

BAB I PENDAHULUAN

1.1. Latar Belakang

Pada era kini, manusia dimudahkan dalam mengakses informasi dan melakukan transaksi secara efektif dan efisien secara online. Di Indonesia, jumlah penduduk yang terkoneksi internet semakin meningkat dari 64,8% penduduk pada tahun 2018, lalu menjadi 73,7% penduduk tahun 2019-2020 dan kemudian meningkat lagi menjadi 77,02% penduduk pada tahun 2021-2022, menurut APJII[1]. Asosiasi Penyelenggara Jasa Internet Indonesia melaporkan alasan terbaik menggunakan internet adalah komunikasi menggunakan messenger online dan media sosial. Media sosial merupakan sebuah sarana yang digunakan untuk berkomunikasi atau memberikan informasi kepada khalayak ramai baik itu berupa suara, gambar maupun video. Dengan adanya media sosial mudah bagi kita mendapatkan informasi dari seseorang ataupun sebuah organisasi. Namun terkadang, media sosial digunakan sebagai sarana dalam menyebarkan konten-konten negatif oleh oknum tertentu. Kementerian Komunikasi dan Informatika Republik Indonesia merilis konferensi pers tentang penggunaan media sosial untuk menyebarkan konten negative, berdasarkan pemaparan ferdinandus setu [2].

Konten negatif adalah konten yang mengandung kebencian. Berdasarkan Buku saku penanganan ujaran kebencian (*hate speech*)[3], dalam UU ITE dan Komisi Nasional Hak Asasi Manusia, ujaran kebencian adalah perbuatan yang dengan sengaja dan melawan hukum, menyebarkan informasi yang ditujukan untuk kebencian dan permusuhan terhadap individu atau kelompok orang tertentu, berdasarkan suku, agama, dan ras. Media sosial dapat menimbulkan efek samping dari kebebasan berpendapat yang didapat hingga kebebasan untuk membenci orang lain. Hal yang dianggap aktivitas kebencian diantaranya berusaha untuk membungkam orang lain; menyudutkan dan menindas orang; atau menimbulkan rasa benci, malu, bersalah, dan berbagai perasaan atau kondisi tidak nyaman lainnya. Kebebasan untuk membenci juga memberikan lahan subur bagi

berkembangnya narasi sektarian dan rasis, menurut Veronika and A. E. Raharjo[4]. Banyak peneliti yang melakukan pendeteksian ujaran kebencian dalam beberapa tahun terakhir untuk mengurangi penyebaran ujaran kebencian di media sosial saat ini. Ada beberapa penelitian terkait ujaran kebencian dalam konten berbahasa Indonesia di media sosial. Selain Twitter, pendeteksian ujaran kebencian juga dilakukan di Instagram. Konten tersebut dapat berisi bahasa yang penuh kebencian, kasar, dan menyinggung. Kajian tersebut terdiri dari pembuatan data penelitian, klasifikasi ujaran kebencian, dan analisis sentiment.

Menurut D. Elisabeth dkk[5] mengatakan bahwa model pendeteksian ujaran kebencian tidak mampu mendeteksi ujaran kebencian secara tepat jika ada kode-kode kebencian baru yang digunakan oleh masyarakat. Contoh postingan yang berisikan kode kebencian adalah sebagai berikut:

“Anak bani kampret sama penista agama pikirannya emang gak beda jauh sama aja”. (Kelelawar kecil sama penghujat agama pemikirannya tidak jauh berbeda sama saja.)

Ada dua kode kebencian pada contoh tersebut. Anak bani kampret adalah julukan para pendukung politisi di Indonesia. Sedangkan kelelawar kecil adalah kelelawar kecil pemakan serangga. Penista agama adalah julukan yang ditujukan pada politisi Indonesia lainnya karena perbuatannya di masa lalu.

Putra dan Nurjanah[6] menggunakan metode word2vec dengan model skip-gram dan TextCNN yang dimodifikasi dan metode random oversampling untuk mempelajari serta mendeteksi teks ujaran kebencian yang ada pada komentar Instagram yang menunjukkan akurasi terbaik, dalam hal F-score. adalah 9370%. Erryan dan Indra[7] melakukan eksperimen untuk mengukur perbedaan kinerja model antara fase pelatihan dan fase pengujian. Hasil percobaan menunjukkan bahwa metode yang diusulkan mengungguli mesin metode berbasis pembelajaran, baik dalam tahap pelatihan, dengan algoritma GRU dengan 85,37% F1-score, dan dalam tahap pengujian, dengan algoritma LSTM dengan skor F1 76,30%. Dan T. Febriana and A. Budiarto[8], menggunakan *Latent Dirichlet Allocation* (LDA) algoritma untuk mengekstrak topik untuk setiap tweet untuk melihat apakah topik ini dapat

dikaitkan dengan tema debat. Analisis sentimen yang telah dilatih sebelumnya juga diterapkan ke dataset untuk menghasilkan skor polaritas untuk setiap tweet.

Berdasarkan kajian-kajian tersebut, peneliti akan melakukan pengujian deteksi text menggunakan algoritma Support Vector Machine, Logistic Regression, K-Nearest Neighbor dan Decision Tree tentang ujaran kebencian pada postingan twitter #indonesia. Pengumpulan data dengan menggunakan API Twitter diimplementasikan kedalam bahasa pemrograman Python. Pelabelan data tweet yang telah didapatkan dilakukan dengan menggunakan *sentistrength* yang diimplementasikan dengan menggunakan bahasa pemrograman Python. Kemudian pada data tersebut akan dilakukan klasifikasi menggunakan metode Support Vector Machine, Logistic Regression, K-Nearest Neighbor dan Decision Tree untuk melebeli data yang nantinya akan digunakan sebagai dasar untuk menentukan sentimennya.

1.2. Ruang Lingkup

Ruang lingkup dari penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya berfokus pada media sosial Twitter.
2. Data yang digunakan berupa postingan Twitter berbahasa indonesia dengan Query “#indonesia” dan akan dibedakan menjadi tiga level yaitu, Negatif (*Hate Speech*), Netral dan Positif.

1.3. Rumusan Masalah

Dengan dasar pemikiran yang tertera pada latar belakang penelitiannya rumusan masalah yang diambil adalah

1. Bagaimana mengklasifikasikan sentimen pada postingan berbahasa indonesia tentang “#indonesia” pada twitter dan mengklasifikasikan postingan tersebut apakah termasuk ujaran kebencian (*hate speech*) atau bukan?
2. Seberapa akurat metode algoritma *Support Vector Machine*, *Logistic Regression*, *K-Nearest Neighbor* dan *Decision Tree* dalam mengklasifikasikan ujaran kebencian dengan analisis sentimen pada postingan twitter?

1.4. Tujuan Penelitian

Peneliti telah menetapkan tujuan dari penelitian ini yaitu:

1. Peneliti dapat mengolah dan menganalisis data postingan berbahasa indonesia tentang “#indonesia” yang ada pada twitter.
2. Melakukan pengklasifikasian postingan twitter menggunakan metode *Support Vector Machine, Logistic Regression, K-Nearest Neighbor* dan *Decision Tree* serta mengetahui seberapa akurat metode-metode tersebut.
3. Mempublikasikan hasil yang diperoleh dalam bentuk jurnal ilmiah sehingga dapat diakses masyarakat umum dan terkhusus pengguna media sosial Twitter.

1.5. Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat membantu pengguna internet dan media sosial indonesia khususnya media sosial Twitter dalam mendapatkan informasi terkait postingan “#indoneisa” yang memiliki unsur-unsur ujaran kebencian. Sehingga dapat menghindari atau mencegah adanya dampak buruk diakibatkan oleh informasi yang mengandung unsur-unsur ujaran kebencian tersebut.

1.6. Sistematika Penulisan

Penulisan ini akan dibagi dalam 5 (lima) bab, dengan sistematika penulisan sebagai berikut:

a. Bab I Pendahuluan

Pada bab ini berisi latar belakang, rumusan masalah, ruang lingkup penelitian, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

b. Bab II Tinjauan Pustaka

Dalam bab ini dijelaskan tentang penelitian terkait dan teori pendukung yang digunakan dalam penelitian ini.

c. Bab III Metode Penelitian

Pada bab ini membahas tentang alat dan bahan penelitian, kerangka kerja yang digunakan, dan tahapan penelitian yang dilakukan.

d. Bab IV Hasil dan Pembahasan

Bab ini menjelaskan hasil penelitian dan pembahasan terkait hasil-hasil tersebut.

e. Bab V Simpulan dan Saran

Bab ini menyimpulkan hasil dari penelitian ini guna mendapatkan hasil analisis dari metode *Support Vector Machine*, *Logistic Regression*, *K-Nearest Neighbor* dan *Decision Tree* yang dijadikan sebagai jawaban dari rumusan masalah yang telah disampaikan.

