

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam dokumen penelitian yang berjudul Analisis Sentimen Dan Pemodelan Topik pada *Platform X* Menggunakan Metode SVM dan LDA (Studi Kasus: Program Makan Bergizi Gratis).

1.1 Latar Belakang

Indonesia masih menghadapi berbagai tantangan sosial dan ekonomi yang berdampak pada kemampuan masyarakat dalam memenuhi kebutuhan dasar seperti pangan. Keterbatasan dalam memperoleh pangan tersebut memicu munculnya permasalahan gizi, seperti stunting, yang banyak dialami oleh bayi, balita, dan anak-anak dari keluarga dengan kondisi ekonomi rendah. Hal ini menyebabkan adanya penurunan kualitas kesehatan serta kualitas sumber daya manusia di Indonesia (Maharani dkk., 2024).

Pemerintah menginisiasi Program Makan Bergizi Gratis (MBG) sebagai upaya untuk mengatasi permasalahan tersebut. Program ini menjadi salah satu kebijakan prioritas yang menarik perhatian publik dan memicu beragam respons dari masyarakat, sejak mulai diimplementasikan pada tahun 2025 (Rahmah dkk., 2025). Pelaksanaan program MBG melibatkan berbagai pemangku kepentingan, seperti Badan Usaha Milik Desa (BUMDes), Usaha Mikro, Kecil, dan Menengah (UMKM), serta koperasi untuk mendukung keberlanjutan dan inklusivitas kebijakan, serta memastikan ketersediaan pangan yang merata (Maharani dkk., 2024). Program ini menjadi salah satu kebijakan prioritas pemerintah yang mendapat perhatian publik secara luas, dengan cakupan dan anggaran yang signifikan.

Beragam opini dan sentimen di kalangan masyarakat muncul seiring dengan berjalannya program. Perbedaan pandangan tersebut banyak disampaikan melalui media sosial, terutama melalui *platform X* (Twitter), yang populer dimanfaatkan sebagai wadah bagi masyarakat untuk berbagi opini dan informasi (Fathoni dkk., 2025).

Opini-opini yang disampaikan pengguna di *platform* tersebut dapat menjadi sumber informasi yang bernilai dalam memahami persepsi masyarakat terhadap kebijakan yang

sedang berlangsung. Namun, opini yang disampaikan melalui media sosial umumnya berbentuk teks yang tidak terstruktur, sehingga sulit untuk dianalisis secara langsung. Oleh karena itu, diperlukan pendekatan *Natural Language Processing* (NLP) untuk mengolah dan menganalisis data secara sistematis. (Fathoni dkk., 2025).

Analisis sentimen merupakan salah satu teknik dalam bidang NLP yang digunakan untuk mengidentifikasi dan mengkaji opini atau sikap yang terkandung dalam suatu teks. Tujuan utama dari analisis sentimen adalah untuk mengetahui kecenderungan pandangan yang disampaikan, baik bersifat positif, negatif, maupun netral. Pendekatan ini dapat diterapkan pada berbagai jenis data teks, seperti ulasan produk, cuitan di media sosial, maupun konten teks lainnya (Amelia dkk., 2023).

Berbagai penelitian terkait analisis sentimen pada media sosial telah banyak dilakukan dengan menggunakan metode *machine learning*. Penelitian oleh Aziz dkk. (2025), berhasil menganalisis sentimen masyarakat terhadap program MBG menggunakan data pada *platform X* dengan menerapkan metode *Support Vector Machine* (SVM) dan *Logistic Regression* pada proses analisis sentimen. Hasil penelitian menunjukkan bahwa kedua metode tersebut mampu mengklasifikasikan sentimen publik dengan baik, sehingga pendekatan *machine learning* dinilai efektif dalam analisis sentimen berbasis media sosial.

Penelitian serupa juga dilakukan oleh Amelia dkk. (2025), yang berhasil mengkaji sentimen masyarakat terhadap program MBG melalui media sosial Twitter dengan menggunakan metode klasifikasi SVM dan *Decision Tree*. Hasil penelitian menunjukkan bahwa kedua metode tersebut mampu menghasilkan tingkat akurasi yang tinggi, namun metode SVM dinilai lebih unggul dalam menangani kelas sentimen prioritas.

Penelitian lain oleh Fathoni dkk. (2025), yang melakukan analisis sentimen terhadap kebijakan pemerintah menggunakan data pada *platform X* dengan menerapkan metode SVM. Hasil penelitian tersebut menunjukkan bahwa metode SVM memiliki performa yang baik dan efektif dalam mengklasifikasikan sentimen berdasarkan data teks dari sosial media.

Keberhasilan penerapan metode SVM dalam berbagai penelitian tersebut mengindikasikan bahwa metode ini memiliki kemampuan yang *robust* dalam menangani data teks berdimensi tinggi dan tidak terstruktur, seperti data dari media sosial.

Proses analisis sentimen masih menghadapi kendala pada tahap pelabelan data. Pelabelan manual membutuhkan waktu dan biaya yang besar serta rentan terhadap subjektivitas, sementara pendekatan berbasis leksikon memiliki keterbatasan dalam memahami konteks bahasa, seperti penggunaan *slang* atau sarkasme. Kondisi ini menyebabkan keterbatasan jumlah data berlabel yang dapat digunakan dalam proses pelatihan model.

Sebagian besar penelitian masih berfokus pada hasil analisis sentimen tanpa menggali topik pembahasan yang melatarbelakangi munculnya sentimen tersebut. Hasil yang diperoleh umumnya hanya menunjukkan distribusi sentimen positif, negatif, dan netral tanpa menguraikan topik spesifik yang melatarbelakangi munculnya sentimen tersebut.

Pemahaman mengenai isu atau topik apa saja yang menjadi sumber respons masyarakat sangat penting dalam konteks kebijakan publik. Analisis sentimen yang tidak disertai identifikasi topik cenderung menghasilkan gambaran yang bersifat umum dan belum mampu menjelaskan secara lebih mendalam alasan di balik kecenderungan sentimen publik (Drašković & Milanović, 2025).

Penelitian yang dilakukan oleh Kambayo dkk. (2025) menganalisis ulasan pengguna aplikasi Threads dengan mengkombinasikan metode LDA untuk pemodelan topik dan DistilBERT untuk analisis sentimen. Hasil pemodelan topik berhasil mengidentifikasi tiga topik utama, yaitu *Usage and Integration*, *Features and Comparisons*, serta *User Experience and Satisfaction*, dengan performa analisis sentimen yang tinggi mencapai akurasi sebesar 96,71%. Temuan penelitian ini menunjukkan bahwa kombinasi pemodelan topik dan analisis sentimen mampu memberikan pemahaman yang lebih komprehensif mengenai opini pengguna terhadap suatu layanan.

Penelitian serupa dilakukan oleh Kustiyahningsih dkk. (2024) yang menganalisis ulasan pengguna aplikasi EdLink menggunakan pendekatan analisis sentimen berbasis aspek, dengan memanfaatkan metode LDA untuk mengidentifikasi topik ulasan, metode *Lexicon-Based* untuk pelabelan sentimen, serta SVM untuk analisis sentimen. Hasil pemodelan topik berhasil mengidentifikasi tiga topik utama, yaitu *Usability*, *Reliability*, dan *Performance Efficiency*, dengan akurasi SVM mencapai 90,00% setelah penanganan ketidakseimbangan data. Temuan penelitian ini menunjukkan bahwa kombinasi pemodelan

topik dan analisis sentimen dapat memberikan gambaran yang lebih terstruktur mengenai persepsi pengguna terhadap suatu layanan.

Penelitian ini mengkombinasikan dua metode, yaitu *Support Vector Machine* (SVM), yang digunakan untuk analisis sentimen, dan *Latent Dirichlet Allocation* (LDA) yang digunakan untuk mengidentifikasi topik. Penelitian ini juga mengusulkan pendekatan analisis sentimen berbasis *semi-supervised learning* untuk mengatasi keterbatasan dalam proses pelabelan data.

LDA merupakan metode yang dipilih karena kemampuannya dalam mengekstraksi topik laten dari kumpulan dokumen teks secara *unsupervised*. Metode ini mampu mengidentifikasi pola topik yang tersembunyi dalam data teks serta banyak digunakan dalam analisis teks, khususnya pada data media sosial.

Support Vector Machine (SVM) juga digunakan karena memiliki kemampuan yang baik dalam menangani data teks berdimensi tinggi dan tidak terstruktur, seperti data media sosial (Fathoni dkk., 2025). Metode ini dikenal memiliki performa yang baik dalam proses klasifikasi, terutama dalam memisahkan data ke dalam kelas yang berbeda secara optimal.

Kombinasi metode LDA untuk identifikasi topik pembahasan dan SVM untuk analisis sentimen, digunakan untuk menghasilkan analisis yang lebih optimal, di mana LDA berperan dalam mengekstraksi topik dari data teks, sedangkan SVM digunakan untuk melakukan analisis sentimen secara akurat. Pendekatan *semi-supervised learning* memungkinkan pemanfaatan data berlabel dalam jumlah terbatas untuk melatih model awal, yang kemudian digunakan untuk membantu pelabelan data dalam jumlah yang lebih besar secara otomatis.

Penelitian ini tidak hanya berfokus pada pengukuran sentimen secara kuantitatif, tetapi juga berusaha memahami bagaimana opini publik berkembang di media sosial. Data Twitter (X) digunakan sebagai sumber data karena mampu merepresentasikan opini masyarakat secara *real-time* dan dinamis. Melalui data tersebut, penelitian ini bertujuan untuk melihat bagaimana persepsi publik terhadap Program MBG sejak awal implementasinya.

Penelitian ini mengkombinasikan metode *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi topik pembahasan dan *Support Vector Machine* (SVM) untuk

mengklasifikasikan sentimen, serta memanfaatkan pendekatan semi-supervised learning untuk mengatasi keterbatasan pelabelan data. Penelitian ini menghipotesiskan bahwa metode SVM dengan variasi kernel (linear, RBF, dan polynomial) mampu menghasilkan performa analisis sentimen yang baik, serta metode LDA mampu mengidentifikasi topik utama yang relevan dalam percakapan publik. Kombinasi kedua metode tersebut diharapkan mampu memberikan analisis yang lebih mendalam mengenai distribusi sentimen pada setiap topik sehingga dapat menggambarkan persepsi publik terhadap Program MBG secara lebih jelas dan terstruktur.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini berfokus pada bagaimana kualitas metode SVM dalam mengklasifikasikan sentimen opini publik terkait Program MBG pada *Platform X* (Twitter) dengan menggunakan beberapa jenis kernel, yaitu linear, RBF, dan *polynomial*. Selain itu, penelitian ini juga berfokus pada bagaimana pemodelan topik menggunakan LDA dapat mengidentifikasi topik-topik utama yang muncul dalam percakapan publik serta bagaimana distribusi sentimen pada setiap topik yang terbentuk.

1.3 Tujuan

Penelitian ini bertujuan untuk mengklasifikasikan sentimen opini publik terkait Program MBG pada *Platform X* (Twitter) dengan menggunakan beberapa jenis kernel, yaitu linear, RBF, dan *polynomial*. Selain itu, penelitian ini juga untuk mengidentifikasi topik-topik utama yang muncul dalam percakapan publik menggunakan metode LDA, serta mengetahui distribusi sentimen pada setiap topik yang terbentuk.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat baik secara akademis maupun praktis. Secara akademis, penelitian ini diharapkan dapat menjadi tambahan referensi dalam bidang analisis sentimen dan pemodelan topik, khususnya dalam penerapan metode LDA untuk identifikasi topik serta SVM dalam analisis sentimen.

Secara praktis, penelitian ini diharapkan dapat memberikan gambaran mengenai pola opini publik terhadap Program MBG berdasarkan topik-topik yang terbentuk serta distribusi sentimennya. Selain itu, hasil evaluasi metode SVM dengan berbagai kernel dapat

memberikan informasi mengenai konfigurasi model yang paling optimal dalam analisis sentimen berbasis media sosial, sehingga dapat menjadi referensi bagi penelitian selanjutnya di bidang analisis teks dan data sosial media.

1.5 Ruang Lingkup

Ruang lingkup pada penelitian berjudul Analisis Sentimen Dan Pemodelan Topik pada Platform X Menggunakan Metode SVM dan LDA (Studi Kasus: Program Makan Bergizi Gratis), dibatasi pada hal-hal sebagai berikut:

1. Data Twitter yang dianalisis merupakan data berbahasa Indonesia.
2. Data diperoleh melalui platform Kaggle, dengan rentang waktu isi *tweet* dari Januari hingga Oktober 2025.
3. Dataset yang digunakan terdiri dari data berlabel sebanyak 656 data yang digunakan untuk melatih model analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM).
4. Metode SVM yang telah dilatih kemudian digunakan untuk mengklasifikasikan sentimen pada data tidak berlabel sebanyak 9.792 data, data tersebut juga digunakan untuk pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA).
5. Analisis sentimen yang dihasilkan dibatasi pada dua kelas, yaitu sentimen positif dan negatif.
6. Metode SVM yang digunakan dalam penelitian ini diuji menggunakan beberapa jenis kernel, yaitu linear, *Radial Basis Function* (RBF), dan *polynomial*, untuk mengetahui kernel yang menghasilkan performa terbaik dalam analisis sentimen.
7. Evaluasi performa metode SVM dilakukan menggunakan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*.
8. Penelitian ini tidak membahas dampak kebijakan secara langsung, melainkan terbatas pada analisis opini publik di media sosial Twitter.
9. Penelitian ini tidak membahas analisis jaringan sosial, *bot detection*, maupun analisis sentimen berbasis multimodal.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran umum mengenai isi pada setiap bab yang disajikan sebagai berikut:

BAB I PENDAHULUAN

Bab ini menyajikan latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian, serta sistematika penulisan skripsi yang digunakan sebagai acuan dalam penyusunan penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan landasan teori mengenai Program MBG, tahapan pengumpulan data, *preprocessing data* teks Twitter, analisis sentimen menggunakan metode SVM, pemodelan topik menggunakan metode LDA, evaluasi performa metode, serta penelitian terdahulu yang berkaitan dengan topik penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan metode penelitian yang digunakan, meliputi tahapan pengumpulan data, *preprocessing data* teks Twitter, analisis sentimen menggunakan metode SVM, pemodelan topik menggunakan metode LDA, serta metode evaluasi performa.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian yang diperoleh dari proses analisis data, meliputi hasil analisis sentimen menggunakan metode SVM, hasil pemodelan topik menggunakan metode LDA, evaluasi performa model, serta pembahasan hasil penelitian berdasarkan tujuan yang telah ditetapkan.

BAB V PENUTUP

Bab ini menyajikan kesimpulan yang diperoleh dari hasil penelitian serta saran yang dapat digunakan sebagai bahan pertimbangan untuk penelitian selanjutnya.