Sentiment Analysis of Twitter's Opinion on The Russia and Ukraine War Using Bert

Muhammad Fahmi Julianto¹, Yesni Malau²*, Wahyutama Fitri Hidayat³

Teknik Informatika Kampus Kota Pontianak¹, Sistem Informasi Kampus Kota Pontianak³
Universitas Bina Sarana Informatika
fahmi.fjl@bsi.ac.id, wahyutama.wfh@bsi.ac.id

Teknik Elektro²
Universitas Bina Sarana Informatika
²*yesni.ymu@bsi.ac.id

(*) Corresponding Author

Abstract

News about the war between Russia and Ukraine can not be denied affecting various aspects of life worldwide. It affects the writings of every citizen of the world on various social media platforms, one of which is Twitter. Sentiment analysis is a process of identifying and making sentiment categories computationally. The sentiment analysis process is also intended to make computers understand the meaning of human sentences by processing algorithms. This research uses the deep learning method of the BERT (Bidirectional Encoder Representation Form Transform) model language to analyze the sentiments in the tweets written about the wars between Russia and Ukraine by Twitter social media users. The sentiment will be divided into three parts: positive, neutral, and hostile. The hyperparameters in this study used ten epochs, with a learning rate of 2e-5 and a batch size of 16. The test used in sentiment analysis was the BERTbase Multilingual-cased-model model, and the accuracy was 97%. Suggestions for further research are the need for a more balanced dataset between positive, neutral, and negative sentiments. They reward the dataset before training so that better results are expected.

Keywords: War, Russia-Ukraine, Sentiment Analysis, BERT

INTRODUCTION

Conflicts inevitably occur in everyday life, big or small. Many things can cause conflict, but conflict often arises because of differences in interests and can also be caused by domination and the desire to dominate.

The conflict between Russia and Ukraine occurred over Crimea, Eastern Ukraine. Crimea itself has been a struggle for centuries. It looks at its history. Crimea was formerly known as Tauris or as
Tavrida. By Russian society, the region is considered home to various ethnic groups. However, the area began to be contested after the Cimmerian tribes invaded Tauris, and this action also triggered Greek colonists to enter the Tauris region in the 6th century BC (BC) (Sudiq & Yustitianingtyas, 2022).

Russia’s invasion began on February 24, 2022, to defend the country’s security from the threat of Ukraine. Russia is trying to limit Ukraine’s proximity to the European Union and NATO to safeguard its security from the threat of the western bloc and influence from the United States and maintain Russia’s close relations with the former Soviet Union. It is related to the position of the territory of the Ukrainian state, which is directly adjacent to Russia, so if Ukraine joins NATO, there will no longer be a barrier between Russia and NATO, and this is a threat to Russia. That is why Russia invaded Ukraine so that Ukraine could not get in touch with NATO. Russia preventing Ukraine from joining NATO is an attitude that should not be done because it leads to intervention. As a sovereign country, Ukraine should have the freedom to make decisions about its membership in NATO, and Russia should not be able to affect that freedom with physical pressure in the form of an invasion (Sudiq & Yustitianingtyas, 2022).

The incident caused much discussion in the community. The response is widely expressed on social media. One of the social media used to conduct these conversations is Twitter. Twitter is considered suitable for conducting a discussion, where on Twitter, and people can express their responses in the form of threads or reply to each other’s tweets from other accounts.

In previous research related to this research, research conducted by (Andraini & Mahdiyah, 2022) conducted research on Twitter users regarding the Russia-Ukraine War using the SVM method with various optimizations, including cross-validation. It split validation resulting in an accuracy of 86.0%. Meanwhile, the research (Aditya & Wibowo, 2022) uses the Naive Bayes method, and the sentiment analysis system is divided into 5 (five) parts, namely crawling, preprocessing, word weighting, model formation, and sentiment classification. Existing data will be grouped into 2 (two) categories, negative and positive, resulting in an accuracy value of 78.2%.

Based on this background, the author will use a deep learning method using a language model and bidirectional encoder representation Form transformers (BERT) to analyze the sentiments of Twitter users regarding the Russia-Ukraine war writings. The sentiments will be divided into three parts: Positive, Neutral, and Negative.

**RESEARCH METHODS**

The method used in the dataset taken from twitter this time uses the Bidirectional Encoder Representations from Transform method and has the following steps as shown in Figure 1.

![General Architecture](image)

Figure 1. General Architecture

Sentiment analysis begins with collecting datasets, where the datasets are taken on the kaggle.com page. Then, the dataset will be labeled with positive, neutral, and negative names. Then after being labeled, it will enter the next stage, namely preprocessing the dataset, where this stage is the stage to prepare the dataset will later be processed by doing data cleaning and tokenization. The dataset that has gone through this process will then be trained to get a positive, neutral and negative classification using the BERT model. The classification results will then be evaluated to obtain results.

Collecting dataset data is the least paid attention component of the ML process (Wiegrefe & Marasović, 2021). Selection of optimal features in determining sentiment where the content in the form of online text greatly influences the superior classification results. Selection of optimal features
can also affect computational performance, which is very difficult and causes the need to design new techniques to improve classification performance (Kumar, Jaiswal, Garg, Verma, & Kumar, 2022). The dataset collected by Kaggle was adopted in this study to create a collection of tweets that match the discussion material.

Dataset labeling marks specific texts or documents with appropriate labels, known as text categorization or classification (El Rifai, Al Qadi, & Elnagar, 2022). At this stage, text lines are extracted as an integration process of the pre-classification method with the segmentation and deep-learning methods so that each text is labeled with a label.

The next stage is data preprocessing, a crucial step before classification (Yutika, Adiwijaya, & Faraby, 2021). There are two preprocessing stages, namely data cleaning and tokenization which can be seen in Figure 1. Data cleaning can also be referred to as scrubbing, which functions as detection and removes incorrect or inconsistent words to improve the quality of the data used (Lomet, 2001). Raw data requires processing first, which serves to improve data quality (Fatima, Nazir, & Khan, 2017). Tokenization is the process of separating the upward flow of text into parts, symbols, or other meaningful elements, while the goal of tokenization is the exploration of words in a sentence.

BERT was first introduced in 2018. BERT is a representation of the word bidirectional. The previous model only applies a one-way representation by looking at the word order from right to left or a mixture of right to left and left to right (Devlin et al., 2019). As the name implies, BERT uses to transform. Transform is a mechanism to determine the contextual relationship between several words in a text (Suresha, 2020). Transform can understand and transform the understanding obtained through a mechanism known as the self-attention mechanism. Transform’s a way of translating the "understanding" of several related words into several words that are processed in the mechanism. The transformation has two mechanisms, namely, Encoder and the Decoder. An illustration of the Encoder and Decoder can be seen in Figure 2.

The steps involved in the encoder and decoder are:
1. Every input word that goes to the encoder is converted into a vector list using embedding.
2. Vector input through the two existing layers in all encoders, the self-attention layer, and the forward-looking neural network.
3. When every process in the encoder is complete, the encoder output, key, and vector value are sent to the decoder. From the explanation of points one to point three, it can be illustrated based on Figure 3.
As the name suggests, BERT uses only one encoder. Therefore, the BERT architecture is similar to Figure 4. BERT differs from directional models in that it displays text strings in right-to-left, left-to-right, or right-to-left and left-to-right combinations. The model used in the two-way training language can be used for understanding in context so that it can be more profound than the one-dimensional language model. Figure 5 shows a comparison of the BERT architecture using OpenAI GPT and ELMo. Of the three architectural models, only BERT controls each level’s left and proper context.

Source: (Alammar, 2019)
Figure 4 BERT Architecture

![BERT Architecture](image)

Source: (Devlin, Chang, Lee, & Toutanova, 2019).
Figure 5. Differences in the Architecture of BERT, OpenAI GPT, and ELMo

Formula:

\[ f_{\text{baud}} = \frac{2^{\text{SMOD}}}{64} \times f_{\text{osc}} \]  

Before the BERT method conducts training, the dataset needs to be adjusted based on the input representation that BERT will receive using Formula 1. That is why we need a tokenizer where the goal is to tokenize sentences so that the input can match. The tokenization process is illustrated based on Figure 6.

Source: (McCormick & Ryan, 2019)
Figure 6. Tokenization Process

![Tokenization Process](image)

Source: (McCormick & Ryan, 2019)
Figure 7. Input and Output on BERT

BERT will later receive a series of words or sentences used as input and follow up on the encoder stack. Self-attention will be sent to each encoder to provide results through a feed-forward network, which is then continued by the next encoder. The selected model in this study is BERTBASE so this process will be repeated 12 times. After passing through all the encoders, tokens at each position will generate a vector value with a hidden size of 768 in BERTBASE, as shown in Figure 7. For the sentiment analysis process, the output considered is the first position, the [CLS] token. This vector is used as the input of the classifier.

Source: (McCormick & Ryan, 2019)
Figure 8. Illustration used to perform sentiment analysis

![Sentiment Analysis](image)

Source: (McCormick & Ryan, 2019)
Figure 9. BERT can achieve good results using only one neural network as a classifier (Devlin et al., 2019). The neural network layer is used for classification with the softmax function shown in Figure 10. Therefore, the BERT output used for classification is derived from the [CLS] token vector. It is because the [CLS] token is supposed to collect the average of the word tokens and take the sentence vector. The last level of the classification level generates a logit. Logit is generated as a result of rough probability prediction of rank sentences. Softmax takes the exponent of each logit value and converts this logit into a probability so that the sum of the probabilities is precisely 1. Therefore, the probability value can be either 0 or a positive number. The illustration used to perform sentiment analysis can be seen in Figure 8.
After that, it enters the training process, where in this process, the dataset will be divided into three parts, namely the dataset for training, testing, and validation. This process aims to reduce the lengthy process due to large datasets. At this stage, a data loader is also needed to help reduce the memory used so that it can speed up the training process. This data loader is a function of PyTorch which will act as an iterator. The data loader will also create three parts: training, validation, and testing. At this stage, the AdamW optimizer will also be used to correct the weight of the sentence.

In the training phase, BERT will perform fine-tuning using predefined hyperparameters. Refers to (Devlin et al., 2019) hyperparameters include Batch size: 16, 32, Learning Rate (Adam): 5e-5, 3e-5, 2e-5, Epoch: 2, 3, 4

RESULTS AND DISCUSSION

Collecting Dataset

The author of this study uses public datasets that all parties can access. The dataset can use EDA ON the Data, Sentimental Analysis on the Data, and Build a Model that can predict the Impact on the Different Nations, To which people are Supporting More, and What People are thinking about the war. The dataset can be accessed via https://www.kaggle.com/datasets/vanamayaswath/russia-vs-ukraine-tweets.

Labelling Dataset

Sentiment analysis uses supervised learning methods, so it requires a dataset that already has a label. Labeling aims to determine reviews into neutral, negative, and positive categories. This labeling is needed because the supervised learning method requires examples to generate generalizations until the model obtained is a prediction that matches the existing label (Goldberg, 2017). The model can see and understand which comments have positive, neutral, and negative labels.

The next stage is labeling the dataset. Labels are divided into 3, namely positive, neutral and negative. Where to get positive, neutral, and negative sentiments are grouped first into 0.1 and 2. The cleaned dataset is then labeled. The results of labeling data using 0, 1, and 2 can be seen in Figure 9. While the results of labeling using negative, neutral, and positive can be seen in Figure 10.
The preprocessing stage at this stage has several processes, namely, tokenization, unique tokens, and choosing a sequence length.

1. **Tokenization**

   BERT's task is to classify two sentences. Tokenization is simple in this graph because BERT uses WordPieces as tokens rather than words --- so some words are broken down into simpler parts. (Alammar, 2021). Some basic operations can convert the text to tokens and tokens to unique integers (ids) shown in Figure 13:

   ![Figure 13. Process tokenization](image)

2. **Special Tokens**

   [SEP] - at the end of the sentence, [CLS] - so a token must be added at the beginning of the sentence so that BERT can run the classification. (Valkov, 2020) The results of removing unique signs are shown in Figure 14, namely cleaning special characters with sep_token of 102 characters, cls_token of 101 characters, pad_token of 0 characters, and unk_token of 100 characters:

   ![Figure 14. result of special tokens](image)
3. Choosing Sequence length

Sequence length is one of the methods used as choosing sequence length based on data that has been cleaned so as to produce a comparison of the taken count and density as shown in Figure 15.

![Figure 15. Result of choosing sequence length](image)

**Figure 15. Result of choosing sequence length**

**BERT Implementation**

To be able to do a test, it takes a data loader for each dataset to be able to iterate. The goal is to maintain memory during training so that all datasets do not need to be entered into memory simultaneously. For that, we also need a block that is useful for creating a data loader that will later generate tokenized comments. Sentiments and comments can be as long as 160 words. Researchers do fine-tune using hyperparameters using the recommendations from BERT as follows: Batch Size: 16, Epoch: 10, Learning Rate 2e-5 (0.0005)

The hyperparameters used are determined for several reasons, so batch 16 is chosen. The bigger the batch size, the longer it will take (Osinga, 2018). Then use epochs as much as 10 (Song, Wang, Liang, Liu, & Jiang, 2020). In addition, the learning rate of 2e-5 was chosen because it can overcome the problem of forgetting disasters (Sun, Qiu, Xu, & Huang, 2019). The problem of understanding obtained from pre-training is removed while examining new data or information.

After determining the Hyperparameters, testing will be carried out using ten epochs. The dataset used in each training, validation, and testing is done without specifying a random seed. So that every time the test is carried out, the data is different. The results of the epochs test are shown in Figure 16.

![Figure 16. The training process and dataset evaluation](image)

**Figure 16. The training process and dataset evaluation**

After repeating the epoch on the model, the result that gets the best value will be used. Figure 18 compares the training and validation of the test shown in Figure 17.

![Figure 17. The curve of training and validation](image)

**Figure 17. The curve of training and validation**

From the experimental observations, it can be concluded that the training data has better results than the validation data. Training has a stable upward curve, while validation has a slightly up-and-down curve.

**Training Dataset**

The dataset is divided into a training, validation, and testing dataset before classification. Train the model using the training dataset. Meanwhile, to minimize overfitting, a validation dataset is used. As a final test to obtain network accuracy that has been trained with the training dataset, dataset testing is used. The results of the dataset training process in the form of a split dataset can be seen in Figure 18.
The experimental results show the testing of the Twitter opinion dataset with the BERT model. The stages that will be carried out in the experiment are testing using a dataset divided into training, validation, and testing, which will later get the classification results from the desired sentiment.

**Evaluation**

BERT is used for sentiment analysis, using a dataset to test it based on the best accuracy value. After the accuracy value is obtained using dataset testing shown in Figure 19, the BERT model of the test data set is used to make predictions resulting in an accuracy value of 97%, as shown in Figure 20.

Based on the confusion matrix diagram in Figure 21, it can be concluded that the neutral detecting system has a value that tends to be more significant.

**CONCLUSIONS AND SUGGESTIONS**

**Conclusion**

Sentiment analysis using Bidirectional Encoder Representations Transform (BERT) produces an accuracy of 97% with batch size 16 selection of hyperparameters, learning rate 2e-5, and epoch 10. These results can be said to be good results. The following conclusions are obtained based on the testing implementation of the BERT model in sentiment analysis. The results show more neutral sentiments than positive and negative sentiments.

**Suggestion**

The authors propose several suggestions for further research from the results, including a more balanced dataset between Positive, neutral, and negative sentiment. They are balancing the dataset before training so that better results are expected.

**REFERENCES**


Yaswanth, V. (2022). Russia Vs Ukraine Tweets: Tweets of the war that is happening between the Nations.
