

Analisis Perbandingan Performansi CNN dan LSTM dalam Mendeteksi Ujaran Kebencian di Twitter

Artisa Bunga Syahputri¹, Yuliant Sibaroni²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹artisabunga@students.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id,

Abstrak

Perkembangan internet pada saat ini menjadi salah satu faktor yang memberi peluang bagi pengguna media sosial untuk meninggalkan komentar dan postingan yang mengandung ujaran kebencian. Mendeteksi ujaran kebencian pada media sosial terutama pada Twitter menjadi topik yang banyak diteliti beberapa waktu belakangan. Penelitian yang telah dilakukan biasanya menerapkan pendekatan machine learning biasa, dan saat ini pendekatan deep learning menjadi populer karena dianggap dapat memberikan hasil yang lebih baik dan lebih efektif, namun masih jarang diterapkan untuk melakukan deteksi ujaran kebencian pada teks yang berbahasa Indonesia. Penelitian ini menunjukkan hasil perbandingan performansi dari pendekatan deep learning dengan menggunakan model arsitektur CNN, LSTM, dan kombinasi CNN+LSTM untuk mendeteksi ujaran kebencian pada postingan Twitter yang menggunakan bahasa Indonesia. Dataset yang digunakan terdiri dari data umum yang merupakan keseluruhan dataset dan data dengan topik yang spesifik membahas mengenai pemerintahan yang diambil dari dataset umum. Dari penelitian yang telah dilakukan menampilkan hasil yang lebih baik saat model arsitektur CNN diimplementasikan pada data tweet berbahasa Indonesia dibandingkan hasil yang diperoleh dari model arsitektur LSTM dan kombinasi CNN+LSTM dengan nilai akurasi dan F1-score mencapai 81%. Selain itu penerapan model deep learning dalam mendeteksi ujaran kebencian memberikan performansi yang lebih baik dibandingkan dengan penelitian sebelumnya dengan dataset yang sama namun menerapkan model machine learning dengan ekstraksi fitur. Penelitian ini juga menunjukkan bahwa data yang spesifik membahas satu topik khusus dan pemilihan parameter model memberikan pengaruh yang signifikan pada performansi model sehingga performa model menjadi lebih meningkat dibandingkan saat diterapkan pada data dengan topik yang umum dan memiliki kamus kata yang lebih banyak.

Kata kunci : ujaran kebencian, Twitter, deep learning, CNN, LSTM

Abstract

The internet's current development has become one factor that gives social media users opportunities to leave comments and posts containing hate speech. Detecting hate speech on social media, particularly on Twitter, has recently become a widely researched topic. Research that has been conducted usually applies a standard machine learning approach. The deep learning approach has become popular because it provides better and more effective results. However, it's still rare to be applied to detect hate speech in Indonesian language texts. This research shows the results of a performance comparison from the deep learning approach using CNN, LSTM, and CNN+LSTM architecture models for detecting hate speech in tweets using the Indonesian language. The dataset used is divided into a general dataset which is the entire dataset and a specific topic dataset that deals with the topic of government, which was taken from the general dataset. The research shows better results when the CNN architecture model is implemented on Indonesian language tweet data compared to the results obtained from the LSTM architecture model and the combination of CNN+LSTM with accuracy and F1 score reaching 81%. Furthermore, the implementation of deep learning models in detecting hate speech performs better than previous research using the same dataset but applying machine learning models with feature extraction. This research also shows that specific data discussing a particular topic and selection of model parameters significantly impact the model's performance. Thus the version of the model becomes better when applied to data with a general topic and a more extensive vocabulary.

Keywords: hate speech, Twitter, deep learning, CNN, LSTM.

1. Pendahuluan

Latar Belakang

Sosial media menjadi salah satu media informasi dan komunikasi yang digunakan hampir seluruh masyarakat di dunia untuk dapat saling berkomunikasi dan memberikan kemudahan dalam mengunggah berbagai postingan berupa teks, gambar, video, dan audio. Hal ini menjadi salah satu peluang bagi beberapa orang untuk dengan bebas memberikan ujaran kebencian dalam bentuk postingan maupun komentar pada orang lain [1]. Twitter menjadi salah satu media sosial yang banyak digunakan saat ini. Twitter membatasi penggunaan karakter dalam sebuah