

**REPRESENTASI BIDIRECTIONAL BERBASIS *DEEP LEARNING*
DENGAN METODE BERT DAN LSTM UNTUK MENGANALISIS
UJARAN KEBENCIAN PADA PENGGUNA MEDIA SOSIAL DI
INDONESIA**

ABSTRAK

Penggunaan media sosial untuk berkomunikasi sesama warga pengguna internet (*netizen*) sudah sangat umum dan menyebar ke seluruh lapisan masyarakat Indonesia. Pada penelitian ini, akan diimplementasikan algoritma Long Short-Term Memory (LSTM) pada kasus ujaran kebencian, khususnya yang berkaitan dengan ujaran kebencian dengan pemerintah dan juga menggunakan BERT. Teknik pencarian data set menggunakan web crawler yang diambil dari Twitter API untuk 5000 kata yang akan dianalisis apakah mengandung Hate Speech (HS) atau Non Hate Speech (NHS) dengan pemrograman python.

Hasil penelitian menunjukkan bahwa Analisis terhadap ujaran kebencian (*hate speech*) dengan metode *Bidirectional Encoder Representations dari Transformers* 86% menghasilkan tingkat akurasi yang lebih tinggi sedikit dibanding Long Short Term Memory yaitu 85,2%. menghasilkan tingkat akurasi yang lebih tinggi sedikit dibanding Long Short Term Memory yaitu 85,2%. Hal ini disebabkan oleh pemilihan *hyperparameter* yaitu *batch size* 32, *learning rate* 3e-6, dan *epoch* 5.

Untuk penelitian selanjutnya, hasil pada penelitian ini berpeluang untuk ditingkatkan dengan menambah jumlah data latih maupun data uji serta khususnya untuk kalimat yang bermakna ambigu, atau berupa bahasa informal dan singkatan yang umum digunakan di media sosial. Penelitian selanjutnya juga dapat menggunakan metode *k-fold cross validation* untuk menentukan akurasinya.

Kata Kunci : representasi *bidirectional*, *deep learning*, BERT dan LSTM, ujaran kebencian, media sosial

**BIDIRECTIONAL REPRESENTATION BASED ON DEEP LEARNING
USING BERT AND LSTM METHODS TO ANALYZE HATE SPEECH ON
SOCIAL MEDIA USERS IN INDONESIA**

ABSTRACT

The use of social media to communicate with fellow internet users (netizens) is very common and has spread to all levels of Indonesian society. In this research, the Long Short-Term Memory (LSTM) algorithm will be implemented in cases of hate speech, especially those related to hate speech against the government and also using BERT. The data set search technique uses a web crawler taken from the Twitter API for 5000 words which will be analyzed whether they contain Hate Speech (HS) or Non Hate Speech (NHS) using Python programming.

The research results show that analysis of hate speech using the Bidirectional Encoder Representations method from Transformers 86% produces a slightly higher level of accuracy than Long Short Term Memory, namely 85.2%. produces a slightly higher level of accuracy than Long Short Term Memory, namely 85.2%. This is caused by the selection of hyperparameters namely batch size 32, learning rate $3e-6$, and epoch 5.

For further research, the results of this study have the opportunity to be improved by increasing the amount of training data and test data, especially for sentences that have ambiguous meanings, or in the form of informal language and abbreviations that are commonly used on social media. Future research can also use the k-fold cross validation method to determine its accuracy.

Keywords: *bidirectional representation, deep learning, BERT and LSTM, hate speech, social media*