

Tweets Sentiment Analysis of Putin's Participation at the G20 Summit in Indonesia

Syifa Auliyah Hasanah[1], Isal Firmansyah[1], Farid Azhar Lutfi Nugraha[1], Anindya Apriliyanti Pravitasari[1], Eka Novita Sari[2], and Tutut Herawan[3,4(✉)]

[1] Department of Statistics, Universitas Padjadjaran, Jl. Ir. Soekarno KM. 21, Jatinangor, Sumedang, West Java 45363, Indonesia
{syifa19010,farid1900}@mail.unpad.ac.id, anindya.apriliyanti@unpad.ac.id

[2] AMCS Research Center, Jalan Griya Taman Asri, Sleman, Yogyakarta, Indonesia
eka@amcs.co

[3] Sekolah Tinggi Pariwisata Ambarrukmo Yogyakarta, Jl. Ringroad Timur No. 52, Bantul, Daerah Istimewa Yogyakarta 55198, Indonesia
tutut@stipram.ac.id

[4] Institute for Big Data Analytics and Artificial Intelligence, UiTM Shah Alam, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia

Abstract. The Group of Twenty (G20) is an international economic cooperation forum established to address global and regional issues. The members are countries with large economies and the ability to have a significant impact on the global, since their GDP (Gross Domestic Product) shares about 80% of the global GDP. The dispute between Russia and Ukraine drew international attention regarding Russia's involvement in the G20. The Russian President, Vladimir Putin, played a major role and became a major source of concern in this dispute. The Western Bloc countries, such as the United States, wanted Russia to be expelled from the forum, but the Eastern Bloc believed the G20 was not a platform for discussing security matters, therefore Russia could still be included. As a result, sentiment analysis has to be conducted to determine the global community's reaction to Putin's presence at the G20 summit in Indonesia. Aside from that, the goal of this study is to automate the forecast of future public sentiments on the same subject. The Bidirectional Gated Recurrent Unit (BiGRU) architecture was employed in this work, using three sentiment labels: positive, negative, and neutral. The model created in this investigation has an accuracy value of 86.76% after being run for 20 epochs. This demonstrates that the model may get a reasonably accurate classification result when categorizing sentiments about Putin's attendance at the G20 Indonesia summit.

Keywords: Tweets · Sentiment analysis · G20 · Conflict · Russia · Vladimir Putin · BiGRU

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14107, pp. 162–178, 2023.
https://doi.org/10.1007/978-3-031-37114-1_12

1 Introduction

The G20 or Group 20 is a multilateral cooperation forum consisting of 19 countries and the European Union (EU), which was formed to help solve global and regional problems with two focuses, i.e., the Finance Track and the Sherpa Track. The G20 is an important world forum because it represent 2/3 of the world's population or around 5.267 billion people, covering 80% of the world's GDP and also representing about 75% of the world economy [1]. The G20 Forum does not have a permanent leader, but it does have a presidency function when one of the country's members will host the event for a full year. In 2022, Indonesia has the opportunity to hold the G20 presidency. This is a great opportunity for Indonesia to host such a big event with countries that have great economic and political influence. The G20 presidency in Indonesia will have a positive impact on Indonesia. According to Finance Minister Sri Mulyani Indrawati, the G20 forum in Indonesia can contribute to Indonesia's GDP of US $533 million, or around Rp7.4 trillion. The G20 Forum is also predicted to increase domestic consumption by up to IDR 1.7 trillion. Meanwhile, in other sectors, such as tourism, the G20 forum will contribute to the increase in foreign tourists by 1.8 million–3.6 million and will also create 600 thousand–700 thousand new jobs [1]. Besides the positive effects, Indonesia still needs to face some problems that will affect the G20 forum. The Russia-Ukraine conflict is one of them. The conflict has given rise to the political stance of the United States and several western bloc countries towards the G20 forum. They asked Indonesia not to invite Russia to the G20. The United States threatened not to attend the G20 summit if Russian President Vladmir Putin was present.

The Russia-Ukraine conflict brings a dilemma to the invitation decision of the President of Russia to the G20 summit in Indonesia. As the host, Indonesia has to invite all G20 members, including Russia, but some countries asked Indonesia not to invite Russia to the G20 forum. Indonesia needs to make the best decision for all parties. Good diplomacy skills are needed in this situation. Various responses have appeared regarding the issue of Putin's participation in the G20 Summit in Indonesia. Some narratives from some media are made to support one side of the two disputing camps. Therefore, sentiment analysis is needed to see the actual public's response based on data. Sentiment analysis will analyze a person's opinions, attitudes, and emotions towards a thing [2]. The benefit of this sentiment analysis is to know how the majority of the public responds to Putin's participation in the G20 summit in Indonesia, which can be used as a public sentiment reference and consideration by the Indonesian government on the decision to invite Putin to the G20 summit. This research is one of the first studies that discusses public sentiment on the issue of Putin's participation in the G20 summit in Indonesia. The methods that will be used in this research are deep learning methods that are expected to provide greater performance than the traditional methods.

This study tries some deep learning algorithms such as BiLSTM, CNN, and BiGRU to know which algorithm has the best performance in classifying public sentiment towards Putin's participation in the G20 issue. Three classification categories are used in this study, i.e., positive, neutral, and negative. In the end, the BiGRU algorithm shows the highest accuracy and the lowest loss, so this study decided to use BiGRU to predict public sentiment towards Putin's participation in the G20 forum.

Sentiment analysis in this study is limited to Twitter data in English with the issue of "Putin's participation at the G20 2022 Summit in Indonesia," taken from April 22 to April 30, 2022. This study focuses on public opinion on Twitter, which is used because this platform allows people to express their feelings and perceptions about things that are happening "right now" [3]. Usually, some information and issues spread faster on Twitter than on any other social media or news platforms. Aside from that, Twitter's character restriction feature influences the use of "to-the-point" sentences.

The remainder of the paper is organized as follows. In Sect. 2, we introduce some traditional machine learning and deep learning architectures. Section 3 describes the deep learning architecture we use in this research. In Sect. 4, we show the results of the sentiment analysis process and performance evaluation of the proposed BiGRU method compared to other deep learning architectures such as BiLSTM and CNN. Finally, we describe our conclusion and future work in Sect. 5.

2 Related Works

The use of Twitter data in sentiment analysis to see the public's response to an issue is a common thing for researchers [3]. Sentiment analysis on Twitter and several social media applications has become one of the trending topics for researchers. The business sector has used a lot of sentiment analysis to find out user responses and reviews related to products from a manufacturer, but now sentiment analysis is not only used to see consumer responses to an item; sentiment analysis is also used to see responses to an issue such as politics, economics, and others. In recent years, researchers have created a number of automated approaches for detecting and classifying abusive, sarcastic, and toxic language [4]. Sentiment analysis helps us to find various opinions on social media [5]. In politics, sentiment analysis is also often used as a means of effective and efficient promotion, campaigning and dissemination related to political policies so that it can generate opinions and comments from the public, in the form of positive comments, negative comments and neutral.

Caprolu et al. [6] gather and analyze data from Twitter. In particular, investigating data volumes and public perception of the Russia-Ukraine conflict by utilizing statistical analysis techniques and Aspect-Based Sentiment Analysis (ABSA). The results obtained state that the claims of some media contradict the actual situation. Therefore sentiment analysis is very important to validate results based on data found in the field.

Previously, there was no research that examined the sentiment analysis of Putin's participation in the G20. The topic that was covered in this study is a new topic that was brought up because of the Russia-Ukraine conflict that coincides with the G20 summit in Indonesia. So the data used here is a new data that will be subjected to a sentiment analysis using the appropriate machine learning model.

The machine learning models' effectiveness is based on two key factors: the availability of a huge amount of labeled data and the intelligent manual construction of a set of features that can be used to distinguish samples [7]. Many different algorithms can be used to perform sentiment analysis. Generally, two types of algorithms are used, namely traditional machine learning algorithms i.e., Naïve Bayes, SVM, KNN, decision tree, and others, as well as deep learning algorithms such as BiLSTM, BiGRU, CNN, and others.

Traditional machine learning algorithms tend to have simpler concepts. In a previous study that conducted sentiment analysis on Twitter by comparing three algorithms, i.e., SVM, Naïve Bayes, and also KNN [8], the traditional machine learning model was still sufficient to handle sentiment on Twitter data by providing an accuracy of 63% for the Naive Bayes model, 60% for the SVM model, and 61% for the KNN model. In another similar study [9], which conducted sentiment analysis by comparing the Naïve Bayes, KNN, and Decision Tree algorithms, the accuracy was around 50% for all models. Although the accuracy of the traditional machine learning algorithms in the two studies above is not high enough, several studies using traditional machine learning algorithms can achieve a higher accuracy, such as the sentiment review research of the Pedulilindungi application, which compares the Naïve Bayes model and SVM [10], which obtained an accuracy of 84.33% for SVM and 81.16% for Naïve Bayes.

To improve the performance of traditional models, deep learning models can be tried, which are conceptually more complex than traditional models but provide higher performance values based on large-scale datasets. One of the studies that uses deep learning algorithms is research with the title CNN for sentence classification [11], this study enables the easy extraction of textual information, and relational research has made significant progress in sentiment analysis. Zhang employed character-level CNN to classify text in 2015 [12]. On the other hand, CNN isn't a perfect algorithm because it can't catch long-range features [4]. Another study that uses a deep learning model is Sentiment Analysis of Comment Texts based on BiLSTM [13]. This study states that the context information is considered by the BiLSTM model, which allows for a better text representation of the comments. This study compares the BiLSTM model with other deep learning models and one traditional algorithm Naive Bayes. The accuracy value obtained in sentiment analysis using the BiLSTM algorithm in this study is 91.54%, while the traditional Naïve Bayes model provides an accuracy of 86.02%, which is smaller than other deep learning models. Not different from BiLSTM, BiGRU deep learning model in the study with a title "Sentiment analysis based on BiGRU information enhancement" [14] with binary sentiment, i.e., positive and negative, shows good accuracy, and the use of two-layer BiGRU based on BERT preprocessing gives a good accuracy value of 82.63%. The researcher says the length of the running model process is the drawback of using this algorithm. Other related studies have also demonstrated that BIGRU achieves higher accuracy. For example, a binary sentiment analysis research paper titled 'Sentiment Analysis Using an Ensemble Approach of BiGRU Model: A Case Study of AMIS Tweets' analyzed various topics on Twitter, including games, football, LGBT, Trump, and Syria, and reported an accuracy of the sentiment analysis is about 84.8% [15].

This study will use deep learning algorithm to predict public sentiment towards Putin's participation on G20 event, the more complex sentiment analysis algorithm (deep learning algorithm) is used in this study with the hope of providing good performance value.

3 Data and Proposed Method

3.1 Data

We use Twitter crawled data using the keywords "Russia G20" and "Putin G20". By using these keywords, 15213 tweets were collected. However, the data is still dirty data that must go through various processes before it can be used for analysis. The first thing to do with the data is to pre-process the data through the case folding process and data cleaning, tokenization, and stopword removal. So, we get 12462 lines of data that are ready to use which are then labeled with the help of the package in python, namely Vader. The dataset is divided into three categories of labels, namely positive, negative, and neutral with the proportions of each category obtained as follows: positive (1675 tweets), negative (4921 tweets), and neutral (5866 tweets).

3.2 Method

As in Fig. 1, this research begins with the data pre-processing step. At this step, the data will be prepared for further analysis using a predetermined algorithm. The next step is to divide the dataset into training and testing data, which will then be analyzed using several algorithms, namely LSTM, BiLSTM, GRU, BiGRU, and Convolution. The algorithm with the highest evaluation value will be used in this analysis.

3.2.1 Data Preprocessing

Pre-processing is at an early stage. It's a very important step to handling raw data. Without this process, the next process will not produce a good output. Pre-processing helps deal with various problems and produces data that is ready for analysis. The following are the steps in pre-processing data.

Case Folding and Data Cleaning
Case folding is the process of converting all existing letters to lowercase. So, words that still have capital letters will automatically turn into lowercase letters, for example, the letter "B" will become "b". This is done because with case folding, data redundancy can be reduced. Data cleaning is the process of preparing data such as detecting and correcting the arrangement of the dataset so that it can be used for analysis. In this process, words that are not needed will be removed, such as "http" and cleaning of duplicated data is also carried out. Data that has the same content even though it comes from different users will be discarded. At this stage, the author also removes several elements such as removing punctuation marks, eliminating numbers, and so on.

Stop-Word Removal
Stopword Removal is a process to remove certain words that often appear but have no meaning that has a major influence on the analysis process. Examples of words that will be deleted are conjunctions, prepositions, and slang words that are usually used but give the impression of being informal in a sentence. This stopword removal process is done by first making a list of what words are included in the stopword itself. The list will be

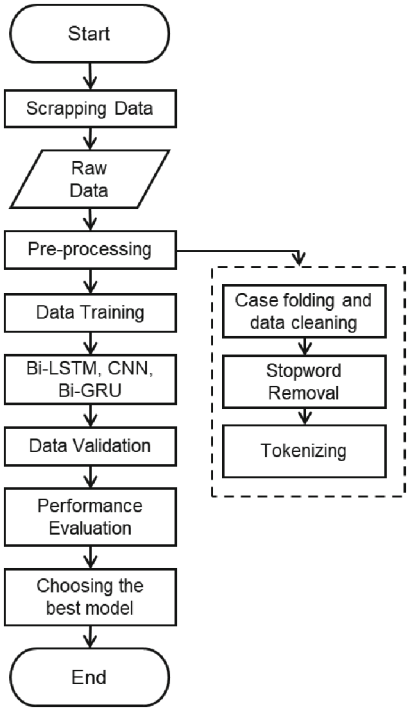


Fig. 1. Flowchart of this research

a benchmark if in the dataset we have these words will be discarded. In this process, emoticons, symbols and pictographs, transport and map symbols, flags, and Chinese characters are usually used in tweeting.

Tokenizing

Tokenizing is the process of separating words in sentences and then collecting them in an array of data which will then be weighted in each of these words. The list of words that make up the sentence will be separated by commas and spaces.

Split Data

In this study, the data will be divided into two groups, namely training data, and testing data. Training data is used to create and train the model to be used. While data testing is used to test the model that has been made.

3.2.2 Sentiment Analysis with BiLSTM

Bidirectional Long Short-Term Memory (BiLSTM) is an extension of the regular LSTM which can improve performance on classification problems. The results of the LSTM output produce global information; it is still difficult to capture the important information contained in it [16]. BiLSTM conducts data training twice, unlike LSTM which only does training once. The raw dataset from the first process is fed back into the second

process. This can provide additional networking and obtain faster and more complete results. From Fig. 2, the layer below it moves forward (forward) to understand the first word to the last, while the other layers move backward (backward) to understand the word from the end to the first word.

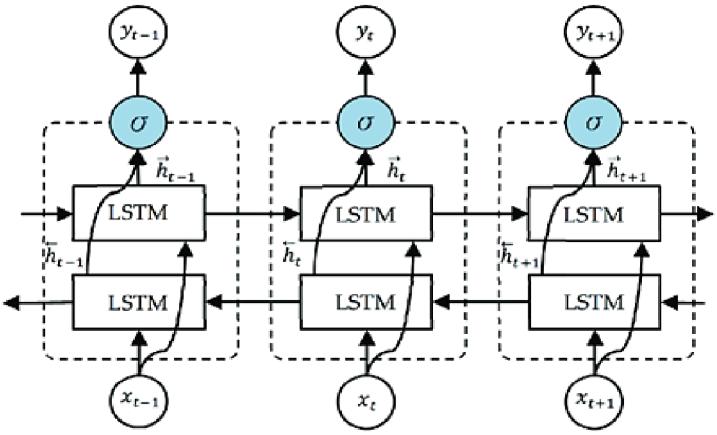


Fig. 2. BiLSTM Architecture [17]

3.2.3 Sentiment Analysis with BiGRU

The Gated Recurrent Unit (GRU) is a development of the RNN architecture because the RNN cannot record a lot of information at the time step, in addition to solving problems considering long dependencies and large text, the GRU architecture is suitable for use. GRU has two gates called Reset Gates and Updates Gates. Update gate is used to determine how much data will be discarded [18]. The two gates are used to control the flow of hidden information.

The GRU architecture has problems in the context that was used previously without considering the future context resulting in the loss of information obtained where the GRU architecture can only arrange the order from front to back [19]. Therefore, this is the basis for developing the Bidirectional Gated Recurrent Unit (BiGRU) architecture, where this architecture consists of two GRUs to take input in a forward direction and another GRU to take input in a backward direction. One of the advantages of this architecture is that the structure is not too complicated so BiGRU can train faster. The BiGRU model is obtained from replacing hidden layer neurons in the Bidirectional Recurrent Neural Network with GRU memory units, with the following structure (see Fig. 3).

The picture above shows the process that occurs in the BiGRU architecture which consists of two GRUs going back and forth. The first one functions as an input reader in the forward direction $(x_1, \ldots, x_n)$ resulting in a forward hidden state sequence $(\vec{h}_1, \ldots, \vec{h}_n)$, while the other functions as an input reader in the reverse order $(x_n, \ldots, x_1)$ and returns a sequence of backward hidden states $(\overleftarrow{h}_n, \ldots, \overleftarrow{h}_1)$. At time t, the hidden layer

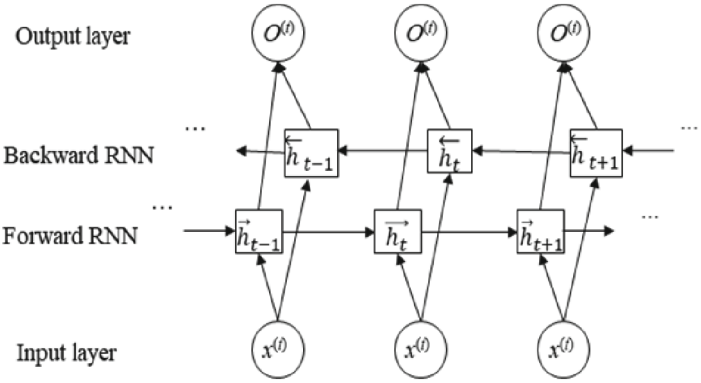


Fig. 3. BiGRU Architecture [20]

BiGRU returns h_t. The calculation process is as follows:

$$\vec{h}_t = \sigma\left(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}\right),$$

$$\vec{h}_t = \sigma(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\overleftarrow{h}_{t-1} + b_{\vec{h}}),$$

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t, \tag{1}$$

where W is the weight matrix connecting the two layers, b is the bias vector, σ is the activation function, $\vec{h}_t$ and $\overleftarrow{h}_t$ are the outputs of the forward and backward GRUs, and is the sum of elements.

3.2.4 Sentiment Analysis with CNN

The Deep Learning method is a method that is commonly used to solve a problem that is more complex than other methods because this method has many hidden layers that can help to solve various complex problems. One of the deep learning methods is the Convolutional Neural Network (CNN) (see Fig. 4). This method is generally used for image-based data analysis. However, now many researchers are using CNN to analyze public sentiment towards an object. Among the advantages when using CNN because it is easy to train and has connections are also fewer parameters compared to other algorithms [21].

3.2.5 Evaluation Matrix

To find out which algorithm is suitable for the problems we have, we need a tool that can calculate how well the performance of our classification model is. In this case, the confusion matrix (see Table 1) is a suitable tool to use. This table can display and compare the actual value with the predicted value of the model. The evaluation matrices commonly used include accuracy, precision, recall, and $F1$-score.

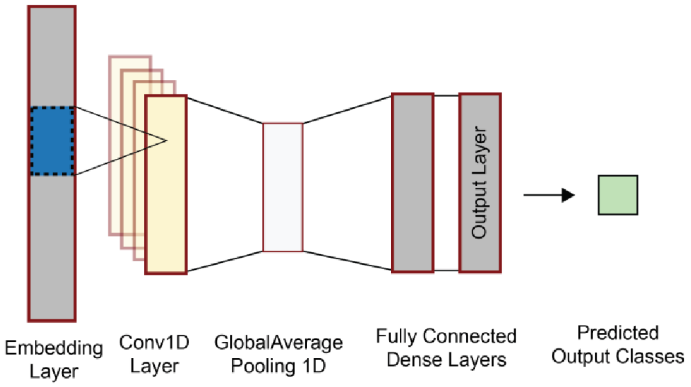


Fig. 4. CNN Architecture

Table 1. Confusion Matrix

	Positive	Negative
Positive	True Positive (*TP*)	False Positive (*FP*)
Negative	False Negative (*FN*)	True Negative (*TN*)

Accuracy

Accuracy is the value obtained from the sum positive data that is predicted to have positive value as well as negative data that is predicted to have a negative value with all the data in the dataset. The calculation of the accuracy value can be seen in the following equation:

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

The accuracy value is only suitable when the comparison of the actual number of data tables is relatively the same.

Precision

Precision is the value obtained from the probability that the data is predicted to have a positive value and, in fact, the data is included in the positive category. Precision may be defined as the degree of dependability of a model when the model provides a positive prediction. In computing accuracy, just the first or second line of the confusion matrix is necessary. Precision is the fraction of positively correct forecasts versus the overall positive prediction. Mathematically precision may be stated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

Recall

Recall or sensitivity is a strategy for determining how effectively a test can detect a

genuine positive; recall represents a model's success in rediscovering knowledge. Recall is calculated by comparing positive accurate predictions to total positive correct data. Mathematically the recall value may be expressed as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

F1-Score
Calculations that summarize accuracy and sensitivity/recall by calculating harmonic averages of each. The following formula may be used to calculate the value of the F1-Score:

$$\text{F1 - Score} = \frac{2\,x\,Precision\,x\,Recall}{Precision\,x\,Recall} \tag{5}$$

4 Results and Discussion

This section describes the data set used for experimentation and analysis of several models. The experiments are performed using the Keras Tensorflow library and the python language. Model comparison is done by comparing the accuracy validation obtained by each model. The network architecture is set homogeneously in each model used with the first layer being a vector embedding layer with dimensions of 1000 words to represent each word. Model tested is used in the second layer, for the CNN model an additional layer is needed for the pooling layer. The third layer uses a dense layer with ReLU activation and the last layer is an output layer with 3 neurons with Softmax activation. The number of epochs is set to 100 and Adam optimizer is used for optimization with learning rate set to 1e-4. Model checkpoint is applied by monitoring validation accuracy with mode max to save the best model.

4.1 Experimental

The data used in this study is the tweets with the keywords putin G20 and Russia G20 from April 22nd–30th, 2022. Dataset is obtained by crawling data from twitter using the tweepy library. This study mined the documents of about 12,462 tweets. Table 2 below shows the comparison of one example of a comment that has done through the text pre-processing stage.

The tweets have undergone changes in the text preprocessing. At this phase, the process of case folding and stopwords, non-alphanumeric and emojis removal is carried out. The results of the text reprocessing phase produce text data that are ready to enter the tokezining stage. The following is a data visualization of the results of text preprocessing.

Table 2. Sample of data preprocessing result based on problem

Raw Data	Data after Preprocessing
RT @AaronParnas: Vladimir Putin should not be allowed at this year's G20 summit	Vladimir putin not allowed g20
Brazil, China, South Africa and others still support Putin in G20 membership in spite of 'Ol Joe Biden. ✅	Brazil china south africa others still support putin g20 membership spite ol joe biden
RT @maythamdk: Indonesia invites Russian President Vladimir Putin to attend the G20 summit	Indonesia invites russian president vladimir putin attend g20

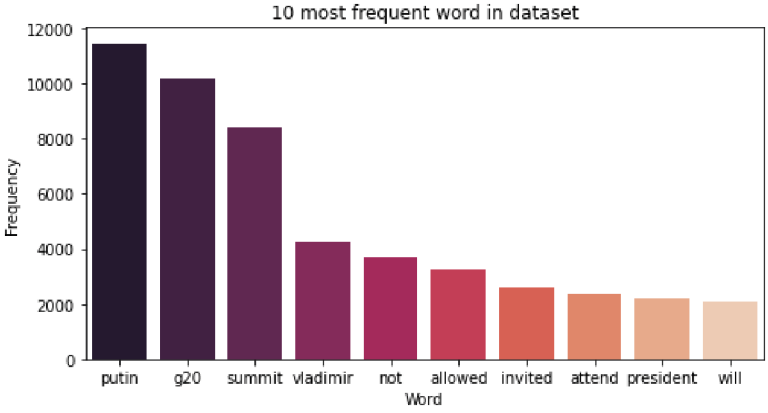


Fig. 5. Bar chart of the number of words that appear in dataset

The Fig. 5 above visualizes the number of words that appear in tweets related to Putin and the G20, the word 'putin' is the word with the highest number of occurrences, namely 11,374 words, followed by the word 'g20' with 9,679 words, "Vladimir" with 4,231 words, 'not' with 3,685 words, 'allowed' with 3,258, 'peak' with 2,863 words, the word 'invites' with 2,614 words, 'will' with 2,114 words, 'president' with 2,069 words, and the word 'invites' with 1,556 words. In this case, we can see that the words 'putin' and 'vladimir' have a different number of words, but in reality, we do not distinguish between the two words as having two different meanings. Because sometimes people write it using the full name "Vladimir Putin" or just "Putin".

The first step we do after collecting the data is labeling. We conducted our labeling using packages in Python, namely Vader. The package helps us designate 15213 tweets, forming three classifications for this case, namely positive, negative, and neutral. However, after that, we also did class validation of the labeling results manually for several sample tweets so that we were able to make clear that each tweet had entered the correct class. After carrying out several lengthy processes, such as the labeling stage on the data, followed by the data pre-processing stage such as data cleaning and case folding,

stopword removal, and tokenizing, the next step is the sentiment analysis process carried out using the BiLSTM, CNN, and BiGRU algorithms on tweets regarding Putin and the G20. All data that is ready to be analyzed is then separated by 70% for training data from all data (8,723 data) and 30% for validation data from all data (3,739 data). The distribution of training data and validation data for each sentiment can be seen in the following Table 3.

Table 3. Result of the dataset labels distribution

Sentiment	Training	Validation	Total
Neutral	4106	1760	5866
Negative	3445	1476	4921
Positive	1172	503	1675

Table 3 above shows the distribution sentiment label for the data. It appears that mostly the tweets' sentiments towards this topic are negative. This information can represent sentiments from the public of twitter regarding the topic of President Putin's participation in the G20 Summit, an international event which is being discussed around the world.

4.2 Performance Analysis

The best model is determined by making comparisons between several models.The Bidirectional Gated Recurrent Unit (BiGRU) is chosen to be the best model to be used in classifying tweets.

4.2.1 BiLSTM Model Performance

The BiLSTM model is used for testing using 100 epochs. The graph of validation loss and validation accuracy in Fig. 6 below shows the graph of validation loss showing a decrease and the graph of validation accuracy showing an increase in each epoch which indicates better results and does not experience overfitting.

The highest validation accuracy was obtained for the BiLSTM model at 0.98583 with a validation loss of 0.0589 which was achieved in the 65th epoch.

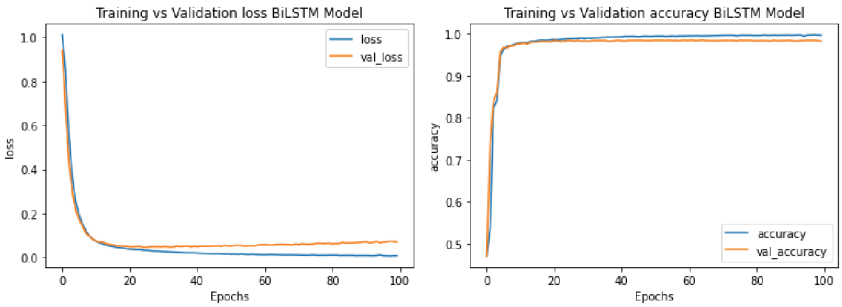


Fig. 6. Training and Validation Loss and Accuracy for BiLSTM

4.2.2 CNN Model Performance

The CNN model was also tested the same as other models with a maximum epoch of 100 epochs. The validation loss graph and the CNN model validation accuracy graph are the same as the previous model which shows good results and does not experience overfitting (see Fig. 7).

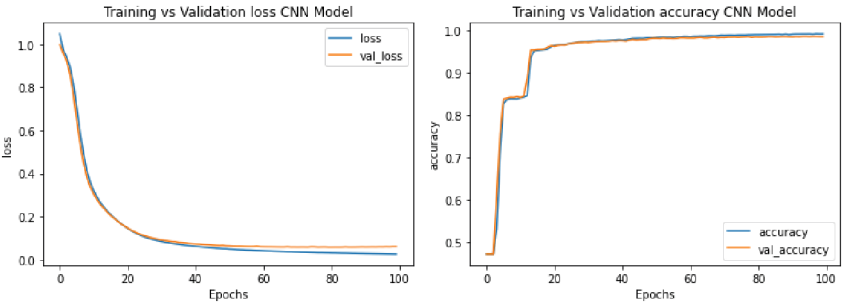


Fig. 7. Training and Validation Loss and Accuracy for CNN

In the 94th epoch, the CNN model got the highest validation accuracy at 0.98583 with a validation loss of 0.0583.

4.2.3 BiGRU Model Performance

Similar to other model tests, the BiGRU model was tested with a maximum epoch of 100 epochs. The graph in Fig. 8 below also shows the validation loss graph and the validation accuracy graph which are quite good and do not experience overfitting.

The BiGRU model gets the highest validation accuracy at 0.98636 with a validation loss of 0.0534 which can be achieved in the 29th epoch.

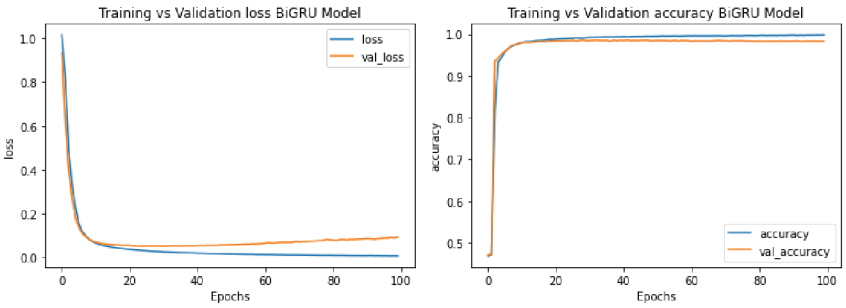


Fig. 8. Training and Validation Loss and Accuracy for BiGRU

4.3 Classification Results

The Confusion Matrix in Table 4 shows that the data that was predicted to be correct with a neutral class amounted to 1743 data, correctly predicted with a negative class totaling 1461 data, and correctly predicted with a positive class totaling 484 data. The labels that are predicted according to the actual label are represented by the main diagonal of the confusion matrix, while the elements outside the diagonal of the confusion matrix are the elements that are mislabeled by the machine learning model. A high diagonal value in the confusion matrix indicates the selected model shows good performance in classifying sentiments.

The comparison of our model with CNN and BiLSTM models to evaluate the model and conclude the results. Results of other models along with our models are tabulated in Table 4.

Table 4. Comparison of different evaluation measures of BiLSTM, CNN, and BiGRU

Model	Evaluation Metric			
	Accuracy	Precision	Recall	F1 – score
BiLSTM	0.98449	0.98	0.976	0.98
CNN	0.98583	0.98	0.98	0.98
BiGRU (our proposed model)	0.98636	0.983	0.98	0.983

The results of the classification of our proposed model as reported in Table 4 are the best among other models, this shows that our proposed model is very promising in this dataset with an accuracy of 0.98636.

The Performance Metrics in Table 5. Shows the evaluation of sentiment classification carried out using the BiGRU architecture. Sentiment classification is performed on the data used with an accuracy score of 98%. The neutral class and the negative class got an f1 score of 99%, higher than the positive class which got an f1 score of 97%. These results mean that the model used can predict the category of tweets with neutral and negative

Table 5. Performance matrix of each classes for BiGRU architecture

	Neutral	Negative	Positive
Precision	0.99	0.99	0.97
Recall	0.99	0.99	0.96
F1 – score	0.99	0.99	0.97
Accuracy		0.98636	

sentiments more accurately than the categories of tweets with positive sentiments. Table 6 is a Confusion matrix which shows data that is correctly predicted with a neutral class of 1743 data, correctly predicted with a negative class of 1461 data, and correctly predicted with a positive class of 484 data.

Table 6. Confusion matrix for BiGRU architecture

Confusion Matrix		Predicted		
		Neutral	Negative	Positive
Actual	Neutral	1743	11	6
	Negative	8	1461	7
	Positive	8	11	484

The labels that are predicted according to the actual label are represented by the main diagonal of the confusion matrix, while the elements outside the diagonal of the confusion matrix are the elements that are mislabeled by the machine learning model. A high diagonal value in the confusion matrix indicates the selected model shows good performance in classifying sentiments.

5 Conclusion and Recommendation

In this study, we have successfully compared three deep learning algorithms, i.e., BiLSTM, CNN, and BiGRU. We got BiGRU as the best algorithm with an accuracy of 98.63% and a loss of 0.053, so in this case, the BiGRU algorithm is considered the most suitable for use in predicting public sentiment on the issue of Putin's participation in the G20 in Indonesia. This study used 12,462 data tweets from April 22, 2022 to April 30, 2022. To improve performance results, additional datasets and data collection ranges can be added. This research can find out the actual public response to Putin's participation in the G20. The results of this study can be a source of reference for the government in considering the decision to invite Putin to the G20 forum.

The recommendation that can be given in future research is to compare other deep learning algorithms with traditional algorithms. We can find out whether traditional algorithms can handle sentiment analysis on the issue of Putin's participation in the G20 forum. Utilizing better labeling methods and collecting datasets over a longer period of time is also more likely to improve this research.

Acknowledgement. The authors thanks to the Department of Statistics and Research Center for Artificial Intelligence and Big Data, Universitas Padjadjaran which supports this research. The work of Tutut Herawan and Eka Novita Sari is supported by AMCS Research Center.

References

1. Ministry of Communication and Information Technology: G20pedia (2022)
2. Ananda, F.D., Pristyanto, Y.: Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine. MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer **20**(2), 407–416, (2021)
3. Qiu, L., Lin, H., Ramsay, J., Yang, F.: You are what you tweet: personality expression and perception on Twitter. J. Res. Pers. **46**(6), 710–718 (2012)
4. Islam, M.S., Ghani, N.A.: A novel BiGRUBiLSTM model for multilevel sentiment analysis using deep neural network with BiGRU-BiLSTM. In: Recent Trends in Mechatronics Towards Industry 4.0: Selected Articles from iM3F 2020, Malaysia, pp. 403–414. Springer Singapore, Singapore (2021)
5. Behrendt, S., Schmidt, A.: The Twitter myth revisited: intraday investor sentiment, Twitter activity and individual-level stock return volatility. J. Bank. Finance **96**, 355–367 (2018)
6. Caprolu, M., Sadighian, A., Di Pietro, R.: Characterizing the 2022 Russo-Ukrainian Conflict Through the Lenses of Aspect-Based Sentiment Analysis: Dataset, Methodology, and Preliminary Findings. arXiv preprint arXiv:2208.04903 (2022)
7. Jabreel, M., Hassan, F., Moreno, A.: Target-dependent sentiment analysis of tweets using bidirectional gated recurrent neural networks. In: Advances in Hybridization of Intelligent Methods: Models, Systems and Applications, pp. 39–55 (2018)
8. Pratama, B.Y., Sarno, R.: Personality classification based on Twitter text using Naive Bayes, KNN and SVM. In: 2015 International Conference on Data and Software Engineering (ICoDSE), pp. 170–174. IEEE (2015)
9. Zerrouki, K., Hamou, R.M., Rahmoun, A.: Sentiment analysis of tweets using Naïve Bayes, KNN, and decision tree. In: Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines, pp. 538–554. IGI Global (2022)
10. Firmansyah, I., Asnawi, M.H., Hasanah, S.A., Novian, R., Pravitasari, A.A. A comparison of support vector machine and naïve Bayes classifier in binary sentiment reviews for PeduliLindungi application. In: 2021 International Conference on Artificial Intelligence and Big Data Analytics, pp. 140–145. IEEE (2021)
11. Chen, Y.: Convolutional neural network for sentence classification. Master's thesis, University of Waterloo (2015)
12. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
13. Xu, G., Meng, Y., Qiu, X., Yu, Z., Wu, X.: Sentiment analysis of comment texts based on BiLSTM. IEEE Access **7**, 51522–51532 (2019)
14. Yin, X., Liu, C., Fang, X.: Sentiment analysis based on BiGRU information enhancement. J. Phys. Conf. Ser. **1748**(3), 032054. IOP Publishing (2021)

15. Hameed, Z., Shapoval, S., Garcia-Zapirain, B., Zorilla, A.M. Sentiment analysis using an ensemble approach of BiGRU model: a case study of AMIS tweets. In: 2020 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) (2021)
16. Jia, L., Jiang, T., Meng, J.H., Zhang, T.: Tibetan text classification method based on BiLSTM model. In; 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), pp. 27–31. IEEE (2020)
17. Li, Y.H., Harfiya, L.N., Chang, C.C.: Featureless blood pressure estimation based on photoplethysmography signal using CNN and BiLSTM for IoT devices. Wirel. Commun. Mob. Comput. **2021**, 1–10 (2021)
18. Yu, Q., Zhao, H., Wang, Z.: Attention-based bidirectional gated recurrent unit neural networks for sentiment analysis. In: Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition, pp. 116–119. ACM Press (2019)
19. Abdelgwad, M.M., Soliman, T.H.A., Taloba, A.I., Farghaly, M.F.: Arabic aspect based sentiment analysis using bidirectional GRU based models. J. King Saud Univ.-Comp. Inf. Sci. **34**(9), 6652–6662 (2022)
20. Ouyang, X., Zhou, P., Li, C.H., Liu, L.: Sentiment analysis using convolutional neural network. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 2359–2364. IEEE (2015)
21. Shi, L., Du, K., Zhang, C., Ma, H., Yan, W.: Lung sound recognition algorithm based on Vggish-Bigru. IEEE Access. **7**, 139438–139449 (2019)