

1. Pendahuluan

1.1 Latar Belakang

Media sosial merupakan alat komunikasi yang dapat mendukung interaksi sosial dari para penggunanya dengan menggunakan teknologi berbasis web menjadi dialog interaktif [1]. Pengguna dapat berpartisipasi di media sosial untuk memberi feedback dan komentar secara terbuka dan dalam waktu yang tidak terbatas [1]. Media sosial juga digunakan sebagai alat penyebaran informasi pada berbagai bidang seperti bidang bisnis, bidang pariwisata, bidang pendidikan, bidang politik, bidang kesehatan, dan lain sebagainya. Namun, dalam penyebaran informasi tersebut tidak sedikit yang menyalahgunakan informasi di media sosial [2]. Pemerintah juga telah mengeluarkan Undang-Undang tentang Informasi dan Transaksi Elektronik yang terdapat pada Pasal 27 ayat 3 UU ITE [3,22,23].

Twitter merupakan salah satu sosial media yang populer di Indonesia dengan pengguna sebanyak 18,45 Juta pengguna di Indonesia pada awal tahun 2022 [4]. Twitter dianggap oleh *virtual police* sebagai sebagai salah satu media sosial yang banyak berkontribusi terhadap penyebaran ujaran kebencian [5,19-21]. Menurut Kepala Bagian Penerangan Umum Polri Kombes Pol Ahmad Ramadhan, pada rentang waktu 23 Februari – 12 April 2021 terdapat sebanyak 195 akun Twitter yang terjaring dalam *virtual police* akibat tweet yang mengandung ujaran kebencian dan SARA [5]. Ujaran kebencian sendiri merupakan sebuah tindakan komunikasi berupa perkataan, perbuatan, maupun tindakan yang dilakukan satu orang atau berkelompok dengan tujuan memprovokasi dan menghina seseorang maupun kelompok lain dalam berbagai aspek seperti ras, warna kulit, etnis, agama, dan sebagainya [6].

Berdasarkan permasalahan tersebut, diperlukan tindakan seperti melakukan deteksi ujaran kebencian pada Twitter Indonesia. Terdapat beberapa penelitian yang melakukan deteksi ujaran kebencian pada Twitter Indonesia. Salah satunya yaitu pada penelitian [7] yang menunjukkan bahwa penelitian tersebut dapat mengidentifikasi *cyberbullying* pada media sosial Twitter Indonesia menggunakan metode LSTM dan BiLSTM. Nilai akurasi yang didapat pada penelitian tersebut yaitu pada LSTM sebesar 93.77% dan pada BiLSTM sebesar 95.24%. Namun, penelitian tersebut hanya membagi data menjadi dua kelas yaitu kelas *cyberbullying* dan kelas *non-cyberbullying*. Adapun penelitian menggunakan klasifikasi multi-label [12] menggunakan metode BiLSTM. Penelitian ini melakukan *labelling* data secara multi-label menjadi 12 label ujaran kebencian. Penelitian ini memperoleh performa yang baik dengan nilai akurasi sebesar 82.31%.

Pada penelitian ini, penulis melakukan penelitian mengenai deteksi ujaran kebencian menggunakan klasifikasi multi-label serta membandingkan metode LSTM dan BiLSTM. Penelitian ini menggunakan metode yang telah digunakan pada penelitian sebelumnya [7] yaitu menggunakan LSTM dan BiLSTM karena pada penelitian tersebut, algoritma LSTM dan BiLSTM cukup baik dalam mengklasifikasi data teks. Pada *labelling* data menggunakan klasifikasi multi-label yaitu memasukkan data ke dalam beberapa kategori dengan harapan dapat meningkatkan keberhasilan pada penelitian ini.

1.2 Topik dan Batasannya

Adapun batasan masalah dari tugas akhir ini yaitu:

1. Deteksi ujaran kebencian pada Twitter Indonesia menggunakan LSTM dan BiLSTM
2. Dataset yang digunakan adalah dataset mengenai Twitter bahasa Indonesia yang mengandung ujaran kebencian.

1.3 Tujuan

Tujuan dari tugas akhir ini yaitu:

1. Mendeteksi ujaran kebencian pada Twitter Indonesia menggunakan LSTM dan BiLSTM
2. Mengetahui performansi dari klasifikasi multi-label pada tweet ujaran kebencian Twitter Indonesia menggunakan metode LSTM dan BiLSTM
3. Mengetahui perbandingan metode LSTM dan BiLSTM pada tweet ujaran kebencian Twitter Indonesia menggunakan klasifikasi multi-label.

Tabel 1. Tabel keterkaitan antara tujuan, pengujian dan kesimpulan

No	Tujuan	Pengujian	Kesimpulan
1	Deteksi ujaran kebencian pada Twitter Indonesia menggunakan LSTM dan BiLSTM	Pengujian dilakukan dengan meng-input tweet pada kumpulan data	Deteksi berhasil dilakukan menggunakan LSTM dan BiLSTM
2	Performansi dari klasifikasi multi-label pada tweet ujaran kebencian Twitter Indonesia menggunakan LSTM dan BiLSTM	Pengujian dilakukan menggunakan metode LSTM dan BiLSTM	Hasil pengujian menunjukkan hasil akurasi testing pada LSTM sebesar 78,67% dan BiLSTM sebesar 80,25%
3	Perbandingan metode LSTM dan BiLSTM pada tweet ujaran kebencian Twitter Indonesia menggunakan klasifikasi multi-label	Pengujian dilakukan dengan menggunakan akurasi pada metode LSTM dan BiLSTM	Analisis hasil pengujian menunjukkan bahwa BiLSTM lebih unggul dibandingkan LSTM dengan melakukan perbandingan akurasi.

1.4 Organisasi Tulisan

Pada jurnal TA ini terdapat beberapa bagian. Bagian pertama yaitu bab 1 pendahuluan yang berisi mengenai gambaran umum tentang latar belakang dan tujuan penelitian yang menjadi topik pada tugas akhir. Bab 2 studi terkait yang berisi teori/studi literatur yang berkaitan dengan tugas akhir. Selanjutnya terdapat bab 3 yaitu sistem yang dibangun berisi tentang penjelasan sistem yang digunakan pada tugas akhir. Bab 4 merupakan evaluasi yang berisi hasil dan analisis pengujian yang dilakukan pada tugas akhir. Bab 5 merupakan kesimpulan dan saran dari penelitian tugas akhir.