# ABSTRAK

Ujaran kebencian adalah sebuah bentuk dari komunikasi yang berisikan kebencian dengan melakukan hal-hal, seperti menghasut, menghina, meremehkan, ataupun merendahkan seseorang atau sebuah kelompok. Masalah ujaran kebencian di Indonesia selalu mempunyai keterkaitan dengan masalah politik. Sebagai contoh, pada tahun 2018 dan 2019, ujaran kebencian terkait dengan pemilihan gubernur dan presiden. Umumnya pelaku ujaran kebencian menggunakan jaringan sosial, seperti Instagram, untuk menyebarkan kata-kata kebenciannya. Sekitar 60% hate speech ini ditemukan di kolom komentar dari sebuah kiriman dan hal ini akan menjadi sebuah ancaman yang nyata jika tidak segera terdeteksi. Penelitian ini bertujuan untuk mendeteksi ujaran kebencian pada komentar di Instagram. Kami mengusulkan penggunaan metode word2vec dengan model *skip-gram*, dan *TextCNN* yang telah dimodifikasi untuk mempelajari dan mendeteksi teks ujaran kebencian. Lebih lanjut, *random oversampling*, *random undersampling*, dan *class weight* juga digunakan untuk menyelesaikan masalah dataset yang tidak seimbang. Hasil dari eksperimen menunjukkan bahwa akurasi terbaik, dalam hal *F-score*, yaitu 99.25% didapatkan oleh metode yang digunakan pada penelitian sebelumnya yaitu kombinasi dari *Fasttext*, *word Bi-gram*, dataset tidak seimbang yang diekstrak dari 33 *hashtag* Instagram dan *metode random oversampling*.

***Keywords—Komentar ujaran kebencian, Instagram, Word2vec, TextCNN, Fasttext, Dataset tidak seimbang***