BAB I
PENDAHULUAN

1.1 Latar Belakang

Media sosial telah menjadi sarana komunikasi utama masyarakat dalam menyampaikan pendapat dan berinteraksi secara daring. Perkembangan teknologi informasi mendorong peningkatan penggunaan media sosial secara signifikan. Laporan Digital 2024 *Global Overview* menunjukkan bahwa lebih dari 5 miliar penduduk dunia aktif menggunakan media social [1]. Di Indonesia, jumlah pengguna media sosial juga terus meningkat dan mencapai lebih dari 167 juta pengguna aktif [2]. Tingginya aktivitas pengguna tersebut menjadikan media sosial sebagai ruang komunikasi publik yang sangat aktif, khususnya pada platform berbasis teks seperti media sosial X.

Namun, tingginya intensitas komunikasi di media sosial juga memunculkan berbagai bentuk komunikasi negatif, seperti *hate speech* dan *abusive language*. *Hate speech* merupakan bentuk ujaran yang menyerang individu atau kelompok berdasarkan identitas tertentu, sedangkan *abusive language* berupa penggunaan kata-kata kasar, hinaan, dan serangan verbal [3] [4]. Kedua bentuk komunikasi negatif tersebut sering muncul dalam percakapan di media sosial dan dapat memicu konflik sosial serta menurunkan kualitas interaksi digital.

Permasalahan komunikasi negatif pada media sosial X semakin kompleks karena jumlah data teks yang sangat besar dan penggunaan bahasa yang tidak terstruktur. Pengguna media sosial sering menggunakan singkatan, bahasa gaul, serta istilah informal yang menyulitkan proses identifikasi secara manual [5]. Selain itu, karakteristik bahasa Indonesia yang beragam juga menjadi tantangan dalam analisis teks otomatis [6]. Oleh karena itu, diperlukan pendekatan komputasional yang mampu mengolah data teks secara efisien melalui metode *machine learning*.

Beberapa penelitian sebelumnya telah menerapkan metode *text mining* untuk mendeteksi komunikasi negatif pada media sosial. Penelitian pada platform Twitter, YouTube, dan aplikasi digital menunjukkan bahwa algoritma *Naive Bayes*

mampu memberikan hasil klasifikasi yang cukup baik [5] [7] [8] [9]. Selain itu, beberapa penelitian juga menggunakan algoritma lain seperti *Support Vector Machine* dan *deep learning* untuk meningkatkan performa klasifikasi teks [10] [11] [12]. Hasil penelitian tersebut menunjukkan bahwa *machine learning* memiliki peran penting dalam proses identifikasi komunikasi negatif pada media sosial.

Meskipun demikian, sebagian besar penelitian sebelumnya masih memisahkan *hate speech* dan *abusive language* sebagai kategori yang berbeda. Padahal, kedua bentuk komunikasi negatif tersebut sering muncul secara bersamaan dalam interaksi media sosial [13]. Selain itu, penelitian terkait komunikasi negatif pada media sosial X dengan dataset berbahasa Indonesia masih terbatas. Beberapa penelitian juga lebih banyak menggunakan algoritma yang kompleks sehingga sulit diterapkan pada penelitian dasar. Oleh karena itu, penggunaan algoritma *Naive Bayes* yang sederhana namun efektif masih relevan untuk diteliti lebih lanjut [14] [15].

Penelitian ini menggunakan dataset Indonesian *Abusive and Hate Speech Twitter Text* yang telah dianotasi untuk kategori *hate speech* dan *abusive language* [16]. Dataset tersebut digunakan untuk menganalisis komunikasi negatif pada media sosial X menggunakan algoritma *Naive Bayes*. Algoritma ini dipilih karena memiliki proses komputasi yang sederhana, efisien, dan cukup baik dalam klasifikasi data teks [17] [18]. Evaluasi model dilakukan menggunakan *accuracy, precision, recall*, dan F1-*score* untuk mengetahui kinerja algoritma dalam mengklasifikasikan komunikasi negatif.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis komunikasi negatif pada media sosial X berdasarkan *hate speech* dan *abusive language* menggunakan algoritma *Naive Bayes*. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan kajian *text mining* berbahasa Indonesia serta menjadi referensi dalam pengembangan sistem deteksi komunikasi negatif di media sosial.

1.2 Rumusan Masalah

1. Bagaimana proses klasifikasi komunikasi negatif pada media sosial X berdasarkan *hate speech* dan *abusive language* menggunakan algoritma *Naive Bayes*?
2. Bagaimana kinerja algoritma *Naive Bayes* dalam mengklasifikasikan komunikasi negatif dan non-negatif pada media sosial X berdasarkan dataset berbahasa Indonesia?

1.3 Batasan Masalah

Penelitian ini difokuskan pada analisis komunikasi negatif di media sosial X dengan pendekatan kuantitatif menggunakan algoritma klasifikasi teks. Mengingat luasnya ruang lingkup komunikasi digital dan beragamnya bentuk interaksi di media sosial, diperlukan batasan yang jelas agar penelitian dapat dilakukan secara terarah, realistis, dan menghasilkan temuan yang fokus. Adanya batasan ini juga disebabkan oleh keterbatasan dalam tujuan penelitian, ketersediaan data, metode yang digunakan, serta waktu dan sumber daya penelitian. Dengan adanya batasan, analisis yang dilakukan diharapkan mampu memberikan hasil yang lebih mendalam dan dapat dipertanggungjawabkan secara ilmiah. Oleh karena itu, penelitian ini hanya mencakup aspek-aspek yang relevan secara langsung dengan tujuan penelitian.

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya menganalas komunikasi negatif yang didefinisikan sebagai cuitan yang berisi *hate speech* dan *abusive language*, sesuai dengan karakteristik dataset yang digunakan. Bentuk komunikasi negatif lain seperti misinformasi, hoaks, atau sarkasme tidak menjadi fokus penelitian ini karena memerlukan pendekatan dan indikator yang berbeda.
2. Objek penelitian terbatas pada media sosial X (*Twitter*), sehingga hasil penelitian tidak dimaksudkan untuk merepresentasikan komunikasi negatif pada platform media sosial lain seperti

Facebook, Instagram, atau TikTok. Pembatasan ini dilakukan karena setiap platform memiliki ciri khas komunikasi dan kebijakan penggunaan yang berbeda.

3. Dataset yang digunakan dalam penelitian ini merupakan data sekunder dalam bahasa Indonesia, khususnya dalam konteks uang yang dimiliki oleh sebagian dari pengaruh yang signifikan. Penelitian ini tidak melibatkan pengambilan data langsung melalui API atau *crawling*. Pembatasan ini bertujuan untuk memastikan konsistensi serta validitas data yang telah melalui proses anotasi sebelumnya.
4. Variabel utama yang dianalisis dalam penelitian ini terbatas pada teks cuitan (*Tweet*) menjadi data input, sementara label komunikasi negatif sebagai variabel output. Label komunikasi negatif dibentuk dari penggabungan variabel *Hate Speech* (HS) dan *Abusive Language*, sehingga tidak dilakukan analisis terhadap subkategori atau tingkat keparahan ujaran kebencian.
5. Metode analisis yang digunakan terbatasi hanya pada algoritma *Naive Bayes*, tanpa melakukan perbandingan dengan algoritma klasifikasi lain seperti *Support Vector Machine, Decision Tree*, atau metode *deep learning*. Pembatasan ini dilakukan agar fokus penelitian tetap pada evaluasi kinerja satu algoritma yang sederhana dan umum digunakan.
6. Evaluasi kinerja model dalam penelitian ini terbatasi pada metrik evaluasi klasifikasi, seperti akurasi, *precision, recall*, dan F1-*score*. Penelitian ini tidak membahas aspek implementasi sistem secara *real-time* maupun pengembangan aplikasi deteksi komunikasi negatif.

Dengan batasan masalah tersebut, penelitian ini diharapkan mampuan lebih terfokus dalam menganalisis komunikasi negatif pada media sosial X serta

menghasilkan penemuan yang relevan dengan tujuan penelitian dan konteks keilmuan yang dikaji.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis komunikasi negatif pada media sosial X dengan pendekatan kuantitatif menggunakan algoritma *Naive Bayes*. Analisis dilakukan berdasarkan keberadaan *hate speech* dan bahasa *abusive* dalam cuitan pengguna media sosial X. Penelitian ini bertujuan memberikan gambaran yang sistematis dan terukur mengenai pola komunikasi negatif yang terjadi di ruang digital, khususnya pada media sosial berbasis teks. Selain itu, penelitian ini juga bertujuan untuk memahami bagaimana perubahan dalam tingkat penggunaan media sosial dapat memengaruhi klasifikasi teks. Dengan tujuan tersebut, penelitian diharapkan dapat menghasilkan data yang objektif dan dapat dipertanggungjawabkan secara ilmiah.

Secara khusus, tujuan penelitian ini adalah sebagai berikut:

1. Untuk mengetahui proses klasifikasi komunikasi negatif pada media sosial X berdasarkan *hate speech* dan *abusive language* menggunakan algoritma *Naive Bayes*.
2. Untuk mengevaluasi kinerja algoritma *Naive Bayes* dalam mengklasifikasikan komunikasi negatif dan non-negatif pada media sosial X menggunakan dataset berbahasa Indonesia.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoretis maupun praktis dalam pengembangan kajian komunikasi digital dan pemanfaatan teknologi analisis teks. Dengan fokus pada komunikasi negatif di media sosial X, penelitian ini memberikan kontribusi terhadap pemahaman fenomena komunikasi di ruang digital yang semakin kompleks. Selain itu, penggunaan pendekatan kuantitatif berbasis algoritma klasifikasi diharapkan mampu menghasilkan temuan yang objektif dan terukur. Objektif penelitian ini juga diharapkan dapat memberikan manfaat dalam berbagai bidang, baik akademik maupun praktis. Oleh

karena itu, manfaat penelitian ini dibagi menjadi manfaat teoretis dan manfaat praktis.

1. Manfaat Teoritis

   Secara teoretis, penelitian ini memberikan kontribusi dalam pengembangan kajian ilmu komputer, khususnya pada bidang *text mining* dan analisis data media sosial. Penelitian ini memperluas pemahaman tentang konsep komunikasi negatif dengan mengintegrasikan *hate speech* dan *abusive language* dalam satu kerangka analisis. Selain itu, penelitian ini menambahkan referensi ilmiah terkait penerapan algoritma *Naive Bayes* dalam klasifikasi teks. Kemampuan penelitian ini juga dapat menjadi dasar bagi penelitian selanjutnya yang mengkaji komunikasi digital dan analisis sentimen. Dengan demikian, penelitian ini diharapkan dapat memperkaya literatur akademik yang relevan dengan topik komunikasi negatif di media sosial.

2. Manfaat Praktis

   Secara praktis, hasil penelitian ini dapat digunakan sebagai referensi bagi pengembang sistem moderasi konten dalam mendeteksi komunikasi negatif secara otomatis di media sosial. Penelitian ini juga dapat menjadi pertimbangan bagi pengelola platform media sosial dan pembuat kebijakan dalam merancang strategi penanganan komunikasi negatif di ruang digital. Selain itu, penelitian ini dapat membantu meningkatkan kesadaran masyarakat mengenai dampak negatif penggunaan bahasa di media sosial. Konsep ini merupakan bagian penting dalam memahami bagaimana perubahan penghasilan mempengaruhi perilaku ekonomi. Penelitian ini diharapkan dapat mendukung terbentuknya lingkungan komunikasi digital yang lebih sehat.

1.6 Sistematika Penulisan

Penulisan skripsi ini disusun secara sistematis ke dalam lima bab agar pembahasan penelitian tersaji secara terstruktur dan mudah dipahami. Adapun sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN, Bab ini membahas gambaran umum penelitian yang meliputi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta sistematika penulisan. Bab ini bertujuan untuk memberikan pemahaman awal mengenai konteks, arah, dan ruang lingkup penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA, Bab ini berisi landasan teori dan kajian pustaka yang relevan dengan penelitian. Pembahasan meliputi konsep komunikasi negatif, *hate speech* dan *abusive language*, media sosial X, *text mining*, algoritma *Naive Bayes*, serta penelitian terdahulu yang berkaitan. Bab ini juga menyajikan kerangka pemikiran yang menjadi dasar dalam pelaksanaan penelitian.

BAB III METODE PENELITIAN, Bab ini menjelaskan metode penelitian yang digunakan, meliputi jenis dan pendekatan penelitian, sumber dan karakteristik data, teknik pengumpulan data, tahapan preprocessing teks, metode klasifikasi menggunakan algoritma *Naive Bayes*, serta teknik evaluasi kinerja model. Bab ini bertujuan untuk menjelaskan secara rinci proses penelitian yang dilakukan.

BAB IV HASIL DAN PEMBAHASAN, Bab ini menyajikan hasil penelitian yang diperoleh dari proses klasifikasi komunikasi negatif pada media sosial X. Pembahasan meliputi hasil pengujian model, evaluasi kinerja algoritma *Naive Bayes*, serta analisis terhadap temuan penelitian. Hasil penelitian juga dibandingkan dengan penelitian terdahulu yang relevan untuk memperkuat pembahasan.

BAB V PENUTUP, Bab ini berisi kesimpulan yang merangkum hasil penelitian sesuai dengan tujuan yang telah ditetapkan serta saran yang dapat digunakan sebagai bahan pertimbangan untuk penelitian selanjutnya atau pengembangan lebih lanjut di bidang terkait.