ABSTRACT

Twitter is a social media platform that allows users to write text-based messages. In addition to being used for positive purposes, Twitter is also often utilized as a means to spread hate speech. Hate speech can result in negative impacts such as creating social pressure, stress, despair, trauma, and can even contribute to suicidal actions. Therefore, there is a need for a system that can automatically detect hate speech. This research aims to develop a text classification model using an improved stemming algorithm in the text preprocessing stage for detecting hate speech on social media, particularly on the Twitter platform. The research results demonstrate that the use of the improved stemming algorithm contributes positively to the accuracy of text classification compared to other scenarios, such as using the Porter algorithm or not using stemming at all. The model implementing the improved stemming algorithm performs well, achieving an accuracy rate of 85.03% in the conducted testing. These findings provide insights into the importance of proper text preprocessing using the appropriate stemming method in addressing hate speech issues on social media. The results of this research can serve as a foundation for further advancements in hate speech detection and improved text processing on social media platforms.

Keywords: hate speech, text classification, text preprocessing, stemming

ABSTRAK

Twitter adalah media sosial yang memungkinkan penggunanya menuliskan pesan berbasis teks. Selain digunakan untuk hal-hal positif, twitter juga sering digunakan sebagai sarana untuk menyebarkan *hate speech*. *Hate speech* dapat menimbulkan dampak negatif seperti menyebabkan tekanan sosial, stres, putus asa, trauma, dan bahkan dapat berkontribusi pada tindakan bunuh diri. Oleh karena itu, diperlukan suatu sistem yang dapat mendeteksi *hate speech* secara otomatis. Penelitian ini bertujuan untuk mengembangkan model klasifikasi teks dengan menggunakan *improved stemming algorithm* pada tahap *text preprocessing* untuk mendeteksi *hate speech* di media sosial, terutama di platform Twitter. Hasil penelitian menunjukkan bahwa penggunaan *improved stemming algorithm* memberikan kontribusi positif terhadap akurasi klasifikasi teks dibanding skenario lainnya yaitu penggunaan algoritma porter dan juga tanpa menggunakan *stemming*. Model yang menerapkan *improved stemming algorithm* memiliki performa yang cukup baik dengan akurasi mencapai 85,03% pada pengujian yang dilakukan. Temuan ini memberikan wawasan tentang pentingnya *text preprocessing* dengan menggunakan metode *stemming* yang tepat dalam mengatasi masalah *hate speech* di media sosial. Hasil penelitian ini dapat menjadi landasan untuk pengembangan lebih lanjut dalam deteksi *hate speech* dan pengolahan teks yang lebih baik di platform media sosial.

Kata Kunci: *hate speech*, klasifikasi teks, prepemrosesan teks, *stemming*