Institute of Research and Publication Indonesia (IRPI)
Public Research Journal of Engineering, Data Technology and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/predatecs
Vol. 3 Iss. 2 January 2026, pp: 138-149
ISSN(P): 3024-921X | ISSN(E): 3024-8043

Sentiment Analysis of Public Opinion on the Gaza Conflict Using Machine Learning

Agil Irman Fadri[1], Nur Futri Ayu Jelita[2*], Diamond Dimas Bagaskara[3], Raudiatul Zahra[4]

[1,2]Department of Information System, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[3]Department of Management Information System, Faculty Uygulamali Bilimler Fakultesi/ Applied Science, Necmettin Erbakan University, Turkiye
[4]Department of İlahiyat Fakültesi, Faculty of Theology, Erciyes Üniversitesi, Turkiye

E-Mail: [1]12250314181@students.uin-suska.ac.id, [2]12250320374@students.uin-suska.ac.id, diamondbagas@gmail.com, [4]raudiatulzahra498@gmail.com

Received Jun 08th 2025; Revised Jan 22th 2026; Accepted Jan 25th 2026; Available Online Jan 31th 2026
Corresponding Author: Nur Futri Ayu Jelita

Abstract

The 2023 escalation of the Gaza conflict triggered widespread public discourse on the X platform, highlighting the importance of sentiment analysis for understanding public opinion on global geopolitical issues. While sentiment analysis has been widely applied to social media data, comparative evaluations of machine learning models on conflict-related datasets remain limited. This study analyzes public sentiment toward the Gaza conflict by comparing the performance of Multi-Layer Perceptron, XGBoost, and Logistic Regression models. A dataset of 2,175 tweets was processed using standard text preprocessing techniques and TF-IDF feature extraction. Model performance was evaluated using multiple train-test split scenarios. The results indicate that Logistic Regression consistently outperformed the other models, achieving the highest accuracy of 73.17% with an 80:20 data split. These findings suggest that simpler linear models may perform more robustly and efficiently than more complex approaches when applied to high-dimensional, noisy social media text data. This study provides practical insights into model selection for sentiment analysis of conflict-related discussions on social media platforms.

Keywords: Gaza Conflict, Logistic Regression, Multi Layer Perceptron, Sentiment Analysis, XGBoost

1. INTRODUCTION

The Israeli-Palestinian conflict is a complex geopolitical issue that has long attracted global attention and stimulated discussions among governments, international organizations, and civil society [1]. In recent years, social media platforms have become important spaces for public expression, allowing individuals to share opinions and emotional responses to political and humanitarian crises in real time. As a result, understanding public sentiment toward the Gaza conflict is increasingly relevant, as it influences media narratives, public discourse, and international responses to the conflict [2] [3]. Sentiment analysis, defined as the computational process of identifying and categorizing opinions expressed in textual data, has been widely used to examine public opinion on social media platforms. Previous studies have demonstrated that machine learning-based sentiment analysis can effectively capture public attitudes toward political events and armed conflicts. Various supervised learning algorithms and feature extraction techniques, such as Term Frequency-Inverse Document Frequency (TF-IDF), have been applied to classify sentiment polarity in social media [4].

However, many existing studies primarily focus on reporting classification performance, with limited discussion on how different machine learning models behave when applied to high-dimensional, noisy, and relatively small datasets that are typical of social media data [5]. In addition, comparative studies that systematically evaluate both simple and complex machine learning models within the context of conflict-related social media discourse remain limited. In particular, there is a lack of research that examines how linear and non-linear models perform when analyzing public sentiment toward the Gaza conflict on contemporary social media platforms. This gap makes it difficult to draw methodological conclusions regarding the most appropriate model choices for sentiment analysis tasks involving conflict-driven public discourse [6] [7][8].

Various machine learning algorithms have been applied for sentiment analysis with varying levels of complexity and performance. In sentiment classification tasks, both linear and non-linear models are commonly employed [8]. Previous studies have demonstrated that non-linear approaches such as Multi-Layer Perceptron and ensemble-based methods like XGBoost are capable of capturing complex patterns in textual data and have shown strong performance across different application domains [9][10][11][12]. At the same time, Logistic Regression remains one of the most widely used linear baseline models in sentiment analysis due to its simplicity and interpretability [13][14]. Despite the successful application of these models in prior studies, their comparative performance in analyzing conflict-related social media data, particularly in the context of the Gaza conflict, has not been sufficiently explored. This gap raises an important methodological question: do more complex models necessarily provide superior performance compared to simpler approaches when dealing with limited, high-dimensional, and noisy social media text data?

To address this gap, this study conducts a comparative analysis of public sentiment related to the Gaza conflict using three supervised machine learning models: Multi-Layer Perceptron, XGBoost, and Logistic Regression. Using a dataset of 2,175 tweets collected from the X platform, the study evaluates model performance under different train–test split scenarios to assess robustness and generalization. The main contribution of this research is to provide empirical insights into model selection for sentiment analysis of conflict-related social media data, particularly by examining whether simpler linear models can outperform more complex approaches under limited data conditions.

2. MATERIAL AND METHOD

Sentiment analysis of social media text requires appropriate feature representation and a structured analytical workflow to ensure reliable results. In this study, Term Frequency–Inverse Document Frequency (TF-IDF) is employed for feature extraction due to its effectiveness in representing high-dimensional and sparse textual data, particularly in short and noisy social media content. Previous studies have demonstrated that TF-IDF provides a strong baseline for sentiment classification and remains competitive when combined with supervised machine learning models [8]. The overall research process consists of data collection, text preprocessing, feature extraction, model training, and evaluation, as illustrated in Figure 1.

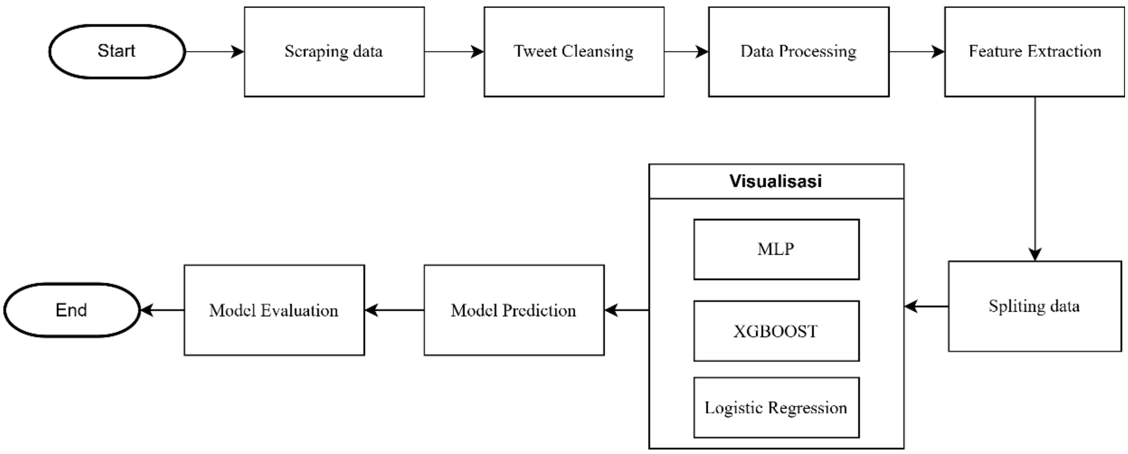

Figure 1. Research Methodology

2.1. Scraping Data

At this stage, we collect raw data from social media, especially Twitter. The data collection process is done through the scraping method by utilising APIs such as Twitter API as well as additional libraries such as snscrape and Tweepy. The data collected are tweets containing certain keywords that are relevant to the topic we are researching.

2.2. Tweet Cleaning

At this stage, the raw data from scraping is cleaned first so that it can be used in the next analysis process. The cleaning process begins by removing duplicate data and empty lines using the drop_duplicates() and dropna() functions. After that, the content of the tweets is cleaned of punctuation, numbers, URLs, and other irrelevant symbols using regular expressions. All text is then converted to lowercasing to maintain uniformity. In addition, common words that do not contribute meaning (stopwords) are removed using a list from the NLTK library. As a final step, emojis and non-alphabetic characters were removed to make the data cleaner and more suitable for analysis. After all these steps were completed, the number of tweets remaining and ready for analysis was 2175.

2.3. Data Processing

After the tweet cleaning process, the next step is to prepare the data for sentiment modelling. The first step is tokenisation to break the text into words. Then stemming is carried out, stemming is the process of mapping and removing inflections in a word into the form of a base word [15], e.g. 'running' and 'jog' become 'run'. Each tweet is then labelled with a positive, negative or neutral sentiment based on its meaning. Finally, the labelled data is converted to numerical format using techniques such as label encoding or one-hot encoding in order to be processed by the classification algorithm. From this process, we obtained 2,175 data distributed into three sentiment categories, as shown in Figure 2.

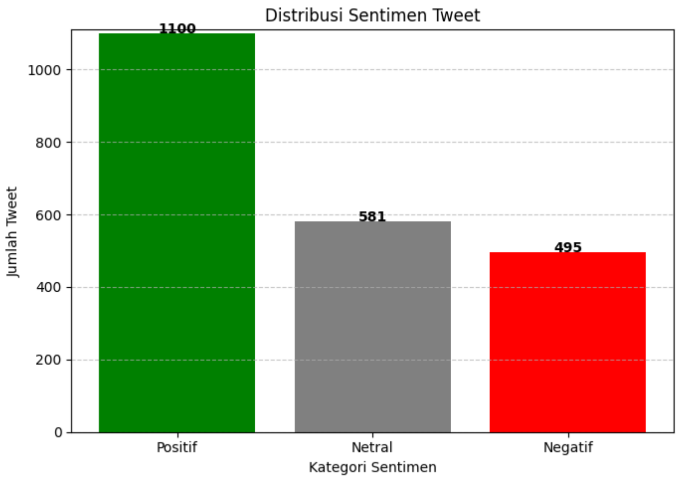

Figure 2. Number of reviews by sentiment

2.4. Feature Extraction

Figures Machines cannot understand text directly as they can only process data in numeric form. Therefore, text data needs to be converted to numerical form via a feature-extraction process. One commonly used method is TF-IDF (Term Frequency-Inverse Document Frequency) [10]. TF measures how often a word appears in a document, while IDF gives greater weight to words that appear infrequently throughout the document. The TF-IDF value is calculated as shown in Equation 1.

$$\mathrm{IDF} = \log\left(\frac{\mathrm{N}}{\mathrm{DF}}\right) \tag{1}$$

TF(k,d) indicates the number of words displayed in document d, while IDF(k) indicates the inverse of document frequency, as shown in Equation 2.

$$\mathrm{TF} - \mathrm{IDF}\ (\mathrm{d}, \mathrm{k}) = \mathrm{TF}\ (\mathrm{d}, \mathrm{k})\ \times \mathrm{IDF}\ (\mathrm{k}) \tag{2}$$

2.5. Multi-Layer Perceptron (MLP)

Multilayer Perceptron (MLP) is a type of artificial neural network that is often utilised in various applications, including machine learning and artificial intelligence. MLP is particularly effective in supervised learning tasks, where it learns to connect inputs to outputs based on examples that have been labelled [8]. This network is composed of several layers of interconnected neurons, including an input layer, one or more hidden layers, and an output layer [16]. Each neuron performs calculations and forwards the results to neurons in the next layer, and the weight values on the connections between neurons are adjusted during the training process to produce accurate predictions [17].

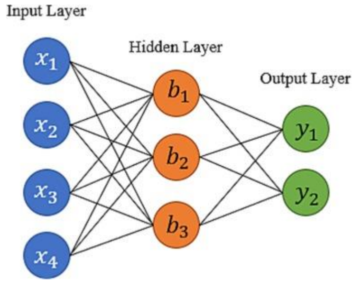

Figure 3. MLP Network Architecture

One crucial aspect of MLPs is the presence of hyperparameters, which are determined before training and affect network performance [18]. These hyperparameters include various settings such as the number of hidden layers, learning rate, activation function, and regularisation parameters. Hyperparameter optimisation is essential to achieve optimal MLP performance. Various methods, including genetic algorithms, have been proposed to optimise MLP hyperparameters, aiming to improve accuracy, generalisation ability, and training speed. Figure 3 depicts the MLP network architecture.

2.6. XGBoost

XGBoost, which stands for Extreme Gradient Boosting, is an efficient and widely used machine learning algorithm, especially for classification and sentiment analysis tasks [11]. As an ensemble method and boosting technique, XGBoost combines predictions from several weaker models, such as decision trees, to produce a stronger and more accurate prediction model [19]. It implements gradient boosting to iteratively minimise the loss function, emphasises computational efficiency, and uses regularization techniques to prevent overfitting [20]. This works well because the model is updated using a robust formula, equation 3.

$$y^{\wedge(t)}_{1} = y^{\wedge(t-1)} + \eta \,.\, f_t\,(X_i) \quad (3)$$

In this formula, $y^{\wedge(t)}_{1}$ is the prediction result for the i-th data at the t-th iteration.. Symbol η shows how large the pace of change is (called the learning rate), and $f_t\,(X_i)$ is the result of the new decision tree created at the tth iteration. This process is iterative, with each prediction improved by adding a new tree. This is the core of XGBoost's strength. It excels because it has several optimization strategies that make it perform better [21].

2.7. Logistic Regression

Logistic Regression is a statistical method used for classification, especially in binary classification situations, intending to predict the probability of the outcome of an event [22]. This technique uses a sigmoid function to convert linear combinations of input variables into probability values [23], which allows data to be grouped into different categories. In logistic regression, the probability of a data point falling into the 'positive' class is calculated by applying a logistic function to the linear sum of the predictor variables. This equation is then mathematically formulated as equation (4).

$$p(x) = \frac{1}{1 + e^{-(\beta_0 \beta_1 x_1 + \ldots + \beta_n x_n)}} \quad (4)$$

In this formula, p(x) means the probability of the data belonging to the positive class, 〖(β〗 _0 is the initial value (intercept), β1\beta_1β1 to β_1 are the effect numbers (coefficients) for each variable x_1 to x_n, and e is the natural log base number. The decision boundary occurs when p(x) =0.5 - anything above 0.5 is 'positive,' anything below 0.5 is 'negative.'

2.8. Model Evaluation

The model evaluation stage is the last step after the model is used to predict the testing data. At this stage, measurements are made of the model's performance to determine how well the model is classified [20]. Some commonly used evaluation methods include accuracy, precision, recall, F1-score, and confusion matrix which helps to see the number of correct and incorrect predictions in each category [24].

2.9. Confusion Matrix

Confusion Matrix is a classification performance evaluation method that compares correct and incorrect data. This matrix calculates accuracy, precision, recall, and error rate to assess the performance of the model based on the correctness and error rate of the classification results [25]. Confusion Matrix is a table that displays the number of correctly classified test data and the number of incorrectly classified test data [26]. Diagram confusion matrix can be seen in Figure 4.

3. RESULTS AND DISCUSSION

In the results section, the dataset will be processed by scraping tweet data, then cleansing and processing the data, and finally extracting features. The data is then split for training three models: MLP, XGBoost, and Logistic Regression. After the models are trained, predictions and evaluations are made to assess the performance of each model.

3.1. Processing Data

Pre-processing is the initial stage in data processing that aims to prepare the data to be more easily processed by the system. The first step in this process is data cleaning, which removes irrelevant data or corrects it to suit the needs of the analysis. Processing can be seen in Table 1.

Table 1. Processing

No	Full Text	Tweet English
1.	tanya memang bikin orang pikir keras aku respect….	Ask indeed makes people think loudly I really respect …
2.	sesuai konteks sih emang lagi bahas kamu sebut atas lagi israel….	according to the context, it is really discussing that you call it again…..
3.	bahas bebegini tambah kena karma lah apa lah padahal waktu palestina israel….	Discussing Beegini, I got more karma even though the Palestinian Israeli….
4.	perang india vs pakistan ubah nuansa geopolitik gkamubal…..	Indian War vs Pakistan Change the geopolitical nuances of Gkamubal
5.	hakamu cartoon nunjukin orang bendera israel ….	Your rights Cartoon shows the Israeli flag people….
6.	lima negara lapor bantu madam bakar israel mei 2025 ….	Five Countries reported Madam Bakar Israeli May 2025….
….	….	….
2173	perlu merdeka buat negara nya jajah laknatullah ….	Need to be independent for its country Jajah laknatullah…
2174	ya btul pakdhe ngutukin israel fatal liat apa beliau alami skg jujur sy sampe gk tega…..	Yes, Btul Pakdhe cursed Israel fatal, seeing that he experienced honestly until I couldn't see the sadness…
2175	bukan masalah orang asing menakusai kak mrk lg juang hak tanah….	It's not a problem for foreigners to defeat, sis, they will be the right …..

Before the data is analysed, several preprocessing stages are performed such as normalisation, stopword removal, tokenisation, stemming, and translation. Table 2 shows the results of each of these stages on one data sample.

In the data preprocessing stage, this research uses Python on the Google Colab platform. The data preparation process starts with cleaning irrelevant elements such as hashtags, emojis, and hashtags. After that, several advanced stages were carried out, namely data normalisation, tokenisation, stemming, and finally data translation into English. This translation is done because the model is better able to recognise sentence structures in English, has more accurate word embeddings, and can better distinguish between positive, negative and neutral sentiments. Word clouds of positive and negative sentiment can be seen in Figures 4 and 5.

The word cloud visualization of positive and negative comments related to the Israeli-Palestinian conflict shows that key words such as conflict, Israel, Palestine, Gaza, and Palestinian appear in both sentiment types, indicating that the main issue and the parties involved remain at the center of attention. However, there are differences in the nuances of the words used. Positive comments tend to contain peaceful and solution words such as support, peaceful, solution, and human, reflecting hope for peace and empathy. Meanwhile, negative comments are dominated by emotional and violent words such as hate, killed, military, attack, and genocide, indicating anger, suffering, and condemnation of violence.

Table 2. Sample Result of Data Preprocessing

Data	Result
Normalization	
tanya memang bikin orang pikir keras aku respect banget sama ingin tetap adil meski asa benci kuat aku paham konflik israelpalestina emang rumit bikin hati panas dengan tindak israel sering bikin	pertanyaanmu ini memang bikin orang berpikir keras dan aku respect banget sama keinginanmu untuk tetap adil meski ada perasaan benci yang kuat aku paham konflik israelpalestina ini emang rumit dan bikin hati panas apalagi dengan tindakan israel yang sering bikin
Stopword Removal	
pertanyaanmu ini memang bikin orang berpikir keras dan aku respect banget sama keinginanmu untuk tetap adil meski ada perasaan benci yang kuat aku paham konflik israelpalestina ini emang rumit dan bikin hati panas apalagi dengan tindakan israel yang sering bikin	pertanyaanmu memang bikin orang berpikir keras aku respect banget sama keinginanmu tetap adil meski perasaan benci kuat aku paham konflik israelpalestina emang rumit bikin hati panas dengan tindakan israel sering bikin
Tokenization	
pertanyaanmu memang bikin orang berpikir keras aku respect banget sama keinginanmu tetap adil meski perasaan benci kuat aku paham konflik israelpalestina emang rumit bikin hati panas dengan tindakan israel	Pertanyaanmu, memang, bikin, orang, berpikir, keras, aku, respect, banget, sama, keinginanmu, tetap, adil, meski, perasaan, benci, kuat, aku, paham, konflik, israelpalestina, emang, rumit, bikin, hati, panas, dengan,

Data	Result
sering bikin	tindakan, israel, sering, bikin
Stemming	
Pertanyaanmu, memang, bikin, orang, berpikir, keras, aku, respect, banget, sama, keinginanmu, tetap, adil, meski, perasaan, benci, kuat, aku, paham, konflik, israelpalestina, emang, rumit, bikin, hati, panas, dengan, tindakan, israel, sering, bikin	tanya memang bikin orang pikir keras aku respect banget sama ingin tetap adil meski asa benci kuat aku paham konflik israelpalestina emang rumit bikin hati panas dengan tindak israel sering bikin
Translate Data	
tanya memang bikin orang pikir keras aku respect banget sama ingin tetap adil meski asa benci kuat aku paham konflik israelpalestina emang rumit bikin hati panas dengan tindak israel sering bikin	Ask indeed makes people think loudly I really respect the same as wanting to remain fair even though I hate strongly I understand the conflict of Israelpalestine it is complicated to make the heart hot with Israeli acts often make

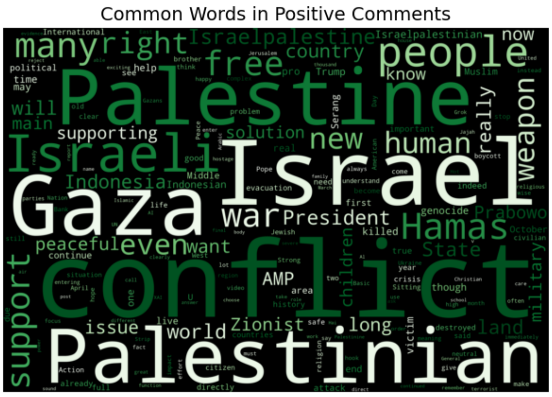

Figure 4. Wordcloud of positive sentiment.

Figure 5. Wordcloud of negative sentiment.

3.2. Term Frequency-Inverse Document Frequency (TF-IDF)

The data after the preprocessing step will be converted to numeric form using TF-IDF weighting. In this step, the calculation is performed in Python using the scikit-learn model. TF-IDF results are in Table 3.

Table 3. TF-IDF

No	About	Accept	Accomodate	Accurate	accusations	abal
1	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0
…	…	…	…	…	…	….
2176	0.0	0.0	0.0	0.0	0.0	0.0

3.3. Multi-Layer Perceptron

The performance of the Multi-Layer Perceptron (MLP) model shows relatively stable results across different data-sharing ratios, such as 70:30, 80:20, and 90:10 (see Table 4). Although the accuracy ranged only between 65% and 69%, the model still showed a balance between precision and recall. The best result was achieved at a ratio of 90:10 with an accuracy of 69.27%, precision of 69.29%, and recall of 69.27%, while at a ratio of 80:20, the model recorded an accuracy of 69.04%, precision of 69.25%, and recall of 68.04%, with an F1-Score of 68.24%. At a ratio of 70:30, Figures 6 show the confusion matrix results which indicate that MLP still shows consistency in mapping sentiment towards the Palestinian and Israeli issues equally. However, with better parameter settings and data training, MLP has the potential to achieve better results in the future.

Table 4. MLP Evaluation Result

Algorithm	Split	Evaluation	Result (%)
		Accuracy	65.85%
	70% : 30%	Precision	65.10%
		Recall	65.85%
		F1-Score	65.11%
		Accuracy	69.04%
MLP	80% : 20%	Precision	69.25%
		Recall	68.04%
		F1-Score	68.24%
		Accuracy	69.27%
	90% : 10%	Precision	69.29%
		Recall	69.27%
		F1-Score	68.07%

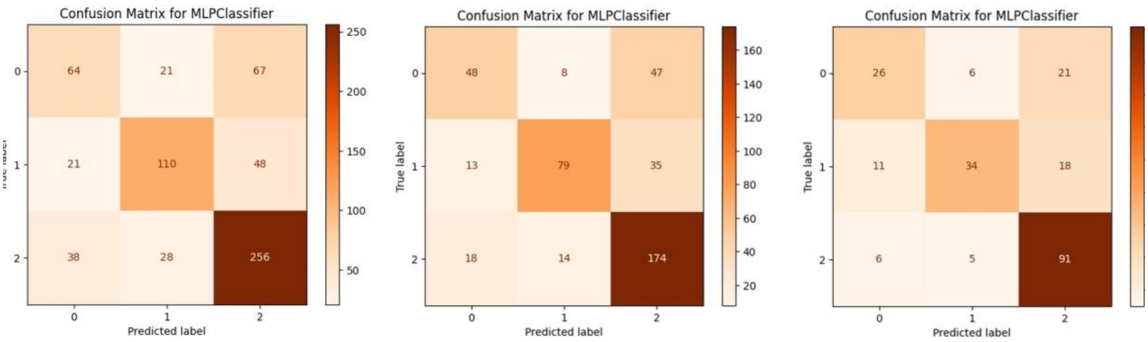

Figure 6. Combined confusion matrices of the MLP algorithm for 70:30, 80:20, and 90:10 data splits (left to right).

3.4. XGBoost

In the XGBoost algorithm, the model's performance shows an increasing trend as the data-sharing ratio changes. The best results were obtained at a ratio of 90:10 with an accuracy of 70.18%, precision of 74.79%, recall of 70.18%, and F1-Score of 67.93%. At a ratio of 80:20, the accuracy achieved was 67.66%, with 71.16% precision, 67.66% recall, and 65.55% F1-Score. Meanwhile, at a 70:30 ratio, the model achieved 67.08% accuracy, 71.11% precision, 67.08% recall, and 64.07% F1-Score. The consistent precision values above 70% across all ratios indicate that XGBoost is quite good at correctly identifying positive sentiment. Although the accuracy is not very high (see Table 5), the model's performance is stable and tends to improve at higher training data ratios. The confusion matrix results shown in Figure 7 show a fairly balanced distribution of predictions. Overall, XGBoost shows competitive performance in sentiment classification towards Palestine and Israel and can be considered a viable model for similar cases.

Table 5. Xgboost Evaluation Result

Algorithm	Split	Evaluation	Result (%)
		Accuracy	67.08%
	70% : 30%	Precision	71.11%
		Recall	67.08%
		F1-Score	64.07%
		Accuracy	67.66%
XGBOOST	80% : 20%	Precision	71.16%
		Recall	67.66%

Algorithm	Split	Evaluation	Result (%)
		F1-Score	65.55%
	90% : 10%	Accuracy	70.18%
		Precision	74.79%
		Recall	70.18%
		F1-Score	67.93%

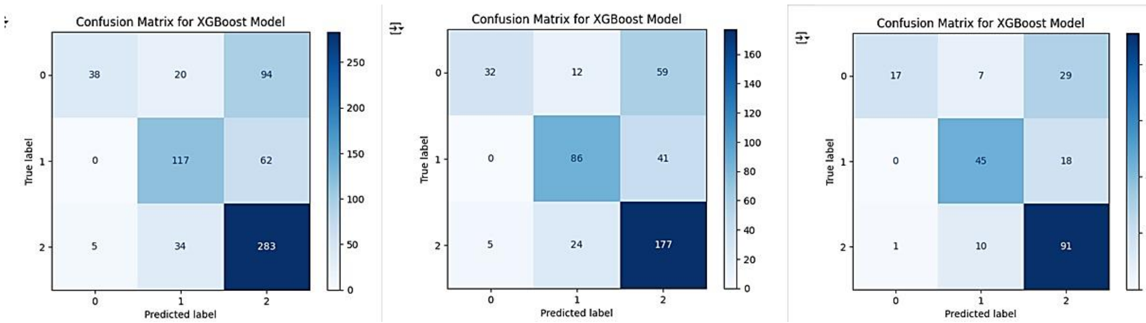

Figure 7. Combined confusion matrices of the XGBoost algorithm for 70:30, 80:20, and 90:10 data splits (left to right)

3.5. Logistic Regression

In the Logistic Regression algorithm, the model shows a fairly good and stable performance at various data sharing ratios, namely 70:30, 80:20, and 90:10. The best results were achieved at a ratio of 80:20 with an accuracy of 73.17%, precision of 73.57%, recall of 73.17%, and F1-Score of 71.89%. At a ratio of 70:30, the model obtained an accuracy of 71.21% with 71.88% precision, 71.21% recall, and F1-Score of 69.92%. While at a ratio of 90:10, the accuracy slightly decreased to 71.10%, with a precision of 73.68%, recall of 71.10%, and F1-Score of 69.70% (see Table 6). The confusion matrix results in Figures 8 show that Logistic Regression is able to maintain a balance between precision and recall, and provide fairly reliable classification results for the case of sentiment towards Palestine and Israel. Overall, the model shows good consistency and can be one of the effective algorithms in handling sentiment classification on this dataset.

Table 6. Logistic Regression Evaluation Result

Algorithm	Split	Evaluation	Result (%)
Logistic Regression	70% : 30%	Accuracy	71.21%
		Precision	71.88%
		Recall	71.21%
		F1-Score	69.92%
	80% : 20%	Accuracy	73.17%
		Precision	73.57%
		Recall	73.17%
		F1-Score	71.89%
	90% : 10%	Accuracy	71.10%
		Precision	73.68%
		Recall	71.10%
		F1-Score	69.70%

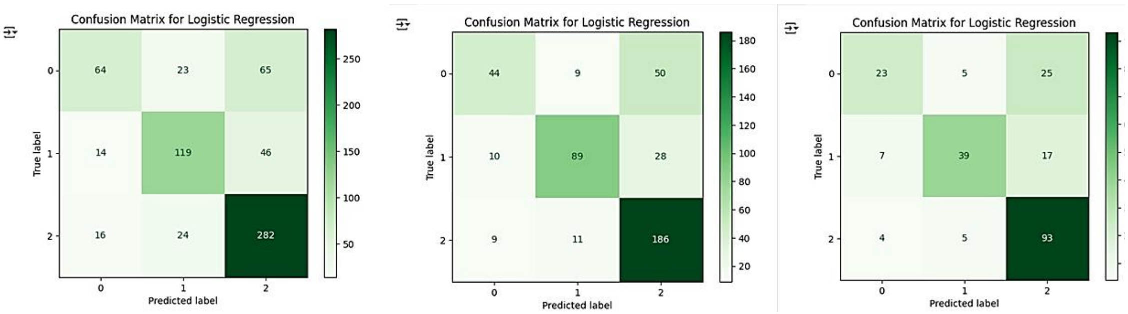

Figure 8. Combined confusion matrices of the Logistic Regression algorithm for 70:30, 80:20, and 90:10 data splits (left to right)

3.6. Algorithm Comparison

The results of the comparison among the three algorithms show that Logistic Regression performs best at all data-sharing ratios. The highest accuracy was obtained at 80:20 ratio of 73.17%, with 73.57% precision and 73.17% recall, showing a good balance in prediction. Meanwhile, XGBoost showed quite competitive results with improved performance as the training data ratio increased, where the highest accuracy was achieved at 90:10 ratio of 70.18% and precision of 74.79%. On the other hand, MLP (Multi-Layer Perceptron) showed the lowest performance among the three, with only 65.85% accuracy at 70:30 ratio, and limited improvement at 80:20 (69.04%) and 90:10 (69.27%) ratios. Overall, Logistic Regression proved to be the most consistent and reliable for sentiment classification towards the Palestinian-Israeli conflict.

Based on comparisons of 2175 Twitter data sets with several data-sharing ratios, the Logistic Regression algorithm performs best, followed by MLP and XGBoost. The highest accuracy was achieved at an 80:20 ratio of 73.17%. Meanwhile, XGBoost achieved 70.18% accuracy at a 90:10 split, and MLP showed the lowest results, with a maximum accuracy of 69.27%. The accuracy comparison graph can be seen in Figure 9.

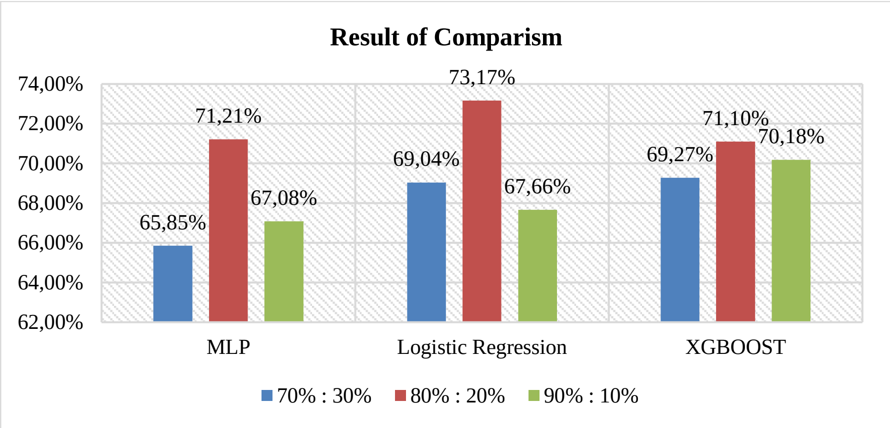

Figure 9. Result of Comparison Models

The dominance of Logistic Regression over MLP and XGBoost is mainly influenced by the dataset's characteristics and the representation of textual features. Conflict-related sentiment texts generally exhibit explicit polarity and linear decision boundaries, which can be efficiently captured by Logistic Regression. Furthermore, the sparsity and high dimensionality of TF-IDF features favor the use of linear classifiers with regularization. MLP, as a neural-based model, requires a larger training set to achieve optimal generalization, whereas XGBoost tends to perform better on dense numerical features rather than sparse textual data. These factors contribute to the consistent and superior performance of Logistic Regression in this study.

3.7. Discussion

Based on the experimental results, Logistic Regression demonstrated the most optimal performance in classifying sentiment on Twitter data related to the Gaza conflict, achieving the highest accuracy of 73.17% at a data split ratio of 80:20. This finding is consistent with previous sentiment analysis studies that reported the effectiveness of linear classifiers on high-dimensional and sparse text data. K. Shah et al. showed that Logistic Regression performs well in sentiment classification tasks involving limited contextual information[24], while M. A. Ullah et al. reported that Logistic Regression can provide competitive or even superior performance compared to more complex models in certain scenarios[27].

The superior performance of Logistic Regression in this study can be attributed to the characteristics of conflict-related social media data and the feature representation employed. Tweets discussing the Gaza conflict are typically short, noisy, and emotionally explicit, with sentiment polarity often conveyed through specific keywords. Such patterns tend to be linearly separable and are therefore well captured by linear classifiers. In addition, the use of TF-IDF–based representations produces high-dimensional and sparse feature spaces, which favor models such as Logistic Regression that are designed to handle sparse data efficiently.

The impact of dataset size further explains the observed performance differences among the models. With a medium-sized dataset consisting of 2,175 tweets, Logistic Regression benefits from strong generalization capability and low variance, resulting in stable performance across different data splitting ratios. In contrast, MLP requires larger training datasets and careful hyperparameter tuning to effectively learn non-linear representations, which may lead to suboptimal performance under limited data conditions.

Meanwhile, XGBoost, although powerful for structured numerical data, shows limited effectiveness when applied to sparse and high-dimensional textual features.

Beyond methodological considerations, the results also have important social and policy implications. Public sentiment toward the Gaza conflict reflects collective emotional responses and public polarization during crises. Accurate sentiment classification can assist policymakers, humanitarian organizations, and media institutions in monitoring public opinion, identifying shifts in discourse, and designing more effective communication strategies. Overall, the findings indicate that simpler linear models, such as Logistic Regression, can outperform more complex approaches in conflict-related sentiment analysis when dealing with limited, noisy, and high-dimensional social media data.

4. CONCLUSION

Based on the evaluation of the three algorithms, Logistic Regression demonstrated the best performance in sentiment classification of Twitter data related to the Gaza conflict, achieving the highest accuracy of 73.17% at an 80:20 data split ratio. The results indicate that Logistic Regression provides more stable and consistent performance across different data splitting scenarios compared to MLP and XGBoost. While XGBoost achieved competitive results, its performance remained below that of Logistic Regression, and MLP consistently produced the lowest accuracy. These findings suggest that simpler linear models can be more effective than complex approaches when applied to conflict-related social media sentiment analysis under limited data conditions.

Despite these promising results, this study has several limitations that should be acknowledged. First, the dataset size is relatively limited, consisting of only 2,175 tweets, which may restrict the generalizability of the findings. Second, the study relies on word-based feature representations using TF-IDF, which may not fully capture contextual and semantic information in short and noisy social media texts. Third, the analysis is constrained to English-language data translated from Indonesian, which may introduce translation bias and affect the accuracy of sentiment classification.

Future research should address these limitations by employing larger and more diverse datasets to improve model robustness and generalization. The use of context-aware deep learning models, such as LSTM or transformer-based architectures including BERT, is recommended to better capture semantic and contextual information in conflict-related discourse. In addition, incorporating multilingual sentiment analysis and analyzing native-language data directly may yield more accurate insights into public opinion across linguistic communities.

Overall, this study provides empirical evidence that model simplicity and dataset characteristics play crucial roles in sentiment analysis performance. The findings highlight the importance of aligning model selection with data properties and application contexts, particularly in analyzing public sentiment toward sensitive and complex geopolitical conflicts.

REFERENCES

[1] P. Kayade, A. Pardeshi, S. Patil, P. Shetkar, P. Raut, and V. Singh, “Sentiment Analysis On Israel-Palestine Reddit Dataset,” J. Emerg. Technol. Innov. Res., vol. 11, no. 11, pp. 1–8, 2024.

[2] A. Gangwar and T. Mehta, “Sentiment Analysis of Political Tweets for Israel Using Machine Learning,” in ICMLBDA-2022, 2023, p. 10.

[3] A. Liyih, S. Anagaw, M. Yibeyin, and Y. Tehone, “Sentiment analysis of the Hamas-Israel war on YouTube comments using deep learning,” Sci. Rep., vol. 14, 2024.

[4] D. Sarah Delva and K. Muslim Lhaksmana, “Sentiment Analysis of the Palestine-Israel Crisis on Social Media using Convolutional Neural Network,” Technol. Sci., vol. 6, no. 1, pp. 334–343, 2024.

[5] J. Rosenberger, F. U. Erlangen-nürnberg, and L. Gasse, “Hate Speech and Sentiment of YouTube Video Comments From Public and Private Sources Covering the Israel-Palestine Conflict,” in Wirtschaftsinformatik 2024 Proceedings, 2024.

[6] K. El Damanhoury, F. Saleh, and M. Lebovic, “Covering the Israeli–Palestinian Conflict: A Critical Discourse Analysis of Al Jazeera English and BBC’s Online Reporting on the 2023 Gaza War,” J. Media Stud., vol. 6, no. 1, pp. 1–16, 2025.

[7] F. Fahrani; J. Aryanto, “Sentiment Analysis of Public Opinion on the Palestinian-Israeli Conflict using Support Vector Machine and Naïve Bayes Algorithms,” JSRET (Journal Sci. Res. Educ. Technol., vol. 3, no. 4, pp. 1890–1900, 2024.

[8] F. Z. El-Hassani, M. Amri, N. E. Joudar, and K. Haddouch, “A New Optimization Model for MLP Hyperparameter Tuning: Modeling and Resolution by Real-Coded Genetic Algorithm,” Neural Process. Lett., vol. 56, no. 2, 2024.

[9] M. Askari and F. Keynia, “Mid-term electricity load forecasting by a new composite method based on optimal learning MLP algorithm,” IET Gener. Transm. Distrib., vol. 14, no. 5, pp. 845–852, 2020.

[10] K. Afifah, I. N. Yulita, and I. Sarathan, “Sentiment Analysis on Telemedicine App Reviews using XGBoost Classifier,” 2021 Int. Conf. Artif. Intell. Big Data Anal. ICAIBDA 2021, pp. 22–27, 2021.

[11] E. Dataset and H. A. Taher, "Dual-Language Sentiment Analysis : A Comprehensive Evaluating SVM , Dual-Language Sentiment Analysis : A Comprehensive Evaluating SVM , Logistic Regression , XGBoost , and Decision Tree Using TF-IDF On Arabic and English Dataset," Wasit J. Pure Sci., vol. 3, no. 4, pp. 59–69, 2023.

[12] I. G. B. A. Budaya and I. K. P. Suniantara, "Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 3, pp. 1077–1086, 2024.

[13] A. Madasu and S. Elango, "Efficient feature selection techniques for sentiment analysis," Multimed. Tools Appl., vol. 79, no. 9–10, pp. 6313–6335, 2020.

[14] F. H. Rachman, "Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regresion," Int. Conf. Inf. Technol. Inf. Syst., pp. 238–242, 2020.

[15] T. Hema et al., "Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier," Procedia Comput. Sci., vol. 197, pp. 660–667, 2022.

[16] G. Airlangga, "Predicting Student Performance Using Deep Learning Models : A Comparative Study of MLP , CNN , BiLSTM , and LSTM with Attention," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 4, pp. 1561–1567, 2024.

[17] A. Al Bataineh and S. Manacek, "MLP-PSO Hybrid Algorithm for Heart Disease Prediction," J. Pers. Med., vol. 12, no. 8, 2022.

[18] G. Airlangga, "Spam Detection in YouTube Comments Using Deep Learning Models : A Comparative Study of MLP , CNN , LSTM , BiLSTM , GRU , and Attention Mechanisms," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 4, pp. 1533–1538, 2024.

[19] L. Duan, "Machine Learning Modeling for Forecasting Repeat Purchases in Online Shopping," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 3, pp. 863–874, 2024.

[20] N. Anggraini, S. J. Putra, L. K. Wardhani, F. D. U. Arif, N. Hakiem, and I. M. Shofi, "A Comparative Analysis of Random Forest, XGBoost, and LightGBM Algorithms for Emotion Classification in Reddit Comments," J. Tek. Inform., vol. 17, no. 1, pp. 88–97, 2024.

[21] G. Airlangga, "Exoplanet Classification Through Machine Learning: A Comparative Analysis of Algorithms Using Kepler Data," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 3, pp. 753–763, 2024.

[22] A. Samih, A. Ghadi, and A. Fennan, "Enhanced sentiment analysis based on improved word embeddings and XGboost," Int. J. Electr. Comput. Eng., vol. 13, no. 2, pp. 1827–1836, 2023.

[23] F. A. K. Murti, "Predictive Model Comparison for Predicting Condom Use: Comparison of Conventional Logistic Regression and Other Machine Learning," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 4, pp. 1320–1329, 2024.

[24] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification," Augment. Hum. Res., vol. 5, no. 1, 2020.

[25] Z. C. Dwinnie et al., "Application of the Supervised Learning Algorithm for Classification of Pregnancy Risk Levels," vol. 1, pp. 26–33, 2023.

[26] Z. C. Dwynne; Mustakim; Mustafa, "Comparison Of Machine Learning Algorithms On Sentiment Analysis Of Elsagate Content," in 2024 International Conference on Smart Computing, IoT and Machine Learning (SIML), 2024, pp. 239–243.

[27] M. A. Ullah, S. M. Marium, S. A. Begum, and N. S. Dipa, "An algorithm and method for sentiment analysis using the text and emoticon," ICT Express, vol. 6, no. 4, pp. 357–360, 2020.