Deteksi Ujaran Kebencian Menggunakan CNN dan BiGRU dengan Mekanisme Attention di Twitter

Qomarudin Sifak[1], Erwin Budi Setiawan[2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]qomaru@students.telkomuniversity.ac.id, [2]erwinbudisetiawan@telkomuniversity.ac.id

---

Abstract
Indonesia is the country that leads the most use of social media in Asia. Twitter is one of the most popular social media platforms in Indonesia. Twitter is a social media that is typically used for opinion exchange, criticism, and storytelling. This is frequently abused by Internet users, including the dissemination of hate speech directed at an individual or groups. Therefore, the aim of this research is to address these issues by building a hate speech detection system for Indonesian Twitter. This research uses 69.484 tweet data and implements hybrid deep learning models with Convolutional Neural Network (CNN) and Bidirectional Gated Recurrent Unit (BiGRU). Other methods applied are Bidirectional Encoder Representation from Transformer (BERT) used as word embedding to help the system better understand the context and meaning of tweets and attention mechanism to help the system to find the essential word from tweets. This research examines eight hybrid approaches in classification process, either CNN-BiGRU or BiGRU-CNN and adds attention mechanisms for those models. The result shows BIGRU-CNN hybrid model with attention mechanism for both layers achieving the highest accuracy of 88.12%.

Keywords: hate speech, twitter, convolutional neural network, bidirectional gated recurrent unit, attention mechanism