

BAB I

PENDAHULUAN

1.1. Latar Belakang

Revolusi industri 4.0 yang ditandai dengan kemajuan teknologi digital telah mendorong transformasi besar-besaran dalam cara manusia berinteraksi dan menyebarkan informasi. Salah satu dampak paling nyata adalah ledakan data yang dihasilkan melalui media sosial. DataReportal mencatat bahwa terdapat lebih dari 4,95 miliar pengguna internet aktif di seluruh dunia, dan Twitter termasuk dalam platform utama yang digunakan untuk menyuarakan opini publik secara real-time. Dengan jumlah pengguna aktif bulanan mencapai lebih dari 450 juta, Twitter telah menjadi sumber data yang kaya dan dinamis, menggambarkan pemahaman masyarakat mengenai beragam isu, termasuk kebijakan di bidang pendidikan (Pramayasa et al., 2023).

Program Merdeka Belajar Kampus Merdeka (MBKM) adalah kebijakan yang diciptakan oleh Kementerian Pendidikan dan Kebudayaan Republik Indonesia dengan tujuan utama memfasilitasi mahasiswa dalam mengembangkan potensi diri secara optimal. Melalui program ini, mahasiswa diberi kesempatan untuk menentukan jalur pembelajaran yang paling sesuai dengan minat, bakat, serta aspirasi karier mereka, baik di dalam maupun di luar lingkungan kampus. Dengan pendekatan yang lebih fleksibel ini, MBKM mampu mencetak lulusan yang tidak hanya kuat secara akademis, tetapi juga adaptif dan siap bersaing dalam dunia kerja yang dinamis dan penuh tantangan. Sejak pertama kali diperkenalkan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, implementasi program MBKM telah menimbulkan beragam tanggapan dari berbagai kalangan masyarakat, baik yang mendukung inovasi tersebut maupun yang masih mempertanyakan efektivitasnya dalam meningkatkan kualitas pendidikan tinggi di Indonesia (Hanantaqi, 2023).

Media sosial, khususnya Twitter, menawarkan potensi besar sebagai cermin persepsi publik terhadap kebijakan semacam MBKM. Pengguna aktif di platform ini sering kali membagikan pengalaman pribadi, kritik, maupun pujian terhadap program-program pemerintah. Konten yang diproduksi secara organik oleh pengguna Twitter ini menjadi sumber data berharga bagi para peneliti untuk menangkap opini publik secara langsung dan dalam waktu nyata. Analisis terhadap data ini dapat mengungkap bagaimana persepsi masyarakat terbentuk dan berkembang, serta memberikan gambaran yang lebih utuh mengenai penerimaan kebijakan MBKM di kalangan mahasiswa (Fajar et al., 2025).

Untuk memahami persepsi publik dari data media sosial, analisis sentimen telah menjadi pendekatan yang banyak digunakan dalam ranah ilmu komputer dan ilmu sosial. Teknik ini memungkinkan identifikasi dan klasifikasi opini masyarakat ke dalam kategori seperti positif, dan negatif. Dalam konteks MBKM, analisis sentimen dapat berfungsi sebagai alat bantu evaluatif berbasis data dalam mengukur seberapa besar dukungan atau penolakan terhadap kebijakan tersebut. Dengan menganalisis ribuan unggahan di Twitter, peneliti dapat menyusun tren opini yang bermanfaat bagi pembuat kebijakan dalam pengambilan keputusan (Indrayuni et al., 2021).

Semangat Kampus Merdeka yang telah diterapkan selama beberapa tahun terakhir, beserta beragam program unggulannya, akan terus dikembangkan. Program ini akan diimplementasikan dengan aksi yang lebih nyata dan memberikan dampak langsung kepada masyarakat, dunia usaha, serta mendukung ekosistem riset dan inovasi untuk pembangunan nasional. Dengan mengusung gerakan baru yang disebut Kampus Berdampak, perguruan tinggi dapat hadir di tengah masyarakat sebagai pendorong peradaban, menerangi jalan, menyatukan harapan, dan menjadi wadah bagi terjadinya perubahan. Di sisi lain, muncul pula kritik dan kontra dari kalangan mahasiswa dan akademisi, antara lain terkait kurangnya kesiapan infrastruktur, beban administratif yang tinggi,

ketimpangan akses antar perguruan tinggi, serta kurangnya kejelasan dalam konversi nilai dan jaminan mutu pembelajaran di luar kampus(Adolph, 2025).

Untuk mengolah dan mengklasifikasikan sentimen secara akurat, pemilihan algoritma klasifikasi sangat penting. Metode seperti *K-Nearest Neighbors* (KNN) dan *Logistic Regression* telah terbukti efisien untuk klasifikasi teks pendek dan tidak berstruktur seperti tweet. *Logistic Regression* menghasilkan akurasi tertinggi dibandingkan algoritma lain pada data sentimen aplikasi digital(Nugroho & Kom, 2021).

Analisis perbandingan performa antara KNN dan *Logistic Regression* masih jarang ditemukan dalam konteks data Twitter yang berbahasa Indonesia informal, khususnya yang berkaitan dengan topik MBKM. Sebagian besar penelitian hanya menggunakan satu algoritma dan tidak menyertakan analisis komprehensif terkait *preprocessing*, *visualisasi*, dan *balancing data*. Ini menciptakan kesenjangan penelitian yang perlu diisi agar diperoleh hasil klasifikasi yang lebih akurat dan dapat digunakan sebagai dasar evaluasi kebijakan(Pramayasa et al., 2023).

Permasalahan utama dalam penelitian analisis sentimen media sosial adalah pada tahap *preprocessing* data. Teks dalam tweet sering kali tidak baku, penuh singkatan, emotikon, atau kode informal lainnya. Menekankan pentingnya tahapan *preprocessing* seperti tokenisasi, normalisasi, stemming, dan TF-IDF agar hasil klasifikasi menjadi optimal. Kegagalan dalam *preprocessing* dapat menyebabkan penurunan akurasi model secara signifikan(Setiawan, 2024).

Penelitian ini menjadi relevan karena menawarkan pendekatan evaluatif berbasis data terhadap program MBKM yang sedang berjalan. Dengan menggunakan algoritma *Logistic Regression* dan KNN secara bersamaan, penelitian ini dapat memberikan wawasan yang lebih luas terkait performa masing-masing metode, baik dari sisi akurasi, efisiensi, maupun interpretabilitas hasil klasifikasi(Setiawan, 2024).

Keunggulan pendekatan ini juga terletak pada visualisasi hasil sentimen secara temporal, sehingga dapat menunjukkan perubahan persepsi publik dari waktu ke waktu. Hal ini memungkinkan pengambil kebijakan untuk menelusuri dinamika respons mahasiswa, baik sebelum maupun sesudah pelaksanaan program MBKM di kampus-kampus (Shabrina Assyifa & Luthfiarta, 2024). Dari segi ilmiah, penelitian ini berkontribusi pada pengembangan metode *machine learning* dalam konteks pemrosesan bahasa alami (NLP) berbahasa Indonesia. Pendekatan gabungan KNN dan Logistic Regression juga dapat dijadikan acuan metodologis dalam penelitian-penelitian sejenis yang berkaitan dengan opini publik terhadap kebijakan pemerintah lainnya.

Tujuan utama penelitian ini adalah untuk membandingkan performa algoritma KNN dan *Logistic Regression* dalam mengklasifikasikan sentimen terhadap program MBKM berdasarkan data dari media sosial Twitter, serta memberikan hasil analisis visual yang dapat dimanfaatkan oleh pihak terkait sebagai alat evaluasi kebijakan (Wafda, 2025).

Manfaat dari penelitian ini bersifat teoritis dan praktis. Secara teoritis, penelitian ini memperkaya khazanah keilmuan dalam bidang analisis sentimen dan NLP berbahasa Indonesia. Secara praktis, hasil penelitian ini dapat dijadikan rujukan oleh Kementerian Pendidikan dan universitas dalam merumuskan kebijakan strategis berbasis masukan nyata dari mahasiswa dan masyarakat luas (Atmaja & Sukmasetya, 2024).

1.2. Rumusan Masalah

Berikut merupakan rumusan masalah pada penelitian ini yang terdiri dari permasalahan, solusi permasalahan, dan pertanyaan penelitian.

1.2.1. Permasalahan

Berdasarkan latar belakang yang telah dijelaskan, dapat diidentifikasi bahwa permasalahannya adalah pada program MBKM

(Merdeka Belajar Kampus Merdeka) belum ada penelitian yang secara khusus menganalisis sentimen publik terhadap program ini khususnya pada media sosial Twitter. Dalam permasalahan diatas, persepsi masyarakat terhadap program MBKM bisa juga menjadikan bahan evaluasi yang sangat penting dibutuhkan bagi pihak akademik khususnya. Oleh karena itu, diperlukan pengujian sentimen menggunakan metode KNN dan *Logistic Regression* terhadap program tersebut secara sistematis, sehingga nantinya output yang dihasilkan bisa menjadi evaluasi pada program MBKM agar meningkatkan program dari persepsi publik.

1.2.2. Solusi Permasalahan

Solusi yang diberikan dalam permasalahan yang telah dijelaskan yaitu dengan penelitian pendekatan analisis sentimen secara sistematis terhadap data yang bersumber dari media sosial Twitter. Metode yang digunakan dalam proses klasifikasi sentimen adalah *K-Nearest Neighbor (KNN)* dan *Logistic Regression*. Hasil dari proses analisis dan klasifikasi ini akan disajikan dalam bentuk visualisasi serta interpretasi data yang dapat menggambarkan opini masyarakat secara objektif. Informasi tersebut dapat memberikan masukan atau evaluasi bagi pihak akademik, penyusun kebijakan, serta pengelola program MBKM dalam rangka mengevaluasi dan mengembangkan program agar semakin relevan dan berdampak positif sesuai kebutuhan dan harapan mahasiswa maupun masyarakat luas.

1.2.3. Pertanyaan Penelitian

Berdasarkan permasalahan dan solusi yang diidentifikasi, maka pertanyaan penelitian yang dapat diusulkan dalam penelitian ini adalah sebagai berikut :

1. Bagaimana persepsi masyarakat terhadap program MBKM berdasarkan analisis sentimen di Twitter?
2. Bagaimana performa algoritma KNN dalam mengklasifikasikan sentimen (positif dan negatif) terhadap

tweet mengenai MBKM?

3. Apakah berpengaruh metode *Logistic Regression* dalam meningkatkan akurasi klasifikasi sentimen pada data Twitter?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Berdasarkan permasalahan, solusi, dan pertanyaan penelitian, maka tujuan penelitian ini adalah sebagai berikut :

1. Menganalisis persepsi masyarakat terhadap program MBKM (Merdeka Belajar Kampus Merdeka) dengan melalui klasifikasi sentimen (positif, negatif) berdasarkan data yang diambil pada Twitter, dengan bertujuan mengetahui opini menurut publik.
2. Mengevaluasi terkait penggunaan metode KNN dalam mengklasifikasikan sentimen (positif, negatif) pada data teks berbahasa Indonesia, khususnya pada tweet yang berkaitan dengan program MBKM, dengan memperhatikan akurasi, presisi, recall, dan F1-score sebagai indikator performa.
3. Menganalisis terkait penggunaan teknik *Logistic Regression* yang gunanya untuk memperkuat metode KNN (*K-Nearest Neighbor*) terkait peningkatan performa model klasifikasi sentimen KNN, khususnya untuk memperkuat akurasi yang dihasilkan dari jumlah data pada masing-masing kelas sentimen. Dari beberapa jurnal yang saya baca, metode yang sering digunakan yaitu KNN dan *Logistic Regression*. Karena itulah saya mengambil kedua algoritma diatas yang menjadikan metode untuk tugas akhir.

1.3.2. Manfaat

Penelitian ini memberikan manfaat sebagai berikut :

1. Mendapatkan gambaran objektif dan data mengenai opini publik yang nantinya membuat keterbukaan dari dua arah

yaitu publik dengan pihak akademik.

2. Bagi Penulis menambahkan wawasan pengetahuan terkait penggunaan metode KNN (*K-Nearest Neighbor*) dengan diperkuat oleh *Logistic Regression* pada studi kasus analisis sentimen pada program MBKM (Merdeka Belajar Merdeka).
3. Bagi Mahasiswa bisa menjadikan pembelajaran atau bisa menjadikan rujukan sebagai penelitian selanjutnya.

1.4. Batasan Masalah

Batasan masalah yang dapat diterapkan agar penelitian ini tidak terlalu luas dan terarah, sebagai berikut :

1. Data yang diambil untuk dilakukan analisis sentimen pada penelitian kali ini yaitu hanya berasal dari Twitter.
2. Penelitian ini menggunakan metode KNN (*K-Nearest Neighbor*) yang diperkuat dengan *Logistic Regression*.
3. Jumlah kelas sentimen ada 2, yaitu positif dan negatif.
4. Penelitian ini berfokus pada klasifikasi sentimen masyarakat terhadap program MBKM berdasarkan data Twitter dan evaluasi kinerja model KNN serta *Logistic Regression*.

1.5. Sistematika Penulisan Proposal

Pada penelitian ini sistematika penulisan skripsi yang digunakan adalah sebagai berikut :

- I. **BAB I PENDAHULUAN**, yang terdiri dari latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, dan sistematika penulisan skripsi.
- II. **BAB II KAJIAN PUSTAKA**, yang terdiri dari teori-teori pendukung seperti *K-Nearest Neighbor* (KNN), *Logistic Regression*, *Text Preprocessing*, Klasifikasi sentimen, *Machine Learning*, MBKM, dan penelitian terkait.
- III. **BAB III METODE PENELITIAN**, bab ini berisi penjelasan tentang data penelitian, diuraikan arsitektur sistem serta skenario uji coba.
- IV. **BAB IV HASIL DAN PEMBAHASAN**, bab ini menjelaskan

hasil dari perbandingan metode *K-Nearest Neighbor* (KNN) dan *Logistic Regression* dalam menganalisis sentimen terkait program Merdeka Belajar Kampus Merdeka pada media sosial X. Pembahasan meliputi proses pengumpulan data, preprocessing teks, pelabelan data, pembagian data latih dan data uji, proses klasifikasi, hingga evaluasi performa model menggunakan metrik yang relevan. Selain itu, bab ini juga membahas hasil perbandingan kedua metode untuk mengetahui model dengan kinerja terbaik dalam analisis sentimen.

- V. **BAB V PENUTUP**, bab ini menjelaskan kesimpulan dari hasil penelitian yang telah dilakukan terkait analisis sentimen mengenai program Merdeka Belajar Kampus Merdeka pada media sosial twitter menggunakan metode *K-Nearest Neighbor* (KNN) dan *Logistic Regression*. Kesimpulan disusun berdasarkan hasil analisis dan evaluasi performa model yang telah dilakukan. Selain itu, bab ini juga memuat saran untuk pengembangan penelitian selanjutnya, baik dari segi metode, pengolahan data, maupun penerapan pendekatan lain yang dapat digunakan dalam penelitian serupa di masa mendatang.