Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

Sentiment Analysis of the Palestine-Israel Crisis on Social Media using Convolutional Neural Network

Dwina Sarah Delva, Kemas Muslim Lhaksmana*

School of Computing and Informatics, Telkom University, Bandung, Indonesia
Email: [1]dwinasarah@student.telkomuniversity.ac.id , [2,*]kemasmuslim@telkomuniversity.ac.id
Corresponding Author Email: kemasmuslim@telkomuniversity.ac.id
Submitted: **04/06/2024**; Accepted: **29/06/2024**; Published**: 29/06/2024**

Abstract−The issue of Palestine and Israel is currently ongoing and is becoming increasingly heated. The struggle for territory and power is the reason for this conflict, thus attracting the world's attention, especially that of the the Indonesians. People actively express various views in the form of opinions via social media platforms such as Twitter. Communities are competing to make posts and tweet as a form of support for either party. Various tweets appear, making it difficult to draw conclusions through manual analysis. Therefore, this study employs automatic sentiment analysis to enable mass data processing. The sentiment analysis process uses a Deep Learning algorithm, specifically the Convolutional Neural Network (CNN). Convolutional Neural Network (CNN) is a Neural Network algorithms designed for visual shape processing and developed for classification tasks. Based on the explanation provided, it is expected to provide high accuracy and achieve the designed goals. This sentiment analysis research needs to be conducted because to understand and classify various forms of public sentiment toward the issue of Palestine and Israel, thereby providing an overview of the fluctuations in public sentiment concerning this matter in Indonesia. Outcomes of this investigation found the highest performance was achieved by the Convolutional Neural Network (Oversampling) algorithm with accuracy of 93.85%, precision of 93.76%, recall of 93.95%, and F1-score of 93.86%.

Keywords: Sentiment Analysis; Palestine; Israel; CNN; RNN

1. INTRODUCTION

Currently, the world is being shocked by the Palestinian and Israeli conflict, which is increasingly heating up and has become a global topic of conversation. For almost 75 years, this conflict has persisted and become an international controversy that has attracted world attention [1]. Differences in views, beliefs, and values have heightened societal tensions in response to this incident, resulting in thousands of lives being lost and the conflict continuing unabated.

The rise of pros and cons in world opinion on social media has prompted Indonesian society to actively participate in voicing their support for Palestine due to human rights (HAM) concerns, while others defend Israel, believing Hamas to be the cause of the conflict [2]. Various social media platforms have become public forums for conveying opinions regarding the issue of Palestine and Israel, with Twitter being a prominent example. Indonesia, the country with the fifth most active Twitter users, had 202.6 million users as of January 2021. Twitter allows users to create status messages called "tweets" that can be accessed by other users, usually containing opinions on various topics. These tweets, limited to 140 characters, make Twitter a platform that collects a vast amount of opinion data from people worldwide [3]. The Palestine-Israel conflict, has elicited strong and divergent opinions worldwide. This underscores the urgency of the research, given the escalating tensions and polarized discussions on social media, particularly on Twitter, where Indonesians are actively expressing their views. Communities compete to make their voices hear by tweeting as a form of support for either side. This tweet data was processed and analyzed to determine user sentiment regarding the Palestinian and Israeli conflict.

Sentiment Analysis, a subset of Natural Language Processing (NLP), Serves the purpose of analyzing and comprehending individuals' perspectives and emotions concerning a particular subject. This process evaluates opinions expressed in various sources such as comments, posts, reviews, and tweets [2]. Sentiment analysis is a research area that is currently quite popular because it is considered capable of providing warnings and insights to the public [3]. Researchers use sentiment analysis to achieve specific goals, such as understanding and classifying various forms of public sentiment on Palestinian and Israeli issues.

In optimizing the process of sentiment analysis, the right algorithm is needed to streamline the system. An algorithm is a set of instructions that provides a detailed explanation of the necessary actions to achieve a desired goal. Algorithms are applied in various domains, such as mathematics, computer science, data science, and engineering, to solve different types of problems, including data processing, optimization, and retrieval. Numerous sentiment analysis studies have been conducted using various algorithms, such as Machine Learning and Deep Learning. For instance, Wang Yue and Lei Li conducted research using a machine learning model and a support vector machine (SVM), Naïve Bayes, and Logistic Regression. Their research achieved accuracy levels ranging from 75% to 85% [4]. In contrast, research using Hybrid Deep Learning achieved an accuracy value of 91.48% [5]. A different study also used a Twitter dataset with product reviews and the Word2Vec, GloVe, and FastText algorithms. The study discovered that CBOW, at 86.3%, and skip-gram, at 91.6%, had the highest performance values [6]. Finally, machine learning and deep learning techniques like CNN, Naïve Bayesian, and RNN were compared using Word2Vec features. The results showed that the RNN method, with the expansion of Word2Vec features, achieved the highest accuracy rate of 91.88%. Following this, CNN with Word2Vec feature expansion achieved an accuracy rate of 89.23%, while Naïve Bayesian reached an accuracy of 44% [7].

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

The Convolutional Neural Network (CNN) stands as a neural network algorithms specifically crafted for image object detection. Its architecture encompasses a Convolutional Layer, tasked with the reduction of calculation complexity, followed by a Pooling Layer dedicated to dimensionality reduction and complexity simplification, among other functions. Following this, a Fully Connected Layer assumes responsibility for classification, leveraging previously extracted features. In the realm of sentiment analysis, input data undergo conversion into an embedded vector via the embedding layer. After this conversion, the data are formatted into parallel structures and directed to the fully connected layer for classification, with the objective of expediting feature dimension reduction for swifter computations. Notably, the data is further structured into 15 parallel forms before being directed to the fully connected layer for sentiment analysis [8]. The Recurrent Neural Network (RNN) emerges as another network component proficient in continuous data processing. Research frequently employing the CNN algorithm has spurred innovation across diverse domains, including word translation and sound categorization, among others [9].

This research aims to enhance the precision of sentiment analysis outcomes by examining the accuracy metrics of both Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) algorithms. CNNs are typically used for visual data analysis, whereas RNNs are employed to understand the context and sequence of data. The primary aim of this research is to develop a system capable of automatically analyzing sentiment regarding the Palestinian and Israeli conflict and determining the levels of accuracy, recall, precision, and F1-score of the Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) methods on Indonesian people's sentiments concerning the Palestinian and Israeli issues. This study is delimited to a dataset comprising 22,537 Indonesian tweets sourced from Twitter, which are segregated into positive and negative sentiment categories.

2. RESEARCH METHODOLOGY

2.1 Research Stages

At this stage, we will elucidate the general scheme of the flowchart design intended for system utilization. The forthcoming research flowchart is delineated in Figure 1.

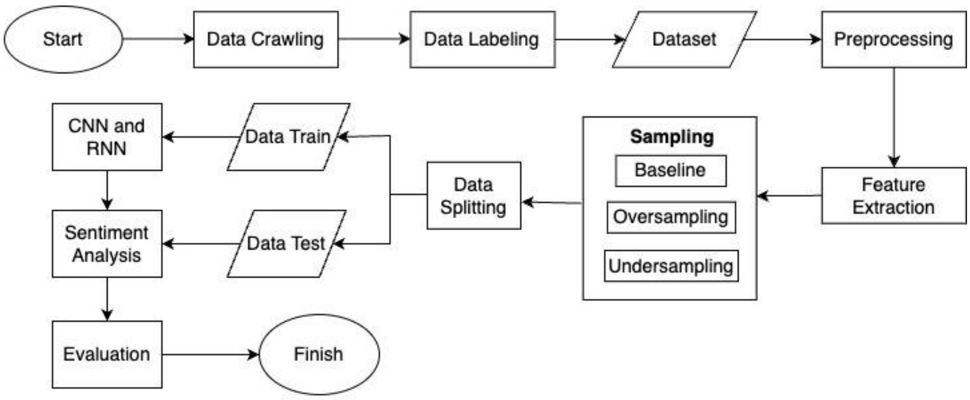

Figure 1. Flowchart Research Stages

As shown in Figure 1, the first stage is to collect data through the Data process. After the data have been collected and aggregated into one, the process of labeling the data is obtained. Once labeled, the dataset is ready for analysis. Following the dataset collection, the next step is Data Preprocessing, which eliminates irrelevant elements from the data. Subsequently, Feature extraction is conducted to extract essential features from sentences used for sentiment classification. Then comes the sampling process, aimed at reducing overfitting and other potential issues. Next is Data Splitting, dividing the dataset into two categories: training data (Train Data) and test data (Test Data). Following this, the CNN and RNN algorithms are implemented, and the accuracy, recall, precision, and F1-score values are computed.

2.2 Data Crawling

Data crawling involves gathering data, which will subsequently form a dataset [6]. The data retrieval process uses the Twitter API to access and collect the required data. The gathered data consists of tweets containing public opinions or comments regarding the Palestinian issue and Israel and are in Indonesian. This research yielded 22,537 raw data entries, obtained from 62 hashtags discussing Palestine and Israel issues.

2.3 Data Labeling

Data Labeling involves assigning labels or categories to text to test the model. Additionally, this procedure aids in classifying the sentiment conveyed in a sentence. The data is divided into three categories: positive, neutral, and negative. The outcomes of the labeling process are described in Table 1.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

Table 1. Data Labeling

Review	Label
Tentara penjajah Israel mencuri uang warga Palestina dari rumah mereka di Gaza. Inilah prestasi terbaik si kera (The Israeli occupation army steals Palestinian money from their homes in Gaza. This is the monkey's best achievement)	Positive
Aceh untuk Palestina (Aceh for Palestine)	Neutral
Kementerian Kesehatan Gaza ini 198 syahid termasuk 58 kanak-kanak dan 35 wanita dan 1300 lain cedera akibat serangan Shitrel (Gaza Ministry of Health: 198 martyrs, including 58 children and 35 women, and 1,300 others injured due to Israeli attacks)	Negative

Based on Table 1, categorization is carried out in two stages, namely automatic and manual. The automatic data labeling stage uses the Python programing language to collect the datasets. This step ensures that the collected text data are accurate and appropriate, thereby reducing errors in the code design of the labeling process [10].

2.4 Dataset

The dataset used in this study was gathered from a Twitter application using Data Crawling techniques. The processed data consists of Indonesian-language tweets containing opinions or views of the Indonesians on the issues of Palestine and Israel. The dataset includes a total of 22,537 tweets, with selected data attributes comprising Tweets, Polarity, and Label.

2.5 Preprocessing Data

Preprocessing is the phase that occurs before the processing stage. The data undergoes processing to eliminate elements that may hinder subsequent analysis. The preprocessing stages in this study are outlined as follow

a. Case Folding

The initial stage in data processing is data cleaning., where errors, inaccuracies, and inconsistencies in the data are corrected or removed. The purpose is to ensure that the data align with the designed model. Data cleaning also involves handling data, eliminating duplicates, and improving data formats or structures to make the data accurate and ready for further stages [10]. In this research, the stages include Remove URL, Remove Username, Remove Symbol, Remove Hashtag, Remove Whitespace, and Remove Punctuation. The outcomes of case folding in this study are shown in Table 2.

Table 2. Data Cleaning

Before	After
#Gaza Anak-anak Gaza melakukan Demonstrasi Kami Ingin Makan" #GazaStarving #FreePalestine https://t.co/wrCAiLErmQ" (Children of Gaza Conducting Demonstration 'We Want to Eat' #GazaStarving #FreePalestine https://t.co/wrCAiLErmQ)	Anak anak Gaza melakukan Demonstrasi Kami Ingin Makan (Children of Gaza Conduct Protest 'We Want to Eat')

As shown in Table 2, we can conclude that the Data Cleaning function effectively removes unimportant elements from each comment.

b. Case Folding

This aims to standardize the representation of the same text written in different cases (upper and lower case). This stage is also performed to reduce issues arising from sentiment analysis, thereby increasing accuracy in text processing. Every letter in a sentence is converted to lowercase (from a to z) [11]. The data of Case Folding in this research are shown in Table 3.

Table 3. Case Folding

Before	After
Anak anak Gaza melakukan Demonstrasi Kami Ingin Makan (Children of Gaza Conduct Protest 'We Want to Eat')	anak anak gaza melakukan demonstrasi kami ingin makan (children of gaza conduct protest we want to eat)

Based on Table 3, we can conclude that every capital letter in the sentence has been changed to lowercase. Thus, it can reduce the risk of problems arising from the sentiment analysis in this study.

c. Tokenization

Tokenization involves the segmentation of text into smaller units. units for each word in a sentence [12]. Tokens are usually in the form of words, characters, or phrases, depending on the desired level of granularity. The

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

goal of this stage is to transform the text into tokens that can be processed by the Neural Network model. The data of tokenization in this research are shown in Table 4.

Table 4. Tokenization

Before	After
Anak anak Gaza melakukan Demonstrasi Kami Ingin Makan	[‘anak’, ‘anak’, ‘gaza’, ‘melakukan’, ‘demonstrasi’, ‘kami’, ‘ingin’, ‘makan’]
(children of gaza conduct protest we want to eat)	[‘children’, ‘of ’, ‘gaza’, ‘conduct’, ‘protest’, ‘we’, ‘want’, ‘to’, ‘eat’]

Based on Table 4, we can conclude that this process changes or separates the words in the comments into smaller units.

d. Stemming

Stemming serves as a procedure for eliminating affixes from a word, aiming to derive its root form. Stemming also simplifies word variations to ensure uniformity in the analysis. The data of stemming in this research are shown in Table 5.

Table 5. Stemming

Before	After
[‘anak’, ‘anak’, ‘gaza’, ‘melakukan’, ‘demonstrasi’, ‘kami’, ‘ingin’, ‘makan’]	[‘anak’, ‘anak’, ‘gaza’, ‘laku’, ‘demonstrasi’, ‘kami’, ‘ingin’, ‘makan’]
([‘children’, ‘of ’, ‘gaza’, ‘conduct’, ‘protest’, ‘we’, ‘want’, ‘to’, ‘eat’])	([‘child’, ‘of ’, ‘gaza’, ‘conduct’, ‘protest’, ‘we’, ‘want’, ‘to’, ‘eat’])

Based on Table 5, we can conclude that this process removes every affix from each word in the comments.

e. Stopword Removal

Stopword Removal is a stage that aims to remove common words such as "and", "which", "or", which do not contribute significantly to the sentiment of the text.

2.6 Feature Extraction TF- IDF

Feature Extraction carried out in this research extracts identifying features from comments, predicting sentiment by identifying text that contains sentiment or opinion, and determining the polarity of sentiment such as positive, negative, and neutral. The outcomes of feature extraction take text as input and produce features that can be extracted in various forms [13]. Term Frequency-Inverse Document Frequency (TF-IDF) is a commonly employed feature extraction method to assess the importance of vocabulary in a document.

2.7 Sampling

Sampling is a process that functions to avoid overfitting, improve or speed up training time in managing datasets, balance minority or majority datasets, and Enhance the model’s capacity for generalization. Sampling is divided into two types as follows:

a. Baseline: The baseline is a model used as a reference for assessing more complex models. This study aims to serve as a comparison and a means of improving the performance of complex models.
b. Oversampling: Oversampling is a method used to tackle data inequality by improving the performance of the minority class.
c. Undersampling: Undersampling is a method used to tackle data inequality by reducing the size of the dominant class.

2.8 Data Splitting

Data Splitting is a crucial stage in dividing the total dataset. The data is divided into two segments: Test Data and Train Data. The objective lies in employing the training data for model training, whereas the test data serves to assess the performance of the trained model. Evaluation is conducted to assess how accurately the model can predict new data [5].

2.9 Convolutional Neural Networks

Convolutional Neural Network (CNN) is a neural network algorithms designed for visual shape processing and developed for classification tasks. CNNs are renowned for their ability to extract hierarchical features from images or visual image data, which makes them particularly effective for applications like image classification, object recognition, and image segmentation [14]. The form of the CNN architecture is depicted in Figure 2.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

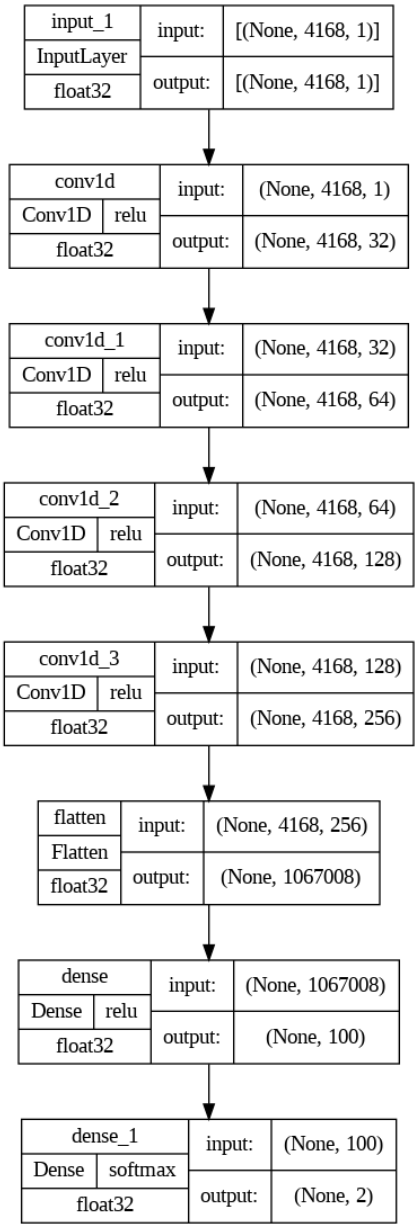

Figure 2. The CNN Architecture

The CNN architecture comprises several components, including the Convolutional Layer, which employs filters (kernels) that move (convolve) across the input to generate a feature map. Each filter learns specific patterns such as edges, textures, and shapes. The Pooling Layer decreases the size of the feature map, maintaining essential information. A frequently used operation is max pooling, which selects the highest value within a subsample region. The Fully Connected Layer is In the final stage, this layer connects all the neurons from the previous layer to each neuron in the subsequent layers, similar to conventional neural networks.

2.10 Recurrent Neural Network

A Recurrent Neural Network (RNN) is a neural network model capable of storing information from previous inputs using sequential memory, which is highly useful for sequential data such as text. The RNN algorithm processes input word by word over certain time steps, thus producing a vector with a fixed size [5]. Previous studies have shown that RNN can replace the pooling layer in sentiment analysis of short texts by repeatedly processing each word to produce a semantic representation in the form of a low-dimensional vector. This setting includes several additional layers, such as batch normalization, return sequences, dropout, and recurrent dropout, whereas the other layers do not return sequences. This architecture also includes a batch normalization layer and two dense layers [7]. Figure 3 illustrates the structure of the RNN architecture employed.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

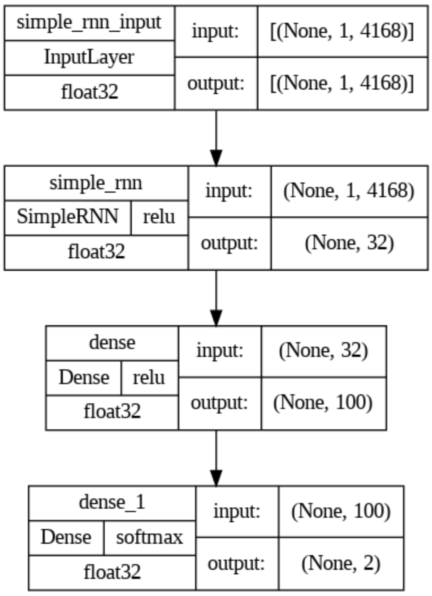

Figure 3. The RNN Architecture

The RNN architecture used consists of the following components. Every unit in the hidden layer of a recurrent neural network takes input from both the input layer and the output of hidden units that came before it. This allows the network to retain a "memory" of its previous state. Weight Sharing is weights in an RNN are shared among all time steps. In other words, the RNN uses the same weights to process each element in the input sequence, making it efficient in recognizing patterns in data sequences.

2.11 Evaluation

Evaluation is a stage aimed at assessing the performance of the processed model. This process provides sentiment predictions from the data it analyzes. The evaluation process also functions as a visualization of the algorithm performance. The matrix consists of

a. True Positive (TP): Accurately, the data expected to be positive and is correctly predicted as positive.
b. False Negative (FN): Accurately the data expected to be negative and is correctly predicted as positive.
c. True Negative (TN): Accurately the data expected to be negative and is correctly predicted as negative.
d. False Positive (FP): Accurately the data expected to be positive and is correctly predicted as negative.

Based on these values, other metrics such as Precision, Accuracy, Recall, and F1-Score can be calculated [15]. The formulations used are explained as follows:

a. Precision

Precision is a measure of how well the model predicts the correct class overall or how often it makes true predictions when making positive predictions [16]. Equation 1 shows the formula for calculating the precision values.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

b. Accuracy

Accuracy measures how accurately the model predicts a document, whether positive or negative. The accuracy value functions as a comparison of data that has been verified as correct to the overall content of the data [17]. Equation 2 shows the formula for calculating the accuracy value.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

c. Recall

Recall measures how completely the model predicts documents [18]. Equation 3 shows the formula for calculating the recall value.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

d. F1-Score

The F1 score integrates both precision and recall [19]. The performance of a classification algorithm is assessed through the F1-Score, calculated as a weighted average of recall and precision values [20]. which measures how well the model predicts documents. Equation 4 shows the formula for calculating the F1 score.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (4)$$

3. RESULT AND DISCUSSION

3.1 Data

The process of data retrieval involves using the Twitter API to access and gather the necessary information from the Twitter platform. The collected data comprise tweets with public opinions or comments on the Palestinian issue and Israel, written in Indonesian. This research resulted in the collection of 22,537 raw data entries, sourced from 62 hashtags related to discussions on the Palestine and Israel issues. Three sentiment categories, positive, neutral, and negative, were applied to the reviews**.** The results shown in Table 6 are based on.

Table 6. Data Distribution

Sentiment	Data
Positive	14255
Neutral	6309
Negative	1971

Based on Table 6, the positive sentiment in this study was 14,255, the neutral sentiment was 6,309, and the negative sentiment was 1971. The labeling process was performed using the Text Blob Python library to obtain the results of the polarity value calculation as a score, which determines the sentiment of the data used.

3.2 Testing Result

This research conducted an evaluation after the modeling process using CNN and RNN algorithms. The data proportion is 75:25. The dataset is divided into two subsets: the Training Set and the Testing Set. The purpose of this division is to assess how well the model generalizes to fresh data that was not used for training. After splitting the data, we have three types of data sampling. First, the Baseline Data refers to the original data and is used for testing the model. Second, Oversampling is used to rectify the majority class's imbalance by improving the performance of the minority class, thereby balancing the data classes. Third, Undersampling is used to address the imbalance of the minority class by reducing the performance of the majority class and balancing the data classes. The evaluation results include accuracy, precision, recall, and F1-score, aiming to achieve optimal results in the classification model. This research resulted in the collection of 22,537 raw data entries, sourced from 62 hashtags related to discussions on the Palestine and Israel issues. The results of this test are presented in Table 7.

Table 7. Test Result

Model	Sampling Type	Accuracy	Precision	Recall	F1-Score
Convolutional Neural Network	Baseline	89.50%	93.57%	94.60%	94.08%
	Oversampling	93.85%	93.76%	93.95%	93.86%
	Undersampling	50.05%	50.05%	100%	66.71%
Recurrent Neural Network	Baseline	88.47%	92.48%	94.63%	93.54%
	Oversampling	92.43%	97.76%	86.86%	91.99%
	Undersampling	48.66%	48.67%	47.02%	47.83%

The test results presented in Table 6 demonstrate that the highest performance was achieved by the Convolutional Neural Network (Oversampling) algorithm, which attained an accuracy of 93.85%, precision of 93.76%, recall of 93.95%, and F1-score of 93.86%. Conversely, the lowest performance was observed with the Recurrent Neural Network (Undersampling) algorithm, which achieved an accuracy of 48.66%, precision of 48.67%, recall of 47.02%, and F1-score of 47.83%. The method employed in this study involves data collection from Twitter using the Twitter API, followed by sentiment analysis using positive, neutral, and negative categories. The dataset is divided into Training Set and Testing Set, with the application of various data sampling techniques such as oversampling and undersampling. Evaluation is conducted on the performance of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), with CNN utilizing oversampling demonstrating the highest performance. In conclusion, CNN with oversampling is more effective in addressing sentiment classification issues in unbalanced datasets, while RNN shows lower performance. This emphasizes the importance of data preprocessing in machine learning and highlights the need for further research in the development of data sampling methods to enhance machine learning model performance.The highest and lowest scores are illustrated in Figures 4 and 5

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

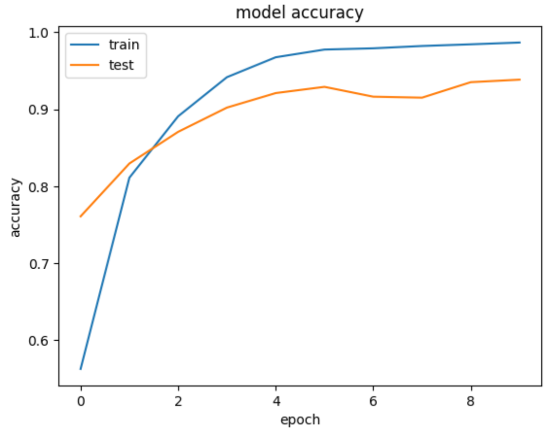

Figure 4. High Accuracy Graph

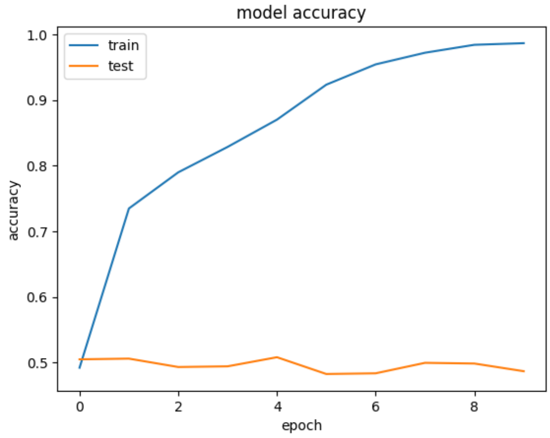

Figure 5. Low Accuracy Graph

Based on Figure 4, we can observe the process of the Convolutional Neural Network (CNN) algorithm with oversampling. The model trains over 10 epochs, and with each epoch, the learning process experiences a significant increase, ultimately outperforming other tests in terms of accuracy. Therefore, the results of this study demonstrate that the CNN algorithm is effective in handling the sentiment classification process, achieving the highest accuracy.

In contrast, Figure 5 shows the process of the Recurrent Neural Network (RNN) algorithm with undersampling, which also trains over 10 epochs. However, in this case, the learning process remains stable across epochs but results in the lowest accuracy value in this study. These findings underscore the substantial disparity in performance between algorithms employing oversampling and undersampling methods. Notably, the CNN with oversampled data markedly surpassed the RNN with undersampled data, indicating that oversampling is an effective approach for enhancing model performance in the context of imbalanced datasets. The high accuracy, precision, recall, and F1-score achieved by CNN highlight its robustness and suitability for the specified task. A closer examination of the CNN's performance reveals that its architecture, which adeptly captures spatial hierarchies through convolutional layers, contributes significantly to its success. This feature is especially beneficial for image recognition tasks, where spatial relationships are critical. Conversely, the RNN, which is designed to capture temporal dependencies, may underperform on datasets where such dependencies are not prominent or are less significant than spatial hierarchies.

The equal values of precision, recall, and F1-score with the accuracy for the CNN with oversampling indicate a balanced performance across all metrics. This balance is essential in applications where both false positives and false negatives carry substantial consequences. For instance, in medical diagnostics, achieving high precision and recall ensures that the model accurately identifies positive cases while minimizing the risk of missing true positive cases. The inferior performance of the RNN with undersampling can be attributed to the limited amount of data available for training, resulting in a less robust model. Undersampling can lead to the loss of critical information necessary for effective model learning. This issue is particularly pronounced for complex models like RNNs, which require a significant amount of data to capture intricate patterns and dependencies.

Figure 4, which displays the highest scores, visually confirms the superior performance of the CNN with oversampling . This table highlights the efficacy of data augmentation methods to enhance the performance of machine learning models. It also underscores the significance of choosing suitable algorithms and data preprocessing methods based on the particulars of the dataset and the issue at hand. In summary, the comparative analysis of CNN and RNN algorithms with different data sampling techniques emphasizes the critical role of data preprocessing in machine

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

learning. The superior performance of the CNN with oversampled data demonstrates that oversampling can significantly improve model accuracy, precision, recall, and F1 score. These results suggest that future research should focus on exploring and optimizing data sampling methods to further enhance the performance of machine learning models, particularly in scenarios involving imbalanced datasets.

3.3 Analysis of the Test Result

According to the outcomes of experiments with the CNN and RNN algorithms, this study applied a data ratio of 75:25, resulting in 75% training data and 25% testing data. The test results for the CNN algorithm showed that the dataset with oversampling achieved the highest performance, with an accuracy of 93.85%, precision of 93.76%, recall of 93.95%, and F1-score of 93.86%. Meanwhile, the RNN algorithm also demonstrated its highest performance on the oversampling dataset, yielding an accuracy of 92.43%, precision of 97.76%, recall of 86.86%, and an F1-score of 91.99%. The performance evaluation results using the Confusion Matrix for the CNN algorithm are presented in Table 8.

Table 8. Confusion Matrix CNN

	Predict Negative	Predict Positive
Actual Negative	3304	220
Actual Positive	213	3311

Based on Table 8, the resulting confusion matrix is:

a. True Positive (TP): 3311 A positive predicted value that is truly positive according to the actual value. There are 3311 cases where the model predicts positive results and is actually positive.
b. True Negative (TN): 3304 A negative predicted value that is truly negative according to the actual value. There are 3304 cases where the model predicts a negative outcome and it is actually negative.
c. False Positive (FP): 220 A positive predicted value that is negative according to the actual value. There are 220 cases where the model predicts positive results but is actually negative.
d. False Negative (FN): 213 A negative predicted value that turns out to be positive according to the actual value. There are 213 cases in which the model predicts negative but is actually positive.

The performance evaluation results using the Confusion Matrix for the RNN algorithm are shown in Table 9.

Table 9. Confusion Matrix RNN

	Predict Negative	**Predict Positive**
Actual Negative	3454	70
Actual Positive	463	3061

Based on Table 9, the resulting confusion matrix is:

a. True Positive (TP): 3061 A positive predicted value that is truly positive according to the actual value. There are 3061 cases where the model predicted positive results and they were indeed positive.
b. True Negative (TN): 3454 A negative predicted value that is truly negative according to the actual value. There are 3454 cases where the model predicted negative results and they were indeed negative.
c. False Positive (FP): 70 A positive predicted value that is negative according to the actual value. There are 70 cases where the model predicted positive results but were actually negative.
d. False Negative (FN): 463 A negative predicted value that turns out to be positive according to the actual value. There are 463 cases where the model predicted negative results but they were actually positive.

4. CONCLUSION

In this research, a sentiment analysis process was conducted using the Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) comparison algorithms to address the Palestine-Israel issue. The findings revealed that the Convolutional Neural Network algorithm, particularly with oversampled data, achieved the highest performance, boasting an accuracy of 93.85%, precision of 93.76%, recall of 93.95%, and F1-score of 93.86%. Following closely, the Recurrent Neural Network algorithm, also employing oversampled data, attained the second-highest scores, with an accuracy of 92.43%, precision of 97.76%, recall of 86.86%, and an F1-score of 91.99%. Consequently, it can be inferred that the Convolutional Neural Network algorithm, coupled with the oversampling process, outperforms other methodologies. The analysis underscores the efficacy of employing advanced neural network models to discern sentiment nuances among complex geopolitical contexts. These findings contribute to the broader discourse on using AI to understand sociopolitical dynamics, particularly in conflict-ridden regions like Palestine-Israel. Furthermore, the meticulous comparison between CNN and RNN algorithms sheds light on the nuanced differences in their performance, offering insights for future research endeavors in sentiment analysis and machine learning.

REFERENCES

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 334-343
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5282

[1] N. Nurlela, “Analisis sentimen terhadap isu konflik Palestina-Israel pada twitter menggunakan algoritma convolutional neural network,” UIN Sunan Gunung Djati Bandung, 2022.
[2] R. B. Sinaga, H. R. Al Fajri, H. Mubarok, A. D. Pangestu, and D. S. Prasvita, “Analisis Sentimen Pengguna Twitter terhadap Konflik antara Palestina dan Israel Menggunakan Metode Na\"\ive Bayesian Classification dan Support Vector Machine,” in Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya, 2021, pp. 166–175.
[3] M. I. Fikri, T. S. Sabrila, and Y. Azhar, “Perbandingan metode na\"\ive bayes dan support vector machine pada analisis sentimen twitter,” SMATIKA Jurnal: STIKI Informatika Jurnal, vol. 10, no. 02, pp. 71–76, 2020.
[4] W. Yue and L. Li, “Sentiment analysis using word2vec-cnn-bilstm classification,” in 2020 seventh international conference on social networks analysis, management and security (SNAMS), 2020, pp. 1–5.
[5] M. F. Avidiansyah and E. B. Setiawan, “The Influence of Sentiment on the Movement of Bank BCA (BBCA) Shares with the CNN-GRU Classification Model and Feature Expansion with GloVe,” in 2023 International Conference on Data Science and Its Applications (ICoDSA), 2023, pp. 105–110.
[6] P. Chen et al., “Editorial Board Members.” [Online]. Available: http://www.springer.com/series/7899
[7] M. H. D. T. A. Bangsa, S. Priyanta, and Y. Suyanto, “Aspect-based sentiment analysis of online marketplace reviews using convolutional neural network,” IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 14, no. 2, pp. 123–134, 2020.
[8] K. Adrian Manurung, “Sentiment Analysis of Tourist Attraction Review from TripAdvisor Using CNN and LSTM,” International Journal on Information and Communication Technology (IJoICT), vol. 9, no. 1, pp. 73–85, Jul. 2023, doi: 10.21108/ijoict.v9i1.756.
[9] M. J. Naufal, D. Richasdy, and M. A. Bijaksana, “Sentiment Analysis of Student Satisfaction on Telkom University Language Center (LaC) Services on Instagram Using the RNN Method,” Journal of Computer System and Informatics (JoSYC), vol. 99, no. 99, 2019.
[10] R. N. Aufa, S. S. Prasetiyowati, and Y. Sibaroni, “The Effect of Feature Weighting on Sentiment Analysis TikTok Application Using The RNN Classification,” Building of Informatics, Technology and Science (BITS), vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3597.
[11] A. Rahmadanisya, E. B. Setiawan, and D. Adytia, “The Influence of Sentiment on Bank Mandiri (BMRI) Stock Movements Using Feature Expansion with Word2vec and Support Vector Machine Classification,” in 2022 10th International Conference on Information and Communication Technology (ICoICT), 2022, pp. 287–292.
[12] G. Kamil and E. B. Setiawan, “Aspect-Level Sentiment Analysis on Social Media Using Gated Recurrent Unit (GRU),” Building of Informatics, Technology and Science (BITS), vol. 4, no. 4, Mar. 2023, doi: 10.47065/bits.v4i4.3105.
[13] S. Thavareesan and S. Mahesan, “Sentiment analysis in Tamil texts: A study on machine learning techniques and feature representation,” in 2019 14th Conference on industrial and information systems (ICIIS), 2019, pp. 320–325.
[14] M. I. Prabha and G. U. Srikanth, “Survey of sentiment analysis using deep learning techniques,” in 2019 1st international conference on innovations in information and communication technology (ICIICT), 2019, pp. 1–9.
[15] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” INSYST: Journal of Intelligent System and Computation, vol. 1, no. 1, pp. 43–49, 2019.
[16] M. Hudha, E. Supriyati, and T. Listyorini, “ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN# MATANAJWAMENANTITERAWAN DENGAN METODE NAÃ VE BAYES CLASSIFIER,” JIKO (Jurnal Informatika dan Komputer), vol. 5, no. 1, pp. 1–6, 2022.
[17] V. A. Permadi, “Analisis sentimen menggunakan algoritma Na\"\ive Bayes terhadap review restoran di Singapura,” Jurnal Buana Informatika, vol. 11, no. 2, pp. 141–151, 2020.
[18] M. Farhan, M. D. Purbolaksono, and W. Astuti, “Sentiment Analysis of Practo App Reviews using KNN and Word2Vec,” Building of Informatics, Technology and Science (BITS), vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3598.
[19] M. R. U. Pulungan, D. E. Ratnawati, and B. Rahayudi, “Analisis Sentimen Ulasan Aplikasi PeduliLindungi dengan Metode Random Forest,” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 6, no. 9, pp. 4378–4385, 2022.
[20] I. W. B. Suryawan, N. W. Utami, and K. Q. Fredlina, “Analisis Sentimen Review Wisatawan Pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine,” Jurnal Informatika Teknologi dan Sains (Jinteks), vol. 5, no. 1, pp. 133–140, 2023.