Journal of Informatics and Web Engineering
Vol. 4 No. 2 (June 2025) eISSN: 2821-370X

Sentiment Analysis of the Israel-Palestine Conflict on X: Insights from the Indonesian Perspective using a Long Short-Term Memory Algorithm

Mohammad Taleb Noori[1], Muhammad Alif Rahman[2], Agus Purnomo[3], Aripin[4*]
[1]Department of Public Health, Maulana Azad University, India
[2,4]Department of Informatics, Dian Nuswantoro University, Semarang, Central Java, Indonesia
[3]Department of Electrical Engineering, Dian Nuswantoro University, Semarang, Central Java, Indonesia
corresponding author: (arifin@dsn.dinus.ac.id; ORCiD: 0000-0001-8434-5775)

Abstract - The Israel-Palestine conflict which has persisted for decades drives mounting global interest that consequently influences public opinion worldwide. This article examines the sentiment analysis of X (Twitter) data pertaining to the conflict using the Long Short-Term Memory (LSTM) model. This study presents public reactions through an analysis of 1,700 tweets collected between May and July 2023 which encapsulate key recent developments. In this study, several steps were conducted, namely 1) crawling process to get raw data; 2) preprocessing: cleansing, case folding, tokenization, stop word removal, and stemming; 3) modelling and validation using the LSTM model; 4) model evaluation based on performance metrics to evaluate the ability of the classification model to distinguish between classes; 5) visualization of experimental results. The LSTM model is a modification of the recurrent neural network (RNN). The LSTM model has many advantages, including being able to remember a collection of information that has been stored for a long period of time, being able to delete information that is no longer relevant, and being more efficient in processing, predicting, and classifying data based on a certain time sequence. Another advantage is that LSTM's ability to identify temporal dependencies and contextual interactions in sequential data makes it suitable for social media text analysis. The model demonstrated success in sentiment classification on geopolitical topics with an impressive accuracy rate of 91%. The findings demonstrate deep learning's potential applications for sentiment analysis and offer insights into public opinion dynamics during times of international crises.

Keywords— Sentiment Analysis, Long Short-Term Memory, Israel-Palestine Conflict, X (Twitter), Indonesian Perspective

Received: 28 February 2025; Accepted: 14 May 2025; Published: 16 June 2025

1. INTRODUCTION

Started in the year 1948 and currently going strong, the conflict between Israel and Palestine has attracted worldwide attention for more than 77 years. One of the most amazing rises took place on May 7, 2021, when the Israeli Defence Forces fired a barrage of more than 4,000 missiles towards Gaza City, Palestine [1]. The growing number of public opinions, for and against the engaged parties, especially on internet channels, with X playing as an important outlet, reflects this escalating conflict. A major social media site, X is vital for allowing people to voice their opinions and

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2025.4.2.27

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

emotional reactions to the Israel-Palestine issue. Reflecting public opinion and viewpoint on this continuous geopolitical catastrophe, it is an evolving repository of user-generated material. For academics looking at changing opinions and the psychological atmosphere around the conflict [2], these tweets provide an invaluable source of data. The use of social media has exploded over the last ten years, changing necessary channels for individuals all around to express opinions, participate in discussions, and discuss subjects of interest. Social media has become a vital tool to predict and grasping public mood [3] because of its large user base, varied conversation topics, and abundance of real-time data.

By means of information extraction, sentiment analysis is the technique of automatically deciding the subjective polarity of a given entity. The goal is to determine whether the user-generated text conveys a positive, negative, or neutral attitude. LSTM is a deep learning technique in the field of artificial intelligence, is well-known for its ability to analyse and interpret sequential inputs, including natural language. Its application in sentiment analysis is particularly beneficial for detecting temporal dependencies and contextual relationships among words in short-form text. Given that messages are short but linguistically rich, LSTM has been particularly appropriate for analysing sentiment on X. In the setting of the Israel-Palestine conflict, in which feelings preconceptions, and ideological biases are inextricably intertwined, understanding public sentiment offers essential insights that can direct more responsive and sophisticated recommendations for policy. The growing impact of social media on influencing public discourse underlines the importance of researching open sentiment dynamics. Particularly X offers a sizable corpus for examining such opinions in real time.

Previous studies have indicated that sentiment investigation on social media can exactly reflect shifts in public feelings and opinion on controversial issues. This paper aims to use advanced sentiment analysis techniques, which is the LSTM model, to examine the complexity and temporal change of public feelings on X regarding the Israel-Palestine conflict. Key topics addressed in this paper are: How does emotions connect to X public opinion regarding this conflict? How can the results of this study clarify and direct more careful policy recommendations? This paper also evaluates the LSTM model's effectiveness in capturing these dynamic opinions on X regarding the Israel-Palestine conflict by means of an examination of the evolution of sentiment patterns over time to reveal changes in public discourse.

2. LITERATURE REVIEW

2.1 Data Minning

Data mining techniques are methodologies employed in the data mining process to extract valuable patterns or insights from a dataset. Data mining is a discipline within computer science and statistics that encompasses several methodologies for the analysis and interpretation of data. Frequently employed data mining techniques encompass intricate data processing, modelling, and analysis. Data mining is typically employed for classification or clustering processes. The primary objective of data mining clustering techniques is to identify commonalities in data pertinent to the professional domain and the employment status of graduates. The data is subsequently categorized into distinct groups [4]. The X social media network is extensively utilised by the Indonesian populace for rapid communication and information access. This has generated numerous sentiments among the Indonesian populace that can serve as case studies, one of which pertains to the 2024 Indonesian presidential contender. Sentiments towards these public personalities will be categorised with the LSTM algorithm with positive and negative classifications [5].

2.2 Analysis Sentiment

Sentiment analysis is a component of natural language processing (NLP) and data processing that emphasizes the identification and classification of views or emotions expressed in text [6]. Social media has been extensively utilised for public attention, particularly in the context of the world issue. Sentiment analysis seeks to evaluate opinionated words expressed in text from the X platform. Utilising the machine learning assistance method, crawled data is acquired and subsequently processed through text preprocessing, which includes cleaning, case folding, normalisation, stop word removal, tokenisation, and stemming [7].

2.2.1 Sentiment Analysist Approach

There are three main approaches to sentiment analysis, namely: (a) Rule-based: this system categorizes sentiment using a set of recognized linguistic principles. These rules may include word patterns, regular expressions, or grammatical conventions [8]. (b) Machine learning: this approach uses machine learning techniques to extract sentiment patterns from annotated training data. Popular algorithms include LSTM and deep learning [9]. (c) Lexicon-based: central to this research is a mechanism that determines a text's sentiment score by looking up words in a dictionary and giving them a positive, negative, or neutral score. Summing up the sentiment scores of all the words in the text is how the sentiment score is determined [10], [11].

2.3 Lexicon-Based: The Most Approach in This Investigate

The lexicon-based approach has two primary sorts: (a) Dictionary-based: this approach assesses sentiment by comparing words in the text to a pre-established sentiment lexicon. This dictionary may consist of a compilation of generic positive and negative terms, or a specialised lexicon designed for a particular field. (b) Corpus-based: this approach constructs a sentiment lexicon from an extensive, unstructured text corpus. The mood of a word is ascertained by the co-occurrence of other words within the corpus. This study selected the lexicon-based strategy due to its simplicity and ease of implementation. This approach necessitates no intricate model training and may be rapidly implemented on extensive text datasets. Furthermore, the analytical outcomes are readily interpretable as they rely on a lexicon of terms with established meanings [10].

2.4 History Israel-Palestine

The origins of the enduring conflict between Israel and Palestine have been contested since its inception in the early 20th century. Numerous research has elucidated the elements contributing to the conflict through diverse instruments and ideas. This research aims to examine the origins of the conflict through the lens of the political culture of each nation, employing the Political Culture theory developed by Gabriel Almond and Sydney Verba. The analysis will be examined through three components. Initially, the cultural framework, cultural dynamics, and cultural policies of Israel and Palestine. The behavioural disposition of each nation about the political system. The reactions and conduct of Israel and Palestine towards foreign and domestic policies within the political framework. This research evaluates that the dispute arose from the divergence of interests and beliefs inherent in the political cultures of the two parties. Israel asserts that the territory formerly inhabited by the Palestinian population is their rightful claim, bestowed upon them by their trust. The mission of the Zionist movement is also a contributing factor to this enduring conflict. Subsequently, these ideals conflict with the political culture of the Palestinian nation within the context of the Arab country. This study asserts that political culture theory adequately elucidates the origins of the Israeli-Palestinian issues, particularly concerning the identities and behaviours of both nations in relation to their political goals. Nevertheless, this theory is less capable of elucidating the degree to which the political culture of the Palestinian nation underpins this enduring struggle.

2.5 Long-Short Term Memory (LSTM)

Earlier RNN architectures have predominantly been replaced by LSTM. Since its inception, other modifications of this basic design have emerged. Nonetheless, it remained extensively utilized, and we are unaware of any gated-RNN design that surpasses LSTM in a general context while maintaining comparable simplicity and efficiency. The work proposed a modified LSTM-like architecture [12]. The architecture remains uncomplicated and demonstrates superior performance on the tasks evaluated. They presented a novel RNN performance benchmark utilizing handwritten digits, emphasizing numerous critical network properties.

To obtain a comprehensive picture of the State-of-the-Art, we examined prior research pertinent to sentiment analysis on social media, particularly focusing on X. The study by [13] employed the LSTM approach to examine public attitude over the rise in fuel costs. Data collection was conducted through the scraping of information from the X social media platform. This data extraction employs the snscrape module. Snscrape is a tool designed for extracting data from Social Networking Services (SNS). This scraper can extract data such as user profiles, hashtags, or search results and return relevant information, including pertinent posts. The Long Short-Term Memory model had the best accuracy evaluation values of 90% in the first and third trials. The predictive pattern in five experiments exhibited a greater prevalence of positive emotion compared to negative sentiment.

2.6 Scrapping

Scraping is the process of obtaining data from web pages using scripts or software for data collection and storage. This process can be executed manually or automatically. Web scraping is conducted to extract specific information from a designated section of a website. Web scraping selectively retrieves specified data from the target site as required, in contrast to web crawling, which navigates all relevant sites. The results of web scraping may be used by other systems or subjected to further analysis [13].

Web scraping or web crawling refers to the process of automatically extracting data from websites using software. It is a process that is particularly important in fields such as Business Intelligence in the modern age. Web scraping is a technology that allows us to extract structured data from text such as HTML. Web scraping is extremely useful in situations where data isn't provided in machine-readable formats such as JSON or XML [14].

3. RESEARCH METHODOLOGY

Research in data mining has a Cross-Industry Standard Process for Data Mining (CRISP- DM) standard that is generally used for structured data [15], [16]. Therefore, for sentiment analysis that uses unstructured data, the preprocessing stage has a very important role as depicted in Figure 1. This is further explained in detail in the following explanation.

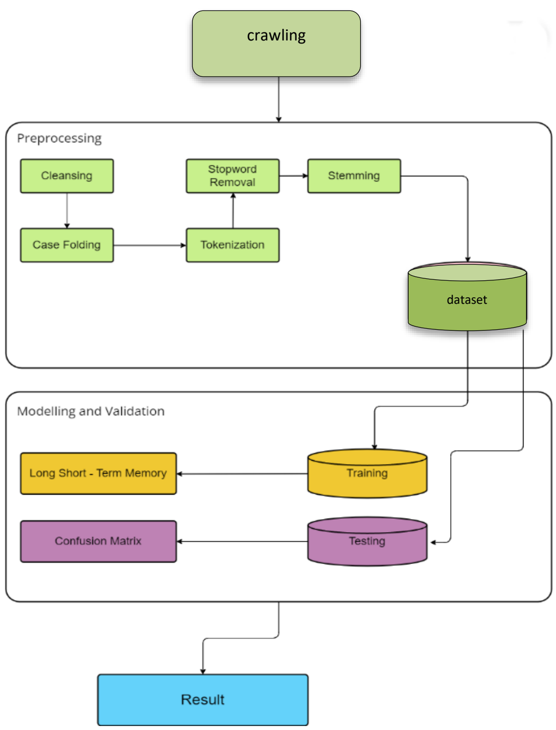

Figure 1. CRISP-DM Diagram for Sentiment Analysis

3.1 Research Objective

This research centres on the flow of thoughts and digital activity expressed on the X network. The choice of Twitter as a inquire about subject is persuaded by its critical impact in changing open supposition and empowering talk on worldwide things, such as the Israeli-Palestinian struggle. X, a microblogging organizes characterised by brief messages, has developed as the essential scene for people to verbalize considerations, spread data, and respond to noteworthy worldwide occasions. X's worldwide reach encourages the quick dispersal of talks on worldwide points, such as the Israeli-Palestinian strife, locks in an assorted cluster of members, from standard citizens to conspicuous open identities. This inquire about question was chosen to investigate the public's viewpoints, demeanours, and responses to the Israeli-Palestinian struggle as communicated in various tweets on X. This continuous strife, which impacts various countries counting Indonesia, has inspired a run of perspectives and sentiments among social media clients. This think about utilizes assumption investigation of circulating tweets to find out the public's expression of conclusions with respect to the struggle, categorising them as transcendently positive, negative, or impartial, and looking at the potential effect of these reactions on open discernment in countries not specifically locked in within the strife. Moreover, concentrating on X facilitates real-time conversational dynamics, wherein viewpoints can swiftly transform over time. This platform's utilisation in this study aims to yield profound insights into the impact of social media on public opinion, particularly during significant political or humanitarian occurrences like the Israeli-Palestinian conflict. This think about looks for to explore the application of assumption investigation methods, counting LSTM, in capturing and classifying opinion inside sincerely charged and modern conversational situations.

3.2 Collecting Data

This study employs data collection methods by extracting tweets from X users concerning the Israeli-Palestinian conflict. Data collecting is conducted with specific keywords like "Israel" and "Palestine," enabling academics to pinpoint tweets pertinent to the subject. The crawling technique is utilised on the X network for the automated and efficient collection of data. This data collection primarily concentrates on tweets from X users in Indonesia, acknowledging the significant influence of public opinion in the nation on the comprehension and analysis of worldwide events. The slithering method empowers the quick and broad collection of information, encouraging analysts in securing an agent test of tweets with respect to the Indonesian populace's reactions and conclusions on the progressing Israel-Palestine strife. The information gotten from these tweets will serve as a profitable asset for opinion examination, focusing on the classification of estimations as positive, negative, or unbiased. The data primarily comprises comments or tweets sourced from the X platform. The data collection process employs crawling techniques to gather information from the comment section on X pertaining to the Israeli-Palestinian Conflict.

3.2.1 Raw Data

The raw data as depicted in Table 1 is obtained directly from the crawling results, utilizing the keyword search “Israel–Palestine Conflict” on X. The data collection focused on tweets written in the Indonesian language to capture regional sentiment and public discourse. To ensure relevance, only tweets posted between May and July were included, aligning with major developments in the conflict during that period.

3.2.2 Data Training

Training data is raw data that has undergone preparation to facilitate subsequent analysis. This study's preparing information comprises a compilation of tweets relating to the Israeli-Palestinian struggle on X, which have experienced different stages of cleansing and change to prepared them for estimation investigation using LSTM calculation. The preprocessing stage is vital as crude tweets frequently incorporate various unessential components that ruin the model's capacity to observe assumption designs, like unnecessary words, images, or orthographic blunders. The preprocessing phases commonly employed in text processing, particularly for sentiment analysis, encompass the elimination of stop words: Stop words are terms frequently seen in text that lack significant meaning for analysis, such as "dan," "atau," "dimana," "di," and so on. The elimination of stop words diminishes data volume and mitigates noise in the analytical process.

Table 1. Example of Raw Data Retrieved

No	Username	Tweet
1	adisatya	Indonesia bela Uighur mati2an, sampai marak anti-China, bahkan selalu jd pemnatik rasisme ke etnis Tionghoa disini. Lalu konflik Israel Palestina Indonesia bela Palestina Uighur bela Israel. Pertanyaannya: siapakah yg bodoh di sini?
2	pikiran_rakyat	EErdogan yg munafik, mendua memanfaatkan konflik palestina Dng penyaluran pupa minyak dan gas dari azerbaizan ke israel, parah nih negara
3	Ammooee	Tapi, apa memang hanya karena itu. Konflik antara Hamas dan Israel sudah berlangsung selama berpuluh- puluh tahun. Hamas berdiri karena Palestina, Gaza di jajah Oleh Kaum Zionis. Penyerangan yang di lakukan Hamas adalah bentuk perlawanan.
4	_MasWis	Yg terjadi di Gaza sekarang bukan lagi konflik, tapi perluasan kompleks perumahan Israel dg pengusiran n pembantaian warga Palestina alias perampokan hak hidup. Dg kata lain: Perebutan Wilayaha udh terjadi sejak 75 tahun lalu, bukan tanggal 7 oktober.
5	wonyucil	capek banget ngehadepin internalized mentality kalau palestina israel = konflik agama. BUKAN. ini PEMBANTAIAN. ETHNIC CLEANSING. A literal second HOLOCAUST. coba deh sesekali hatinya dipake jangan sok edgy.

3.2.3 Preprocessing

During the preparation phase, text preprocessing techniques are utilized to refine the sentiment dataset pertaining to the Israeli-Palestina conflict on X. This procedure seeks to enhance dataset quality, eliminate superfluous components, and guarantee the precision of the classification model. Each text in the dataset will undergo a cleansing process in which ambiguous or non-contributory terms will be eliminated. For instance, generic or filler terms devoid of substantial meaning in the context of sentiment analysis will be eliminated. Furthermore, symbols and emoticons within sentences will be eliminated to guarantee that the dataset is pristine and prepared for classification utilizing the LSTM as depicted in Figure 2.

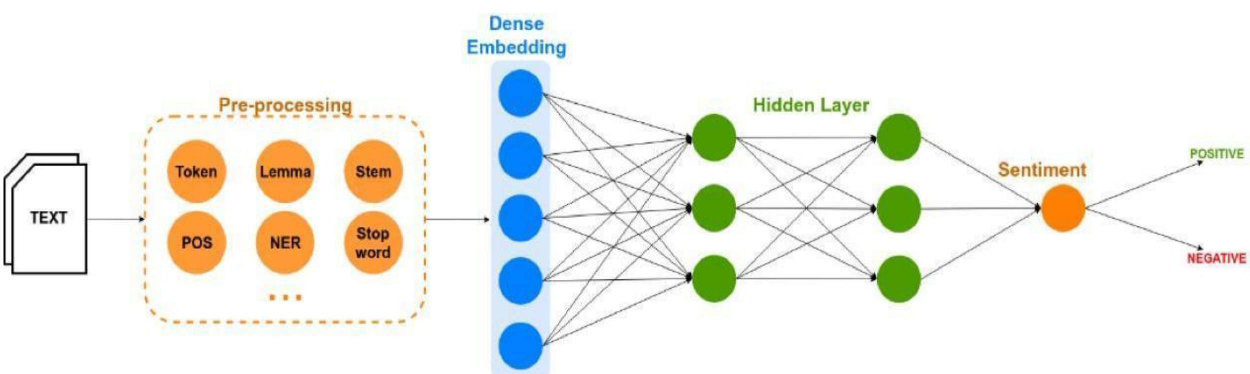

Figure 2. LSTM Model Architecture

4. RESULTS AND DISCUSSIONS

4.1 Data Collecting

This research utilises data acquired via a crawling procedure from X concerning the Israeli-Palestinian conflict. The crawling was conducted via the tweet-harvest package in Python. Data gathering was conducted via obtaining the X Auth Token. To access the X API and get an authentication token, log in with an existing X account, right-click, select 'inspect', locate the application menu, retrieve your authentication token, and input it into the pre-existing Python code. A tweet search is conducted using keywords including "Israel," "Palestine," and "Israeli-Palestinian Conflict." The quantity of tweets collected can be modified as required, and in this study, approximately 500 data points were gathered for each keyword. This procedure generated a total of 500 tweets in the Indonesian language. The gathered material encompasses popular reactions to the war, political opinions, and overall perceptions of the current situation. This technique generated 1,700 Indonesian-language tweets pertinent to the Israeli-Palestinian conflict within a certain timeframe. The gathered information encompasses diverse viewpoints of the dispute, including public reactions, political opinions, and overall perceptions of the current situation. The raw data comprises elements including created_at, username, tweet text, retweet count, reply count, tweet URL, and others. Prior to additional analysis, superfluous columns are eliminated, retaining only created_at, username, and tweet text. The gathered tweets exhibit a spectrum of responses to the Israeli-Palestinian conflict, encompassing support for one faction, appeals for peace, and denunciations of activities deemed to exacerbate the situation. Certain tweets also encompass information regarding recent occurrences, perspectives of political figures, and appeals for humanitarian assistance. The subsequent results pertain to data retrieval using the keyword "Israeli-Palestinian Conflict". The data distribution throughout the collection period was relatively uniform; however, there was a notable surge in tweet volume during and following significant tournaments or major updates from the game developer.

4.2 Preprocessing

Upon securing the crude information from X, a preprocessing strategy is executed to dispose of clamour from the tweet information, rendering it appropriate for handling and examination. The preprocessing stages include cleansing, case collapsing, stop word disposal, standardization, tokenisation, and eventually stemming [17], [18]. The preprocessing stage is pivotal for improving the quality of content information earlier to investigation with the LSTM. Dispensing with unessential parts, normalizing phrasing, disposing of stop words, and actualizing stemming upgrades the data's cleanliness and consistency. These procedures enhance the model's ability to comprehend and evaluate sentiment with greater precision, hence rendering the analysis results more dependable.

4.3 Data Labelling

Data annotation is the next stage. Data labelling in this work is done using a lexicon-based sentiment analysis tool. This process calls for looking at a pre-existing text file's lexicon of positive and negative attitudes. The lexicon calculates the sentiment score of every tweet about the Israeli-Palestinian conflict gathered from X. Using the sentiment calculation tool, every tweet is assessed based on the existence of keywords in the positive and negative lexicons. Tweets with a positive score are labelled 'Positive', those with a zero score are labelled 'Neutral', and those with a negative score are labelled 'Negative'. Each tweet's sentiment label is determined by the results of this calculation.

After labelling all tweets, a sentiment distribution study was conducted to determine the percentage of positive, neutral, and negative tweets. The labelled data's sentiment distribution showed that negative sentiment was most common, followed by positive sentiment, while neutral sentiment was minimal. This distribution shows a difference in the amount of data between positive emotion and alternative feelings. This can help to clarify the public's view on the Israeli-Palestinian conflict. Data annotation employing a lexicon-based approach shown in Figure 3 produced the following sentiment results.

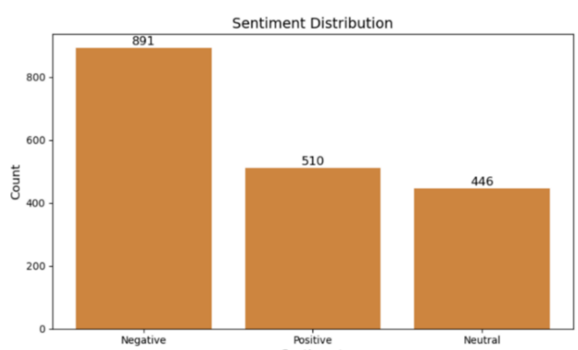

Figure 3. Data Labelling Result

To address the evident class imbalance observed in the original sentiment distribution as depicted in Figure 3, this study employed random oversampling as a data augmentation technique. By duplicating samples from the minority classes (positive and neutral), the dataset was balanced to ensure equal representation across all sentiment categories. This approach aimed to prevent the model from being biased toward the majority class (negative), thereby improving fairness in prediction. Oversampling also helped enhance the generalization capability of the LSTM model by allowing it to better learn minority class patterns

4.4 LSTM Implementation

This work uses LSTM approach to train the model utilizing the previously processed crawling dataset. The objective is to enhance the model's efficacy in aspect classification and sentiment classification, along with predicting aspects and sentiments of a tweet [19], [20]. The LSTM model is employed to forecast elements and sentiments of the gathered tweets. The implementation procedure comprises many key steps: integrating the crawling dataset, data preprocessing, constructing the LSTM model, training the model, evaluating the model, and analysing the results.

After merging the above-mentioned crawl data sets, the following phase involves generating a Word Cloud visualization of the aggregated tweet text. Word Cloud as depicted in Figure 4 offers a visual representation of the most prevalent words within the examined X corpus. This Word Cloud aids in recognizing prevalent themes or subjects frequently addressed in tweets concerning the Israeli-Palestinian conflict. Words displayed in a larger font size in the visualization signify a greater frequency of occurrence in the examined tweet content.

Figure 4. Word Cloud Result

In the subsequent phase, a statistical graph was generated to analyse word frequency, thereby supporting a more accurate and effective preprocessing workflow. The process began by reading the merged dataset and applying essential preprocessing operations. Irrelevant columns were removed to streamline the analysis, retaining only the relevant fields: full_text, username, and created_at. Following the sentiment labelling process, exploratory data analysis and visualization were conducted to derive meaningful insights from the pre-processed data. The resulting sentiment distribution revealed a dominance of negative sentiment, followed by positive and neutral sentiments. These findings reflect the prevailing public perspectives on the Israel-Palestine conflict as expressed on X.

The classification report indicates that the sentiment classification model demonstrates strong overall performance. The model was trained and evaluated using an 80:20 train-test split, ensuring a fair assessment of its generalization capabilities. The model achieved an overall accuracy of 91%, correctly classifying 91% of all samples in the test dataset. For the Negative class, the model obtained a precision of 94%, a recall of 91%, and an F1-Score of 93%, based on 93 instances. The Neutral class showed a precision of 82%, a recall of 91%, and an F1-Score of 86%, evaluated over 79 samples. Meanwhile, the Positive class achieved a precision of 96%, a recall of 90%, and an F1-Score of 92%, across 96 samples. The macro-average and weighted-average scores for precision, recall, and F1-Score were all 91%, highlighting consistent model performance across all categories. Although the Neutral class had slightly lower precision than the other classes, the model still performed reliably and effectively in classifying emotional sentiment overall (see Table 2).

Table 2. Table Classification

Label	Precision	Recall	F1-Score	Support
Negative	0.94	0.91	0.93	93
Neutral	0.82	0.91	0.86	76
Positive	0.96	0.90	0.92	96
Accuracy	-		0.91	268
Macro Avg	0.91	0.91	0.91	268
Weighted Avg	0.91	0.91	0.91	268

4.5 Evaluation Model

The confusion matrix in the Figure 5 demonstrates the classification performance of the LSTM model across three sentiment categories: Negative, Neutral, and Positive. Out of the 93 actual Negative instances, the model correctly classified 85 tweets, misclassifying 7 as Neutral and 1 as Positive. For Neutral sentiments, 72 out of 79 were accurately predicted, with minor misclassifications of 4 as Negative and 3 as Positive. The Positive class shows strong performance as well, with 86 out of 96 correctly classified, 9 misclassified as Neutral, and 1 as Negative. These results reflect high precision and recall for each class, especially for Negative and Positive sentiments. The model demonstrates balanced performance without heavy bias toward any class. Misclassifications were most common between adjacent sentiment categories, such as Neutral being confused with Positive or Negative. This is expected due to the often-subtle semantic differences between these classes in social media language. The overall distribution of predictions indicates that the model is effective at distinguishing clear sentiment signals. However, the few misclassifications suggest that incorporating context-aware mechanisms like attention layers or larger training data may further improve accuracy.

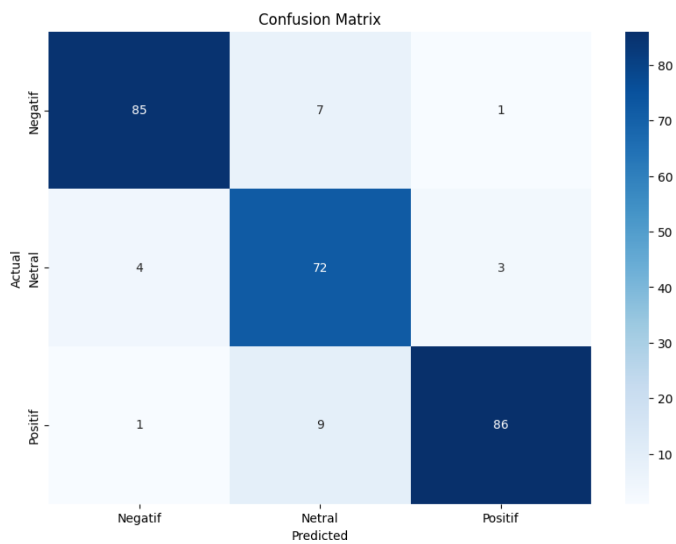

Figure 5. Confusion Matrix

The Receiver Operating Characteristic - Area Under Curve (ROC-AUC) is a performance metric used to evaluate the ability of a classification model to distinguish between classes. It calculates the area under the ROC curve, which plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. A higher AUC indicates better model performance. ROC-AUC is computed using Equation (1).

$$\text{ROC-AUC}_{\text{macro}} = \frac{1}{\binom{C}{2}} \sum_{i<j} \text{AUC}\,(i,j) \quad (1)$$

Where:

$C = 3$ classes: Negative, Neutral, Positive

$\binom{C}{2} = 3$ total unique class pairs

Class Pairs:

a. Negative vs Neutral
b. Negative vs Positive
c. Neutral vs Positive

- Binary Class Pairs and their Confusion Matrices

  - *Negative vs Neutral*: TP = 85, FP = 4, FN = 7, TN = 72

    $\text{TPR}_{\text{Neg}} = \frac{TP}{TP+FN} = \frac{85}{85+7} = 0.9247$; $\text{FPR}_{\text{Neg}} = \frac{FP}{FP+TN} = \frac{4}{4+72} = 0$

    AUC (Neg, Neut) ≈ 0.936 (based on trapezoidal integration using TPR/FPR)

  - *Negative vs Positive*: TP = 85, FP = 1; FN = 1, TN = 86

    $\text{TPR}_{\text{Neg}} = \frac{85}{85+1} = 0.9884, \quad \text{FPR}_{\text{Neg}} = \frac{1}{1+86} = 0.0114$

    AUC (Neg, Pos) ≈ 0.989

  - *Neutral vs Positive*: TP = 72, FP = 9; FN = 3, TN = 86

    $\text{TPR}_{\text{Neut}} = \frac{72}{72+3} = 0.96, \quad \text{FPR}_{\text{Neut}} = \frac{9}{9+86} = 0.0947$

    AUC (Neut, Pos) ≈ 0.932

- Compute ROC-AUC

  - $\text{ROC-AUC} = \frac{1}{3}(0.936 + 0.989 + 0.932) = \frac{2.857}{3} = 0.952$

Referring to the ROC-UAC score, that is 0.952, it means that the model has very strong discriminative ability in distinguishing between the classes being evaluated.

5. CONCLUSION

This work built a sentiment analysis model utilizing LSTM method to categorize tweets into three sentiment classifications: negative, neutral, and positive. Significant measures have been implemented, including data preparation, partitioning data into training and testing sets, training the model with regularization to mitigate overfitting, and assessing the model through accuracy, precision, recall, and F1-Score metrics. The assessment findings indicate that the constructed sentiment analysis model demonstrates strong performance, with an overall accuracy of 91%. The high precision, recall, and F1-Score values for each sentiment class indicate that this model can classify tweets with an adequate level of accuracy and reliability. The incorporation of embedding layers, SpatialDropout1D, LSTM with regularization, and early stopping has demonstrated efficacy in enhancing model performance and mitigating overfitting.

Case studies of various recent tweets demonstrate that this model can reliably predict sentiment, indicating its reliability for sentiment analysis applications including Indonesian text data. Following the findings of the conducted research and evaluation, various recommendations can be proposed for the advancement of this sentiment analysis application. Investigation of Alternative Models: While LSTM has demonstrated efficacy, the examination of other models, such as Transformer or advanced BERT, may yield substantial enhancements in performance, particularly in comprehending intricate textual contexts. Hyperparameter Optimization: Additional study may be conducted to identify a more ideal combination of hyperparameters utilizing approaches such as grid search or random search to enhance model performance. Implementation of Multi-task Learning: Integrating sentiment analysis with additional tasks, such as topic modelling or aspect-based sentiment analysis, can enhance contextual richness and augment the precision of sentiment prediction.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for the careful review of our manuscript.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Mohammad Taleb Noori: Conceptualization, Writing – Original Draft Preparation;
Muhammad Alif Rahman: Data Curation, Methodology, Writing – Review & Editing;
Agus Purnomo: Project Administration, Supervision;
Aripin: Methodology, Validation, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests was disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

REFERENCES

[1] "The Military Lessons of the Gaza War of May 2021". Accessed: May 09, 2025. [Online]. Available: https://trendsresearch.org/insight/the-military-lessons-of-the-gaza-war-of-may-2021/

[2] Md. E. Sharkar, M. J. Hosen, Md. Abdullah, S. Islam, S. Rana, and N. Sultana, "Sentiment Analysis of Israel-Palestine Conflict Comments Using Sentiment Intensity Analyzer and TextBlob," in 2024 *15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jun. 2024, pp. 1–7. doi: 10.1109/ICCCNT61001.2024.10724497.

[3] J. Choi and J. and Lee, "Enthusiasm" toward the other side matters: Emotion and willingness to express disagreement in social media political conversation," *The Social Science Journal*, vol. 0, no. 0, pp. 1–17. doi: 10.1080/03623319.2021.1949548.

[4] P. Gaur, S. Vashistha, and P. Jha, "Twitter Sentiment Analysis Using Naive Bayes-Based Machine Learning Technique", In *Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022, Eds., Singapore: Springer Nature*, 2023, pp. 367–376. doi: 10.1007/978-981-19-5443-6_27.

[5] M. H. Al-Areef, and K. Saputra S, "Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM," *J. SAINTIKOM J. Sains Manaj. Inform. Dan Komput.*, vol. 22, no. 2, pp. 270, Aug. 2023. doi: 10.53513/jis.v22i2.8680.

[6] X. He, "Sentiment Classification of Social Media User Comments Using SVM Models," in 2024 *5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, Mar. 2024, pp. 1755–1759. doi: 10.1109/AINIT61980.2024.10581547.

[7] S. Kumari, and M. P. Singh, "Machine Learning-Based Election Results Prediction Using Twitter Activity," *SN COMPUT. SCI.*, vol. 5, no. 7, p. 819, Aug. 2024. doi: 10.1007/s42979-024-03180-x.

[8] V. Adarsh, V. Mohla, R. K. Mahto, and R. P, "Sentiment Classification of Product Reviews using Machine Learning," in *2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*, Apr. 2024, pp. 1–6. doi: 10.1109/RAEEUCCI61380.2024.10547858.

[9] K. M. K. Kumar, K. Ullas, V. S. Reddy, M. S. Snehitha, M. E. Malkhed, and K. D. Kumar, "Detection of Bullying Text: A Multi-faceted Approach using Machine Learning and Natural Language Processing," in *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, Sep. 2024, pp. 180–186. doi: 10.1109/ICOSEC61587.2024.10722479.

[10] K. Machova, M. Mikula, X. Gao, and M. Mach, "Lexicon-based Sentiment Analysis Using the Particle Swarm Optimization," *Electronics*, vol. 9, no. 8, p. 1317, Aug. 2020. doi: 10.3390/electronics9081317.

[11] O. Manullang, C. Prianto, and N. H. Harani, "ANALISIS SENTIMEN UNTUK MEMPREDIKSI HASIL CALON PEMILU PRESIDEN MENGGUNAKAN LEXICON BASED DAN RANDOM FOREST," *J. Ilm. Inform.*, vol. 11, no. 02, pp. 159–169, Sep. 2023. doi: 10.33884/jif.v11i02.7987.

[12] A. Pulver, and S. Lyu, "LSTM with working memory," in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA: IEEE, May 2017, pp. 845–851. doi: 10.1109/IJCNN.2017.7965940.

[13] M. Khder, "Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application," *Int. J. Adv. Soft Comput. Its Appl.*, vol. 13, no. 3, pp. 145–168, Dec. 2021. doi: 10.15849/IJASCA.211128.11.

[14] R. Arboretti *et al.*, "An Integrated Framework for Automated Web Scraping and Sentiment Analysis of Product Reviews," *Journal of Machine Intelligence and Data Science*, vol. 5, 2024. doi: 10.11159/jmids.2024.006.

[15] J. S. Saltz, "CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps," in *2021 IEEE International Conference on Big Data (Big Data)*, Dec. 2021, pp. 2337–2344. doi: 10.1109/BigData52589.2021.9671634.

[16] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, "The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks," *SBC Reviews on Computer Science*, vol. 4, no. 1, Art. no. 1, Oct. 2024. doi: 10.5753/reviews.2024.3757.

[17] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organizational Research Methods*, vol. 25, no. 1, pp. 114–146, Jan. 2022. doi: 10.1177/1094428120971683.

[18] S. Feuerriegel et al., "Using natural language processing to analyse text data in behavioural science," *Nat Rev Psychol*, vol. 4, no. 2, pp. 96–111, Feb. 2025. doi: 10.1038/s44159-024-00392-z.

[19] N. Mughal, G. Mujtaba, S. Shaikh, A. Kumar, and S. M. Daudpota, "Comparative Analysis of Deep Natural Networks and Large Language Models for Aspect-Based Sentiment Analysis," *IEEE Access*, vol. 12, pp. 60943–60959, 2024. doi: 10.1109/ACCESS.2024.3386969.

[20] A. K. Zarandi and S. Mirzaei, "A survey of aspect-based sentiment analysis classification with a focus on graph neural network methods," *Multimed Tools Appl*, vol. 83, no. 19, pp. 56619–56695, Jun. 2024. doi: 10.1007/s11042-023-17701-y.

BIOGRAPHIES OF AUTHORS

Mohammad Taleb Noori holds a master's degree from Maulana Azad College, India. He brings over a decade of experience working with governmental, national, and international organizations, including serving as Head of Information and Monitoring & Evaluation (M&E) for Polio Eradication initiatives. His commitment to improving healthcare outcomes is evident through his professional roles and research interests. He is also a polyglot, fluent in five languages. He can be contacted via email at: noori2016@gmail.com

Muhammad Alif Rahman is a research assistant at the Department of Informatics Engineering in Dian Nuswantoro University. His research focuses on recommendation Artificial Intelligent Technology and Neural Language Processing for text processing. He can be contacted at email: 111202013119@mhs.dinus.ac.id.

	**Agus Purnomo** is a Student Department of Electrical Engineering in Dian Nuswantoro University. His research focuses on IoT and robotics, as well as AI for data acquisition processing. He can be contacted at email: 511202101105@mhs.dinus.ac.id.
	**Aripin** is a researcher and lecturer at Dian Nuswantoro University specializing in computer vision, natural language processing, and audiovisual synthesis. His work includes modelling Indonesian visemes, emotion classification using machine learning, and educational studies on language and identity. He has contributed to both technical and pedagogical fields, with publications in international journals and conferences. His current research focuses on pronunciation learning and sentiment analysis using machine learning more focus on deep learning. He can be contacted via email at: aripin@dsn.dinus.ac.id.