BAB I
PENDAHULUAN

1.1 Latar Belakang

Media sosial telah menjadi salah satu *platform* utama untuk berinteraksi, berbagi informasi, dan berkomunikasi dalam masyarakat saat ini (Gellysa Urva dkk., 2022). Berdasarkan *data* yang ditemukan dari (*Digital 2023*, 2023) pengguna *platform* media sosial di Indonesia sebanyak 167 juta pengguna dimana *platform* yang mendominasi di Indonesia yaitu *facebook* sebanyak 119.9 juta pengguna, *youtube* sebanyak 139 juta pengguna, *instagram* sebanyak 89.15 juta pengguna, *tiktok* sebanyak 109.9 juta pengguna, *facebook messenger* sebnayak 27.30 juta pengguna, *linkedin* sebanyak 23 juta pengguna, *snapchat* sebanyak 3.55 juta pengguna dan *twitter* sebanyak 24 juta pengguna. Namun, penggunaan media sosial yang tidak bijaksana dapat mengarah pada penyebaran bahasa *toxic* yang dapat merugikan individu (Aslan & Samet, 2017) , kelompok (Jonathan dkk., 2022), atau masyarakat secara umum (Putra, 2022).

Berdasarkan survei oleh lembaga donasi *anti-bullying*, *Ditch The Label*, *Instagram* merupakan platform media sosial yang paling sering digunakan untuk melakukan *cyberbullying*. Survei ini melibatkan 10.020 remaja Inggris dengan rentang usia 12 hingga 20 tahun, dan hasilnya menunjukkan bahwa sekitar 42 persen dari mereka pernah menjadi korban *cyberbullying* di *Instagram*. Persentase korban di Facebook dan Snapchat masing-masing sebesar 37 persen dan 31 persen. Di sisi lain, *WhatsApp* (12 persen), *YouTube* (10 persen), dan *Twitter* (9 persen) memiliki persentase kasus *cyberbullying* yang lebih rendah. Ini mengindikasikan bahwa *Instagram* memiliki andil tertinggi dalam masalah *cyberbullying* di antara platform media sosial lainnya. Hal ini menunjukkan bahwa masalah *cyberbullying* tetap signifikan, dengan banyak remaja yang mengalami perundungan secara online (Pratama & Nistanto, 2021).

Kalimat *toxic* adalah bahasa yang dapat merendahkan, menghina, atau merugikan pihak lain (Ferdy dkk., 2021). Contoh kalimat *toxic* meliputi penghinaan, pelecehan, ujaran kebencian, atau tindakan *cyberbullying* (Alika dkk., 2022). Penghinaan dalam kalimat *toxic* mencakup kata-kata atau ungkapan yang merendahkan individu atau kelompok, sementara pelecehan melibatkan konten merendahkan secara seksual, rasial, atau emosional. Ujaran kebencian adalah kalimat yang menyebarkan kebencian atau diskriminasi berdasarkan karakteristik seperti ras, agama, orientasi seksual, atau gender. Kalimat *toxic cyberbullying* adalah pesan yang digunakan dalam perilaku penindasan daring, termasuk serangan pribadi, penghinaan, penyebaran rumor palsu, dan penindasan berkelanjutan. Tujuannya adalah menyakiti, menghina, mengancam, atau merendahkan harga diri seseorang secara *online*.

Berdasarkan paparan di atas, penggunaan kalimat *toxic* yang mencakup bahasa yang merendahkan, menghina, atau merugikan individu atau kelompok, memiliki dampak negatif yang signifikan dalam sejumlah aspek. Fenomena ini telah mendapatkan perhatian peneliti sebelumnya, seperti yang terungkap dalam hasil penelitian sebelumnya. Secara psikologis dan emosional, kalimat *toxic* dapat menyebabkan stres, kecemasan, dan depresi pada individu yang menjadi sasaran, karena kata-kata merendahkan dapat merusak harga diri dan memicu perasaan malu (Ramadhani, 2022). Dalam hubungan sosial, penggunaan bahasa tersebut dapat mengganggu komunikasi, menciptakan konflik, dan menghambat pertukaran ide yang sehat (Wardana, 2023). Lingkungan *online* pun terpengaruh, di mana kalimat *toxic* menciptakan ruang yang tidak aman dan tidak ramah (Chairunnisa, 2021). Selain itu, penggunaan kalimat *toxic* dapat memperkuat siklus kebencian dan diskriminasi, serta merendahkan kualitas komunikasi dan diskusi (Mardianto, 2023).

Penting untuk peningkatan metode deteksi bahasa *toxic* dalam bahasa Indonesia untuk mengidentifikasi dan mengurangi penyebaran bahasa *toxic* di media sosial. Algoritma *Decision Tree* adalah salah satu metode yang dapat digunakan dalam pengenalan pola bahasa *toxic*. *Decision Tree* adalah model pembelajaran mesin

yang dapat digunakan untuk menggambarkan keputusan atau keputusan berdasarkan aturan yang ada (Romadloni dkk., 2022). *Decision Tree* dapat mengenali pola bahasa *toxic* berdasarkan atribut-atribut tertentu yang ditemukan dalam teks, seperti kata-kata atau frase yang cenderung bersifat *toxic*.

Dalam upaya untuk menghadapi tantangan ini, penelitian telah mengarah pada pemanfaatan teknik pemrosesan bahasa alami dan kecerdasan buatan, termasuk algoritma *Decision Tree*, untuk mengidentifikasi dan mendeteksi kalimat *toxic* secara otomatis. Sebagai contoh, dalam penelitian sebelumnya, para peneliti menerapkan algoritma *Decision Tree* pada kumpulan *data* besar berisi kalimat-kalimat dari berbagai sumber *online*. Mereka mengembangkan model yang dapat mengidentifikasi kata-kata atau pola kalimat yang umumnya terkait dengan bahasa *toxic*, seperti penghinaan atau ujaran kebencian. *Model* ini diuji menggunakan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* untuk mengukur kemampuannya dalam mengklasifikasikan kalimat *toxic* dan *non-toxic*. Penelitian yang dilakukan oleh (Kusuma & Pamungkas, 2023) mengaplikasikan algoritma *Support Vector Machine (SVM)* dan *Decision Tree* pada *dataset Twitter* berbahasa Indonesia dengan tujuan mendeteksi bahasa kasar. Hasilnya menunjukkan bahwa model *SVM* mencapai akurasi 83%, sedangkan *Decision Tree* memiliki nilai Presisi 84%, *Recall* 89%, Akurasi 83%, dan *F1-Score* 86%. Penelitian lainnya yang dilakukan oleh (Septiawan & Chairani, 2023) memperoleh hasil penelitian dari penggunaan metode *SVM* dan *Decision Tree*s yang ditingkatkan dengan Adaboost dianalisis pada *dataset Twitter* yang mengandung *tweet* ujaran kebencian dan netral. Akurasi deteksi dengan algoritma *SVM* setelah dioptimalkan *Adaboots* mencapai 90,03%, sedangkan algoritma *Decision Tree*s setelah dioptimalkan *Adaboots* mencapai 89,53%. Selanjutnya penelitian yang dilakukan oleh (Ihsan dkk., 2021) dimana penelitian ini mengembangkan sistem klasifikasi untuk mengidentifikasi *twitter* yang mengandung ujaran kebencian dan kata-kata kasar. Dengan menggunakan algoritma *Decision Tree* dan word embedding sebagai fitur teks, penelitian ini menemukan bahwa penggunaan fitur leksikon dalam klasifikasi *Decision Tree* memberikan akurasi tertinggi untuk deteksi kelas ujaran kebencian,

kata-kata kasar, dan tingkat ujaran kebencian. Akurasi rata-rata dari ketiga kelas meningkat dari 69,77% menjadi 70,48% pada komposisi *data* latih-uji 90:10.

Dengan mengembangkan algoritma *Decision Tree* untuk deteksi bahasa *toxic* pada media sosial berbahasa Indonesia, penelitian ini diharapkan dapat membantu mengurangi penyebaran bahasa *toxic* dan meningkatkan kesadaran tentang pentingnya berbicara dengan sopan di media sosial. Penelitian ini juga dapat menjadi dasar untuk pengembangan sistem atau aplikasi yang dapat secara otomatis mengidentifikasi dan menghapus konten berbahasa *toxic* di media sosial berbahasa Indonesia, sehingga dapat memberikan kontribusi positif terhadap keberagaman dan keharmonisan komunikasi di dunia maya.

Penelitian ini melibatkan aspek tambahan, yakni penggunaan *platform* media sosial *Instagram*. *Instagram* menduduki posisi keempat sebagai *platform* dengan jumlah pengguna terbanyak di Indonesia, mencapai 89,15 juta pengguna. Sebagai *platform* media sosial yang populer, *Instagram* memberikan kesempatan yang signifikan untuk menggali *data* dalam upaya mendeteksi kalimat *toxic*. Komentar yang muncul di *Instagram* dapat memberikan wawasan lebih komprehensif tentang kalimat *toxic* yang mungkin muncul dalam lingkungan ini. Dengan interaksi sosial yang aktif, pengumpulan *data* dari komentar dapat memberikan gambaran yang lebih akurat tentang pola kalimat *toxic* dalam konteks interaksi pengguna. Pemanfaatan *Instagram* sebagai *platform* penelitian diharapkan dapat memberikan kontribusi positif dalam menciptakan lingkungan digital yang lebih aman dan positif, serta mengurangi prevalensi kalimat *toxic*.

Berdasarkan uraian di atas maka telah dibuat sebuah model dengan menggunakan algoritma *Decision Tree* yang berfungsi untuk mendeteksi kalimat berbahasa Indonesia yang ada pada media sosial yang mangandung kalimat *toxic* dengan judul ***"DETEKSI KALIMAT TOXIC DALAM POSTINGAN MEDIA SOSIAL BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA DECISION TREE".***

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah pada penelitian ini adalah bagaimana mendeteksi kalimat yang ada pada media sosial *Instagram* apakah mengandung unsur kalimat *toxic* atau tidak?

1.3 Batasan Penelitian

Berdasarkan rumusan masalah, peneliti membatasi permasalahan dari penelitian yang dilakukan, diantaranya adalah sebagai berikut:

1. Penelitian ini hanya mendeteksi kalimat yang ada pada media sosial *instagram* apakah mengandung kalimat *toxic* atau tidak.
2. *Data* yang digunakan hanya menggunakan *data* pada media sosial *instagram* yang berupa komentar dari *posting*an-*posting*an pengguna *platform Instagram*.
3. Pengumpalan *data* diambil berdasarkan *posting*an gambar, *reels* dan video terbaru yaitu *posting*an yang telah di*posting* pada tahun 2023. Hal ini didasari oleh pola atau trend Bahasa yang digunakan dalam bermedia sosial yang selalu berbubah untuk setiap periode waktu tertentu.

1.4 Tujuan Penelitian

Dari rumusan masalah yang telah diberikan, penelitian ini dilakukan dengan tujuan:

1. Mempelajari pola atau struktur kalimat untuk mendeteksi apakah susunan kalimat tersebut mengandung kalimat *toxic* atau tidak.
2. Membuat model *machine learning* yang dapat digunakan untuk mendeteksi kalimat *toxic* dengan menggunakan algortima *Decision Tree*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Dalam konteks pengembangan teknologi untuk meningkatkan keamanan dan kenyamanan pengguna media sosial berbahasa Indonesia.
2. Diperolehnya hasil dari pembuatan model untuk mendeteksi dan analisa terhadap kalimat-kalimat yang menggunakan kalimat *toxic*.

1.6 Sistem Penulisan

Sistem penulisan dibuat untuk mempermudah dalam penyusunan penelitian ini maka perlu ditentukan sistem penulisan yang baik. Berikut adalah sistem penulisan penelitian ini:

BAB I PENDAHULUAN

Bab ini berisi latar belakang mengenai dampak dari penggunaan bahsa atau kalimat *toxic* dan mengapa perlu dibuatnya model deteksi *Machine Learning* untuk mendeteksi penggunaan bahasa atau kalimat pada media sosial apakah memiliki unsur bahasa atau kalimat *toxic* didalamnya, rumusan masalah, batasan penelitian, tujuan penelitian, manfaat penelitian dan sistem penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang teori – teori yang mendukung penelitian yang dilakukan oleh penulis antara lain *Artificial Intelligence*, *Decision tree*, *confusion matrix*, kebutuhan *software* dan bahasa pemrograman dan metode pembuatan model deteksi, alat yang di gunakan berdasarkan pustaka dan sumber-sumber dari internet serta literatur terkait penelitan-penelitan terdahulu untuk menunjang penelitian ini.

BAB III METODE PENELITIAN

Bab ini berisi tentang metode pembuatan model deteksi *Machine Learning* dalam penelitian ini meliputi *model requirements, data collection, data cleaning, data labeling, feature extraction, training* dan *testing model dan evaluation model*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang proses pembuatan model *decision tree* menggunakan metode penelitian *machine learning life cycle* serta hasil yang didapatkan dari setiap proses yang dilakukan dan analisa hasil performa model yang dibuat berdasarkan sistem pengujian model dengan menggunakan matriks evalusai akurasi.

BAB V KESIMPULAN DAN SARAN

Bab ini memberikan kesimpulan dari penelitian yang telah di lakukan dan saran yang berguna untuk penelitian selanjutnya.