total number of positive words is fewer than the negative words, the review is labelled as negative sentiment. If the total number of positive and negative words is equal, the review is labelled as neutral sentiment.

$$Sentence_{sentiment} = \begin{Bmatrix} positive\ if\ S_{positive} > S_{negative} \\ netral\ if\ S_{positive} = S_{negative} \\ negative\ if\ S_{positive} < S_{negative} \end{Bmatrix} \tag{2}$$

Word Weighting

After the labeling process, the next step is word weighting, or term weighting, using TF-IDF across one or more documents [24]. TF-IDF is a weighting method that combines term frequency and inverse document frequency. Term frequency refers to the total occurrence of a term within a document, while inverse document frequency measures the importance of a word within a document [25]. Because machines can only process numbers, so each word in the document is assigned a weight or frequency value. The TF-IDF formula is shown below:

$$Tf = \begin{cases} 1 + log_{10}(f_{t,d}),\ f_{t,d} > 0 \\ 0, f_{t,d} = 0 \end{cases} \tag{3}$$

$$IDF = log_{10} \frac{N}{DF_t} \tag{4}$$

$$W_{t,d} = TF\ .\ IDF \tag{5}$$

Tf = word weight in each document
ft,d = number of times a term appears in a document
IDF = inverse weight in the document frequency
DF = number of documents containing the term
Wt,d = TF – IDF weight

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a widely used algorithm in machine learning for classification and is classified as supervised learning. SVM performs well with large datasets and yields high-quality text classification results [13]. The core concept of this algorithm is to find the most optimal hyperplane to separate the two classes [25]. The illustration of the SVM algorithm is shown in Figure 2.

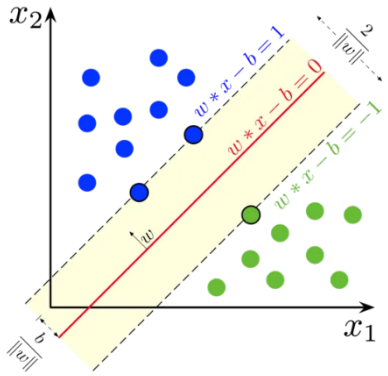

Figure 2. Ilustration of The SVM Hyperplane [26]

The hyperplane is obtained using the following equation (6).

$$(w \,.\, x_i) + b = 0 \tag{6}$$

For the data x_i belonging to class -1, it can be formulated as equation (7).

$$(w \,.\, x_i + b) \leq 1, y_i = -1 \tag{7}$$

and for the data x_i belonging to class +1, it can be formulated as equation (8).

$$(w \,.\, x_i + b) \geq 1, \text{yi} = 1 \tag{8}$$

The classification step in the SVM algorithm may encounter a situation where the linear kernel does not perform optimally, resulting in poor classification results. To address this, one approach is to use a non-linear kernel by utilizing the kernel trick [25]. The kernel functions in SVM include linear, polynomial, RBF, and sigmoid [27]. The formulas or equations for these kernels are presented in Table 6 below.

Table 6. The Equations for Each SVM Kernel Are as Follows

Kernel	Equation
Linear	$K(x_i, x) = x_i^T x$
Polynomial	$K(x_i, x) = (\gamma . x_i^T x + r)^p, \gamma > 0$
RBF	$K(x_i, x) = exp(-\gamma \lvert x_i - x \rvert^2), \gamma > 0$
Sigmoid	$K(x_i, x) = tanh(\gamma x_i^T x + r)$

Confusion Matrix

The final step after completing all processes is to evaluate the method used. The instrument for evaluating the employed method is the confusion matrix. The confusion matrix generally has a 2x2 structure for binary classification, as shown in Table 7 below.

Table 7. Confusion Matrix 2x2

Actual Data	Data Prediction	
	Positive	**Negative**
Positive	TP (*True Positive*)	FP (*False Positive*)
Negative	FN (*False Negative*)	TN (*True Negative*)

In non-binary classification, the structure is adjusted based on the number of labels or classes used. In this study, a classification with three labels or classes—positive, negative, and neutral—is applied. Therefore, the structure of the confusion matrix used is 3x3, as shown in Table 8.

Table 8. Confusion Matrix 3x3

Actual Data	Data Prediction		
	Negative	**Neutral**	**Positive**
Negative	TN (*True Negative*)	FN (*False Netral*)	FP (*False Positive*)
Neutral	FN (*False Negative*)	TN (*True Netral*)	FP (*False Positive*)
Positive	FN (*False Negative*)	FN (*False Netral*)	TP (*True Positive*)

The confusion matrix is necessary to assess the performance of the classification method, including accuracy, recall, precision, and F1-score. The formulas for calculating these performance metrics are as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

$$F1\text{-}Score = 2 \text{ x } \frac{Precision\ x\ Recall}{Precision + Recall} \tag{12}$$

Explanation:

- TP (True Positive): Indicates the instances where positive results are correctly detected.
- TN (True Negative): Indicates the instances where negative results are correctly detected.
- FP (False Positive): Indicates the instances where negative results are incorrectly detected as positive.
- FN (False Negative): Indicates the instances where positive results are incorrectly detected as negative.

Testing Scenario

This study applies a 20:80 testing ratio, meaning that 20% of the data is used for testing while 80% is allocated for training. The 20:80 ratio was chosen because previous studies [18] [20] [27] have demonstrated that it yields the highest accuracy. Additionally, K-Fold Cross Validation is not utilized in this study, as prior research [28] indicates that SVM achieves superior accuracy without it. Grid search for parameter tuning is also not applied. Given the sufficiently large dataset, this study focuses on evaluating the accuracy of SVM across different kernels without the need for parameter tuning.

RESULT AND DISCUSSION

In this study, a dataset consisting of 63,146 review comments was collected from the comment section of the YouTube channel @DirtyVote using the crawling method. After data filtering, a final dataset of 59,480 review comments was obtained, ready for further processing, from pre-processing to the evaluation stage. The evaluation results are presented in terms of accuracy and F1-score. The F1-score is an evaluation metric used to measure the balance or stability between precision and recall. It represents the harmonic mean of precision and recall. The F1-score describes how well the model classifies labels accurately. Therefore, the evaluation method used in this study is the F1-score. A higher F1-score indicates a better-performing model.

Test Results and Analysis of Scenario One
The testing in the first scenario uses a lexicon-based approach and the SVM algorithm with a linear kernel. The confusion matrix for the SVM linear kernel is shown in Table 9 below.

Table 9. Confusion Matrix SVM Linear Kernel

Actual Data	Data Prediction		
	Negative	Neutral	Positive
Negative	7159	103	117
Neutral	348	933	161
Positive	228	92	2755

The results from the confusion matrix in the table above are used to assess the model's performance. The obtained results can be seen in Table 10.

Table 10. Results of Testing Scenario One

Kernel	*Accuracy*	*Recall*	*Precision*	*F1-score*
Linear	**91.1%**	**91.1%**	**90.9%**	**90.8%**

Table 10 shows the evaluation results based on the implementation of the F1-score calculation. A higher F1-score indicates a better-performing model. The values in Table X were obtained with an 80:20 split for training and testing data, meaning 20% of the data was used for testing and 80% for training, with a random state set to 42. The table demonstrates that the F1-score achieved is 90.8%.

Test Results and Analysis of Scenario Two
The testing in the second scenario also uses the lexicon-based approach. In this scenario, the kernel used in the SVM algorithm is the polynomial kernel. The confusion matrix for the SVM polynomial kernel is shown in Table 11 below.

Table 11. Confusion Matrix for SVM Polynomial Kernel

Actual Data	Data Prediction		
	Negative	Neutral	Positive
Negative	7255	35	89
Neutral	916	428	98
Positive	1561	28	1486

The results from the confusion matrix in the table above are used to assess the model's performance. The obtained results can be seen in Table 12.

Table 12. Result of Testing Scenario Two

Kernel	*Accuracy*	*Recall*	*Precision*	*F1-score*
Polynomial	**77%**	**77%**	**79.7%**	**74.1%**

Table 12 displays the evaluation results from the testing, calculated using the F1-score. The testing data and training data composition remains the same as in the first scenario. The table shows that the F1-score achieved is 74.1%.

Test Results and Analysis of Scenario Three
The testing in the third scenario also uses the lexicon-based approach. In this scenario, the Radial Basis Function (RBF) kernel is used in the SVM algorithm. The confusion matrix for the SVM Radial Basis Function (RBF) is shown in Table 13 below.

Table 13. Confusion Matrix SVM Kernel Radial Basis Function (RBF)

Actual Data	Data Prediction		
	Negative	Neutral	Positive
Negative	7187	69	123
Neutral	450	841	151
Positive	372	76	2627

The results from the confusion matrix in the table above are used to assess the model's performance. The obtained results can be seen in Table 14.

Table 14. Result of Testing Scenario Three

Kernel	*Accuracy*	*Recall*	*Precision*	*F1-score*
RBF	**89.5%**	**89.5%**	**89.4%**	**89%**

Table 14 displays the evaluation results from the testing, calculated using the F1-score. The data testing and training composition remains the same as in the first and second scenarios. The table shows that the F1-score achieved is 89%.

Test Results and Analysis of Scenario Four

The testing in the fourth scenario also uses the lexicon-based approach. In this scenario, the Sigmoid kernel is used in the SVM algorithm. The confusion matrix for the SVM Sigmoid kernel is shown in Table 15 below.

Table 15. Confusion Matrix SVM Kernel Sigmoid

Actual Data	Data Prediction		
	Negative	Neutral	Positive
Negative	7033	179	167
Neutral	476	767	199
Positive	303	138	2634

The results from the confusion matrix in the table above are used to assess the model's performance. The obtained results can be seen in Table 16.

Table 16. Result of Testing Scenario Four

Kernel	*Accuracy*	*Recall*	*Precision*	*F1-score*
Sigmoid	**87.7%**	**87.7%**	**87.1%**	**87.2%**

Table 16 displays the evaluation results from the testing, calculated using the F1-score. The data testing and training composition remains the same as in the first, second, and third scenarios. The table shows that the F1-score achieved is 87.2%.

Comparison of Testing Results for Each SVM Kernel

The testing for each SVM kernel, using the lexicon-based approach with a 20:80 split (20% for testing and 80% for training) and a random state of 42, resulted in different F1-scores for each kernel. Table 17 below shows the performance results for each kernel.

Table 17. Result of Testing Comparison for Each SVM Kernel

Kernel	Evaluation			
	Accuracy	*Recall*	*Precision*	*F1-score*
Linear	**91.1%**	**91.1%**	**90.9%**	**90.8%**
Polynomial	**77%**	**77%**	**79.7%**	**74.1%**
RBF	**89.5%**	**89.5%**	**89.4%**	**89%**
Sigmoid	**87.7%**	**87.7%**	**87.1%**	**87.2%**

In Table 17, it can be seen that the linear kernel outperforms in all evaluation metrics. In this study, the linear kernel provided the best and most optimal results for the case or dataset used. In this case, the linear kernel performed better than the non-linear kernels.

In the study conducted by Saputra et al [27] on the performance comparison of SVM algorithm kernels in rice disease classification, it was found that the linear kernel with a 20% testing data and 80% training data split resulted in the highest accuracy of 89%. The data testing and training composition used in that study is the same as in the "Dirty Vote" case study, and the results are consistent, with the linear kernel achieving the highest accuracy. However, in that study, data embedding was performed because the dataset used consisted of image data, which was then numerically represented.

Another research conducted by Muhammadi et al [18], they researched on the combination of SVM and lexicon-based approaches for sentiment analysis on Twitter, it was found that the RBF kernel produced the highest accuracy of 81.73%. This study also used a 20% testing data and 80% training data composition. Additionally, the study in [18] employed hyperparameter tuning for each kernel.

CONCLUSION

This study used a dataset of 59,480 reviews about the documentary film titled "Dirty Vote." After undergoing the preprocessing process, labeling was performed using a lexicon-based approach with the InSet dictionary. The labeling results yielded 36,701 negative, 15,462 positive, and 7,317 neutral data points. The combination of lexicon-based methods and SVM was performed sequentially, with lexicon-based used to determine sentiment values, and the lexicon results serving as labeled data for SVM. Four kernels were tested to find the best kernel, comparing linear and non-linear options (polynomial, RBF, and sigmoid). The results, with a 20% testing data and 80% training data split, showed that the linear kernel outperformed the others in every evaluation metric, achieving an accuracy of 91.1%, recall of 91.1%, precision of 90.9%, and an F1-score of 90.8%. The results of this study differ from the findings in [18] which could be attributed to factors such as the features used, parameter settings, and the version of the lexicon applied. Based on these values, the linear kernel model demonstrates good performance in classifying sentiment in textual data, such as comments or reviews. This model can be used to determine whether a comment or review is positive, negative, or neutral.

Several recommendations for future research include: (1) Using the latest version of lexicons or more comprehensive dictionaries to capture a wider range of words; (2) Exploring different data testing and training split ratios; (3) Implementing hyperparameter tuning for each kernel to identify the best-performing SVM model.

REFERENCES

[1] Z. Amatahir, “Peran Mahasiswa Dalam Mencegah Politik Uang Dan Kecurangan Pemilu,” *Jurnal Media Hukum (JMH)*, vol. 11, no. 2, pp. 87–98, Sep. 2023, doi: 10.59414/jmh.v11i2.577.

[2] F. N. Heriani, “Mengenal Masa Tenang Pemilu Beserta Aturan dan Larangannya,” Hukumonline. Accessed: Mar. 29, 2024. [Online]. Available: https://www.hukumonline.com/berita/a/mengenal-masa-tenang-pemilu-beserta-aturan-dan-larangannya-lt65ca970c207b5/

[3] CNBC Indonesia, “Viral Dirty Vote, Begini Komentar Positif-Negatif Netizen di Medsos,” CNBC Indonesia. Accessed: Mar. 31, 2024. [Online]. Available: https://www.cnbcindonesia.com/tech/20240213102942-37-513711/viral-dirty-vote-begini-komentar-positif-negatif-netizen-di-medsos

[4] A. Thanzani, A. D. P. Sari, L. T. Yulia, and S. Fikri, “Black Campaign Melalui Media Elektronik Dari Perspektif Hukum Pemilu,” *Journal Evidence Of Law*, vol. 1, no. 3, pp. 42–51, Sep. 2022, doi: https://doi.org/10.59066/jel.v1i3.103.

[5] BBC News Indonesia, “Dirty Vote: Film ‘tentang kecurangan pilpres’ tuai pro-kontra, bagaimana publik harus menyikapinya?,” BBC News Indonesia. Accessed: Mar. 31, 2024. [Online]. Available: https://www.bbc.com/indonesia/articles/c72g1x45gj4o

[6] A. A. Pratiwi, “Pakar Komunikasi Politik Respon Perilisan Film ‘Dirty Vote’ Jelang Pemilu 2024,” UNAIR NEWS. Accessed: Mar. 31, 2924. [Online]. Available: https://unair.ac.id/pakar-komunikasi-politik-respon-perilisan-film-dirty-vote/

[7] R. A. W. S. Rizky, R. A. Rajagede, and R. Rahmadi, “Analisis Sentimen Politik Berdasarkan Big Data dari Media Sosial Youtube: Sebuah Tinjauan Literatur,” *Automata*, vol. 2, no. 1, 2021.

[8] B. Liu, *Sentiment Analysis and Opinion Mining*. in Synthesis Digital Library of Engineering and Computer Science. Morgan & Claypool, 2012.
[9] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 5, no. 1, pp. 122–130, Feb. 2022, doi: https://doi.org/10.32672/jnkti.v5i1.4004.
[10] C. A. A. Soemedhy, N. Trivetisia, N. A. Winanti, D. P. Martiyaningsih, T. W. Utami, and S. Sudianto, "Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube 'Kekerasan Seksual')," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 7, no. 2, pp. 80–84, May 2022, doi: https://doi.org/10.30591/jpit.v7i2.3547.
[11] C. Jonathan, T. H. Rochadiani, and T. Sofian, "Analisis Sentimen Komentar Video Youtube Flat Earth Theory Dengan Menggunakan Metode Unsupervised Dan Supervised Learning," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 378–387, Sep. 2023, doi: 10.51454/decode.v3i2.210.
[12] A. S. J. Anvar and K. P. K. Krishna, "A Literature Review on Application of Sentiment Analysis Using Machine Learning Techniques," *International Journal of Applied Engineering and Management Letters (IJAEML)*, vol. 4, no. 2, pp. 41–77, 2020, doi: http://doi.org/10.5281/zenodo.3977576.
[13] M. D. Purbolaksono, D. T. B. Pratama, and F. Hamzah, "Perbandingan Gini Index dan Chi Square pada Sentimen Analsis Ulasan Film menggunakan Support Vector Machine Classifier," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 9, no. 3, pp. 528–534, 2023, doi: https://dx.doi.org/10.26418/jp.v9i3.68845.
[14] G. R. Ditami, E. F. Ripanti, and H. Sujaini, "Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 3, pp. 508–516, 2022, doi: https://dx.doi.org/10.26418/jp.v8i3.56478.
[15] S. E. Fadila and S. Y. Irianto, "Analisis Sentiment Review Aplikasi Mobile Legend Menggunakan Support Vector Machine Dan AdaBoost," *Jurnal Komputer Multidisipliner*, vol. 7, no. 2, pp. 36–43, 2024.
[16] D. F. Salsabillah, D. E. Ratnawati, and N. Y. Setiawan, "Analisis Sentimen Ulasan Rumah Makan Menggunakan Perbandingan Algoritma Support Vector Machine dengan Naive bayes (Studi Kasus: Ayam Goreng Nelongso Cabang Singosari, Malang)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 11, no. 1, pp. 107–116, 2024, doi: https://doi.org/10.25126/jtiik.20241117584.
[17] N. E. Oktaviana, Y. A. Sari, and Indriati, "Analisis Sentimen Terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features Dan Support Vector Machine," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 9, no. 2, pp. 357–362, 2022, doi: 10.25126/jtiik.202295625.
[18] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon Based Algorithm in Twitter Sentiment Analysis," *khazanah informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 8, no. 1, pp. 59–71, 2022, doi: https://doi.org/10.23917/khif.v8i1.15213.
[19] A. R. Maulana, S. H. Wijoyo, and Y. T. Mursityo, "Analisis Sentimen Kebijakan Penerapan Kurikulum Merdeka Sekolah Dasar dan Sekolah Menengah pada Media Sosial Twitter dengan Menggunakan Metode Word Embedding dan Long Short-Term Memory Networks (LSTM)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 3, pp. 523–530, Jun. 2023, doi: 10.25126/jtiik.2023106977.
[20] R. A. Lestari, A. Erfina, and W. Jatmiko, "Penerapan Algoritma Support Vector Machine Pada Analisis Sentimen Terhadap Identitas Kependudukan Digital," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 5, pp. 1063–1070, Oct. 2023, doi: 10.25126/jtiik.2023107264.
[21] Z. Abidin, A. Junaidi, and Wamiliana, "Text Stemming and Lemmatization of Regional Languages in Indonesia: A Systematic Literature Review," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 2, pp. 217–231, Jun. 2024, doi: 10.20473/jisebi.10.2.217-231.
[22] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 3, Dec. 2019, doi: https://dx.doi.org/10.26418/jp.v5i3.34368.
[23] S. A. R. Manaf, A. Fitrianto, and A. M. Soleh, "Perbandingan Algoritma Pohon dengan Beberapa Skenario Pelabelan untuk Analisis Sentimen pada Aplikasi Milik Pemerintah/BUMN," *JEPIN*

(Jurnal Edukasi dan Penelitian Informatika), vol. 10, no. 1, Apr. 2024, doi: https://dx.doi.org/10.26418/jp.v10i1.73512.
[24] S. H. Kusumahadi, H. Junaedi, and J. Santoso, "Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 4, no. 1, pp. 54–60, Jan. 2019, doi: 10.30591/jpit.v4i1.1125.
[25] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, Mar. 2021, doi: 10.34148/teknika.v10i1.311.
[26] C. D. García, "Visualizing the effect of hyperparameters on Support Vector Machines," Towards Data Science, February 8, 2021. [Online]. Available: https://towardsdatascience.com/visualizing-the-effect-of-hyperparameters-on-support-vector-machines-b9eef6f7357b. [Accessed September 9, 2024].
[27] K. Saputra, Zuriati, and Sriyanto, "Perbandingan Kinerja Fungsi Kernel Algoritma Support Vector Machine Pada Klasifikasi Penyakit Padi," *JURNAL TEKNIKA*, vol. 17, no. 1, pp. 119–131, 2023, doi: https://doi.org/10.5281/zenodo.7996483.
[28] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *JURNAL INFORMATIKA*, vol. 6, no. 2, pp. 226–235, Sep. 2019, doi: https://doi.org/10.31294/ji.v6i2.5129.