

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era transformasi digital saat ini, media sosial telah berevolusi dari sekadar media interaksi personal menjadi instrumen strategis bagi institusi pendidikan tinggi untuk membangun citra publik (*branding*) dan layanan informasi. Instagram, sebagai platform visual utama, menjadi kanal komunikasi paling aktif yang menghubungkan universitas dengan stakeholder, terutama mahasiswa yang didominasi oleh Generasi Z. Fenomena ini menciptakan volume data tekstual yang masif dalam bentuk komentar, yang sejatinya merupakan aset data berharga bagi institusi untuk memahami persepsi dan tingkat kepuasan publik secara *real-time* [1].

Namun, lonjakan pertumbuhan data (*big data*) pada akun media sosial resmi instansi sering kali tidak diimbangi dengan kemampuan pengolahan data yang mumpuni. Berdasarkan tren global, organisasi yang tidak memanfaatkan *Social Listening* cenderung lambat dalam merespons krisis opini. Universitas Malikussaleh (Unimal), sebagai salah satu perguruan tinggi negeri terbesar di Aceh, memiliki intensitas interaksi yang sangat tinggi pada akun Instagram resminya. Setiap informasi mengenai kebijakan akademik, UKT, hingga prestasi mahasiswa memicu ribuan komentar yang mengandung sentimen beragam mulai dari apresiasi hingga kritik tajam [2].

Permasalahan utama yang dihadapi saat ini adalah pengelolaan komentar tersebut masih dilakukan secara konvensional atau manual. Keterbatasan sumber daya manusia untuk membaca dan mengklasifikasikan ribuan komentar mengakibatkan banyak aspirasi atau keluhan mahasiswa yang terabaikan. Padahal, sentimen negatif yang tidak terdeteksi sejak dini berisiko merusak reputasi digital universitas. Oleh karena itu, diperlukan sebuah sistem cerdas yang mampu melakukan klasifikasi sentimen secara otomatis untuk memetakan opini publik secara akurat [3].

Efektivitas penggunaan algoritma *machine learning* dalam menangani data media sosial telah dibuktikan oleh beberapa penelitian terdahulu. Algoritma NBC

terbukti mampu mengklasifikasikan sentimen pengguna aplikasi BLU BCA di PlayStore dengan akurasi yang cukup tinggi, sehingga membuktikan kehandalan NBC dalam mengolah data teks ulasan berbasis platform digital [4]. Efektivitas NBC juga terbukti dari performanya yang stabil dan konsisten dalam pengolahan sentimen ulasan dan komentar di berbagai platform digital [5].

Dalam perbandingan algoritma, meskipun SVM memiliki akurasi yang kompetitif, NBC terbukti lebih unggul dalam kecepatan komputasi yang krusial bagi sistem berbasis web [6]. Keunggulan serupa juga ditunjukkan dalam penelitian yang membuktikan bahwa model *machine learning* mampu mengklasifikasikan sentimen komentar pada berbagai platform digital, termasuk YouTube [7].

Setiap algoritma memiliki keunggulan tersendiri bergantung pada karakteristik dan topik data yang dianalisis, sehingga pemilihan algoritma perlu disesuaikan dengan kebutuhan penelitian [8]. Efektivitas NBC juga dikonfirmasi pada domain teknologi seperti *Metaverse*, di mana NBC terbukti mampu mengklasifikasikan sentimen dengan hasil yang akurat [9].

Pendekatan *text mining* berbasis NBC dan SVM terbukti efektif digunakan untuk menganalisis opini publik di Twitter, menghasilkan klasifikasi sentimen yang relevan dan bermakna [10]. NBC juga berhasil diterapkan langsung pada data komentar dari platform Instagram dan menghasilkan klasifikasi sentimen yang akurat, sehingga memperkuat relevansi pemilihan NBC dalam penelitian ini [11].

Keandalan NBC dalam mengklasifikasikan sentimen berbasis teks bahasa Indonesia juga terbukti pada analisis ulasan aplikasi digital, di mana NBC mampu mengklasifikasikan sentimen masyarakat terhadap aplikasi TikTok dari 100.000 data ulasan dengan akurasi 88,2%, membuktikan efektivitasnya dalam menangani data teks berskala besar [12]. NBC juga berhasil diterapkan pada data komentar Instagram institusi berbasis komunitas, di mana perbandingan empat algoritma *machine learning* menunjukkan bahwa setiap algoritma memiliki karakteristik performa yang berbeda bergantung pada jenis dan distribusi data komentar yang dianalisis [13].

Penerapan NBC dalam analisis sentimen pada ranah pembayaran digital terbukti efektif pada penelitian analisis sentimen QRIS di luar negeri yang menghasilkan tingkat klasifikasi yang cukup baik [14]. NBC juga mampu mengklasifikasikan sentimen

dengan akurasi yang memadai pada domain konten media sosial bisnis, seperti yang ditunjukkan pada penelitian terhadap caption Instagram *for Business*. [1]

Pada konteks isu kebijakan publik, analisis sentimen terhadap RUU Pilkada 2024 di media sosial Instagram membuktikan bahwa NBC efektif menangkap opini masyarakat dari data teks tidak terstruktur [2]. Kemampuan NBC juga meluas pada domain deteksi informasi palsu, di mana penelitian mendeteksi berita hoaks menggunakan NBC menunjukkan hasil yang menjanjikan dalam pengenalan pola teks manipulatif [14].

Analisis sentimen pada akun Instagram tokoh publik menggunakan NBC pun memberikan hasil yang akurat dalam mengidentifikasi polaritas opini masyarakat, sehingga semakin memperkuat relevansi dan kesesuaian pemilihan NBC dalam penelitian ini [15].

Meskipun penelitian mengenai analisis sentimen sudah banyak dilakukan, terdapat beberapa celah penelitian (*research gap*) yang menjadi dasar urgensi penelitian ini. Sebagian besar penelitian masih berfokus pada objek data yang bersifat umum, sementara karakteristik komentar Instagram institusi pendidikan memiliki tantangan tersendiri berupa penggunaan emoji yang intensif dan ragam bahasa yang lebih bervariasi. Penelitian yang menggunakan SVM pada akun Instagram Gojek pun belum menyentuh konteks institusi pendidikan di daerah dengan karakteristik bahasa campuran seperti yang terdapat pada basis pengguna Universitas Malikussaleh [16].

Selain pada ranah analisis sentimen media sosial, algoritma *Naive Bayes* juga terbukti efektif diterapkan pada domain klasifikasi berbasis kriteria yang berbeda, seperti klasifikasi tingkat kelayakan lahan tanaman padi berdasarkan karakteristik lahan (temperatur udara, curah hujan, kelembaban, tekstur tanah, pH, drainase, dan tinggi tanah), dengan tingkat akurasi sebesar 80% [17]. Hal ini menunjukkan fleksibilitas algoritma *Naive Bayes* dalam menangani berbagai jenis data terstruktur, tidak terbatas pada data teks komentar media sosial saja. Di sisi lain, algoritma *Support Vector Machine* juga terbukti mampu mengklasifikasikan data perilaku yang kompleks, seperti tingkat ketergantungan mahasiswa terhadap penggunaan ChatGPT ke dalam tiga kategori (rendah, sedang, dan tinggi) dengan akurasi mencapai 95% [18]. sehingga

memperkuat relevansi kedua algoritma tersebut sebagai metode klasifikasi yang andal pada berbagai jenis dataset.

Tahapan *preprocessing* teks merupakan kunci utama dalam meningkatkan akurasi model klasifikasi, sehingga penelitian ini akan mengembangkan modul *preprocessing* yang lebih adaptif terhadap bahasa "Gen Z" di lingkungan kampus. Dengan mengintegrasikan keunggulan kecepatan komputasi NBC ke dalam sebuah platform berbasis *website*, penelitian ini akan memberikan kontribusi baru berupa sistem *monitoring* opini publik yang praktis dan efisien [19].

Meskipun beberapa penelitian terdahulu telah mencoba mengkaji sentimen pada berbagai *platform digital*, namun sebagian besar belum mengintegrasikan hasil analisis tersebut ke dalam sebuah sistem *dashboard* berbasis web yang dapat digunakan secara langsung oleh pihak manajemen institusi. Di sisi lain, tren pengolahan data saat ini telah bergeser ke arah otomatisasi menggunakan teknik *Text Mining*. Penelitian ini hadir untuk menutup celah tersebut dengan menerapkan algoritma *Naive Bayes Classifier* yang mampu bekerja secara objektif berdasarkan probabilitas kata langsung dari data komentar. Keunggulan penelitian ini terletak pada pengembangan sistem berbasis web yang mengintegrasikan data *real-time* Instagram Universitas Malikussaleh dengan modul *preprocessing* teks yang disesuaikan dengan karakteristik bahasa mahasiswa di lingkungan Unimal [3].

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana cara melakukan scraping dan preprocessing data komentar Instagram Universitas Malikussaleh agar dapat diolah oleh algoritma *Naive Bayes*?
2. Bagaimana tingkat akurasi algoritma *Naive Bayes Classifier* dalam mengklasifikasikan sentimen opini publik (positif, negatif, dan netral) pada akun Instagram resmi Universitas Malikussaleh?
3. Bagaimana merancang dan membangun sistem analisis sentimen berbasis web yang mampu menyajikan visualisasi data opini publik secara efektif bagi pihak manajemen Universitas Malikussaleh?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah, maka diberikan batasan masalah sebagai berikut:

1. Objek penelitian berfokus pada akun Instagram resmi Universitas Malikussaleh @univ.malikussaleh.
2. Kategori Sentimen klasifikasi opini dibatasi menjadi tiga kelas, yaitu: Positif, Negatif, dan Netral.
3. Metode klasifikasi yang digunakan adalah *Naive Bayes Classifier*.
4. Sistem dikembangkan berbasis *website* menggunakan bahasa pemrograman Python dengan *framework* Flask dan MySQL sebagai sistem manajemen basis data (*database*).
5. Output: Output Sistem hanya menyajikan visualisasi statistik sentimen dalam bentuk *dashboard web* dan tidak melakukan interaksi otomatis (seperti membalas komentar) ke *platform* Instagram.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengimplementasikan teknik *scraping* dan *preprocessing* teks yang optimal untuk menangani karakteristik bahasa campuran (formal, informal, dan bahasa daerah) pada komentar Instagram Universitas Malikussaleh.
2. Mengukur performa dan tingkat akurasi algoritma *Naive Bayes Classifier* dalam mengklasifikasikan opini publik mahasiswa terhadap Universitas Malikussaleh.
3. Menghasilkan sebuah sistem analisis sentimen berbasis web yang dapat digunakan sebagai alat bantu monitoring bagi pihak humas atau manajemen universitas.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi berbagai pihak yang berkepentingan, baik secara teoretis maupun praktis. Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan *referensi* tambahan dalam bidang *text mining* dan *Natural Language Processing* (NLP), khususnya mengenai penerapan algoritma *Naive Bayes*

Classifier pada data komentar media sosial Instagram dengan karakteristik ragam bahasa daerah.

2. Memberikan masukan berbasis data (*data-driven*) kepada pihak birokrasi Universitas Malikussaleh untuk mengevaluasi kebijakan maupun informasi yang dipublikasikan melalui akun Instagram resmi kampus.
3. Mempermudah tugas pengelola media sosial Humas Universitas Malikussaleh dalam memantau tren opini publik secara cepat tanpa harus membaca komentar secara manual.
4. Menyediakan sistem berbasis web yang dapat diakses oleh pihak kampus untuk memantau dan menganalisis sentimen opini publik terhadap Universitas Malikussaleh secara berkala.
5. Meningkatkan pemahaman dan keterampilan peneliti dalam pengembangan sistem berbasis web serta pengolahan data teks menggunakan metode *machine learning*.
6. Mendukung pemanfaatan teknologi informasi dalam pengelolaan layanan institusi pendidikan, khususnya dalam bidang analisis sentimen dan pengambilan keputusan berbasis data.