

INTRODUCTION

Hate speech is a phenomenon that often occurs on various social media platforms. Hate speech can trigger divisions, misunderstandings, and even acts of violence between individuals or groups, especially because of the prejudice it causes. Social media allows users to share opinions and information freely. [1]. Social media is also useful for obtaining data to see people's perceptions of a topic that is circulating. For example, a topic that is quite popular on twitter regarding the Russian and Ukrainian wars, then covid-19 which spread throughout the world. During the Covid-19 pandemic, social media has been helpful for policymakers to strategize by understanding the responses given by the public. [2], [3]. Social media can also illustrate the rating of a product or service. This is useful for companies to evaluate their products based on public opinion[4], [5], [6]. However, this freedom of speech and lax regulations have increased the spread of hate speech [7], which can have a significant negative impact on certain individuals and groups of people

The research adopts BiLSTM and Word2Vec models with continuous bag of words (CBOW) architecture. There are also studies that combine global vector algorithm with deep belief network. The use of GloVe in previous studies showed good accuracy; this study combined long short-term memory (LSTM) with GloVe word embedding for hate speech classifier, with accuracy reaching 81.5% in one-layer LSTM and 80.9% in two-layer LSTM. However, there are difficulties in identifying slang words and local languages. In addition, the model also misinterpreted some words in the context of animal names, rather than as hateful statements. Another study that combined fastText with GRU in classifying offensive and inoffensive texts showed a fairly good accuracy of 84%. The drawback of this model is that it does not fully capture the context of the sentence, and its performance is highly dependent on proper hyperparameter tuning, which can take significant time and computational resources. [8], [9], [10], [11], [12].

Although many studies have been conducted, there is still room for improvement in hate speech detection. In this study, the use of gated recurrent unit (GRU) and word embedding global vector (GloVe) methods is expected to provide good accuracy in hate speech detection. GRU is a type of recurrent neural network architecture designed to overcome several problems in traditional recurrent neural networks, especially problems related to vanishing gradient and exploding gradient. GRU introduces a "gate" structure to control the flow of information in and out of the hidden units which allows the model to better retain which relevant information to retain or ignore[13], [14]. And GloVe is a word representation learning model that combines the strengths of count-based and predictive methods. It uses global word co-occurrence statistics to generate word vectors that capture meaning and semantic relationships. [15].

Based on previous research with the combination of fasttext-GRU and GloVe-LSTM, This research focuses on the use of deep learning methods GRU, GloVe as word embedding and implementing an informal word converter into a formal word that is expected to have better performance in detecting hate speech.