

INTISARI

Pada tahun 2013 konflik Rusia dan Ukraina sudah sering kali terjadi dan pemberitaannya sudah ada di berbagai macam media sosial di seluruh dunia, salah satu media sosial dalam penelitian ini yaitu media sosial *twitter*, banyak sekali komentar masyarakat di media sosial *twitter* yang membahas tentang konflik Rusia dan Ukraina dengan menggunakan Bahasa Inggris. Namun, sentimen pada komentar di *twitter* bervariasi dari komentar positif, netral, maupun negatif membutuhkan suatu otomatisasi dalam mengklasifikasikan komentar berdasarkan polaritas sentimennya.

Analisis sentimen dengan penggunaan algoritma machine learning tradisional seperti Naïve Bayes, SVM, dan lain-lain tidak dapat memahami konteks dari komentar secara mendalam tentang semantik kata yang ada karena hanya mempelajari pola-pola yang diberikan seperti frekuensi kemunculan kata. Untuk itu dibutuhkan sebuah pendekatan *transfer learning* seperti BERT (*Bidirectional Encoder Representation from Transformers*) yang menghasilkan sebuah model bahasa dua arah (*bidirectional*) dan LDA (*Latent Dirichlet Allocation*) yang menghasilkan daftar topik. Dataset yang digunakan melalui tahap *preprocessing* yang terdiri dari *case folding*, *data cleaning*, *tokenisasi*, *stopword removal*, dan *stemming* sebelum dilakukan analisis sentimen. Dalam penelitian ini *hyperparameters* yang digunakan adalah 10 *epoch* dan *batch size* 16. Setelah dataset sudah di *preprocessing* kemudian dilakukan pembentukan *corpus*, *dictionary*, pembentukan model LDA, dan hasil model LDA.

Berdasarkan pengujian algoritma BERT didapatkan hasil prediksi akurasi analisis sentimen dengan perbandingan jumlah data benar dengan jumlah data total dari masing-masing kelas yaitu netral (64%), positif (28%), dan negatif (8%). Kemudian hasil evaluasi akurasi dengan *hyperparameters* yang digunakan 10 *epoch* dan *batch size* 16 didapatkan akurasi sebesar 71% dan hasil LDA didapatkan daftar topik sebanyak 7 topik. Dengan melakukan tahapan *corpus*, *dictionary*, dan pembentukan LDA.

Kata Kunci: Analisis Sentimen, Algoritma BERT, Pemodelan Topik LDA

ABSTRACT

In 2013, the conflict between Russia and Ukraine occurred frequently and the news was available on various social media throughout the world, one of the social media in this research was Twitter social media, there were lots of public comments on Twitter social media discussing the Russian conflict. and Ukrainian using English. However, the sentiment of comments on Twitter varies from positive, neutral, or negative comments requiring automation in classifying comments based on the polarity of their sentiment.

Sentiment analysis using traditional machine learning algorithms such as Naïve Bayes, SVM, etc. cannot understand the context of comments in depth about the semantics of existing words because they only learn given patterns such as the frequency of word occurrences. This requires a transfer learning approach such as BERT (Bidirectional Encoder Representation from Transformers) which produces a two-way language model (bidirection) and LDA (Latent Dirichlet Allocation) which produces a list of topics. The dataset used goes through a preprocessing stage consisting of casefolding, data cleaning, tokenization, stopword removal, and stemming before sentiment analysis is carried out. In this research, the hyperparameters used are 10 epochs and batch size 16. After the dataset has been preprocessed, the corpus, dictionary, LDA model formation and LDA model results are carried out.

Based on testing the BERT algorithm, the accuracy prediction results of sentiment analysis were obtained by comparing the amount of correct data with the total amount of data from each class, namely neutral (64%), positive (28%), and negative (8%), then the results of the accuracy evaluation using hyperparameters. 10 epochs were used and a batch size of 16 obtained an accuracy of 71% and the LDA results obtained a list of 7 topics. By carrying out the corpus, dictionary and LDA formation stages.

Keywords: *Sentiment Analysis, BERT Algorithm, LDA Topic Modeling*