

---

## Abstrak

Seiring dengan meningkatnya jumlah pengguna media sosial, kemungkinan penyebaran ujaran kebencian di media sosial juga meningkat secara tidak langsung. Ujaran kebencian merupakan salah satu kasus yang paling banyak ditemukan di media sosial. Penyebaran ujaran kebencian dapat menimbulkan kerusuhan yang dapat menimbulkan konflik, pemusnahan kelompok, bahkan menimbulkan korban jiwa. Beberapa kontroversi terbaru di Indonesia terkait ujaran kebencian adalah ujaran kebencian yang diucapkan kepada pemerintah yang menimbulkan polemik dan demonstrasi di tanah air. Bersamaan dengan hal tersebut, pendeteksian ujaran kebencian penting dilakukan untuk menghindari terjadinya konflik. Seiring meningkatnya penyebaran ujaran kebencian di media sosial, diperlukan upaya manual yang signifikan dan mahal untuk melakukan pendeteksian. Oleh karena itu, percobaan ini dibangun untuk mendeteksi pendeteksian ujaran kebencian pada teks twitter berbahasa Indonesia menggunakan beberapa algoritma *machine learning* konvensional dan berbasis *deep learning* BiGRU dengan berbagai fitur. Pendekatan *machine learning* yang digunakan adalah SVM dan RFDT, sedangkan metode berbasis *deep learning* yang digunakan adalah BiGRU dan IndoBERT-BiGRU. Beberapa metode yang digunakan adalah Word2vec dan fastText. Hasil percobaan menunjukkan bahwa metode BiGRU dengan IndoBERT dan tanpa penghapusan *stop word* menghasilkan kinerja terbaik dengan akurasi 84.77 %. BiGRU memiliki keunggulan dalam menyimpan informasi penting dari teks, sehingga memberikan hasil yang lebih baik daripada algoritma *machine learning* konvensional.

**Kata kunci :** ujaran kebencian, klasifikasi teks, BiGRU, Word2vec, fastText