VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i3.6270

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

THE IMPACT OF WORD EMBEDDING ON CYBERBULLYING DETECTION USING HYBIRD DEEP LEARNING CNN-BILSTM

Moh. Hilman Fariz[1*]; Erwin Budi Setiawan[2]

Informatics[1, 2]
Telkom University, Bandung, Indonesia[1, 2]
http://telkomuniversity.ac.id/[1, 2]
mohhilmanfariz@student.telkomuniversity.ac.id[1*], erwinbudisetiawan@telkomuniversity.ac.id[2]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Cyberbullying can be perpetrated by anyone, whether children or adults, with the primary aim of belittling or attacking specific individuals. Social media platforms like X (formerly Twitter) often serve as the primary medium for cyberbullying, where interactions frequently escalate into retaliatory attacks, intimidation, and insults. In detecting these actions, short tweets are often difficult to understand without context, making specialized approaches like word embedding important. This research uses GloVe feature expansion, utilizing a corpus generated from the IndoNews dataset containing 127,580 entries to enhance vocabulary understanding in tweets that include the use of Indonesian language in both formal and informal forms. This data was then classified using the Hybrid Deep Learning method, which combines Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) with used 30,084 tweets taken from platform X as the dataset. The analysis results show that the application of expansion features using GloVe can improve the performance of the BiLSTM-CNN hybrid model, with the highest accuracy reaching 83.88%, an increase of +3.65% compared to the hybrid model without GloVe. This research successfully detected cyberbullying on platform X, making a significant contribution to efforts to create a safer and more positive social media environment for users.*

Keywords: *BiLSTM, CNN, cyberbullying, GloVe, hybrid.*

Intisari— *Cyberbullying dapat dilakukan oleh siapa saja, termasuk orang dewasa dan anak-anak, dengan tujuan utama merendahkan atau menyerang individu tertentu. Platform media sosial seperti X (sebelumnya Twitter) sering kali menjadi media utama untuk perundungan siber, di mana interaksi sering kali meningkat menjadi serangan timbal balik, intimidasi, dan penghinaan. Dalam mendeteksi tindakan-tindakan ini, tweet pendek sering kali sulit dipahami tanpa konteks, sehingga pendekatan khusus seperti word embedding menjadi penting. Penelitian ini menggunakan perluasan fitur GloVe, memanfaatkan korpus yang dihasilkan dari dataset IndoNews yang berisi 127.580 entri untuk meningkatkan pemahaman kosakata dalam tweet yang mencakup penggunaan bahasa Indonesia dalam bentuk formal dan informal. Data ini kemudian diklasifikasikan menggunakan metode Hybrid Deep Learning, yang menggabungkan Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory. (BiLSTM), dengan menggunakan 30.084 tweet yang diambil dari platform X sebagai dataset. Hasil analisis menunjukkan bahwa penerapan fitur ekspansi menggunakan GloVe dapat meningkatkan kinerja model hybrid BiLSTM-CNN, dengan akurasi tertinggi mencapai 83,88%, meningkat sebesar +3,65% dibandingkan dengan model hybrid tanpa GloVe. Penelitian ini berhasil mendeteksi cyberbullying di platform X, memberikan kontribusi signifikan terhadap upaya menciptakan lingkungan media sosial yang lebih aman dan positif bagi pengguna.*

Kata Kunci: *BiLSTM, CNN, cyberbullying, GloVe, hybrid.*

INTRODUCTION

Internet users in Indonesia, according to an Indonesian Internet Service Providers Association (APJII) study from 2024, reached 221 million people [1]. Social media is one of the platforms frequently used by internet users in the modern era to interact in the virtual world. The increasing number of social media users in the virtual world is also accompanied by freedom of expression. They seem to find a space to express whatever they think and feel. However, this has created new problems, one of which is cyberbullying.

Cyberbullying, or what is often referred to as online bullying, refers to the act of using the internet to hurt, intimidate, or scare others [2]. This is generally done through the sending of messages, comments, or digital content that is negative, degrading, or filled with hate . This behavior can be carried out by anyone, regardless of age, whether children, teenagers, or adults. Features such as anonymous comments, the ease of spreading information, and the lack of strict regulations on some platforms are often exploited to spread hate or attack individuals or groups.

Social media platforms with a wide reach, such as X (formerly Twitter), often become the primary space for cyberbullying. On this platform, users easily express opinions or criticisms that sometimes escalate into retaliatory attacks, fault-finding, threats, intimidation, and belittling each other, which contain elements of cyberbullying. Cyberbullying actions often occur through the sending of tweets or short messages by social media users, such as harsh comments or statements that belittle certain individuals or groups. As a result, it is critical to create a cyberbullying detection system on platform X that uses word embedding-based expansion features.

The sending of tweets in acts of cyberbullying often contains very short words, incorrect grammar, and word variations that cause difficulties in understanding and errors in vocabulary usage. Therefore, word embedding can be used to expand features and reduce vocabulary errors. Word embedding is a method for determining the vector of a word and its context in a corpus, which can then be matched with specific criteria. Some examples of word embedding feature expansion techniques include Word2Vec, FastText, and GloVe[3].

In the study [4], the feature expansion technique using FastText integrated with Hybrid Deep Learning achieved a high accuracy value of 80.55%. Another study that has been conducted [5] used Word2Vec feature expansion with the CNN-LSTM deep learning model, achieving an accuracy value of 79.26%. In addition to these methods, Global Vectors (GloVe) can be used to expand features for word representations. GloVe is regarded as an efficient and effective method for teaching word vector representations. GloVe is an unsupervised word representation learning model that uses global log-bilinear regression. In the research [6], the results of the GloVe feature expansion classified using Artificial Neural Network (ANN), Random Forest (RF), and Logistic Regression (LR) models show that the GloVe feature expansion is quite effective in improving performance with an accuracy rate of 88.59%.

Based on the study [7], a hybrid model for sentiment analysis is put out that incorporates Bidirectional Long Short-Term Memory (Bi-LSTM) and Convolutional Neural Networks (CNN). leveraging both Bi-LSTM's capacity to identify long-term dependencies in text data sequences and CNN's capacity to extract local features through convolution and pooling techniques. This model employs pre-trained word embeddings using GloVe and the term frequency-inverse document frequency (TF-IDF) approach for text representation. GloVe aids in the creation of vector representations that offer more profound contextual meaning, whereas TF-IDF is utilised to gauge a word's significance inside a document.

The application of hybrid deep learning combined with other word embeddings was conducted by [8] to detect cyberbullying text in the form of tweets by utilizing the field of Natural Language Processing and the hybrid deep learning model Res-CNN-BiLSTM with GloVe used as feature extraction. This study achieved an accuracy of 92%. However, this research has limitations regarding the dataset from tweets on Twitter. The existing tweets are often abbreviated due to character limits, making them difficult to understand, and there is a wide variation in the tweets used, leading to vocabulary inconsistency. As a result, creating a cyberbullying detection system on X with enhanced word embedding capabilities is required [3].

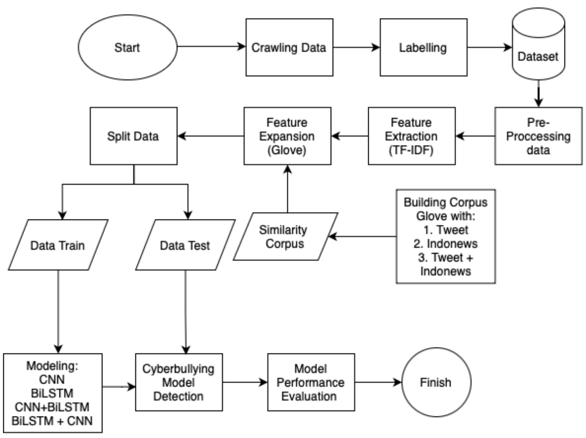
Based on the research [4] on classification algorithms, it can be concluded that integrating Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) has the advantages of both techniques in capturing local patterns and long-range dependencies[9]. The advantages of the GloVe word embeddings method have also been proven in research compared to other methods. Additionally, GloVe provides better and faster results, and also yields the best results regardless of its speed [10].

This study's primary contribution is the Hybrid Deep Learning model and feature expansion, like GloVe, which are used to identify cyberbullying in Indonesian on X. The study of cyberbullying detection using Indonesian language datasets is still restricted to Supervised Learning or Deep Learning and has not yet made use of feature expansion, according to the related research previously described. As previously said, the Indonesian Twitter dataset has its problems, thus selecting Glove as a feature expantion will help with language ambiguities. So, many scenario will be attempted, such choosing different data ratios and employing the best n-grams to extract features. The application of GloVe as a feature expansion and the absence of GloVe will be compared to examine the impact of GloVe as a feature expansion for CNN and BiLSTM models. The combination of Hybrid Deep Learning models, CNN-BiLSTM and BiLSTM-CNN, will also be tested.

Section 2 will offer the materials and method for detecting cyberbullying utilizing Hybrid Deep Learning and GloVe as feature extension, and Section 3 will discuss the results and discussion. Finally, Section 4 will present the conclusions of the conducted research.

MATERIALS AND METHODS

Figure 1. Showing the architecture of this study system the proposed way for creating a cyberbullying detection system on platform X using a hybrid model and GloVe feature expansion as word embedding. This system consists of several stages, namely crawling data, preprocessing data, feature extraction using TF-IDF, feature expansion with GloVe, and classification using CNN, BiLSTM, hybrid CNN-BiLSTM, and BiLSTM-CNN model



Source : (Research Results,2024)
Figure 1. Proposed system design

Crawling Data

Crawling data is a technique for extracting data from a source and generating a dataset. The data for this study was gathered from tweets on social network X. The data to be extracted from X is data that has the potential to contain cyberbullying. The crawling data process will utilize the API provided by X [11]. Data will be collected manually in order to gather the necessary information. The list of keywords utilized during the data collection procedure includes words that are likely to appear in tweets about cyberbullying.

Table 1 . Keyword's Tweet

Keyword	Total
J*l*k	3,663
B*nc*	2,855
L*nt*	3,610
G*bl*k	6,572
B*ngs*t	634
B*d*h	2,367
G*nd*t	1,630
B*g*	926
T*l*l	6,874
K*nt*l	953
Total	30,084

Source: (Research Results,2024)

Labelling

After the data collection process, the next step is to label the data. This is important to help the developed model understand the types of data that contain cyberbullying and those that do not. The majority voting method of labelling is a popular technique for settling disputes over truth labels. Each tweet will be labeled by 3 people to determine whether it constitutes cyberbullying or not, based on the principle of majority vote. Majority vote in the dataset labeling process is a technique for determining the final label for a sample based on the most votes [12].The labeling process is carried out in binary form, where data identified as cyberbullying will be given a value of "1", while those that do not will be given a value of "0". The number of data in the cyberbullying and non-cyberbullying categories is balanced in the analyzed dataset, as shown in Table 2. with 15,005 data identified as cyberbullying and 15,079 data identified as non-cyberbullying.

Table 2 . Labelling Data Distribution

Label	Class	Total	Percentage
Cyberbullying	1	15,005	49.88%
Non-Cyberbullying	0	15,079	50.12%
Total		30,084	100%

Source: (Research Results,2024)

Data Preprocessing

The data obtained from the initial crawling data process is still in the form of unstructured raw

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i3.6270

data [13]. Therefore, a series of data preprocessing steps are needed to organize the data into a more structured form and ready for processing in the classification stage. Preprocessing in this research will involve several stages, including cleaning data, case folding, normalization, tokenization, stopwords, removal, and stemming [14].

1. Cleaning data
This section aims to clean the data of irrelevant elements or symbols used in the classification model development, such as numbers, username symbols (@), hashtags (#), URLs, emoticons, spaces, repeated characters, and so on.

2. Case folding
After cleaning the data, the next step is to convert the words to lowercase. This seeks to simplify classification by removing differences in letter casing within a sentence.

3. Normalization
Normalization is the process of transforming non-standard words so that the model does not treat them as distinct terms. For example, words like “kalo”, “kalau”, “aja”, “saja”, “doang” that have the same meaning.

4. Tokenization
Next, perform Tokenization, which is the process of dividing sentences in the text into pieces of tokens or words.

5. Stopword removal
Stopwords Removal is the process of removing words that generally do not have significant meaning in text analysis, such as conjunctions, pronouns, or frequently occurring words that do not contribute significantly to understanding the content.

6. Stemming
Stemming is a process in text processing aimed at removing prefixes, suffixes, or word affixes so that only the root word remains. The goal is to reduce the variation of words that have the same meaning but are written with different variations.

Feature Extraction

After processing the data in the preprocessing stage, the next step is feature extraction. Feature extraction is the initial step in the text classification process. The goal is to transform the text into a vector representation, where each word is assigned a specific weight [4]. The initial step in representing tweets is through N-gram feature modeling, which includes unigrams, bigrams, and trigrams. The feature weighting method applied in this research is Term Frequency Inverse Document Frequency (TF-IDF). This approach is often used in the Information Retrieval community because it is considered efficient, easy to implement, and provides accurate results [6]. TF-IDF is the combination of Term Frequency (TF), which counts the frequency of terms in a document, and Inverse Document Frequency (IDF), which weights those words. Thus, words that occur frequently in a document will have a lower value than those that appear infrequently. Here is the TF-IDF formula:

$$W_{td} = TF_{td} \times IDF_t \tag{1}$$

$$IDF_t = (\log(\frac{n}{df})) \tag{2}$$

In the Eq.(1) and Eq.(2), W represents the weight of the d word against the t word, while N is the total number of documents in the dataset, and df indicates the number of documents containing the word being identified.

Table 3 shows an example of a document that has gone through the preprocessing stage. The document will be calculated for its TF-IDF value, and the calculation results can be seen in Table 4.

Table 3. Example Document

No	Document
1	"dasar b*d*h otak"
2	"nyinyir orang dasar"
3	"b*d*h otak orang"

Source: (Research Results,2024)

Next, calculate the TF-IDF for the stemmed words: [dasar], [b*d*h], [otak], [nyinyir], [orang]. As in the table below:

Table 4. TF-IDF Process

No	“dasar”	“b*d*h”	“otak”	“nyinyir”	“orang”
1	0.577	0.577	0.577	0.000	0.000
2	0.518	0.000	0.000	0.681	0.518
3	0.000	0.577	0.577	0.000	0.577

Source: (Research Results,2024)

From the TF-IDF results in the table above, it can be seen that there are still several 0 values in the matrix, which results in a mismatch of information from the resulting vector. Therefore, in this study, feature expansion is used to extract information contained in the tweet. Thus, we can expand the scope of features used to better represent tweets, thereby improving the model's ability to capture richer and more varied information from the analyzed text.

Feature Expansion

This method aims to find missing words in the representation of tweets and replace them with semantically related words . This is a process that involves rearranging words by adding new words that have been previously stored [6]. The GloVe word embedding method, which is an unsupervised learning technique, generates vector representations of words to show interesting linear relationships between words in the vector space. This method allows for feature extension. Training uses statistics on the occurrence of terms worldwide from the corpus taken [15]. GloVe gains an understanding of the relationships between words by calculating the frequency of words appearing together in the corpus.

Table 5 . Corpus Build With GloVe

Corpus	Tweet	Indonews	Tweet+Indonews
Total	30,084	127,580	157,664

Source: (Research Results,2024)

To create a similarity corpus, each word found in Tweets and/or Indonews data is combined. Then, the GloVe method is used to calculate the similarity between words using three types of data: Tweet data, Indonews data, and their combination. The final similarity corpus is created by testing each dataset, including the Tweet dataset, Indonews dataset, and combined dataset. By using GloVe to generate the similarity corpus, Table 6. shows the fifteen words that have meanings most similar to the word "b*b*".

Table 6. Example Similarity Corpus Word to "Babi"

Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
B*j*ng	Ng*nt*t	J*nc*k	B*ngs*t	B*g*

Source: (Research Results,2024)

Classification Algorithm

After feature expansion, the generated vectors will be used as input to create a model using Hybrid Deep Learning, which combines Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM). Convolutional Neural Networks (CNN) are types of neural networks that use convolution operations to extract local characteristics in the form of vectors [16]. CNN has a multilayer network where the output of one layer becomes the input for the next layer. Thus, CNN consists of an input layer, several hidden layers, and an output layer [17].

The architecture of the CNN model used in this study consists of one Conv1D layer with 64 filters and a kernel size of 5, followed by the ReLU activation function and using same padding to keep the input dimensions constant. After that, a MaxPooling1D layer with a pool size of 5 is applied to reduce the spatial feature dimensions, followed by a Dropout layer with a ratio of 0.3 to prevent overfitting. Next, the Flatten layer is used to convert the output from spatial features into a one-dimensional vector so that it can be processed by the dense layers. In the final stage, there are two Dense layers, the first layer with 32 neuron units activated by the ReLU function and the second layer with 1 output units using the sigmoid activation function to produce classification probabilities for two classes. This architecture is designed to capture local patterns from the input data, reduce model complexity, and prevent overfitting through the application of Dropout, thereby achieving optimal performance in classification tasks.

Bi-directional Long Short-Term Memory (BiLSTM) is a combination of forward LSTM and backward LSTM [18]. LSTM is used to model sentences by learning what needs to be remembered and ignored during the training process. However, the LSTM model cannot encode information from back to front [16]. Therefore, BiLSTM is needed to capture bidirectional semantic dependencies.

The architecture of the BiLSTM model used in this study consists of one Bidirectional LSTM layer with 64 neuron units. This layer processes input data both forward and backward to capture information from both sequential directions, making it more effective in understanding the temporal context. The LeakyReLU activation function with an alpha parameter of 0.001 is applied after the LSTM layer to address the issue of dying neurons. To prevent overfitting, a Dropout layer with a ratio of 0.8 is used, followed by GlobalMaxPool1D to take the maximum value of the extracted features and reduce the output dimension. Next, there are two Dense layers, the first layer with 32 neuron units and ReLU activation function to capture feature representations more deeply. The final layer uses Dense with one output unit and sigmoid activation function to produce binary classification probabilities. This architecture is designed to optimize the model's performance in understanding the context of sequential data.

Therefore, an experiment was conducted by combining CNN and BiLSTM to leverage the advantages of each model. CNN is used to extract features from word embeddings by detecting significant local patterns in the text data. Subsequently, the generated features are used as input for BiLSTM, which processes the text sequences by considering the contextual relationships between words in both directions (forward and backward) [9].

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i3.6270

The architecture of the CNN-BiLSTM hybrid model in this study is designed to combine the advantages of CNN and BiLSTM. The model begins with a Conv1D layer that has 64 filters with a kernel size of 5 and ReLU activation to extract spatial features, followed by MaxPooling1D to reduce the feature dimensions and Dropout with a ratio of 0.8 to prevent overfitting. Next, the output is processed by a Bidirectional LSTM layer with 64 units that captures sequential context deeply, supported by the LeakyReLU activation function. The output from the BiLSTM is then flattened through a Flatten layer and processed by two Dense layers, with the final layer using the sigmoid activation function to produce binary classification.

Performance Evaluation

Performance measurement is conducted using a confusion matrix, a tool for measuring performance in classification that produces output in the form of two or more classes. There are four possible combinations of actual and expected values, namely False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN) [20]. One of the metrics that can be used to assess performance is accuracy, which is measured by calculating the ratio of true values from all data using the formula shown in Eq. (3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

RESULTS AND DISCUSSION

This study uses 4 classification models, namely CNN, Bi-LSTM, CNN-BiLSTM, and BiLSTM-CNN to detect cyberbullying in Indonesian language texts and conducts 4 scenarios to see the differences in accuracy produced by each model. The process begins with preprocessing which includes cleaning, case folding, normalization, tokenization, stopword removal, and stemming. For example in the Table 7. Tweet: “Dasar b*d*h! Lo nggak ada otak, Cuma bisa nyinyirin orang doang. #sampahmasyarakat.

Table 7. Stages Data Preprocessing

Stages	Output
Cleaning data	Dasar b*d*h Lo nggak ada otak cuma bisa nyinyirin orang doang
Case folding	dasar b*d*h lo nggak ada otak cuma bisa nyinyirin orang doang
Normalization	dasar b*d*h kamu tidak ada otak cuma bisa nyinyirin orang doang
Tokenization	[dasar], [b*d*h], [kamu], [tidak], [ada], [otak], [cuma], [bisa], [nyinyirin], [orang], [doang]
Stopword removal	[dasar], [b*d*h], [otak], [nyinyirin], [orang]
Stemming	[dasar], [b*d*h], [otak], [nyinyir], [orang]

Source: (Research Results,2024)

So as to improve data quality so that the model can learn more effectively and accurately. With good preprocessing, the model can focus more on important information, improving efficiency and accuracy in classification or prediction.

Scenario 1 (Baseline)

In the first scenario, this aims to determine the best data split ratio using a maximum feature count of 10,000 and a min_df parameter of 25 on Unigrams. The results of this scenario will be used as a baseline for comparison with the next scenario, to evaluate its impact on the accuracy value of each scenario. The models tested in this scenario are CNN and BiLSTM with data split ratios of 90:10, 80:20, and 70:30. In previous research, 5 epochs were used because in tests with 10 epochs, an increase in loss value was found after the 6 to the 10 epoch [4], which resulted in a decrease in accuracy. Therefore, the limitation of 5 epochs was used to avoid a decrease in accuracy.

Table 8 . Scenario 1 (Baseline)

Split Ratio	Max Feature	min_df	Accuracy (%)	
			CNN	BiLSTM
90:10	10.000	25	79,26	79,30
80:20	10.000	25	78,44	78,66
70:30	10.000	25	78,61	78,43

Source: (Research Results,2024)

The test results in Table 8 show that the 90:10 (0.1) data split ratio produce the highest accuracy compared to other ratios, with a maximum feature count of 10,000 and a min_df parameter of 25. At this ratio, the CNN model achieves an accuracy of 79.26%, while the BiLSTM model produce a slightly higher accuracy of 79.30%. In this first scenario, the best scenario is achieved with the 90:10 data split, which will be used as the baseline for future model scenarios.

Scenario 2 (N-Gram Test)

The second scenario will involve experiments testing several parameters of N-Gram in TF-IDF as feature extraction. One of the important parameters in TfidfVectorizer is the N-gram range, which indicates the length of the word sequence (n-gram) used as features. By adjusting the N-gram, the model can analyze text patterns and context more accurately. Unigram (1, 1), Bigram (2, 2), Trigram (3, 3), Unigram-Bigram (1, 2), Unigram-Trigram (1, 3), and Bigram-Trigram (2, 3) are six variations of N-gram used in this study for feature extraction.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

This approach allows the model to enhance the combination of single words and the relationships between words at various scales. Table 9 provides an example of using the N-gram parameter in the features extraction process using TF-IDF, and Table 10 is the result of the accuracy of using N-Gram parameters in TFIDF including Unigram (baseline), Bigram, Trigram, Uni-Bi, Uni-Tri, and Bi-Tri.

Table 9. N-Gram Parameter

N-Gram	Tweet
Unigram	[kenapa], [kamu], [g*bl*k]
Bigram	[kenapa, kamu], [kamu, g*bl*k]
Trigram	[kenapa, kamu, g*bl*k]
Unigram – Bigram	[kenapa], [kamu], [g*bl*k], [kenapa, kamu], [kamu, g*bl*k]
Unigram - Trigram	[kenapa], [kamu], [g*bl*k],[kenapa, kamu, g*bl*k]
Bigram - Trigram	[kenapa, kamu], [kamu, g*bl*k], [kenapa, kamu, g*bl*k]

Source: (Research Results,2024)

Table 10. Scenario 2 (N-Gram Test)

N-Gram	Accuracy (%)	
	CNN	BiLSTM
Unigram(Baseline)	79,26	79,30
Bigram	65.30	65.77
Trigram	51.31	50.25
Uni-Bi	**79.40**	**79.36**
Uni-Tri	79.10	79.26
Bi-Tri	65.57	65.74

Source: (Research Results,2024)

The results of the analysis in Table 10 using the N-Gram combination with the maximum number of features and the same min_df parameter as in the first scenario show that the Unigram-Bigram combination produce the highest accuracy. In this combination, the CNN model achieved an accuracy of 79.40%, while the BiLSTM model obtained an accuracy of 79.36%. The results of this test will serve as the basis for further experiments that will be combined with the expansion of Glove features as word embedding.

Scenario 3 (GloVe)

Next, in the third scenario, testing is conducted by adding feature expansion using GloVe as word embedding on the CNN and BiLSTM classification models generated from the previous scenario. GloVe (Global Vectors for Word Representation) works by studying the vector representation of words based on the co-occurrence matrix from the corpus, thus being able to capture the semantic relationships between words [15].

The testing was conducted using a corpus built from three data sources: Tweets, IndoNews, and a combination of Tweets + IndoNews. The evaluation was conducted by considering the influence of the top ranks 1, 5, 10, and 15 on the model's accuracy. The analysis results are used to identify the best configuration in this scenario.

Table 11. Scenario 3 (GloVe)

Model	Rank	Accuracy (%)			
		Base	Tweet	IndoNews	Tweet+ IndoNews
CNN	Top 1	79.40	79.28 (-0.12)	78.99 (-0.41)	79.22 (- 0.18)
	Top 5		79.50 (+0.10)	79.59 (+0.19)	80.11 (+0.71)
	Top 10		81.70 (+2.30)	79.23 (-0.17)	81.00 (+1.60)
	Top 15		81.46 (+2.06)	79.50 (+0.10)	**82.98 (+3.58)**
BiLSTM	Top 1	79.36	79.10 (-0.26)	79.22 (-0.14)	79.21 (-0.15)
	Top 5		78.98 (-0.38)	79.40 (+0.04)	**80.34 (+0.98)**
	Top 10		78.92 (-0.44)	79.39 (+0.03)	79.85 (+0.49)
	Top 15		79.11 (-0.25)	79.16 (-0.20)	79.90 (+0.54)

Source: (Research Results,2024)

After adding feature expansion using GloVe, as shown in Table 11, it is evident that the CNN and BiLSTM classification models experienced a significant increase in accuracy compared to the previous scenario. The highest accuracy improvement in this scenario was achieved by the CNN model at 82.98 and the BiLSTM at 80.34, which used TF-IDF as the feature extraction method and GloVe as the feature expansion, with the tweet + IndoNews corpus.

Scenario 4 (Model Hybrid)

The fourth scenario consists of two parts, Table 12 (hybrid without GloVe) and Table 13 (hybrid with GloVe). This experiment uses a hybrid classification model, namely the combination of CNN+BiLSTM and BiLSTM+CNN. The testing was conducted using the same corpus as in the previous scenario, namely tweets, IndoNews, and a combination of Tweets + IndoNews.

The test results show a comparison of performance between the hybrid model without feature expansion (GloVe) and with feature expansion (GloVe), to identify the best configuration.

Table 12. Scenario 4 (Hybrid Without GloVe)

Model	Baseline	CNN+BiLSTM	BiLSTM + CNN
CNN	79.40	79.87 (+ 0,47)	
BiLSTM	79.36		80.23 (+ 0,83)

Source: (Research Results,2024)

Table 13. Scenario 4 (Hybrid With GloVe)

Model	Rank	Accuracy (%) Base	Tweet	IndoNews	Tweet+ IndoNews
CNN +BiLSTM	Top 1	79.40	80.09 (+0.69)	80.00 (+0.60)	79.88 (+0.48)
	Top 5		81.26 (+1.86)	79.50 (+0.10)	81.16 (+1.76)
	Top 10		**82.62 (+3.22)**	78.93 (−0.47)	81.40 (+2.00)
	Top 15		81.81 (+2.41)	80.10 (+0.70)	81.25 (+1.85)
BiLSTM +CNN	Top 1	79.36	80.00 (+0.60)	80.09 (+0.69)	80.01 (+0.61)
	Top 5		81.92 (+2.52)	80.12 (+0.72)	81.64 (+2.24)
	Top 10		81.73 (+2.33)	79.83 (+0.43)	83.15 (+3.75)
	Top 15		82.41 (+3,01)	80.64 (+1.24)	**83.88** (+4.48)

Source: (Research Results,2024)

The test results in scenario 4 in Table 13 show that the use of GloVe word embedding has a significant impact on the performance of the hybrid model. The hybrid model without feature expansion (without GloVe) was only able to achieve an accuracy of 80.23% in the BiLSTM-CNN configuration. In contrast, when feature expansion using GloVe was applied, there was a substantial improvement in performance.

The hybrid BiLSTM-CNN model with GloVe as the word embedding achieved the highest accuracy of 83.88%. This best result was obtained in the test with a top rank of 15 using the tweet + IndoNews corpus. These findings confirm that the use of GloVe as a feature expansion can significantly improve the accuracy of the hybrid classification model.

Table 14. Significant Test

Parameters	Scenario $S1 \rightarrow S2$	$S2 \rightarrow S3$	$S3 \rightarrow S4.1$	$S4\_1 \rightarrow S4.2$	$S1 \rightarrow S4.2$
P-Value	0.012	0.079	0.283	1.151	7.605
Z-Value	2.499	1.753	-1.073	7.111	6.508
Significant?	True	False	False	True	True

Source: (Research Results,2024)

According to the statistical significance evaluation shown in the Table 14, several situations demonstrate a significant increase in accuracy. The comparison S1→S2 shows a significant increase (P-Value < 0.05 and Z-Value > 1.96), and the scenarios S4_1→S4_2 and S1→S4_2 each show a significant increase with Z-Values of 7.111 and 6.508, respectively, indicating a strong influence of the proposed method. However, the comparisons S2-S3 and S3-S4_1 do not show a significant increase because the P-Value is greater than 0.05. These results indicate that in certain situations, a hybrid method with feature expansion can significantly improve accuracy.

Discussion

Each testing scenario in this study shows the best values for each model. In the first scenario, in Table 8, the CNN model achieved the highest accuracy of 79.26%, while the BiLSTM model reached an accuracy of 79.30%, with a data split ratio of 90:10, using a maximum of 10,000 features and a min_df value of 25. After obtaining the best accuracy in the first scenario, this configuration was then applied to the second scenario, where further testing was conducted to find the best parameters for N-Gram TF-IDF. The results of the tests showed an increase in accuracy, with the CNN model using a combination of unigram and bigram achieving an accuracy of 79.40%, while the BiLSTM model produced an accuracy of 79.36%, as shown in Table 10.

In the third scenario, testing was conducted by adding feature expansion using GloVe as word embedding to evaluate its impact on the accuracy produced by each model. This test uses a corpus built from three data sources: Tweets, IndoNews, and a combination of Tweets + IndoNews. The evaluation considers the impact of the top ranks 1, 5, 10, and 15 on the model's accuracy. If compared to the baseline model tested in the second scenario, the results in the third scenario show a quite significant increase in accuracy. The CNN model showed an accuracy improvement of 82.98, while the BiLSTM model achieved an accuracy improvement of 80.34. This confirms that the use of GloVe as a feature expansion can have a positive impact on the accuracy performance of both models.

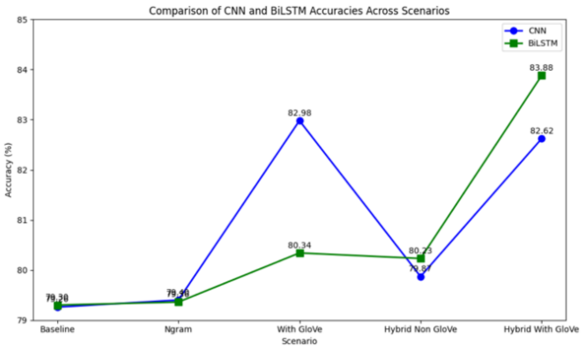
In the fourth scenario, the experiment is divided into two parts, Table 12 (hybrid without GloVe) and Table 13 (hybrid with GloVe). In this experiment, a hybrid classification model combining CNN+BiLSTM and BiLSTM+CNN is used. The purpose of this test is to compare the performance between the hybrid model that does not use feature expansion (without GloVe) and the one that uses GloVe feature expansion, in order to identify the best configuration.

The test results in the fourth scenario show that the use of GloVe word embedding has a significant impact on the performance of the hybrid model. The hybrid model without feature expansion (GloVe) resulted in a decrease in accuracy of -2,75% for CNN-BiLSTM and -3,65% for BiLSTM-CNN%. However, when GloVe feature expansion was

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i3.6270

applied, there was a significant increase in accuracy. The BiLSTM-CNN hybrid model using GloVe as word embedding achieved the highest accuracy of 83.88%, with the best results obtained in the test with a top rank of 15 using the tweet + IndoNews corpus. This comparison shows that feature expansion with GloVe significantly contributes to the increase in accuracy. Without GloVe, the accuracy of the BiLSTM-CNN hybrid model only reaches 80.23%, while after GloVe is applied, the accuracy increases to 83.88%. This proves that the use of GloVe as a feature expansion can significantly enhance the performance of the hybrid classification model, making it more effective in processing and classifying data. Here are the comparison results from several scenarios.



Source : (Research Results,2024)
Figure 2. Comparison Result All Scenario

Table 15. Comparison With Related Study

Authors	Model	Feature Extraction	Feature Expantion	Accuracy
Laxmi et al. [21]	CNN	Doc2vec	-	65.00%
Asqolani et al. [5]	LSTM+ CNN	TF-IDF	word2vec	79.48
Nasution et al. [4]	CNN + BiLSTM	TF-IDF	FastText	80.55 %
Proposed Method	CNN + BiLSTM	TF-IDF	Glove	82.62
Proposed Method	BiLSTM + CNN	TF-IDF	Glove	83.88

Source: (Research Results,2024)

Table 15. Presents a comparison of the proposed method with related research on cyberbullying detection using Indonesian language Twitter data. The proposed method achieved the highest accuracy compared to other studies. These results show that the combination of CNN and BiLSTM as a hybrid model in deep learning is quite effective in understanding the context of sentences. In addition, the application of GloVe word embeddings enhances the model's performance by strengthening accuracy through the utilization of semantic relationships between words, which is a key factor in the success of this hybrid model.

CONCLUSION

This research proposes the use of a hybrid CNN and BiLSTM classification model to detect cyberbullying in Indonesian-language Twitter data. For training and evaluation, this research uses a dataset of 30,084 tweets from platform X, while a GloVe-based corpus was built using an additional IndoNews dataset with 127,580 entries. Through a series of experiments covering various scenarios, this research successfully demonstrated that the proposed method can achieve higher accuracy compared to previous studies.

In the first scenario, the highest accuracy was obtained with a 90:10 data split ratio and using a maximum of 10,000 features and min_df 25. The second scenario, which tested various N-Gram combinations, showed an increase in accuracy, with the use of Unigram and Bigram yielding the best results. The third scenario, which tests the influence of feature expansion using GloVe as word embedding, shows a significant improvement in classification model performance, with the highest accuracy achieved in the CNN model using the tweet corpus + IndoNews top rank 15. In the fourth scenario, the application of GloVe feature expansion provides an even more significant improvement, with the hybrid model using GloVe achieving the highest accuracy of 83.88%, much higher than the model without feature expansion.

Overall, the results of this study affirm that the combination of CNN and BiLSTM models as a hybrid model, along with the use of GloVe word embeddings, is capable of improving performance in detecting cyberbullying in Indonesian language text data. By utilizing GloVe to strengthen the semantic relationships between words, this model can better understand the context of sentences. For further research, testing of combinations of other deep learning models and exploration of feature expansion using various word embedding methods besides GloVe, such as Word2Vec or FastText, can be conducted to analyze their impact on model performance. Additionally, the research can also be expanded by using data from various different social media platforms, considering that cyberbullying has many variations and forms, depending on the context and user characteristics on each platform. This can help in understanding the patterns and dynamics of cyberbullying more broadly and improve the generalization of the model.

REFERENCE

[1] "APJII Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang." Accessed: Apr. 12, 2024. [Online]. Available: https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang

[2] T. Nazir and L. Thabassum, "Cyberbullying: Definition, Types, Effects, Related Factors and Precautions to Be Taken During COVID-19 Pandemic," vol. 9, 2021, doi: 10.25215/0904.047.

[3] S. J. Johnson, M. R. Murty, and I. Navakanth, "A detailed review on word embedding techniques with emphasis on word2vec," *Multimed Tools Appl*, vol. 83, no. 13, pp. 37979–38007, Apr. 2024, doi: 10.1007/s11042-023-17007-z.

[4] M. A. S. Nasution and E. B. Setiawan, "Enhancing Cyberbullying Detection on Indonesian Twitter: Leveraging FastText for Feature Expansion and Hybrid Approach Applying CNN and BiLSTM," *Revue d'Intelligence Artificielle*, vol. 37, no. 4, pp. 929–936, Aug. 2023, doi: 10.18280/ria.370413.

[5] I. A. Asqolani and E. B. Setiawan, "A Hybrid Deep Learning Approach Leveraging Word2Vec Feature Expansion for Cyberbullying Detection in Indonesian Twitter," *Ingenierie des Systemes d'Information*, vol. 28, no. 4, pp. 887–895, Aug. 2023, doi: 10.18280/isi.280410.

[6] Febiana Anistya and Erwin Budi Setiawan, "Hate Speech Detection on Twitter in Indonesia with Feature Expansion Using GloVe," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1044–1051, Dec. 2021, doi: 10.29207/resti.v5i6.3521.

[7] M. Kamyab, G. Liu, and M. Adjeisah, "Attention-Based CNN and Bi-LSTM Model Based on TF-IDF and GloVe Word Embedding for Sentiment Analysis," *Applied Sciences (Switzerland)*, vol. 11, no. 23, Dec. 2021, doi: 10.3390/app112311255.

[8] R. Joshi, A. Gupta, and N. Kanvinde, "Res-CNN-BiLSTM Network for overcoming Mental Health Disturbances caused due to Cyberbullying through Social Media," Apr. 2022, [Online]. Available: http://arxiv.org/abs/2204.09738

[9] A. Toktarova *et al.*, "Hate Speech Detection in Social Networks using Machine Learning and Deep Learning Methods," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, 2023, doi: 10.14569/IJACSA.2023.0140542.

[10] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, Association for Computational Linguistics (ACL), 2014, pp. 1532–1543. doi: 10.3115/v1/d14-1162.

[11] S. S. Sohail *et al.*, "Crawling Twitter data through API: A technical/legal perspective," May 2021, [Online]. Available: http://arxiv.org/abs/2105.10724

[12] A. M. Davani, M. Díaz, and V. Prabhakaran, "Dealing with Disagreements: Looking Beyond the Majority Vote in Subjective Annotations," Oct. 2021, [Online]. Available: http://arxiv.org/abs/2110.05719

[13] M. Lamba and M. Madhusudhan, "Text Pre-Processing," in *Text Mining for Information Professionals*, Springer International Publishing, 2022, pp. 79–103. doi: 10.1007/978-3-030-85085-2_3.

[14] R. A. Rudiyanto and E. B. Setiawan, "Sentiment Analysis Using Convolutional Neural Network (CNN) and Particle Swarm Optimization on Twitter," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 2, pp. 188–195, Feb. 2024, doi: 10.33480/jitk.v9i2.5201.

[15] G. P and S. Juliet, "An Enhanced Approach for Fake News Detection using Ensemble Techniques," in *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, Mar. 2023, pp. 1957–1962. doi: 10.1109/ICACCS57279.2023.10112856.

[16] R. Ben Said, Z. Sabir, and I. Askerzade, "CNN-BiLSTM: A Hybrid Deep Learning Approach for Network Intrusion Detection System in Software-Defined Networking With Hybrid Feature Selection," *IEEE Access*, vol. 11, pp. 138732–138747, 2023, doi: 10.1109/ACCESS.2023.3340142.

[17] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif Intell Rev*, vol. 57, no. 4, p. 99, Mar. 2024, doi: 10.1007/s10462-024-10721-6.

[18] S. Mishra, N. Bhatnagar, P. Prakasam, and S. T. R, "Speech emotion recognition and

classification using hybrid deep CNN and BiLSTM model," *Multimed Tools Appl*, vol. 83, no. 13, pp. 37603–37620, Apr. 2024, doi: 10.1007/s11042-023-16849-x.

[19] S. Narkhede, "Understanding Confusion Matrix," Towards Data Science. Accessed: May 05, 2024. [Online]. Available: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

[20] S. T. Laxmi, R. Rismala, and H. Nurrahmi, "Cyberbullying Detection on Indonesian Twitter using Doc2Vec and Convolutional Neural Network," in *2021 9th International Conference on Information and Communication Technology (ICoICT)*, IEEE, Aug. 2021, pp. 82–86. doi: 10.1109/ICoICT52021.2021.9527420.