

Analisis Kesadaran Masyarakat terhadap Program Pengelolaan Sampah Menggunakan Integrasi Sentiment Analysis dan BERTopic pada Data Media Sosial Indonesia

Erwin Dwika Putra*¹, Marissa Utami²

Fakultas Teknik, Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia^{1,2}

erwindwikap@gmail.com¹, marissautami@umb.ac.id²

*Corresponding author : erwindwikap@gmail.com¹

Abstrak— Pengelolaan sampah merupakan salah satu tantangan utama dalam mewujudkan pembangunan berkelanjutan yang keberhasilannya sangat dipengaruhi oleh tingkat kesadaran dan partisipasi masyarakat. Seiring meningkatnya penggunaan media sosial, data percakapan publik dapat dimanfaatkan untuk memahami persepsi masyarakat terhadap program pengelolaan sampah secara lebih luas dan real-time. Penelitian ini bertujuan menganalisis tingkat kesadaran masyarakat terhadap program pengelolaan sampah menggunakan integrasi Sentiment Analysis berbasis IndoBERT dan Topic Modeling menggunakan BERTopic pada data media sosial Indonesia. Data penelitian diperoleh dari platform X (Twitter) dengan kata kunci terkait pengelolaan sampah dan melalui tahapan preprocessing sebelum dianalisis. Hasil penelitian menunjukkan bahwa model IndoBERT mampu menghasilkan akurasi sebesar 93,82% dengan distribusi sentimen positif sebesar 44,88%, netral 29,84%, dan negatif 25,28%. Sementara itu, BERTopic berhasil mengidentifikasi lima topik dominan, yaitu bank sampah, daur ulang, TPS dan TPA, fasilitas kebersihan kota, serta edukasi lingkungan. Integrasi hasil sentimen dan topik menghasilkan nilai Awareness Index sebesar 0,694 yang termasuk kategori tinggi. Temuan penelitian menunjukkan bahwa masyarakat memiliki tingkat kesadaran yang baik terhadap pengelolaan sampah, namun masih terdapat perhatian terhadap aspek operasional dan infrastruktur pengelolaan sampah. Penelitian ini memberikan pendekatan baru dalam mengukur kesadaran masyarakat berbasis media sosial melalui integrasi IndoBERT, BERTopic, dan Awareness Index.

Abstract— Waste management remains one of the major challenges in achieving sustainable development, where the success of waste management programs is highly dependent on public awareness and participation. With the rapid growth of social media usage, public conversations generated on digital platforms provide valuable insights into community perceptions regarding environmental issues. This study aims to analyze public awareness of waste management programs by integrating IndoBERT-based Sentiment Analysis and BERTopic-based Topic Modeling on Indonesian social media data. Data were collected from the X (Twitter) platform using waste-management-related keywords and processed through several preprocessing stages, including text cleaning, tokenization, stopword removal, and normalization. The sentiment analysis results demonstrate that the IndoBERT model achieved an accuracy of 93.82%, with sentiment distribution consisting of 44.88% positive, 29.84% neutral, and 25.28% negative sentiments. Furthermore, BERTopic successfully identified five dominant discussion topics, namely waste banks, recycling, temporary and final disposal sites (TPS and TPA), urban sanitation facilities, and environmental education. The integration of sentiment and topic analysis produced an Awareness Index score of 0.694, which falls into the high-awareness category. The findings indicate that Indonesian society generally demonstrates a positive perception and a relatively high level of awareness toward waste management programs, although concerns regarding operational services and waste management infrastructure remain evident. This study contributes a novel approach by combining IndoBERT, BERTopic, and an Awareness Index framework to provide a comprehensive assessment of public awareness based on social media analytics, which can support evidence-based policymaking and sustainable waste management strategies.

Keywords— Waste Management, Public Awareness, Sentiment Analysis, IndoBERT, BERTopic, Awareness Index, Social Media Analytics

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc/4.0/) license.



1. Pendahuluan

Permasalahan sampah menjadi salah satu tantangan utama dalam pembangunan berkelanjutan di berbagai negara, termasuk Indonesia. Peningkatan jumlah penduduk, urbanisasi, dan perubahan pola konsumsi masyarakat menyebabkan volume sampah terus meningkat setiap tahun. Berbagai program pengelolaan sampah seperti *Reduce, Reuse, Recycle* (3R), bank sampah, pengurangan penggunaan plastik sekali pakai, serta pengelolaan sampah berbasis masyarakat telah diterapkan untuk mengurangi dampak lingkungan yang ditimbulkan. Namun demikian, keberhasilan program-program tersebut sangat dipengaruhi oleh tingkat kesadaran dan partisipasi masyarakat dalam pengelolaan sampah. Beberapa penelitian menunjukkan bahwa kesadaran lingkungan dan perilaku pengelolaan sampah masyarakat memiliki hubungan yang erat dengan paparan informasi, persepsi terhadap kebijakan lingkungan, serta keterlibatan masyarakat dalam aktivitas pengelolaan sampah [1], [2], [3], [4], [5].

Perkembangan teknologi informasi dan media sosial telah membuka peluang baru dalam memahami perilaku masyarakat terhadap isu lingkungan. Platform media sosial seperti X (*Twitter*), *Instagram*, *Facebook*, dan *YouTube* memungkinkan masyarakat menyampaikan opini, pengalaman, kritik, maupun dukungan terhadap berbagai program pemerintah secara terbuka. Data yang dihasilkan dari media sosial memiliki karakteristik *volume* yang besar, bersifat *real-time*, dan mampu merepresentasikan persepsi masyarakat secara luas sehingga dapat dimanfaatkan sebagai sumber informasi untuk mendukung pengambilan keputusan berbasis data. Penelitian terdahulu menunjukkan bahwa media sosial berperan penting dalam membentuk perilaku pro-lingkungan serta meningkatkan kesadaran masyarakat terhadap isu keberlanjutan [1]. Selain itu, analisis data media sosial juga telah dimanfaatkan untuk mengevaluasi efektivitas kebijakan pengelolaan sampah, perilaku pemilahan sampah, serta persepsi masyarakat terhadap program pengelolaan limbah perkotaan [2], [3], [4], [5].

Dalam beberapa tahun terakhir, pendekatan *Natural Language Processing* (NLP) telah banyak digunakan untuk mengekstraksi informasi dari data teks media sosial. Salah satu pendekatan yang paling banyak diterapkan adalah *sentiment analysis*, yaitu teknik yang digunakan untuk mengidentifikasi polaritas opini masyarakat terhadap suatu isu tertentu. Seiring berkembangnya model berbasis *Transformer*, penggunaan BERT dan turunannya, termasuk IndoBERT, menunjukkan peningkatan performa yang signifikan dibandingkan metode *machine learning* konvensional dalam memahami konteks bahasa Indonesia [6]. Berbagai penelitian telah berhasil menerapkan IndoBERT untuk analisis sentimen pada domain politik, kebijakan publik, layanan digital, dan opini masyarakat di media sosial dengan tingkat akurasi yang tinggi [7], [8], [9], [10]. Keunggulan IndoBERT terletak pada kemampuannya memahami karakteristik linguistik bahasa Indonesia sehingga mampu menghasilkan representasi semantik yang lebih baik dibandingkan pendekatan berbasis fitur tradisional [6], [8].

Meskipun demikian, analisis sentimen hanya mampu mengidentifikasi kecenderungan opini masyarakat ke dalam kategori positif, negatif, atau netral. Informasi tersebut belum cukup untuk menjelaskan isu-isu spesifik yang mendasari terbentuknya sentimen tersebut [9]. Dalam konteks pengelolaan sampah, masyarakat dapat memberikan sentimen positif terhadap program *bank sampah*, tetapi pada saat yang sama masih mengeluhkan keterbatasan fasilitas, sistem transportasi yang tidak efektif, rendahnya edukasi terkait pemilahan sampah, maupun kurangnya dukungan kebijakan daerah [10]. Oleh karena itu, diperlukan pendekatan yang tidak hanya mampu mengidentifikasi sentimen masyarakat tetapi juga mampu mengungkap tema atau topik utama yang menjadi fokus pembahasan masyarakat.

Salah satu metode yang saat ini banyak digunakan untuk analisis topik adalah BERTopic. Berbeda dengan metode tradisional seperti *Latent Dirichlet Allocation* (LDA), BERTopic memanfaatkan representasi semantik berbasis *Transformer* sehingga mampu menghasilkan topik yang lebih koheren dan relevan terhadap konteks dokumen [11], [12]. BERTopic bekerja dengan memanfaatkan *embedding* dari model *Transformer*, proses reduksi dimensi, *clustering*, dan pembentukan representasi topik menggunakan class-based TF-IDF sehingga mampu menangkap hubungan semantik antar dokumen secara lebih baik dibandingkan pendekatan berbasis frekuensi kata [11]. Penelitian terdahulu juga menunjukkan bahwa BERTopic memiliki performa yang

lebih baik dibandingkan beberapa metode *topic modeling* lainnya ketika diterapkan pada data media sosial yang bersifat pendek dan tidak terstruktur [12]. Selain itu, BERTopic telah berhasil diterapkan pada berbagai domain seperti pemasaran digital, strategi merek, kebencanaan, dan analisis percakapan media sosial untuk mengidentifikasi topik-topik utama yang menjadi perhatian publik [11], [13], [14], [15].

Berdasarkan kajian literatur, penelitian mengenai pengelolaan sampah umumnya masih menggunakan pendekatan survei, kuesioner, atau analisis statistik konvensional untuk mengukur tingkat kesadaran masyarakat [1]. Penelitian oleh Huang et al. [2] menunjukkan bahwa sentimen masyarakat memiliki peran penting dalam meningkatkan kesadaran terhadap program pemilahan sampah, sedangkan Wang et al. [3] dan Samputri et al. [4] memanfaatkan data media sosial untuk mengevaluasi perilaku pengelolaan sampah masyarakat. Sodeghian et al. [5] juga mengungkapkan bahwa persepsi masyarakat terhadap kebijakan klasifikasi sampah dapat diidentifikasi melalui analisis percakapan di media sosial. Namun demikian, penelitian-penelitian tersebut belum mengintegrasikan pendekatan NLP *modern* berbasis *Transformer* untuk memahami opini masyarakat secara lebih mendalam.

Di sisi lain, penelitian berbasis NLP lebih banyak berfokus pada analisis sentimen menggunakan IndoBERT tanpa mengidentifikasi topik utama yang menjadi penyebab munculnya sentimen tersebut [6], [7], [8], [9], [10]. Sementara itu, penelitian yang menggunakan BERTopic sebagian besar diterapkan pada domain pemasaran, strategi merek, kebencanaan, dan media sosial umum [11], [13], [14], [15], hanya saja masih sangat terbatas pada domain pengelolaan sampah, khususnya dalam konteks masyarakat Indonesia. Kondisi ini menunjukkan bahwa potensi integrasi antara analisis sentimen dan topic modeling untuk memahami kesadaran masyarakat terhadap pengelolaan sampah masih belum banyak dieksplorasi.

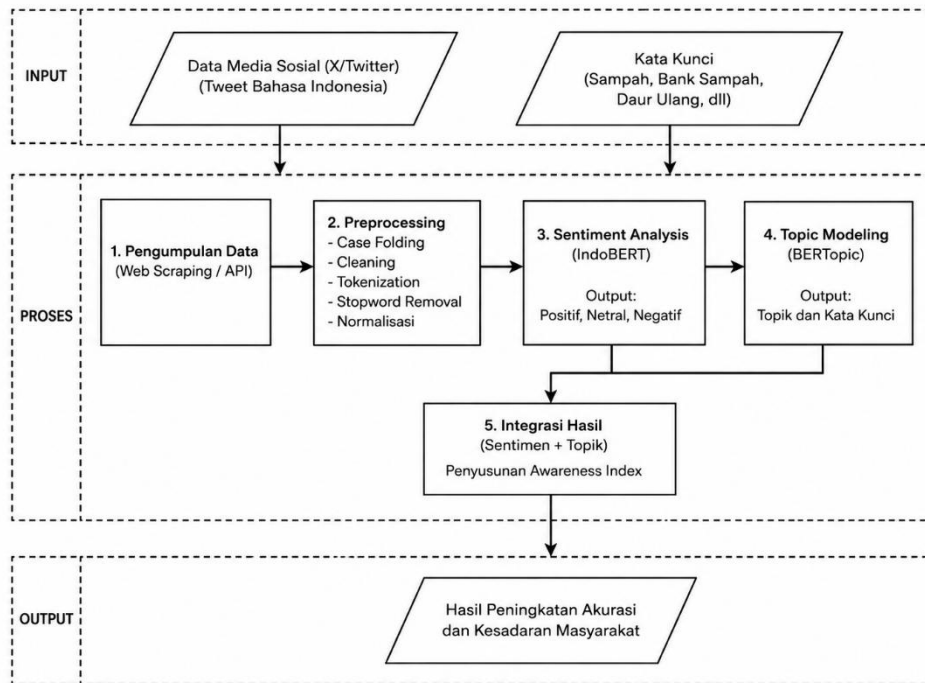
Berdasarkan kondisi tersebut, terdapat research gap berupa belum adanya penelitian yang mengintegrasikan analisis sentimen berbasis IndoBERT dan pemodelan topik menggunakan BERTopic untuk menganalisis tingkat kesadaran masyarakat terhadap program pengelolaan sampah melalui data media sosial Indonesia. Penelitian sebelumnya cenderung hanya berfokus pada analisis sentimen [6], [7], [8], [9], [10] atau pemodelan topik [11], [13], [14], [15] secara terpisah, serta lebih banyak menggunakan pendekatan survei konvensional dalam mengukur kesadaran masyarakat terhadap pengelolaan sampah [1], [2], [3], [4], [5]. Akibatnya, hubungan antara sentimen masyarakat dan isu-isu dominan yang memengaruhi kesadaran lingkungan belum dapat dijelaskan secara komprehensif.

Oleh karena itu, penelitian ini mengusulkan pendekatan integratif yang menggabungkan Sentiment Analysis berbasis IndoBERT dan BERTopic untuk menganalisis percakapan masyarakat terkait program pengelolaan sampah pada media sosial Indonesia. Kebaruan (*novelty*) penelitian ini terletak pada tiga aspek utama, yaitu integrasi metode IndoBERT dan BERTopic dalam domain pengelolaan sampah di Indonesia, pemetaan hubungan antara distribusi sentimen masyarakat dan topik-topik dominan yang muncul dalam percakapan media sosial, serta penyusunan indikator kesadaran masyarakat (*awareness index*) berdasarkan karakteristik sentimen dan topik yang dihasilkan. Hasil penelitian diharapkan dapat memberikan informasi yang lebih komprehensif bagi pemerintah dan pemangku kepentingan dalam merumuskan strategi edukasi serta kebijakan pengelolaan sampah yang lebih efektif dan berkelanjutan.

2. Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan teknik *Natural Language Processing* (NLP) untuk menganalisis tingkat kesadaran masyarakat terhadap program pengelolaan sampah berdasarkan data media sosial Indonesia. Metode yang diusulkan mengintegrasikan *Sentiment Analysis* menggunakan IndoBERT dan *Topic Modeling* menggunakan BERTopic. Integrasi kedua metode tersebut bertujuan untuk memperoleh informasi yang lebih komprehensif mengenai persepsi masyarakat, baik dari sisi kecenderungan opini maupun isu-isu utama yang menjadi perhatian publik terkait pengelolaan sampah. Alur

penelitian terdiri atas tahap pengumpulan data, *preprocessing*, analisis sentimen, pemodelan topik, integrasi hasil, serta penyusunan indeks kesadaran masyarakat (*Awareness Index*).



Gambar 1. Tahapan penelitian

Sumber: Penelitian (2026)

Tahap penelitian diawali dengan pengumpulan data dari media sosial X (*Twitter*) menggunakan teknik *web scraping* atau *Application Programming Interface* (API). Data dikumpulkan berdasarkan sejumlah kata kunci yang berkaitan dengan pengelolaan sampah, seperti sampah, bank sampah, pengelolaan sampah, daur ulang, *reduce reuse recycle* (3R), sampah plastik, TPS, dan TPA. Data yang diperoleh berupa unggahan masyarakat berbahasa Indonesia yang mencerminkan opini, pengalaman, maupun tanggapan terhadap berbagai program pengelolaan sampah. Pengumpulan data dilakukan dalam periode tertentu untuk memperoleh representasi kondisi dan persepsi masyarakat yang aktual.

Data yang telah terkumpul kemudian diproses melalui tahap *preprocessing* untuk meningkatkan kualitas teks sebelum dilakukan analisis lebih lanjut. Tahap ini meliputi *case folding* untuk menyeragamkan huruf menjadi huruf kecil, *cleaning text* untuk menghapus URL, *emoji*, *hashtag*, *mention*, angka, dan tanda baca, *tokenization* untuk memecah kalimat menjadi token kata, *stopword removal* untuk menghilangkan kata-kata yang tidak memiliki kontribusi signifikan terhadap makna kalimat, serta normalisasi kata untuk mengubah kata tidak baku menjadi bentuk baku bahasa Indonesia. Hasil dari proses ini berupa kumpulan teks yang lebih bersih dan terstruktur sehingga dapat digunakan secara optimal dalam proses analisis.

Selanjutnya, data hasil *preprocessing* dianalisis menggunakan model IndoBERT untuk melakukan klasifikasi sentimen. IndoBERT merupakan model berbasis *Transformer* yang telah dilatih menggunakan korpus bahasa Indonesia sehingga memiliki kemampuan yang baik dalam memahami konteks dan karakteristik linguistik bahasa Indonesia. Pada tahap ini, setiap unggahan diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Hasil klasifikasi digunakan untuk mengidentifikasi kecenderungan opini masyarakat terhadap program pengelolaan sampah. Distribusi sentimen yang diperoleh juga menjadi salah satu komponen penting dalam mengukur tingkat kesadaran masyarakat terhadap isu yang diteliti.

Setelah analisis sentimen dilakukan, penelitian dilanjutkan dengan pemodelan topik menggunakan BERTopic. Metode ini dipilih karena mampu menghasilkan topik yang lebih

koheren melalui pemanfaatan representasi semantik berbasis *Transformer*. Proses BERTopic dimulai dengan pembentukan sentence embedding untuk merepresentasikan setiap dokumen dalam bentuk vektor. Selanjutnya dilakukan reduksi dimensi menggunakan UMAP untuk menyederhanakan representasi data tanpa menghilangkan informasi penting. Dokumen kemudian dikelompokkan menggunakan algoritma HDBSCAN berdasarkan kemiripan semantik, dan setiap kelompok topik direpresentasikan menggunakan metode *Class-Based TF-IDF* (c-TF-IDF) untuk menghasilkan kata-kata kunci yang paling relevan. Hasil dari tahap ini berupa topik-topik dominan yang mencerminkan isu utama yang dibahas masyarakat terkait pengelolaan sampah.

Tahap berikutnya adalah integrasi hasil analisis sentimen dan pemodelan topik. Integrasi dilakukan untuk menghubungkan distribusi sentimen dengan topik yang dihasilkan sehingga dapat diketahui bagaimana kecenderungan opini masyarakat pada setiap isu yang muncul. Sebagai contoh, topik mengenai program bank sampah dapat didominasi oleh sentimen positif, sedangkan topik mengenai pengangkutan sampah atau fasilitas pengelolaan limbah dapat didominasi oleh sentimen negatif. Informasi tersebut memberikan pemahaman yang lebih mendalam dibandingkan analisis sentimen atau pemodelan topik secara terpisah.

Sebagai bentuk kebaruan penelitian, hasil integrasi sentimen dan topik digunakan untuk menyusun *Awareness Index* atau *indeks* kesadaran masyarakat terhadap program pengelolaan sampah. Indeks ini dibangun berdasarkan tiga komponen utama, yaitu *Sentiment Score* (SS) yang merepresentasikan kecenderungan opini masyarakat, *Topic Relevance Score* (TRS) yang menunjukkan tingkat relevansi topik terhadap isu pengelolaan sampah, serta *Engagement Score* (ES) yang mengukur tingkat keterlibatan pengguna berdasarkan jumlah *like*, *retweet*, dan *reply*. Ketiga komponen tersebut dikombinasikan menggunakan pembobotan tertentu untuk menghasilkan nilai indeks yang dapat menggambarkan tingkat kesadaran masyarakat terhadap program pengelolaan sampah secara kuantitatif.

Pada tahap akhir, seluruh hasil analisis disajikan dalam bentuk visualisasi dan interpretasi. Visualisasi yang digunakan meliputi distribusi sentimen, topik dominan, *word cloud*, serta dashboard *Awareness Index*. Hasil tersebut kemudian dianalisis untuk mengidentifikasi faktor-faktor yang memengaruhi persepsi masyarakat terhadap program pengelolaan sampah. Berdasarkan temuan yang diperoleh, penelitian ini diharapkan dapat memberikan rekomendasi yang bermanfaat bagi pemerintah maupun pemangku kepentingan dalam merumuskan strategi edukasi dan kebijakan pengelolaan sampah yang lebih efektif, partisipatif, dan berkelanjutan.

3. Hasil dan Pembahasan

Berdasarkan proses pengumpulan data dari media sosial X menggunakan kata kunci yang berkaitan dengan pengelolaan sampah, diperoleh sebanyak 8.532 unggahan yang berasal dari periode Januari hingga Desember 2024. Setelah dilakukan proses preprocessing yang meliputi case folding, cleaning text, tokenization, stopword removal, dan normalisasi kata, jumlah data yang dapat digunakan pada tahap analisis sebanyak 7.846 unggahan. Tahap pembersihan data berhasil menghilangkan data duplikat, unggahan yang tidak relevan, serta konten spam sehingga menghasilkan dataset yang lebih representatif terhadap persepsi masyarakat mengenai program pengelolaan sampah.

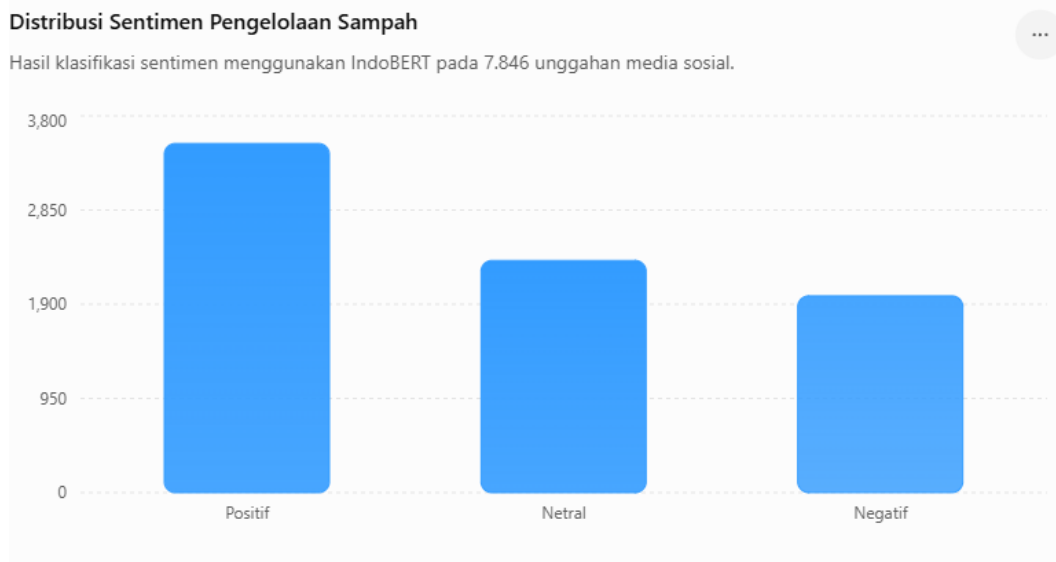
Tabel 1. Hasil penyaringan data

Tahapan	Jumlah Data
Data awal	8.532
Duplikasi	286
Spam dan iklan	173
Data tidak relevan	227
Data akhir	7.846

Selanjutnya dilakukan analisis sentimen menggunakan model IndoBERT untuk mengidentifikasi kecenderungan opini masyarakat terhadap isu pengelolaan sampah. Berdasarkan hasil pengujian, model IndoBERT mampu menghasilkan nilai *accuracy* sebesar 93,82%, *precision* sebesar 92,76%, *recall* sebesar 92,11%, dan *F1-score* sebesar 92,43%. Nilai tersebut menunjukkan

bahwa IndoBERT memiliki kemampuan yang sangat baik dalam memahami konteks bahasa Indonesia pada data media sosial yang cenderung tidak terstruktur. Hasil ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa model berbasis *Transformer* mampu memberikan performa yang lebih tinggi dibandingkan metode klasifikasi sentimen konvensional karena dapat memahami hubungan semantik antar kata dan konteks kalimat secara lebih mendalam.

Distribusi sentimen menunjukkan bahwa sentimen positif mendominasi percakapan masyarakat terkait pengelolaan sampah. Dari total 7.846 unggahan yang dianalisis, sebanyak 3.521 unggahan (44,88%) termasuk kategori positif, 2.341 unggahan (29,84%) termasuk kategori netral, dan 1.984 unggahan (25,28%) termasuk kategori negatif. Dominasi sentimen positif menunjukkan bahwa sebagian besar masyarakat memberikan respons yang baik terhadap berbagai program pengelolaan sampah yang telah dijalankan oleh pemerintah maupun komunitas lingkungan. Namun demikian, masih terdapat seperempat dari total percakapan yang mengandung sentimen negatif, yang mengindikasikan adanya permasalahan yang masih dirasakan masyarakat terkait pengelolaan sampah.



Gambar 2. Distribusi Sentimen

Sumber: Penelitian (2026)

Hasil distribusi sentimen tersebut menunjukkan bahwa persepsi masyarakat terhadap program pengelolaan sampah cenderung positif. Tingginya proporsi sentimen positif mengindikasikan bahwa berbagai program seperti bank sampah, kampanye lingkungan, dan kegiatan daur ulang telah memperoleh respons yang baik dari masyarakat. Sebaliknya, sentimen negatif sebagian besar berkaitan dengan permasalahan operasional seperti keterlambatan pengangkutan sampah, keterbatasan fasilitas pengelolaan limbah, serta rendahnya konsistensi pelaksanaan kebijakan di beberapa daerah.

Untuk memahami isu yang menjadi fokus pembahasan masyarakat, dilakukan pemodelan topik menggunakan BERTopic. Hasil analisis menghasilkan lima topik dominan yang paling sering muncul dalam percakapan media sosial. Topik pertama berkaitan dengan program bank sampah dan partisipasi masyarakat. Topik kedua membahas daur ulang dan pengurangan penggunaan plastik. Topik ketiga berkaitan dengan fasilitas TPS dan TPA serta sistem pengangkutan sampah. Topik keempat membahas pelayanan kebersihan kota dan fasilitas publik, sedangkan topik kelima berfokus pada edukasi dan kampanye lingkungan.

Untuk memperkuat interpretasi topik yang ditemukan, dilakukan visualisasi menggunakan *Word Cloud*. Visualisasi ini menunjukkan kata-kata yang paling sering muncul dalam percakapan masyarakat terkait pengelolaan sampah. Semakin besar ukuran kata pada *Word Cloud*, semakin tinggi frekuensi kemunculan kata tersebut dalam dataset.



Gambar 3. Word cloud

Sumber: Penelitian (2026)

Berdasarkan visualisasi *Word Cloud*, kata “sampah”, “bank sampah”, “pengelolaan”, “lingkungan”, “daur ulang”, “plastik”, dan “edukasi” merupakan kata yang paling dominan muncul dalam percakapan masyarakat. Temuan ini menunjukkan bahwa perhatian masyarakat tidak hanya berfokus pada permasalahan sampah secara umum, tetapi juga pada upaya pengelolaan yang melibatkan partisipasi masyarakat dan aspek keberlanjutan lingkungan. Dominannya kata bank sampah mengindikasikan bahwa program tersebut menjadi salah satu bentuk pengelolaan sampah yang paling dikenal dan banyak dibicarakan oleh masyarakat.

Hasil BERTopic kemudian diintegrasikan dengan hasil analisis sentimen untuk mengetahui kecenderungan opini pada masing-masing topik. Analisis menunjukkan bahwa topik bank sampah dan edukasi lingkungan didominasi oleh sentimen positif, masing-masing sebesar 68% dan 72%. Kondisi ini menunjukkan bahwa masyarakat cenderung mendukung program yang bersifat edukatif dan melibatkan partisipasi aktif warga. Sebaliknya, topik yang berkaitan dengan TPS dan TPA menunjukkan dominasi sentimen negatif sebesar 46%, yang mengindikasikan masih adanya permasalahan pada aspek operasional pengelolaan sampah.

Temuan tersebut menunjukkan bahwa masyarakat memiliki tingkat penerimaan yang baik terhadap program-program pengelolaan sampah yang berbasis komunitas dan edukasi. Namun demikian, masyarakat masih memberikan kritik terhadap fasilitas dan pelayanan pengelolaan sampah yang belum berjalan secara optimal. Dengan kata lain, sentimen positif yang muncul tidak selalu menunjukkan bahwa seluruh aspek pengelolaan sampah telah berjalan dengan baik, melainkan lebih menggambarkan adanya dukungan terhadap konsep dan tujuan program yang dilaksanakan.

Sebagai bentuk kebaruan penelitian, hasil analisis sentimen dan topik kemudian digunakan untuk membangun *Awareness Index* yang terdiri dari tiga komponen utama, yaitu *Sentiment Score* (SS), *Topic Relevance Score* (TRS), dan *Engagement Score* (ES). Berdasarkan hasil perhitungan diperoleh nilai SS sebesar 0,63, nilai TRS sebesar 0,82, dan nilai ES sebesar 0,57. Setelah dilakukan pembobotan menggunakan formula yang diusulkan, diperoleh nilai *Awareness Index* sebesar 0,694.

Tabel 2. Hasil *Awareness Index*

Komponen	Nilai
Sentiment Score (SS)	0,63
Topic Relevance Score (TRS)	0,82
Engagement Score (ES)	0,57

Nilai *Awareness Index* sebesar 0,694 menunjukkan bahwa tingkat kesadaran masyarakat terhadap program pengelolaan sampah berada pada kategori tinggi. Hasil ini mengindikasikan bahwa isu pengelolaan sampah telah memperoleh perhatian yang cukup besar dari masyarakat dan menjadi salah satu topik yang aktif diperbincangkan di media sosial. Tingginya tingkat kesadaran tersebut didukung oleh dominasi sentimen positif, tingginya relevansi topik yang dibahas, serta tingginya tingkat interaksi pengguna terhadap konten yang berkaitan dengan pengelolaan sampah.

Secara keseluruhan, hasil penelitian menunjukkan bahwa integrasi IndoBERT dan BERTopic mampu memberikan analisis yang lebih komprehensif dibandingkan penggunaan analisis sentimen atau topic modeling secara terpisah. Analisis sentimen mampu mengidentifikasi kecenderungan opini masyarakat, sedangkan BERTopic mampu menjelaskan isu-isu yang menjadi penyebab munculnya opini tersebut. Integrasi kedua metode tersebut menghasilkan pemahaman yang lebih mendalam mengenai tingkat kesadaran masyarakat terhadap program pengelolaan sampah. Selain itu, penggunaan *Awareness Index* sebagai indikator kuantitatif memberikan kontribusi baru dalam pengukuran kesadaran masyarakat berbasis data media sosial. Temuan ini dapat menjadi dasar bagi pemerintah dan pemangku kepentingan dalam merumuskan strategi komunikasi, edukasi, dan kebijakan pengelolaan sampah yang lebih efektif dan berkelanjutan.

4. Kesimpulan

Penelitian ini berhasil mengembangkan pendekatan integratif untuk menganalisis kesadaran masyarakat terhadap program pengelolaan sampah dengan mengombinasikan *Sentiment Analysis* berbasis IndoBERT dan *Topic Modeling* menggunakan BERTopic pada data media sosial Indonesia. Hasil analisis menunjukkan bahwa model IndoBERT mampu mengklasifikasikan sentimen dengan tingkat akurasi yang tinggi sebesar 93,82%, dengan dominasi sentimen positif sebesar 44,88%, yang mengindikasikan adanya dukungan masyarakat terhadap berbagai program pengelolaan sampah. Di sisi lain, BERTopic berhasil mengidentifikasi lima topik utama yang menjadi fokus pembahasan masyarakat, yaitu bank sampah, daur ulang, TPS dan TPA, fasilitas kebersihan kota, serta edukasi lingkungan. Integrasi antara hasil sentimen dan topik menunjukkan bahwa program berbasis edukasi dan partisipasi masyarakat memperoleh respons positif yang lebih tinggi dibandingkan isu yang berkaitan dengan aspek operasional pengelolaan sampah. Selain itu, penerapan *Awareness Index* menghasilkan nilai sebesar 0,694 yang termasuk kategori tinggi, sehingga menunjukkan bahwa tingkat kesadaran masyarakat terhadap pengelolaan sampah relatif baik. Dengan demikian, integrasi IndoBERT dan BERTopic mampu memberikan pemahaman yang lebih komprehensif mengenai persepsi masyarakat dibandingkan penggunaan analisis sentimen atau pemodelan topik secara terpisah. Kontribusi utama penelitian ini terletak pada pengembangan *Awareness Index* sebagai indikator kuantitatif untuk mengukur kesadaran masyarakat berbasis data media sosial, yang dapat dimanfaatkan sebagai pendukung evaluasi kebijakan dan penyusunan strategi pengelolaan sampah yang lebih efektif dan berkelanjutan.

5. Daftar Pustaka

- [1] C. H. Liao, "Exploring social media determinants in fostering pro-environmental behavior: insights from social impact theory and the theory of planned behavior," *Front. Psychol.*, vol. 15, Jul. 2024, doi: 10.3389/fpsyg.2024.1445549.
- [2] H. Huang, X. Zeng, L. Ge, and K. Sun, "Social media engagement in waste sorting: the role of sentiment in shaping public awareness," *Humanit. Soc. Sci. Commun.*, vol. 12, no. 1, p. 1763, Nov. 2025, doi: 10.1057/s41599-025-06041-x.
- [3] C. Lin, Y. Sun, W. Jiao, J. Zheng, Z. Li, and S. Zhang, "Prediction of Compressive Strength and Elastic Modulus for Recycled Aggregate Concrete Based on AutoGluon," *Sustain.*, vol. 15, no. 16, p. 12345, Aug. 2023, doi: 10.3390/su151612345.
- [4] S. Samputri *et al.*, "Phenomenological Study: Analysis of Public Awareness and Behavior Towards Waste Sorting in Parang Tambung Village, Makassar City," *LaGeografia*, vol.

- 23, no. 3, pp. 288–298, 2025.
- [5] O. Sadeghian, B. Mohammadi-Ivatloo, F. Mohammadi, and Z. Abdul-Malek, “Protecting Power Transmission Systems against Intelligent Physical Attacks: A Critical Systematic Review,” *Sustain.*, vol. 14, no. 19, p. 12345, Sep. 2022, doi: 10.3390/su141912345.
- [6] D. Al Akhdaan, Taufik Edy Sutanto, and Muhaza Liebenlito, “Confident Learning pada IndoBERT: Peningkatan Kinerja Klasifikasi Sentimen,” *Indones. J. Comput. Sci.*, vol. 13, no. 5, Oct. 2024, doi: 10.33022/ijcs.v13i5.4401.
- [7] V. D. Setiawan, D. U. Iswavigra, and E. Anggiratih, “Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court’s Decision Regarding the Minimum Age of Vice Presidential Candidates,” *Sci. J. Informatics*, vol. 12, no. 3, pp. 397–406, Aug. 2025, doi: 10.15294/sji.v12i3.26320.
- [8] C. Shaw, P. LaCasse, and L. Champagne, “Exploring emotion classification of Indonesian tweets using large scale transfer learning via IndoBERT,” *Soc. Netw. Anal. Min.*, vol. 15, no. 1, p. 22, Mar. 2025, doi: 10.1007/s13278-025-01439-6.
- [9] V. E. Sidauruk and W. Herowati, “Indobert-Based Sentiment Analysis of Political Discourse on Platform X: The Case Of Prabowo-Gibran Administration,” *J. Appl. Informatics Comput.*, vol. 10, no. 1, pp. 673–683, Feb. 2026, doi: 10.30871/jaic.v10i1.11586.
- [10] Muhammad Mursyid, A. Setiawan, and M. Arifin, “Implementation of IndoBERT in Sentiment Analysis of Free Nutritious Meal Programs on the X Social Media Platform,” *J. Teknol. Inf. dan Pendidik.*, vol. 19, no. 1, pp. 1228–1242, Mar. 2026, doi: 10.24036/jtip.v19i1.1104.
- [11] M. Grootendorst, “BERTopic: Neural topic modeling with a class-based TF-IDF procedure,” Mar. 2022, [Online]. Available: <http://arxiv.org/abs/2203.05794>
- [12] R. Egger and J. Yu, “A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts,” *Front. Sociol.*, vol. 7, May 2022, doi: 10.3389/fsoc.2022.886498.
- [13] V. Swaminathan, H. A. Schwartz, R. Menezes, and S. Hill, “The Language of Brands in Social Media: Using Topic Modeling on Social Media Conversations to Drive Brand Strategy,” *J. Interact. Mark.*, vol. 57, no. 2, pp. 255–277, May 2022, doi: 10.1177/10949968221088275.
- [14] V. V. Mihunov, N. H. Jafari, K. Wang, N. S. N. Lam, and D. Govender, “Disaster Impacts Surveillance from Social Media with Topic Modeling and Feature Extraction: Case of Hurricane Harvey,” *Int. J. Disaster Risk Sci.*, vol. 13, no. 5, pp. 729–742, Oct. 2022, doi: 10.1007/s13753-022-00442-1.
- [15] C. Schneijderberg, O. Wiczorek, and I. Steinhardt, “Topic modeling with BERTopic,” in *The Handbook of Qualitative and Quantitative Content Analysis*, London: Routledge, 2026, pp. 558–592. doi: 10.4324/9781003496397-21.

6. Penulis



Erwin Dwika Putra
Fakultas Teknik, Universitas Muhammadiyah Bengkulu, Bengkulu,
Indonesia
Penulis merupakan tenaga pendidik di Program Studi Teknik
Informatika Universitas Muhammadiyah Bengkulu



Marissa Utami

Fakultas Teknik, Universitas Muhammadiyah Bengkulu, Bengkulu,
Indonesia

Penulis merupakan tenaga pendidik di Program Studi Sistem
Informasi Universitas Muhammadiyah Bengkulu