Ekspansi Fitur dengan *GloVe* untuk Deteksi Ujaran Kebencian Menggunakan Metode *Convolutional Neural Network (CNN)* dan *Recurrent Neural Network (RNN)* di Twitter

Vincent Williams Jonathan[1], Erwin Budi Setiawan [2]

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

Abstract

In this day and age, people have easy access to social media, allowing them the freedom to express their opinions. Consequently, hate speech against individuals or groups can be easily found on social media platforms. This paper proposes a system to classify hate speech in Indonesian language tweets. Three distinct deep learning techniques were employed: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Hybrid RNN-CNN. The dataset used for training and testing was collected from Twitter. The collected tweets were categorized into hate speech and non-hate speech. Feature extraction and expansion were performed using Term Frequency – Inverse Document Frequency (TF- IDF) and Global Vectors (GloVe) methods. Several scenarios were considered to compare the effectiveness of various approaches and determine the model with the highest accuracy. Among these approaches, the RNN method with feature expansion on the top 5 word similarity achieved the highest accuracy rate of 91.34%. Following closely, the Hybrid RNN-CNN method with feature expansion on the top 5 word similarity obtained an accuracy of 90.69%, while the CNN method with feature expansion on the top 5 word similarity yielded an accuracy of 90.64%. Notably, all three highest models utilized a corpus constructed by combining the tweet and news datasets.

Keywords: hate speech, hybrid, CNN, RNN, GloVe