Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

Performance Comparison of SVM in Sentiment Analysis of Israel-Palestine Comments Using Lsa and Word2vec

Muh. Arsan Akbar[1], Abd. Azis Syam[2], Muh. Nur Hidayat Al Amanah[3], Andi Akram Nur Risal*[4], Dewi Fatmarani Surianto[5], Nur Azizah Eka Budiarti[6], Abdul Wahid[7]

[1234567]Department of Computer Engineering, State University of Makassar, Indonesia

Email: *[4]akramandi@unm.ac.id

Received : Apr 13, 2025; Revised : Jun 13, 2025; Accepted : Jun 23, 2025; Published : Feb 15, 2026

Abstract

This study compares two feature extraction techniques, namely Latent Semantic Analysis (LSA) and Word2Vec, in the sentiment classification of comments related to the Israeli-Palestinian conflict using Support Vector Machine (SVM). The dataset consists of 1000 YouTube comments and 158 news paragraphs, categorized into pro and con Palestinian sentiments. The preprocessing process includes casefolding, special character and stopword removal, lemmatization, and tokenization. The results show that SVM with Word2Vec has better performance than SVM with LSA in the classification of positive and negative comments. SVM model with Word2Vec recorded a precision value of 92% and F1-Score of 93% on negative comments. Meanwhile, SVM with LSA recorded 90% precision and 92% F1-Score. On positive comments, SVM with Word2Vec recorded 92% recall and 93% F1-Score. While SVM with LSA recorded 89% recall and 91% F1-Score. Word2Vec's strength lies in its ability to capture word context and nuance more effectively, thanks to training using richer contextualized comment and news data. In conclusion, although both methods show good ability in sentiment classification, the use of Word2Vec provides more consistent and accurate results. This research contributes to the advancement of sentiment classification methods in the context of complex socio-political issues and can serve as a reference for applying machine learning to more accurate and contextual public opinion analysis.

Keywords : *Israel-Palestinian Conflict, Latent Semantic Analysis, Sentiment Analysis, Support Vector Machine, Word2Vec.*

1. INTRODUCTION

The conflict between Palestine and Israel began decades ago and has yet to be resolved. Tensions began when Jews wanted their own state in Palestine, which was considered the promised land. This claim sparked conflict with Palestinians who felt it was unfounded [1], [2]. As a result of the conflict that occurred, on October 7, 2023 there were 1200 Israelis who died and several people became hostages [3]. Because of this, Israel carried out massive attacks on the Gaza region, even attacking hospitals and children. On January 24, 2024, the number of child victims reached at least 10,000 [4]. As of March 24, 2024 the total death toll in Gaza reached 30,000 [5].

Israel's massive attack on Gaza attracted a lot of attention and became a frequently discussed topic on the Internet. The rapid development of technology has resulted in the rapid dissemination of information related to the conflict, resulting in the proliferation of positive and negative comments or sentiments on the Internet related to the Israeli-Palestinian conflict. With so many sentiments, it is necessary to analyze the positive and negative comments [6]. A sentiment analysis of the Israeli-Palestinian conflict is necessary to understand the underlying dynamics and reasons for its existence. The conflict involves not only religious ideology but also international law, with reports of human rights violations against civilians [7]. Sentiment analysis is an important process of extracting opinion data

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

from large texts, through automatic extraction, processing, and understanding of messages in opinions [8], [9], [10]. Sentiment analysis also known as opinion mining, involves extracting opinions, attitudes, and emotions from text data, categorizing them as positive, negative, or neutral [11], [12]. Feature extraction plays an important role in sentiment analysis [13] [14] [15] [16]. It involves the selection and transformation of relevant information from text data to improve the accuracy of sentiment classification models. Various techniques such as Bag of Words (BOW), N-gram, TF-IDF, Word2vec, GLove, FastText, Count-Vectorization, Term Frequency-Inverse Document Frequency (TF-IDF), Word Embedding, and Hashing-Vectorizer are commonly used for feature extraction in sentiment analysis [17]. The results of feature extraction are used for input from classification methods, one of which is the Support Vector Machine (SVM) method. Support Vector Machine (SVM) is a supervised machine learning algorithm that aims to find the optimal boundary, or hyperplane, to separate different sample classes by a maximum margin [18] [19]. In sentiment analysis using SVM, the feature extraction methods used such as TF-IDF and Word2Vec greatly affect the accuracy of the model [20].

Several studies have been conducted on sentiment analysis with different methods and research objects. Research related to sentiment analysis on KAI Twitter tweet data uses the Multiclass SVM method with the One Against All (OAA) approach to classify three sentiment classes. The results showed that the TF-IDF unigram model combined with OAA Multiclass SVM achieved the highest accuracy of 80.59%, providing a better view of public opinion on Twitter for better services in the future [21].

Further research on the classification of human emotions from text by comparing TF-IDF and Word2Vec methods, using commuter line and transjakarta tweet data, as well as SVM and MNB methods. The results show that SVM with TF-IDF has the highest accuracy, reaching 93.45%, outperforming MNB with TF-IDF (81.78%) and SVM with Word2Vec (83.07%) [22]. The next study also used Word2Vec and SVM on user reviews of Gojek and Grab applications in the Google Play Store for sentiment data classification. The test results showed satisfactory performance, where the Gojek application achieved accuracy, precision, recall, and f1-score values of 89%, 94%, 86%, and 90%, respectively, while the Grab application achieved values of 87%, 94%, 85%, and 89% [23].

Another study also used Word2Vec for feature expansion with two classification methods namely SVM and ANN on 11,395 tweet data. The results show that ANN is better than SVM, where ANN without feature expansion gets an accuracy of 68.89% and with feature expansion reaches 72.58%. As for SVM, the accuracy without feature expansion is 63.95% and with feature expansion reaches 68.56%. This research indicates that feature expansion can improve the final accuracy in Twitter Sentiment Analysis [24].

The next research is related to sentiment analysis on the discourse of moving the capital city of Indonesia using SVM. The data used was 1,116 tweets, of which half were positive and half were negative. After preprocessing and weighting using TF-IDF, the SVM algorithm achieved 96.68% accuracy, 95.82% precision, 94.04% recall, and 0.979 AUC [9]. The next study also used SVM with TF-IDF to understand the public response to the acceptance of the Covid-19 vaccine program on Twitter data in 2021. The results showed that SVM divided the sentiment into 56.80% positive, 33.75% neutral, and 9.45% negative. The RBF kernel gives the highest accuracy, which is 92%, while the linear and polynomial kernels reach 90%, and the sigmoid kernel reaches 89% [25].

Another study used SVM and Naïve Bayes to analyze restaurant reviews in Jakarta. Data was preprocessed and features were selected using TSI. As a result, SVM is better than Naïve Bayes with recall, precision, and accuracy values around 0.79, while Naïve Bayes reaches around 0.77 for recall, 0.78 for precision, and 0.77 for accuracy [26].

Previous research has shown progress in sentiment analysis using feature extraction techniques such as TF-IDF and Word2Vec as well as classification techniques using SVM algorithms. However,

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

there has been no direct comparison between LSA, and Word2Vec as feature extraction methods, especially in the context of the Israeli-Palestinian conflict. Therefore, the objective of this study is to directly compare the performance of LSA and Word2Vec in extracting sentiment features from comments and news content related to the Israeli-Palestinian conflict, using the SVM classifier.

Thus, this research is expected to provide a better understanding of the advantages and disadvantages of each approach in terms of accuracy, efficiency, and interpretation of sentiment classification results, as well as make a significant contribution in the domain of text-based sentiment analysis of social conflicts.

2. METHOD

In this study, the developed model processes data with several test scenarios to determine the best sentiment analysis method on Israeli-Palestinian conflict comment data. Figure 1 shows the overall sentiment analysis process performed.

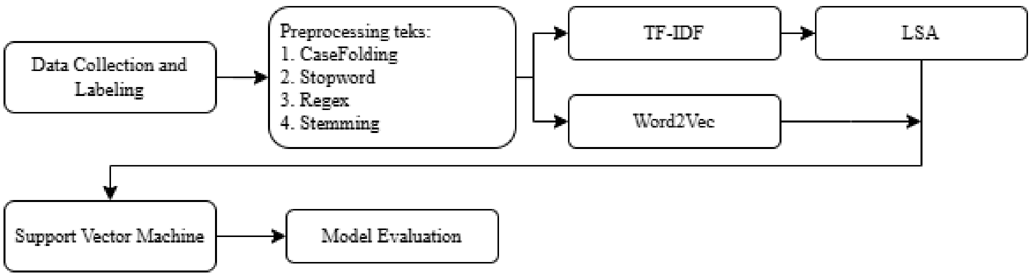

Figure 1. Research Method

2.1. Data Collection and Labelling

Sentiment data was collected via web scraping on YouTube, focusing on comments from videos discussing support for Palestine and Israel, such as “Free Palestine” and “Israel will win.” YouTube was chosen for its vast content on social and political conflicts. The data was labeled as positive (pro-Palestine) or negative (anti-Palestine). In addition, data from news sites, including detikcom, was gathered to supplement the Word2Vec model training. Table 1 below shows an example of the data labeling results.

Table 1. Sample data

Comment Text	Label
Mau donasi lewat mana yah spya tepat sasaran ke warga palestina? Mohon infonya	Positive
Semoga Allah SWT selalu melindungi kalian pejuang2 utk palestina	Positive
Alhamdulillah semoga bang husein dan penduduk palestina selalu di beri kesehatan di selamat kan dari segala bahaya zionis israel terkutuk	Positive
Saya Mendukung Israel IL Israel IL Memiliki hak untuk Membela diri terhadap siapa pun dan semua penyerangnya	Negative
Maju trus Israel hancur dan hilangkan negara Palestina karena nama Palestina adalah nama wilayah yang disebut Kanaan bukan negara	Negative

2.2. Text Preprocessing

Text data from YouTube requires preprocessing to ensure its suitability for analysis. The process involves cleaning noise, standardizing the text format, and preparing the data for sentiment analysis and predictive modeling. Initially, casefolding is applied to convert all characters to lowercase, improving text consistency and reducing variations. Following this, stopword removal eliminates non-informative

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

words such as "and," "or," and "which," as they do not contribute significantly to sentiment analysis. The next step involves using regex to remove special characters, punctuation, symbols, or other irrelevant elements from the comment data. Finally, stemming is performed to convert words into their base forms, such as changing "shooting" to "shoot" or "fight" to "fight." This comprehensive preprocessing enhances the effectiveness and accuracy of sentiment analysis on YouTube comment text. Table 2 provides an example of the final processed text.

Table 2. Example of text preprocessing results.

Comment Text	Label
donasi yah spya sasar warga palestina mohon info	Positive
moga allah swt lindung pejuang2 utk palestina	Positive
alhamdulillah moga bang husein duduk palestina sehat selamat bahaya zionis israel kutuk	Positive
dukung israel israel milik hak bela serang	Negative
maju trus israel hancur hilang negara palestina nama palestina nama wilayah kana negara	Negative

2.3. Feature Extraction using LSA

Latent Semantic Analysis (LSA) is a corpus-based approach in Natural Language Processing that evaluates the similarity of texts based on the semantic relationships between words [27]. LSA is an analytical technique used to identify patterns of hidden meaning in text by reducing the dimensionality of the data [28]. The main concept in LSA is the representation of the document-word matrix (Document-Term Matrix) which is then decomposed into a singular value matrix (SVD). The stages of LSA implementation are as follows:

2.3.1. Text Vectorization using TF-IDF

First, it uses the Term Frequency-Inverse Document Frequency (TF-IDF) method to convert the comment data into a numerical representation based on the occurrence weight of the words in the text. The maximum number of features is limited to 1000 to control the complexity of the model.

2.3.2. SVD Decomposition

After the term-document matrix is formed using TF-IDF, the next step is to perform SVD on the matrix using the equation (1).

$$A = U \Sigma V^T \tag{1}$$

Di mana:

- A is the term-document matrix.
- U is an orthogonal matrix whose columns are left singular vectors (term-concepts).
- Σ is a diagonal matrix with singular values ordered in descending order.
- V^T is an orthogonal matrix whose rows are right singular vectors (document-concept).

After SVD decomposition, data dimensionality reduction is performed by retaining essential information and removing irrelevant noise. The number of LSA topics is limited to 27 to represent the dimensions of meaning contained in the text. Dimensionality reduction is performed using the equation (2).

$$A_k = U_k \Sigma_k V_k^T \tag{2}$$

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

2.3.3. Splitting Training Data and Test Data using LSA Results

LSA transforms the train data into a data frame where each column represents a vector for the selected LSA topics. These vector representations serve as input for the SVM model.

2.4. Feature Extraction using Word2Vec

There are many Word2Vec models available for use. However, it is important to understand that the results of the Word2Vec model are heavily influenced by the training data used [29]. Existing models may not fully address the Israeli-Palestinian conflict, so this study developed a Word2Vec model using relevant training data. Word2Vec converts text into vector representations based on word context. The process starts with tokenizing the training data, where texts are broken into tokens and converted to lowercase for consistency. These tokenized results are then used to build a Word2Vec model using YouTube comments and news data. Word2Vec has two models: continuous bag-of-words (CBOW) and skip-gram. This study uses the skip-gram model, as it better represents infrequent words than CBOW [30]. Figure 2 is the architecture of the skip-gram model.

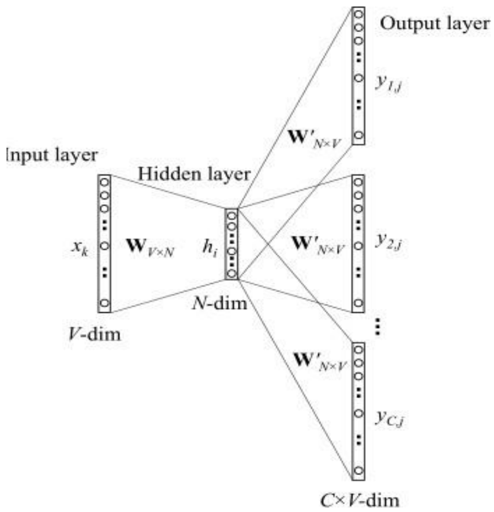

Figure 2. Skip-Gram architecture.

In the Skip-gram architecture, the model uses the current word to predict the surrounding context, learning the probability distribution of words within a predefined window. It consists of an input layer, a hidden layer, and an output layer [22].

The Word2Vec input layer is a one-hot vector, with one word from the vocabulary set to 1 and others set to 0. Each neuron in this layer represents a word. The hidden layer's neuron count corresponds to the word vector's dimensionality. The hidden layer uses a linear activation function, where the neuron value is the input multiplied by the weight. This activation is shown in (3). The hidden layer's output is then multiplied by a different weight in the output layer, as described in (4).

$$h = W^T x \tag{3}$$

Where x is the input one-hot vector and h is the hidden vector.

$$u_j = W'^T h \tag{4}$$

Where u_j is the j output line to the hidden layer and W'^T is the transpose of the weights from the hidden layer to the output layer.

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

In the output layer, the number of neurons used is the same as the number of neurons representing the target word in the input layer. The output layer uses the softmax activation function shown in (5).

$$y_j = \frac{exp\,(u_j)}{\sum_{j'=1}^{V} exp\,(u_j')} \quad (5)$$

Where y_j softmax output line j, u_j' is the output of all lines, and V is the number of vocabularies. After model generation, training and testing texts are vectorized using Word2Vec by calculating word vectors and combining them to represent entire texts. These vectors and corresponding labels are then prepared for SVM classification.

2.5. Support Vector Machine

LSA and Word2Vec text vector representations serve as input features for SVM models, enabling them to detect hidden patterns and separate sentiment classes. SVM excels in sentiment analysis by identifying the optimal hyperplane with the largest margin to distinguish between sentiment classes, such as positive and negative, in high-dimensional text data. This study employs an SVM model with a linear kernel, suited for the often linear nature of text data, and sets the random_state to 42 for reproducibility.

2.6. Model Evaluation

In the model evaluation stage, evaluation metrics such as True Negative (TN), False Positive (FP), False Negative (FN), and True Positive (TP) are calculated using the confusion matrix for SVM results with both feature extraction schemes. In addition, precision, recall, and F1-score, and accuracy using classification report are also calculated using the following formulas [22] [31].

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

3. RESULT

The dataset in this study comprises YouTube comments and news data on the Palestine-Israel conflict. A total of 1,000 comments were collected, evenly split between 500 pro-Palestine (positive) and 500 anti-Palestine (negative) comments. Additionally, 158 paragraphs of news data were gathered from various news sites.

The comment and news data underwent preprocessing, including casefolding (lowercasing), removal of special characters and punctuation, elimination of stopwords, stemming, and tokenization to enhance vector representation quality. The results of this preprocessing can be seen in Table 3, which illustrates each transformation stage from raw text to stemmed form.

Table 3. Text Preprocessing Stages.

Data	CaseFolding	Stopword	Regex	Stemmed
apa yang terjadi di Palestina, berbagai episode	apa yang terjadi di palestina, berbagai episode	palestina, episode kemenangan	palestina episode kemenangan	palestina episode menang juang g z tentara ker b bi

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

Data	CaseFolding	Stopword	Regex	Stemmed
kemenangan para pejuang G@z@ atas para tentara ker@ dan b@bi jelas mengindikasikan kehancuran entitas penjajah itu semakin dekat di depan mata...	kemenangan para pejuang g@z@ atas para tentara ker@ dan b@bi jelas mengindikasikan kehancuran entitas penjajah itu semakin dekat di depan mata...	pejuang g@z@ tentara ker@ b@bi mengindikasikan kehancuran entitas penjajah mata...	pejuang g z tentara ker b bi mengindikasikan kehancuran entitas penjajah mata	indikasi hancur entitas jajah mata
NKRI Harga MATI ngapain nama lu nkri tapi dukung zionis israel, nggak kapok Itu Kristen greja nya di hancurin dan di ludahi yahudi masihh tetep aja ngedukung cuihhhh	nkri harga mati ngapain nama lu nkri tapi dukung zionis israel, nggak kapok itu kristen greja nya di hancurin dan di ludahi yahudi masihh tetep aja ngedukung cuihhhh	nkri harga mati ngapain nama lu nkri dukung zionis israel, nggak kapok kristen greja nya hancurin ludahi yahudi masihh tetep aja ngedukung cuihhhh	nkri harga mati ngapain nama lu nkri dukung zionis israel nggak kapok kristen greja nya hancurin ludahi yahudi masihh tetep aja ngedukung cuihhhh	nkri harga mati ngapain nama lu nkri dukung zionis israel nggak kapok kristen greja nya hancurin ludah yahudi masihh tetep aja ngedukung cuihhhh
Mau donasi lewat mana yah spya tepat sasaran ke warga palestina ? Mohon infonya	mau donasi lewat mana yah spya tepat sasaran ke warga palestina ? mohon infonya	donasi yah spya sasaran warga palestina ? mohon infonya	donasi yah spya sasaran warga palestina mohon infonya	donasi yah spya sasar warga palestina mohon info
Semoga Allah SWT selalu melindungi kalian pejuang2 utk palestina	semoga allah swt selalu melindungi kalian pejuang2 utk palestina	semoga allah swt melindungi pejuang2 utk palestina	semoga allah swt melindungi pejuang2 utk palestina	moga allah swt lindung pejuang2 utk palestina

LSA was employed to extract semantic information from documents, while Word2Vec generated word-based vector representations to capture semantic relationships. The SVM model utilized LSA and Word2Vec-based vector representations as input features for sentiment classification on the comment data. Table 4 presents the SVM model's confusion matrix results using LSA.

Table 4. Confusion Matrix SVM with LSA.

		Prediction	
		Negative	Positive
Actual	Negative	95	6
	Positive	11	88

The SVM evaluation using LSA on 200 test data correctly classified 183 instances: 95 negative (anti-Palestine) and 88 positive (pro-Palestine) comments. However, the model exhibited shortcomings, including 6 false positives (misclassifying negative comments as pro-Palestine) and 11 false negatives (misclassifying pro-Palestine comments as negative). Table 5 presents the confusion matrix for the SVM model with LSA.

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

Table 5. Confusion Matrix SVM with Word2Vec.

		Prediction	
		Negative	Positive
Actual	Negative	98	6
	Positive	8	88

Testing classification performance with SVM using Word2Vec features yielded better results compared to LSA. However, some shortcomings remain 6 false positives (misclassifying negative comments as pro-Palestine) and 8 false negatives (misclassifying pro-Palestine comments as negative). Figure 3 compares the accuracy of SVM models using LSA and Word2Vec, showing both achieve high accuracy, with the Word2Vec-based SVM outperforming the LSA-based model.

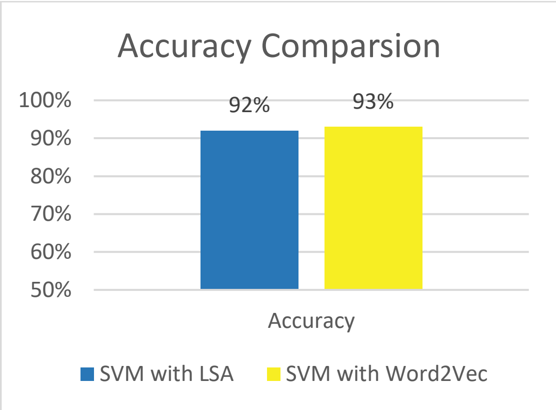

Figure 3. Model accuracy comparison.

This study evaluates the precision, recall, and F1-Score of SVM models with LSA and Word2Vec, reflecting their effectiveness in identifying positive and negative comments. Figure 4 compares the performance metrics of both models.

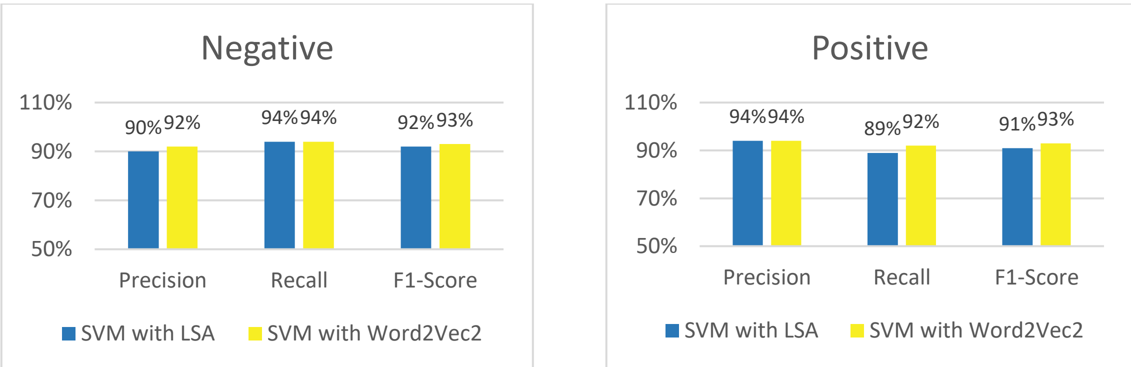

Figure 4. Comparison of precision, recall, and F1-Score of SVM model with LSA and SVM model with Word2Vec.

The negative comment classification results reveal that SVM with Word2Vec achieves higher precision (92%) than SVM with LSA (90%), indicating Word2Vec more accurately identifies negative comments with fewer false positives. Both methods exhibit identical recall (94%), demonstrating equal capability in detecting negative comments. However, the F1-Score, a balance of precision and recall, is slightly higher for Word2Vec (93%) compared to LSA (92%). This suggests that Word2Vec offers

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

improved overall performance by capturing contextual nuances in negative comments more effectively than LSA, providing a slight edge in prediction accuracy and classification consistency.

The positive comment classification results indicate that both methods achieve identical precision (94%), demonstrating equal accuracy in identifying positive comments. However, recall is higher for SVM with Word2Vec (92%) compared to SVM with LSA (89%), indicating that Word2Vec detects positive comments more effectively. This difference is also reflected in the F1-Score, with Word2Vec achieving 93% compared to LSA's 91%. These findings show that while both methods perform equally well in precision, Word2Vec offers better detection ability and consistency in classifying positive comments, capturing the context and nuances of words more effectively than LSA.

To further compare the performance of SVM with LSA and SVM with Word2Vec, classification results on 200 preprocessed test data were analyzed. Table 6 presents the counts of correct and false predictions for both models.

Table 6. Confusion Matrix

	True		False	
	SVM with LSA	SVM with Word2Vec	SVM with LSA	SVM with Word2Vec
Positive	75	79	21	17
Negative	97	101	7	3
Total	172	180	28	20

Table 5 highlights that SVM with Word2Vec outperforms SVM with LSA, as the latter produces more classification errors. To identify the weaknesses of both models, a comparison of the word distribution in misclassified comments was conducted. Figure 5 displays the wordclouds for the False Positive and False Negative categories from the classification results.

False Positive SVM with LSA

False Negative SVM with LSA

False Positive SVM with W2V

False Negative SVM with W2V

Figure 5. Wordcloud of False Positive and False Negative categories from both models.

Figure 5 shows that sentences with “israel,” “negara,” and “hancur” are often misclassified as positive by the SVM model with LSA. Similarly, in the Word2Vec model, sentences with “israel,” “negara,” and “dunia” are negative but classified as positive. For the False Negative category in the LSA-based SVM, sentences with “palestina,” and “negara” are positive but misclassified as negative.

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

In the Word2Vec SVM model, False Negatives include sentences with “palestina,” “israel,” and “negara,”. Tables 7 and 8 show classification examples from both models.

Table 7. Example of SVM model classification with LSA

Prediction Class	Actual Class	Comment
Negative	Positive	moga hancur dgn kaum israel yahudi nya israel bkn negara tp kaum yg benci alloh swt gak aku ada negara israel kaum yg bodoh dunia
Negative	Positive	israel hebat hamas yg keok wkwk sampe frustasi nya sampe nyerang warga sipil krna ga dpt nyerang target nya
Negative	Postitive	ndak sabar pingin lihat hancur zionis israel
Negative	Negative	urus negara kalian campur kalian demo pun israel peduli

Table 8. Example of SVM model classification with Word2Vec

Prediction Class	Actual Class	Comment
Positive	Positive	moga hancur dgn kaum israel yahudi nya israel bkn negara tp kaum yg benci alloh swt gak aku ada negara israel kaum yg bodoh dunia
Negative	Positive	israel hebat hamas yg keok wkwk sampe frustasi nya sampe nyerang warga sipil krna ga dpt nyerang target nya
Negative	Postitive	ndak sabar pingin lihat hancur zionis israel
Positive	Negative	urus negara kalian campur kalian demo pun israel peduli

The classification results in Tables 6 and 7 reveal performance differences between the models. LSA frequently misclassifies positive comments as negative, highlighting its limitations in capturing complex sentiment contexts. Conversely, Word2Vec demonstrates greater accuracy in cases like correctly identifying positive comments, attributed to its ability to capture word meanings in broader contexts. However, Word2Vec also faces challenges in interpreting ambiguous negative sentiments. Overall, Word2Vec proves superior to LSA in handling nuanced and complex sentiment.

4. DISCUSSION

Word2Vec outperformed LSA in classifying negative and positive comments due to its richer and more contextual feature representation. Unlike LSA, which uses only comment data and loses information during dimensionality reduction, Word2Vec leverages both comment and news data, enabling it to better capture word nuances and contexts. Additionally, the larger training dataset enhances Word2Vec's accuracy in generating vector representations, resulting in more consistent and precise classification outcomes.

Previous research in the field of sentiment analysis has applied other feature extraction methods for input to SVMs. Study by [9] used TF-IDF for feature extraction in sentiment analysis of the discourse on the relocation of Indonesia's capital city, while the research by [22] One of the methods uses TF-IDF feature extraction in the classification of human emotions from text. Previous research has shown that an appropriate feature extraction method can improve the accuracy of classification models. However, these studies did not use comment data related to the Israeli-Palestinian conflict, which certainly has different characteristics from the data used in previous studies. This study compares two feature extraction techniques, Latent Semantic Analysis (LSA) and Word2Vec, using SVM for sentiment classification on Israeli-Palestinian conflict comments. The results of this study show lower accuracy compared to previous studies where the study by [9] achieved an accuracy of 96.68% and research by

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

[22] get an accuracy of 93.45%. By comparing the results with previous research, it is found that the feature extraction method and the use of different datasets greatly affect the performance of SVM classification.

A study by [32] employed TF-IDF as a feature extraction technique for sentiment classification related to the Israel-Palestine conflict, achieving an accuracy of 97%. The model recorded a precision, recall, and F1-Score of 95%, 91%, and 93% respectively for negative comments, and 97%, 99%, and 98% for positive comments. In contrast, the present study utilizes Word2Vec, which is capable of capturing word context and semantic relationships more effectively. Although the model's overall accuracy is slightly lower at 93%, its performance on negative comments remains competitive, with a precision of 92%, recall of 94%, and F1-Score of 93%. For positive comments, it achieved a precision of 94%, recall of 92%, and F1-Score of 93%.

While these results are slightly lower compared to the previous study using TF-IDF, the Word2Vec approach offers advantages in capturing semantic context and inter-word relationships, particularly in more complex and contextual data. These performance differences underscore the significant impact of feature extraction methods and dataset characteristics on SVM classification results. The urgency of this research lies in the need for sentiment analysis models that can effectively process and interpret data related to socially and politically sensitive topics, such as the Israel–Palestine conflict. Such topics often involve high emotional intensity, polarized language, and subtle contextual cues that conventional methods may fail to capture. By comparing two widely used semantic representation techniques, this study contributes to the development of more robust, context-aware sentiment analysis approaches and advances the scientific understanding of how language functions within complex and sensitive domains.

5. CONCLUSION

This study compares the performance of SVM with Word2Vec and SVM with LSA to classify positive and negative comments on the Israeli-Palestinian conflict. Classification is done with two different schemes, namely SVM with feature extraction using Word2Vec and SVM with feature extraction using LSA. The results show that SVM with Word2Vec provides an advantage over LSA in the classification of positive and negative comments. For negative comments, Word2Vec recorded higher precision compared to LSA. Although the recall of both methods has the same value, Word2Vec has a better F1-Score value. Then in the classification of positive comments, the precision of both methods has the same value, but Word2Vec has a higher recall and produces an F1-Score that is also higher than LSA. This shows that Word2Vec is more effective in capturing the context and nuances of words, thus improving prediction accuracy and classification consistency. The conclusions of this study can be summarized as follows,

1. Word2Vec outperforms LSA in negative comment classification, achieving higher precision and F1-Score, while recall remains equal for both methods.
2. In positive comment classification, Word2Vec and LSA have equal precision, but Word2Vec achieves higher recall and F1-Score, indicating better overall performance.
3. Word2Vec proves more effective in capturing word context and semantic nuances, leading to improved sentiment prediction accuracy and consistency.
4. The urgency of this research lies in its relevance to current social and political issues, where accurate sentiment analysis can assist in early detection of hate speech, propaganda, and misinformation on social media.
5. Future research is needed to expand the dataset and explore other classification methods (e.g., KNN, Naïve Bayes) using both LSA and Word2Vec to further enhance the effectiveness and generalizability of sentiment analysis models.

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

REFERENCES

[1] F. Firdaus, J. Septian Putra, R. Saaulia, and S. Adnis, "Yasser Arafat dan Konflik Palestina-Israel (Tinjauan Sejarah)," Khazanah J. Sej. Dan Kebud. Islam, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.15548/khazanah.v10i1.265.

[2] A. Syari'ah, N. Nabilah, and R. Wijayanti, "Kekejaman Israel terhadap Rakyat Palestina: Telaah Berita-Berita CNN Indonesia Tahun 2019-2021," vol. 1, no. 1, pp. 58–80, 2022.

[3] R. Picheta and S. McCarthy, "What did Israel know about Hamas' October 7 attack?," CNN World.

[4] M. Haddad and M. Hussein, "Know Their Names: Palestinian Children Killed in Israeli Attacks on Gaza." [Online]. Available: https://interactive.aljazeera.com/aje/2024/israel-war-on-gaza-10000-children-killed/

[5] "Video: Korban Tewas di Gaza Capai Lebih Dari 32 Ribu Jiwa." [Online]. Available: https://www.cnbcindonesia.com/news/20240324150254-8-524917/video-korban-tewas-di-gaza-capai-lebih-dari-32-ribu-jiwa

[6] 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," J. Nas. Komputasi Dan Teknol. Inf., vol. 5, no. 1, 2022, doi: 10.32672/jnkti.v5i1.4004.

[7] T. Mahwati and A. R. Nanda, "Analysis of the Palestinian and Israeli Conflict in the Perspective of International Humanitarian Law," Int. Law Discourse Southeast Asia, vol. 1, no. 1, pp. 23–42, Jan. 2022, doi: 10.15294/ildisea.v1i1.56873.

[8] A. B. Sasmita, B. Rahayudi, and L. Muflikhah, "Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes," J. Pengemb. Teknol. Inf. Dan Ilmu Komput., vol. 6, no. 3, 2022.

[9] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," J. Teknol. Inf. Dan Ilmu Komput., vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[10] A. A. N. Risal, F. Fathahillah, and D. R. A. Sulaiman, "Classification of Sentiment Analysis and Community Opinion Modeling Topics for Application of ICT in Government Operations," Int. J. Environ. Eng. Educ., vol. 5, no. 1, pp. 35–44, Apr. 2023, doi: 10.55151/ijeedu.v5i1.99.

[11] S. C H, "Exploring Sentiment Analysis: Applications, and Challenges —A Comprehensive Survey," INTERANTIONAL J. Sci. Res. Eng. Manag., vol. 07, no. 07, Jul. 2023, doi: 10.55041/IJSREM24621.

[12] R. Koli and S. Redekar, "A Review on Sentiment Analysis Methodologies, Practices and Applications with Machine Learning," Int. J. Comput. Sci. Mob. Comput., vol. 12, no. 6, pp. 64–70, Jun. 2023, doi: 10.47760/ijcsmc.2023.v12i06.007.

[13] G. Kaur and A. Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," J. Big Data, vol. 10, no. 1, p. 5, Jan. 2023, doi: 10.1186/s40537-022-00680-6.

[14] F. Es-sabery, K. Es-sabery, H. Garmani, J. Qadir, and A. Hair, "Evaluation of Different Extractors of Features at the Level of Sentiment Analysis," Infocommunications J., vol. 14, no. 2, pp. 85–96, 2022, doi: 10.36244/ICJ.2022.2.9.

[15] S. Saifullah, R. Dreżewski, F. Andika Dwiyanto, A. Sasmito Aribowo, and Y. Fauziah, "Sentiment Analysis Using Machine Learning Approach Based on Feature Extraction for Anxiety Detection," presented at the Computational Science – ICCS 2023, ICCS 2023, 2023, pp. 365–372.

[16] J. Liu, Z. Yan, S. Chen, X. Sun, and B. Luo, "Channel Attention TextCNN with Feature Word Extraction for Chinese Sentiment Analysis," ACM Trans. Asian Low-Resour. Lang. Inf. Process., vol. 22, no. 4, pp. 1–23, 2023, doi: 10.1145/3571716.

[17] S. Ouaari, T. M. Tashu, and T. Horvath, "Multimodal Feature Extraction for Memes Sentiment Classification," presented at the 2022 IEEE 2nd Conference on Information Technology and Data Science (CITDS), IEEE, 2022, pp. 285–290. doi: 10.1109/CITDS54976.2022.9914260.

[18] P. E. Romero, O. Rodriguez-Alabanda, E. Molero, and G. Guerrero-Vaca, "Use of the Support

Vol. 7, No. 1, February 2026, Page. 31-43
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4601

Vector Machine (SVM) Algorithm to Predict Geometrical Accuracy in the Manufacture of Molds via Single Point Incremental Forming (SPIF) Using Aluminized Steel Sheets," J. Mater. Res. Technol., vol. 15, pp. 1562–1571, 2021, doi: 10.1016/J.JMRT.2021.08.155.

[19] B. S. Waluyo Poetro, E. Maria, H. Zein, E. Najwaini, and D. H. Zulfikar, "Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification," Indones. J. Data Sci., vol. 5, no. 1, pp. 15–22, Mar. 2024, doi: 10.56705/ijodas.v5i1.123.

[20] T. Sabri, S. Bahassine, O. El Beggar, and M. Kissi, "An Improved Arabic Text Classification Method Using Word Embedding," Int. J. Electr. Comput. Eng. IJECE, vol. 14, no. 1, pp. 721–721, 2024, doi: 10.11591/ijece.v14i1.pp721-731.

[21] D. Nur Fitriana and Y. Sibaroni, "Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM)," J. RESTI Rekayasa Sist. Dan Teknol. Inf., vol. 4, no. 5, pp. 846–853, 2020, doi: 10.29207/resti.v4i5.2231.

[22] D. E. Cahyani and I. Patasik, "Performance Comparison of TF-IDF and Word2Vec Models for Emotion Text Classification," Bull. Electr. Eng. Inform., vol. 10, no. 5, pp. 2780–2788, 2021, doi: 10.11591/eei.v10i5.3157.

[23] S. Styawati, A. Nurkholis, A. Ari Aldino, S. Samsugi, E. Suryati, and R. Puji Cahyono, "Sentiment Analysis on Online Transportation Reviews Using Word2Vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm," IEEE, pp. 163–167, 2021, doi: 10.1109/ISMODE53584.2022.9742906.

[24] N. Adi Nugroho and E. Budi Setiawan, "Implementation Word2Vec for Feature Expansion in Twitter Sentiment Analysis," J. RESTI Rekayasa Sist. Dan Teknol. Inf., vol. 5, no. 5, pp. 837–842, 2021, doi: 10.29207/resti.v5i5.3325.

[25] M. Rahardi, A. Aminuddin, F. F. Abdulloh, and R. A. Nugroho, "Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 6, 2022, doi: 10.14569/IJACSA.2022.0130665.

[26] L. Gunawan, M. S. Anggreainy, L. Wihan, Santy, G. Y. Lesmana, and S. Yusuf, "Support vector machine based emotional analysis of restaurant reviews," Procedia Comput. Sci., vol. 216, pp. 479–484, 2023, doi: 10.1016/j.procs.2022.12.160.

[27] R. M. Suleman and I. Korkontzelos, "Extending Latent Semantic Analysis to Manage its Syntactic Blindness," Expert Syst. Appl., vol. 165, pp. 114130–114130, 2021, doi: 10.1016/j.eswa.2020.114130.

[28] T. Deguchi, S. Seo, and N. Ishii, "Meaning of the Clusters on Dimensionality Reduction by Word Clustering," IEEE, pp. 325–330, 2022.

[29] N. Yang, G. Li, H. Ding, and C. Gong, "Study on Tibetan Word Vector based on Word2vec," J. Phys. Conf. Ser., vol. 1187, no. 5, pp. 052074–052074, 2019, doi: 10.1088/1742-6596/1187/5/052074.

[30] X. Zhang and L. Zhang, "The comparison and analysis of Skip-gram and CBOW in creating financial sentimental dictionary," Appl. Comput. Eng., vol. 44, no. 1, pp. 56–67, Mar. 2024, doi: 10.54254/2755-2721/44/20230155.

[31] S. Pandey, D. Heuer, A. Deore, K. Fender, B. Schlomann, and R. Reynolds-Haertle, "Results of Machine Learning Models".

[32] Abd. A. Syam, G. Hardy M, A. Salim, D. F. Surianto, and M. Fajar B, "ANALISIS TEKNIK PREPROCESSING PADA SENTIMEN MASYARAKAT TERKAIT KONFLIK ISRAEL-PALESTINA MENGGUNAKAN SUPPORT VECTOR MACHINE," JIPI J. Ilm. Penelit. Dan Pembelajaran Inform., vol. 9, no. 3, pp. 1464–1472, Aug. 2024, doi: 10.29100/jipi.v9i3.5527.