1. Pendahuluan

Direktorat Tindak Pidana Siber Bareskrim Polri mendapati 89 konten media sosial terverifikasi mengandung ujaran kebencian selama periode 23 Februari hingga 11 Maret 2021. Konten terbanyak berasal dari Twitter. Berdasarkan data Virtual Police (Dit Tipisiber) Bereskrim Polri pada periode itu 125 konten diajukan untuk diberikan peringatan Virtual Police didominasi platform Twitter 79 konten, Facebook 32 konten, Instagram 8 konten, Youtube 5 konten, dan Whatsapp satu konten[1].

Tindak ujaran kebencian hadir dengan peningkatan yang signifikan dalam interaksi sosial di jejaring sosial online, peningkatan yang terjadi memanfaatkan infrastruktur media sosial. Di Twitter, ujaran kebencian adalah kicauan yang berisi kata-kata kasar yang ditujukan untuk menyerang individu (perundungan siber, politisi, selebriti, produk) atau kelompok tertentu (negara, LGBT, agama, gender, organisasi, dll.). Mendeteksi kebencian seperti pada *tweet* itu penting untuk menganalisis sentimen publik suatu kelompok pengguna terhadap kelompok lain, dan untuk mencegah aktivitas yang salah terkait tentang ujaran kebencian[2]. Hal ini menjadi acuan untuk mengidentifikasi ujaran kebencian di media sosial, khususnya di Twitter, karena pengguna dapat menggunakannya sebagai sarana untuk mengungkapkan ujaran kebencian dengan berbagai fitur yang ditawarkan Twitter.Dikarenakan pada media sosial Twitter dinilai dapat menggiring opini masyarakat dengan mudah, maka dengan adanya penelitian ini diharapkan dapat membantu masyarakat untuk mengidentifikasi ujaran kebencian pada Twitter, apakah hasilnya mengandung ujaran kebencian atau tidak pada hastag yang sedang ramai diperbicarakan di media sosial Twitter.

Pembelajaran *Deep Learning* dan penggunaan fitur seperti *bag of words*, *word,* dan *n-grams* sangat efektif untuk mendeteksi ujaran kebencian dan hasilnya menunjukkan bahwa kinerja yang lebih baik daripada metode *Machine Learning* dalam mendeteksi ujaran kebencian[3]. Penerapan metode Convolutional Neural Network (CNN) untuk klasifikasi teks juga menghasilkan performasi yang lebih baik dari metode *Decision Tree, Support Vector Machine,* dan *Naïve Bayes*[4]. Pada penelitian ini akan digunakan metode klasifikasi deep learning Convolutional Neural Network (CNN). CNN merupakan salah satu deep neural network yang telah terbukti secara efektif menyelesaikan beberapa pemrosesan Bahasa seperti penandaan kalimat, analisis sentimen dan pengenalan entitas nama. Dengan berbagai fitur CNN dapat digunakan untuk kategorisasi ujaran kebencian berdasarkan vektor kata informasi semantik yang dibuat untuk semua token menggunakan algoritma *unsupervised learning* dan *word2vec*[5].

Pada penelitian [3] membandingkan antara metode *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN) untuk mendeteksi ujaran kebencian dalam Tweet berbahasa Arab, yang menghasilkan model CNN memberikan kinerja terbaik dengan F1 skor 0,79. Lalu, pada penelitian [5] membahas tentang membangun sistem klasifikasi teks ujaran kebencian menggunakan CNN dengan dibagi menjadi empat kategori yaitu rasisme, seksisme, netral, dan non ujaran kebencian. Sistem dibangun dengan menggunakan *word2vec* yang diuji 10 kali yang menghasilkan bahwa *word2vec* embeddings berkinerja dengan baik dengan nilai presisi lebih tinggi dari recall dan F1-skor 78,3% daripada menggunakan *fast text embedding*. Kekurangan pada penelitian [5] adalah hanya fokus pada membandingkan *word embedding* antara *word2vec* dengan *fast text* tetapi tidak menerapkan perbandingan *hyperparameter tuning* agar mendapatkan akurasi tertinggi. Pada penelitian [6] digunakan model *deep learning* yaitu metode *Convolutional Neural Network* (CNN) dan *Reccurent Neural Network* (RNN) untuk klasifikasi kata menghasilkan bahwa metode RNN dapat menangani klasifikasi kata ini dengan baik. Meskipun demikian, metode RNN tidak dapat memparalelkan kata dengan baik dan lebih cocok untuk pemrosesan teks yang pendek. Waktu pelatihan juga lama saat memproses teks dengan lebih dari puluhan kata. CNN telah banyak digunakan dalam model klasifikasi teks panjang karena dapat memparalelkan kata dengan baik.

Dibandingkan RNN, CNN menggunakan beberapa *channel*, CNN memilih untuk menggunakan ukuran *filter* yang berbeda dan *max pooling* untuk memilih kata yang berpengaruh dan karakteristik klasifikasi *low-latitude* yang lebih rendah kemudian menggunakan *dropout* dari lapisan *full connection* dengan fitur ekstraksi sebagai hasil klasifikasi akhir.

Berdasarkan kelebihan metode CNN tersebut, dalam tugas akhir ini diusulkan menggunakan metode *Convolutional Neural Network* untuk mengidentifikasi ujaran kebencian pada Twitter dengan menggunakann *hyperparameter learning rate* [0.01,0.001,0,0001] dan *batch size* 256, nilai *hyperparameter* yang dipilih memberikan solusi yang lebih baik[25]. Diharapkan hasil dari penelitian ini dapat mengetahui performansi metode klasifikasi CNN untuk mengidentifikasi ujaran kebencian pada Twitter dan juga dapat mengetahui hasil *hyperparameter tuning* model terbaik pada CNN.

. Batasan masalah pada penelitian ini adalah menggunakan dataset yang bersumber dari hasil data *crawling* yang dilakukan berdasarkan teks yang mengandung ujaran kebencian pada *twitter*, lalu teks yang dapat di identifikasi hanya menggunakan Bahasa Indonesia.