



Public Sentiment Analysis of the Public Housing Savings Program Using the IndoBERT Lite Model on YouTube Comments

Analisis Sentimen Publik Terhadap Program Tabungan Perumahan Rakyat Menggunakan Model IndoBERT Lite pada Komentar YouTube

Mutiara Puspita Firdaus¹, Dedi Trisnawarman²

^{1,2}Department of Information System, Faculty of Information Technology,
Tarumanagara University, Indonesia

E-Mail: ¹mutiara.825210061@stu.untar.ac.id, ²dedit@fti.untar.ac.id

Received Oct 28th 2024; Revised Jan 7th 2024; Accepted Jan 11th 2025; Available Online Jan 18th 2025, Published Jan 30th 2025

Corresponding Author: Mutiara Puspita Firdaus

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

In the digital era, social media has become the primary platform for the public to express opinions on public policies, including Public Housing Savings (TAPERA), a government program aimed at providing housing access for low-income communities. This study analyzes public sentiment toward TAPERA using the IndoBERT Lite Large model, optimized for handling large datasets with efficient computational resources. Out of 14,618 YouTube comments collected, 13,766 comments were processed after the preprocessing stage. Sentiment labeling results revealed a dominance of negative sentiment with 9,571 comments, reflecting concerns about transparency, implementation, and program communication. Positive sentiment reached 2,485 comments, indicating limited appreciation for the program, while neutral sentiment totaled 1,710 comments, suggesting a need for clearer information. Visualization using bar charts and word clouds highlighted sentiment patterns and frequently occurring keywords. Based on evaluation using a confusion matrix, the model achieved an accuracy of 78%. While effective in handling large-scale data, this study has limitations in performing a more in-depth evaluation of the model's performance. Future research is recommended to expand the analysis with data from various social media platforms to enhance overall sentiment analysis.

Keywords: IndoBERT Lite Large, Sentiment Analysis, Social Media, TAPERA.

Abstrak

Di era digital, media sosial menjadi platform utama bagi masyarakat menyampaikan opini terhadap kebijakan publik, termasuk Public Housing Savings (TAPERA), program pemerintah untuk menyediakan akses perumahan bagi masyarakat berpenghasilan rendah. Penelitian ini menganalisis sentimen masyarakat terhadap TAPERA menggunakan model *IndoBERT Lite Large*, yang dioptimalkan untuk data besar dengan efisiensi sumber daya. Dari 14.618 komentar *YouTube* yang dikumpulkan, 13.766 komentar diproses setelah tahap *preprocessing*. Hasil pelabelan sentimen menunjukkan dominasi sentimen negatif dengan 9.571 komentar, mencerminkan keresahan terhadap transparansi, implementasi, dan komunikasi program. Sentimen positif mencapai 2.485 komentar, menunjukkan apresiasi terbatas terhadap program, sementara sentimen netral sebanyak 1.710 komentar mengindikasikan kebutuhan informasi yang lebih jelas. Visualisasi menggunakan grafik batang dan *word cloud* menyoroti pola sentimen dan kata kunci yang sering muncul. Berdasarkan evaluasi menggunakan *confusion matrix*, model ini mencapai akurasi sebesar 78%. Meskipun efektif menangani data besar, penelitian ini memiliki keterbatasan dalam evaluasi performa model lebih mendalam. Penelitian selanjutnya disarankan untuk memperluas analisis dengan data dari berbagai *platform* media sosial agar meningkatkan analisis sentimen secara keseluruhan.

Kata Kunci: Analisis Sentimen, *IndoBERT Lite Large*, Media Sosial, TAPERA

1. PENDAHULUAN

Di era digital yang berkembang pesat, aplikasi dan platform *online* telah menjadi bagian integral dari kehidupan sehari-hari. Media sosial merupakan platform digital yang memungkinkan individu untuk

berkomunikasi, berinteraksi, dan membangun jaringan sosial secara online, sehingga berperan penting dalam menciptakan hubungan interpersonal yang dinamis di era digital [1]. Pengguna sering memberikan komentar dan ulasan yang tidak hanya berfungsi sebagai umpan balik berharga, tetapi juga memberikan wawasan penting mengenai kepuasan atau ketidakpuasan mereka terhadap berbagai layanan, termasuk program-program pemerintah. Analisis sentimen, yaitu proses analisis opini, emosi, dan sikap yang diekspresikan dalam berbagai jenis data teks, sangat penting untuk menentukan orientasi emosional terhadap produk, organisasi, individu, atau topik tertentu [2]. Dalam beberapa tahun terakhir, analisis sentimen semakin digunakan untuk memahami persepsi publik dalam skala yang lebih luas, termasuk dalam konteks kebijakan publik [3].

Salah satu program yang menarik perhatian publik adalah Tabungan Perumahan Rakyat (TAPERA), sebuah inisiatif yang bertujuan untuk mempermudah akses masyarakat terhadap perumahan dengan menyediakan tabungan perumahan. Program ini dirancang untuk memberikan bantuan kepada pekerja berpenghasilan rendah agar mereka dapat memperoleh perumahan yang layak huni. Meski begitu, terdapat tantangan yang dihadapi TAPERA, seperti kurangnya pemahaman masyarakat mengenai cara kerja program, ketidakjelasan prosedur pendaftaran, serta masalah dalam akses informasi dan transparansi pengelolaan dana [4].

Masalah-masalah tersebut sering kali memunculkan kritik yang tercermin dalam komentar-komentar negatif di platform seperti *YouTube*. Dalam sebuah studi, platform *YouTube* diidentifikasi sebagai sumber penting untuk mendapatkan umpan balik masyarakat terkait kebijakan publik [5]. Komentar-komentar ini menawarkan data yang kaya untuk melakukan analisis sentimen, yang dapat memberikan wawasan mengenai aspek-aspek program yang dianggap efektif atau yang perlu diperbaiki [6]. Tantangan spesifik yang dihadapi adalah banyaknya komentar yang tidak terstruktur dan beragamnya respons dari masyarakat, yang membuatnya sulit untuk mengidentifikasi tema utama atau faktor-faktor spesifik yang menyebabkan kritik tersebut. Sebagai contoh, apakah kritik tersebut lebih berkaitan dengan transparansi kebijakan, ketidakpuasan terhadap implementasi, atau dampak sosial-ekonomi yang ditimbulkan. Tanpa pemahaman yang jelas tentang faktor-faktor ini, akan sulit untuk mengukur sejauh mana kebijakan tersebut diterima oleh publik dan apa yang perlu diperbaiki agar kebijakan tersebut lebih efektif.

IndoBERT Lite adalah model bahasa yang dikembangkan untuk pemrosesan bahasa alami dalam bahasa Indonesia, yang menggunakan arsitektur *ALBERT* untuk efisiensi parameter dan waktu pelatihan. Model ini dilatih menggunakan dataset besar yang mencakup sekitar 4 miliar kata dalam bahasa Indonesia, memungkinkan pemahaman konteks dan nuansa bahasa yang lebih baik [7]. Beberapa penelitian menunjukkan bahwa varian *IndoBERT-lite*, menunjukkan kinerja yang unggul dalam analisis teks berbahasa Indonesia.

Berbagai penelitian telah menunjukkan bahwa varian *IndoBERT Lite* memiliki kinerja yang unggul dalam tugas-tugas analisis teks berbahasa Indonesia, meskipun lebih ringan dibandingkan dengan model besar seperti *IndoBERT Base* [8]. Dalam analisis sentimen multilabel, kombinasi *IndoBERT Lite* dengan *BiLSTM-CNN* mampu meningkatkan akurasi dibandingkan model tradisional, tetapi performanya masih lebih rendah dibandingkan *IndoBERT Base*, terutama dalam tugas yang lebih kompleks. Selain itu, penggunaan dataset yang terbatas pada platform *Twitter* membuat hasil penelitian kurang dapat digeneralisasikan ke platform lain [9].

Penelitian lain membandingkan kinerja *IndoBERT Base* dan *IndoBERT Lite* dalam klasifikasi sentimen pada aplikasi TikTok Tokopedia. Hasilnya menunjukkan bahwa *IndoBERT Base* lebih unggul dalam akurasi (97% dibandingkan 94%), tetapi *IndoBERT Lite* lebih efisien dalam penggunaan sumber daya komputasi. Meski demikian, penelitian ini hanya berfokus pada satu aplikasi spesifik, sehingga kurang mengeksplorasi penerapan model dalam domain lain yang lebih luas [10].

Dalam konteks penambangan teks, *IndoBERT Lite* juga menunjukkan keunggulan dibandingkan *RoBERTa*, terutama dalam tugas pertanyaan dan jawaban menggunakan dataset terjemahan *TyDi QA* dan *SQuAD*. Namun, penggunaan dataset terjemahan dapat mengurangi akurasi dalam memahami konteks bahasa asli Indonesia, dan penelitian ini terbatas pada satu jenis aplikasi saja [11].

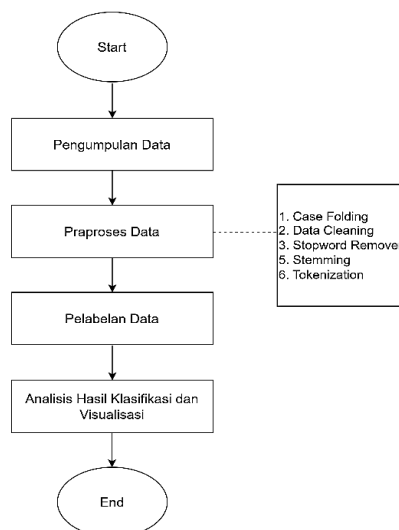
Penelitian lain tentang *IndoBERT* berfokus pada pengoptimalan *hyperparameter* untuk mendeteksi berita palsu, serta kombinasi *IndoBERT* dengan *CNN* untuk mendeteksi aspek ulasan di *e-commerce*. Meskipun menunjukkan hasil yang akurat, penelitian ini hanya diterapkan pada domain tertentu, sehingga belum mengeksplorasi fleksibilitas model dalam domain yang lebih berdampak, seperti kebijakan publik [12][13].

Dengan latar belakang tersebut, penelitian ini berusaha mengatasi keterbatasan dengan menggunakan *IndoBERT Lite* untuk menganalisis sentimen publik terhadap kebijakan pemerintah, khususnya program TAPERA. Penelitian ini memanfaatkan dataset besar dan representatif berupa 14.618 komentar dari *YouTube*, yang mencerminkan opini masyarakat terhadap kebijakan publik secara lebih kaya dan relevan. Pendekatan ini juga menonjolkan efisiensi model *IndoBERT Lite* dalam analisis data besar tanpa memerlukan sumber daya komputasi yang tinggi. Selain itu, penelitian ini menyertakan analisis temporal dan visualisasi data dalam bentuk grafik batang dan *word cloud*, memberikan wawasan yang intuitif dan aplikatif bagi pengambil kebijakan. Dengan pendekatan ini, penelitian diharapkan dapat berkontribusi pada peningkatan kualitas program TAPERA serta kepuasan masyarakat secara keseluruhan.

2. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program TAPERA di Indonesia menggunakan data komentar pengguna di *platform YouTube*. Data diambil dari komentar pada video dengan jumlah penonton tertinggi, menghasilkan 14.618 komentar. Proses analisis sentimen dilakukan secara sistematis menggunakan model *IndoBERT-lite*. Varian *IndoBERT Lite Large (tyqiangz/indobert-lite-large-p2)* merupakan versi yang lebih besar dari *IndoBERT Lite*, dengan jumlah parameter yang lebih banyak, yaitu sekitar 17,7 juta parameter. Meskipun masih lebih kecil dibandingkan dengan model-model besar seperti *IndoBERT Base* (124,5 juta parameter) dan *IndoBERT Large* (335,2 juta parameter), model ini dirancang untuk memberikan kinerja yang lebih baik dalam tugas-tugas tertentu [7]. Model ini dapat digunakan dalam berbagai aplikasi *Natural Language Processing* (NLP) seperti klasifikasi teks, ekstraksi informasi, dan pertanyaan-jawaban. Dengan arsitektur yang efisien dan pelatihan yang berbasis pada dataset bahasa Indonesia yang kaya, *IndoBERT Lite Large* mampu memberikan representasi kontekstual yang kuat untuk analisis teks.

Penelitian yang dilakukan oleh Tyqiangz mengkaji perbandingan kinerja antara model *pre-trained IndoBERT-base* dan *IndoBERT Lite* dalam tugas analisis sentimen terhadap ulasan aplikasi TikTok Tokopedia *Seller Center*. Hasil penelitian menunjukkan bahwa meskipun *IndoBERT Lite* memiliki jumlah parameter yang lebih sedikit, model ini tetap memberikan performa yang memadai dengan akurasi yang kompetitif. Namun, akurasi *IndoBERT Lite* sedikit lebih rendah dibandingkan dengan model *IndoBERT-base* yang memiliki kapasitas lebih besar. Penelitian ini mengungkapkan bahwa meskipun model yang lebih besar memiliki keunggulan dalam hal akurasi, model yang lebih kecil seperti *IndoBERT Lite* menawarkan efisiensi dalam penggunaan sumber daya komputasi. Hal ini menjadi penting dalam konteks aplikasi dunia nyata di Indonesia, di mana keterbatasan sumber daya sering kali menjadi kendala utama. Selain itu, penelitian ini juga menyoroti pentingnya pengembangan model NLP yang efisien namun tetap dapat diakses untuk memenuhi kebutuhan berbagai sektor, terutama yang berfokus pada pemrosesan bahasa Indonesia [10]. Berikut merupakan langkah-langkah penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Langkah-langkah Penelitian

2.1 Pengumpulan Data

Data diambil dari video yang relevan dengan diskusi atau informasi mengenai program TAPERA. Data dikumpulkan menggunakan teknik *crawling*. *Crawling* merupakan program otomatis yang berfungsi untuk mengumpulkan data atau informasi yang dibutuhkan [14]. *Crawling* ini menggunakan *Python* di *Google Colab*. Data yang dikumpulkan meliputi komentar, tanggal, dan *username*.

2.2 Preprocessing Data

Preprocessing merupakan langkah yang dilakukan untuk mengatasi masalah data yang tidak sempurna, mengurangi gangguan pada data, serta memastikan konsistensi data [15]. Proses ini bertujuan untuk meningkatkan kualitas data sebelum dianalisis. Proses ini meliputi beberapa langkah:

1. *Case Folding*: *Case folding* adalah proses mengubah semua huruf dalam sebuah dokumen menjadi huruf kecil sebagai bentuk standar [16].
2. *Data Cleaning*: Merupakan proses penghapusan elemen-elemen tertentu dari kalimat, atau karakter khusus seperti, tanda baca, angka, dan *URL* yang tidak diperlukan agar menciptakan dataset yang konsisten [17].

3. *Stopword Removal*: Menghilangkan kata-kata umum yang tidak memiliki makna signifikan dalam konteks analisis teks, seperti 'di', 'atau', 'yang' [18].
4. *Stemming*: Mengubah kata ke bentuk dasarnya menggunakan algoritma stemming dengan cara menghilangkan imbuhan seperti awalan atau akhiran untuk menyatukan berbagai bentuk kata yang memiliki makna yang sama ke dalam satu bentuk dasar, sehingga memudahkan analisis [19].
5. *Tokenization*: Memecah teks menjadi kata-kata atau token untuk memudahkan proses analisis [20].

2.3 Pelabelan Data

Pelabelan data dilakukan dengan menggunakan model *tyqiangz/indobert-lite-large-p2-smsa*. Model ini dilatih untuk memahami konteks dan melakukan pelabelan data secara otomatis [7]. *IndoBERT Lite* adalah varian dari model *IndoBERT* yang dirancang untuk efisiensi dalam pemrosesan bahasa alami atau NLP untuk bahasa Indonesia. Model ini memiliki 11,7 juta parameter dan dirancang untuk memberikan performa yang baik dengan penggunaan sumber daya yang lebih sedikit. Dengan arsitektur yang lebih ringan, *IndoBERT Lite* dapat diimplementasikan pada perangkat dengan keterbatasan komputasi, menjadikannya pilihan ideal untuk aplikasi mobile dan *real-time* [21]. Efisiensi ini juga didukung oleh kinerja unggul yang telah dibuktikan melalui berbagai evaluasi pada tugas-tugas NLP.

Dalam penelitian yang membandingkan kinerja berbagai model *pre-trained*, *IndoBERT Lite* menunjukkan akurasi yang kompetitif dalam klasifikasi sentimen. Model ini mampu mencapai akurasi sekitar 94% pada dataset ulasan produk. Metrik evaluasi seperti presisi, *recall*, dan *F1-score* juga menunjukkan hasil yang baik, menjadikannya efektif untuk berbagai tugas NLP [22]. Selain itu, *IndoBERT Lite* juga memiliki varian yang lebih besar, yaitu *IndoBERT Lite Large*, yang dirancang untuk menangani tugas-tugas yang lebih kompleks.

IndoBERT Lite Large merupakan varian yang lebih besar dari *IndoBERT Lite*, dengan 17,7 juta parameter. Model ini dirancang untuk memberikan performa yang lebih baik dalam menangkap konteks yang lebih kompleks dalam teks berbahasa Indonesia. Dalam evaluasi, *IndoBERT Lite Large* mampu mencapai akurasi hingga 95% dalam klasifikasi sentiment [10]. Keunggulan ini tidak terlepas dari arsitektur transformer yang digunakan oleh kedua model, meskipun terdapat optimasi khusus untuk menjaga efisiensi.

IndoBERT Lite Large mengadopsi arsitektur transformer yang sama dengan *BERT*, namun dengan pengurangan jumlah parameter. Hal ini memungkinkan model untuk beroperasi lebih cepat tanpa mengorbankan akurasi secara signifikan. Model ini dilatih menggunakan teknik *masked language modeling* (MLM) dan *next sentence prediction* (NSP), yang membantu dalam memahami konteks dan hubungan antar kalimat dalam teks berbahasa Indonesia [23]. Optimasi ini juga membuat kedua varian *IndoBERT Lite* mampu bersaing dengan model lainnya dalam hal akurasi dan kecepatan inferensi.

Dalam perbandingan antara *IndoBERT Lite* dan *IndoBERT Lite Large* dengan model lain seperti *BERT-base* dan *RoBERTa*, hasil menunjukkan bahwa kedua model ini memberikan performa yang sangat baik dalam hal akurasi dan kecepatan inferensi. Penelitian menemukan bahwa meskipun ukuran parameter berbeda, keduanya mampu bersaing secara efektif berkat optimasi arsitektur yang tepat [11]. Dengan demikian, *IndoBERT Lite* dan varian besarnya menjadi solusi yang ideal untuk aplikasi NLP dalam berbagai konteks, terutama untuk bahasa Indonesia.

2.4 Klasifikasi dan Visualisasi Sentimen

Tahap terakhir adalah analisis hasil klasifikasi. Pada tahap ini, hasil klasifikasi sentimen dianalisis untuk melihat tren atau pola sentimen terhadap TAPERA. Analisis dilakukan dengan visualisasi data seperti grafik batang untuk perbandingan hasil sentimen positif, negatif, dan netral, serta *word cloud* untuk melihat kata-kata yang paling sering muncul dalam komentar [24]. Hasil ini diinterpretasikan untuk memberikan wawasan mendalam tentang persepsi publik dan menyusun rekomendasi bagi pemerintah dalam meningkatkan program TAPERA.

2.5 Evaluasi Model

Evaluasi model untuk mengukur kinerja dari model yang digunakan menggunakan *Confusion Matrix*. *Confusion Matrix* merupakan metode evaluasi yang digunakan untuk mengukur performa model klasifikasi. Matriks ini membantu untuk memahami seberapa baik model dalam membedakan kategori sentimen seperti positif, negatif, dan netral. Matriks ini menyajikan perbandingan antara prediksi yang dihasilkan oleh model dan label asli dari data yang dianalisis [25]. Contoh struktur *Confusion Matrix* untuk tiga kelas sentimen (positif, negatif, netral) dapat dilihat pada Tabel 1:

Tabel 1 *Confusion Matrix* Tiga Kelas

	Positif	Negatif	Netral
Positif (Asli)	TP (True Positives)	FN (False Negatives)	FN (False Negatives)
Negatif (Asli)	FP (False Positives)	TN (True Negatives)	FP (False Positives)
Netral (Asli)	FP (False Positives)	FN (False Negatives)	TN (True Negatives)

Beberapa metrik evaluasi dapat dihitung untuk mengukur performa model, yaitu:
Mengukur proporsi prediksi yang benar dibandingkan total data

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Mengukur ketepatan prediksi positif

$$Precision = \frac{TP}{TP+FN} \tag{2}$$

Mengukur kemampuan model mendeteksi semua kasus positif

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Mengukur keseimbangan antara *precision* dan *recall*

$$F1 - Score = 2x \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Metrik-metrik ini memberikan wawasan tentang seberapa efektif model dalam mengklasifikasikan data. *Confusion Matrix* tidak hanya memberikan gambaran umum tentang akurasi model, tetapi juga mendetailkan jenis kesalahan yang dibuat oleh model. Ini sangat berguna dalam situasi di mana kelas-kelas tidak seimbang, karena dapat membantu dalam memahami bagaimana model berperilaku terhadap setiap kelas secara individual [26].

3. HASIL DAN PEMBAHASAN

3.1. Preprocessing dan Pemodelan

Data komentar dikumpulkan dengan menggunakan teknik *YouTube crawling*. *YouTube Crawling* merupakan proses pengambilan data dari platform *YouTube* yang bertujuan untuk mengumpulkan berbagai informasi dari video *Youtube* [27]. *YouTube crawling* ini menyediakan *Application Programming Interface* (API) untuk mempermudah scraping data komentar di *YouTube* menggunakan bahasa pemrograman Python di *Google Colab*. Data yang dikumpulkan yaitu data komentar dari video mengenai program pemerintah TAPERA (Tabungan Perumahan Rakyat) yang difilter berdasarkan jumlah tayangan terbanyak di tahun 2024. Judul video yang dipilih untuk pengumpulan data adalah video dengan jumlah penonton tertinggi, yang mencakup judul-judul seperti "BP Tapera 100% Gagal dalam Matematika", "Tanggapan Netizen & Review TAPERA ft. Guru Gembul - Aduan Masyarakat", "Motong Gaji Karyawan", "Tapera Yakin Bikin Warga Punya Rumah", "Dukung Tapera Program JENIUS Pemerintah", "TAPERA Program Aneh", dan "Tabungan Tapi Penuh Paksaan". Video-video ini dipilih karena popularitasnya yang tinggi dan relevansinya terhadap program TAPERA, sehingga memberikan gambaran yang luas mengenai persepsi publik terhadap program tersebut. Hasil yang didapat dari *scraping* sebanyak 14.618 data ulasan. Pada tahapan *preprocessing* meliputi 5 proses yaitu: *Cleaning*, *Case folding*, *Tokenisasi*, *stopword removal*, dan *Stemming*. Total data setelah dilakukannya tahap *preprocessing* sebanyak 13.766 data. Hasil tahapan *preprocessing* data dapat dilihat pada Tabel 2.

Tabel 2. *Preprocessing Data*

Tahapan	Hasil
Data Awal	Sebenarnya tapera klo terlaksana dgn baik itu bagus cuma masyarakatnya udah gk percaya sama pemerintah, kemarin aja ada kasus korupsi zakat kan kocak
Case Folding	sebenarnya tapera klo terlaksana dgn baik itu bagus cuma masyarakatnya udah gk percaya sama pemerintah, kemarin aja ada kasus korupsi zakat kan kocak
Data Cleaning	sebenarnya tapera klo terlaksana dgn baik itu bagus cuma masyarakatnya udah gk percaya sama pemerintah kemarin aja ada kasus korupsi zakat kan kocak
Tokenizing	['sebenarnya', 'tapera', 'klo', 'terlaksana', 'dgn', 'baik', 'itu', 'bagus', 'cuma', 'masyarakatnya', 'udah', 'gk', 'percaya', 'sama', 'pemerintah', 'kemarin', 'aja', 'ada', 'kasus', 'korupsi', 'zakat', 'kan', 'kocak']
Stopword	['sebenarnya', 'tapera', 'klo', 'terlaksana', 'baik', 'bagus', 'cuma', 'masyarakatnya', 'udah', 'gk', 'percaya', 'pemerintah', 'kemarin', 'ada', 'kasus', 'korupsi', 'zakat', 'kan', 'kocak']
Stemming	['sebenar', 'tapera', 'klo', 'laksana', 'baik', 'bagus', 'cuma', 'masyarakat', 'sudah', 'gk', 'percaya', 'pemerintah', 'kemarin', 'ada', 'kasus', 'korupsi', 'zakat', 'kan', 'kocak']

Setelah data melalui tahap *preprocessing*, langkah selanjutnya adalah memberikan bobot pada setiap kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Bobot suatu kata akan semakin besar ketika kata tersebut sering muncul dalam sebuah dokumen, namun nilai bobotnya akan

berkurang jika kata tersebut tersebar di banyak dokumen. Proses perhitungan *TF-IDF* ini dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka *Scikit-Learn*. Hasil dari penerapan *TF-IDF* dapat dilihat pada Tabel 3 berikut ini.

Tabel 3. Hasil *TF-IDF*

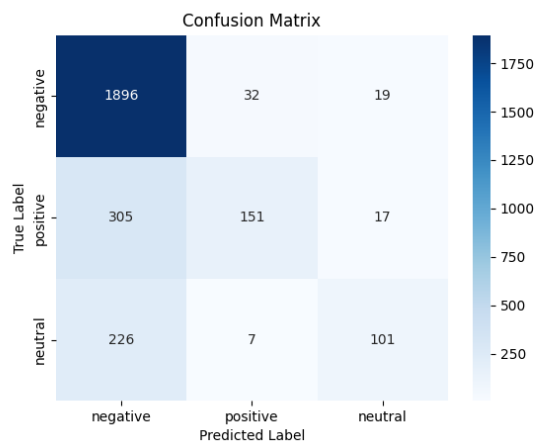
No	tabung	rumah	tapera	korupsi	bohong	rakyat	...	program	jujur	penipu	bangun
0	0.000	0.000	1.000	0.000	0.000	0.000	...	0.000	0.640	0.273	0.000
1	0.000	0.000	0.000	0.561	0.000	0.000	...	0.635	0.433	0.000	0.000
2	0.840	0.615	0.000	0.000	0.000	0.542	...	0.500	0.000	0.000	0.000
3	0.000	0.000	0.000	0.655	0.000	1.000	...	0.650	0.000	0.000	0.000
4	0.811	0.584	0.203	0.000	0.579	0.000	...	0.305	0.000	0.000	0.000
...
13761	0.000	0.535	0.320	0.000	0.000	0.000	...	0.000	0.430	0.449	0.289
13762	0.000	0.000	1.000	0.000	0.000	0.000	...	0.000	0.000	0.462	0.000
13763	0.811	0.584	0.000	0.000	0.000	0.000	...	0.000	0.000	0.346	0.699
13764	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000
13765	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000

Proses pelabelan data dilakukan dengan menggunakan model *tyqiangz/indobert-lite-large-p2-smsa* untuk melakukan pelabelan data secara otomatis, mendeteksi sentimen positif, negatif, atau netral dari komentar yang telah di praproses. Hasil dari *labeling data* dapat dilihat pada Tabel 4.

Tabel 4. Hasil pelabelan data

Aplikasi	Jumlah dan Presentase		
	Positif	Negatif	Netral
Youtube	2.485 (18%)	9.571 (69.6%)	1.710 (12.4%)

Tahap selanjutnya adalah proses evaluasi model menggunakan *confusion matrix* yang dapat dilihat dalam Gambar 2. *Confusion matrix* ini menunjukkan performa model klasifikasi untuk tiga kelas: negatif, positif, dan netral. Model memiliki kinerja yang baik pada kelas negatif, dengan 1896 prediksi yang benar, sementara hanya 32 data negatif salah diprediksi sebagai positif dan 19 salah diprediksi sebagai netral. Namun, performa model untuk kelas positif dan netral kurang memuaskan. Untuk kelas positif, hanya 151 prediksi yang benar, sementara 305 data salah diprediksi sebagai negatif dan 17 salah diprediksi sebagai netral. Hal serupa terjadi pada kelas netral, dengan hanya 101 prediksi yang benar, sementara 226 data salah diprediksi sebagai negatif dan 7 salah diprediksi sebagai positif. Secara keseluruhan, model tampaknya lebih condong untuk memprediksi kelas negatif, yang dapat disebabkan oleh ketidakseimbangan data (*imbalance*) di mana data negatif lebih banyak dibandingkan kelas lainnya. Hasil ini menunjukkan bahwa meskipun akurasi pada kelas negatif tinggi, model memerlukan perbaikan dalam mengenali kelas positif dan netral untuk mencapai performa yang lebih baik secara keseluruhan.



Gambar 2. Hasil dari *Confusion Matrix*

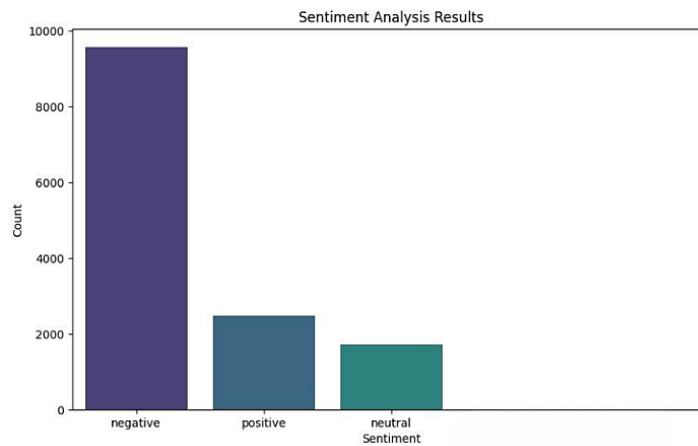
Tabel 5. Hasil Evaluasi Model

	Precision	Recall	F1-Score	Support
Negative	0.78	0.97	0.87	1947
Neutral	0.74	0.3	0.43	334
Positive	0.79	0.32	0.46	473
Accuracy			0.78	2754
Macro avg	0.77	0.53	0.58	2754
Weighted avg	0.78	0.78	0.74	2754

Pada Tabel 5 menunjukkan model ini memiliki akurasi keseluruhan sebesar 78%, yang menunjukkan bahwa 78% dari total prediksi model sesuai dengan label sebenarnya. Performa terbaik terlihat pada kelas negatif, dengan *F1-score* tinggi (0.87) dan *recall* mencapai 97%, menandakan bahwa model sangat andal dalam mengidentifikasi data negatif. Namun, performa pada kelas netral dan positif masih kurang memadai, terutama pada *recall* yang rendah (masing-masing 30% dan 32%), yang menunjukkan bahwa model sering gagal mengenali data dari kedua kelas tersebut.

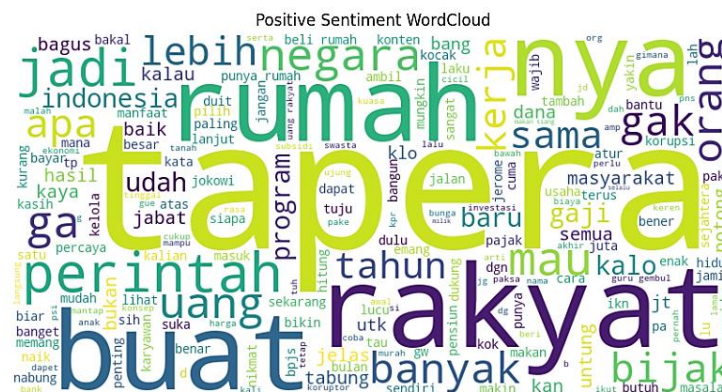
3.2. Klasifikasi dan Visualisasi Sentimen

Analisis sentimen terhadap program TAPERA dilakukan untuk mengidentifikasi tren dan pola respons publik. Hasil visualisasi perbandingan sentimen positif, negatif, dan netral ditampilkan melalui grafik batang, sementara *word cloud* digunakan untuk memvisualisasikan kata-kata yang sering muncul dalam setiap kategori sentimen.



Gambar 4. Hasil Perbandingan Sentimen Positif, Negatif, dan Netral

Pada Gambar 4 menampilkan grafik batang hasil analisis sentimen dari 13.667 data terkait program TAPERA. Grafik tersebut menunjukkan bahwa sentimen negatif mendominasi dengan jumlah 9.571 data, diikuti oleh sentimen positif sebanyak 2.485 data dan sentimen netral sebanyak 1.710 data. Dominasi sentimen negatif ini mencerminkan tingginya keresahan dan ketidakpuasan publik terhadap program TAPERA dibandingkan dengan respons positif maupun netral.



Gambar 5. Hasil Word Cloud Sentimen Positif

Keterkaitan antara hasil penelitian ini dengan studi kasus TAPERA sangat jelas. Program ini dirancang untuk membantu masyarakat berpenghasilan rendah mendapatkan akses ke perumahan yang layak. Namun, dominasi kritik menunjukkan bahwa implementasi program belum sepenuhnya memenuhi ekspektasi masyarakat. Dengan menggunakan model *IndoBERT Lite Large* untuk analisis sentimen, penelitian ini memberikan wawasan mendalam tentang persepsi publik yang dapat memengaruhi keberhasilan kebijakan pemerintah.

Temuan ini dapat menjadi dasar bagi rekomendasi perbaikan dalam pengelolaan TAPERA, terutama dalam aspek peningkatan komunikasi, transparansi pengelolaan dana, dan kualitas layanan. Dengan pendekatan ini, diharapkan TAPERA dapat lebih diterima oleh masyarakat dan memenuhi tujuan utamanya. Penelitian ini tidak hanya berkontribusi pada pemahaman terhadap sentimen publik, tetapi juga memberikan arahan praktis untuk meningkatkan kepercayaan dan kepuasan masyarakat melalui kebijakan yang lebih responsif terhadap kebutuhan mereka.

4. KESIMPULAN

Berdasarkan hasil analisis sentimen terhadap 14.618 komentar mengenai program TAPERA, ditemukan bahwa sentimen negatif mendominasi dengan 9.571 komentar, mencerminkan ketidakpuasan atau kekhawatiran publik terhadap program tersebut. Sementara itu, sentimen positif hanya mencakup 2.485 komentar, menunjukkan adanya apresiasi terhadap upaya pemerintah, meskipun jumlahnya jauh lebih sedikit dibandingkan kritik yang diterima. Sentimen netral, sebanyak 1.710 komentar, mencerminkan adanya kebutuhan masyarakat akan informasi yang lebih jelas dan transparan terkait program TAPERA. Dalam penelitian ini, model *IndoBERT Lite Large* digunakan untuk mengklasifikasikan sentimen, yang terbukti mampu menangani volume data besar dan menghasilkan klasifikasi yang cukup akurat. Selain itu, berdasarkan evaluasi awal, model menunjukkan akurasi sebesar 78%, dengan performa yang sangat baik pada sentimen negatif namun masih lemah dalam mengenali sentimen positif dan netral. Ketidakseimbangan data menjadi salah satu tantangan utama yang memengaruhi hasil klasifikasi. Oleh karena itu, penelitian selanjutnya perlu menangani ketidakseimbangan data, melakukan evaluasi metrik yang lebih mendalam seperti *presisi*, *recall*, dan *F1-score*, serta memperluas cakupan data untuk mendapatkan hasil yang lebih representatif. Langkah-langkah ini diharapkan dapat meningkatkan keandalan analisis sentimen dan memberikan wawasan yang lebih akurat untuk mendukung pengembangan program TAPERA.

REFERENSI

- [1] S. Gilbert, K. Shilton, and J. Vitak, "When research is the context: Cross-platform user expectations for social media data reuse," *Big Data Soc.*, vol. 10, no. 1, 2023, doi: 10.1177/20539517231164108.
- [2] M. Naeem, "Understanding the role of social media in organizational change implementation," *Manag. Res. Rev.*, vol. 43, no. 9, pp. 1097–1116, 2020, doi: 10.1108/MRR-07-2019-0336.
- [3] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [4] R. Abidin and A. Herawati, "Analisis Sentimen Publik Terhadap Kebijakan Program Tabungan Perumahan Rakyat (Tapera)," *J. Inf. Syst. Comput.*, vol. 4, no. 1, pp. 13–19, 2024, doi: 10.34001/jister.v4i1.1002.
- [5] S. Thomas, Yuliana, and Noviyanti, P., "Study Analisis Metode Analisis Sentimen pada YouTube," *J. Inf. Technol.*, vol. 1, no. 1, pp. 1–7, 2021, doi: 10.46229/jifotech.v1i1.201.
- [6] J. E. Tang, V. Arvind, C. Dominy, C. A. White, S. K. Cho, and J. S. Kim, "How Are Patients Reviewing Spine Surgeons Online? A Sentiment Analysis of Physician Review Website Written Comments," *Glob. Spine J.*, vol. 13, no. 8, pp. 2107–2114, 2023, doi: 10.1177/21925682211069933.
- [7] T. D. Purnomo, J. Sutopo, and A. History, "Comparison Of Pre-Trained Bert-Based Transformer Models For Regional," vol. 3, no. 3, pp. 11–21, 2024, doi: 10.56127/ijst.v3i3.1739.
- [8] Anugerah Simanjuntak *et al.*, "Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 1, pp. 60–67, 2024, doi: 10.22146/jnteti.v13i1.8532.
- [9] B. A. Setyawan, *Analisis sentimen multilabel terhadap ujaran kebencian menggunakan gabungan metode indo bert lite dan bidirectional lstm-cnn dengan grid search hyperparameter optimization*, no. 11180910000102. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2024. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/77089>
- [10] I. D. A. N. I. Pada, "Perbandingan Kinerja Pre-Trained Tokopedia Seller Center," vol. 11, no. 2, pp. 13–20, 2024, doi: 10.30656/jsii.v11i2.9168.
- [11] B. Richardson and A. Wicaksana, "Comparison of Indobert-Lite and Roberta in Text Mining for Indonesian Language Question Answering Application," *Int. J. Innov. Comput. Inf. Control*, vol. 18, no. 6, pp. 1719–1734, 2022, doi: 10.24507/ijicic.18.06.1719.
- [12] S. Imron, E. I. Setiawan, and J. Santoso, "Deteksi Aspek Review E-Commerce Menggunakan

- IndoBERT Embedding dan CNN,” *J. Intell. Syst. Comput.*, vol. 5, no. 1, pp. 10–16, 2023, doi: 10.52985/insyst.v5i1.267.
- [13] R. W. Hardian, P. E. Prasetyo, U. Khaira, and T. Suratno, “Analisis Sentiment Kuliah Daring Di Media Sosial Twitter Selama Pandemi Covid-19 Menggunakan Algoritma Sentistrength,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 2, pp. 138–143, 2021, doi: 10.57152/malcom.v1i2.15.
- [14] C. A. A. Soemedhy, N. Trivetisia, N. A. Winanti, D. P. Martiyansih, T. W. Utami, and S. Sudianto, “Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube ‘Kekerasan Seksual’),” *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 80–84, 2022, doi: 10.30591/jpit.v7i2.3547.
- [15] P. Y. Saputra, D. H. Subhi, and Fahmi Zain Afif Winatama, “Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes,” *J. Inform. Polinema*, vol. 5, no. 4, pp. 209–213, 2019, doi: 10.33795/jip.v5i4.259.
- [16] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [17] D. Rifaldi, Abdul Fadlil, and Herman, “Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet ‘Mental Health,’” *Decod. J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 161–171, 2023, doi: 10.51454/decode.v3i2.131.
- [18] A. B. Putra Negara, “The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 14, no. 3, p. 172, 2023, doi: 10.24843/lkjiti.2023.v14.i03.p05.
- [19] M. U. Albab, Y. Karuniawati P, and M. N. Fawaiq, “Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic,” *J. Transform.*, vol. 20, no. 2, pp. 1–10, 2023, doi: 10.26623/transformatika.v20i2.5374.
- [20] C. T. Yuang, R. E. Banchs, and C. E. Siong, “An Empirical Evaluation of Stop Word Removal in Statistical Machine Translation,” *Jt. Work. Exploit. Synerg. between Inf. Retr. Mach. Transl. ESIRMT 2012 Hybrid Approaches to Mach. Transl. HyTra 2012 Eur. Chapter Assoc. Comput. Linguist. EACL 2012 - Procee*, no. January, pp. 30–37, 2012, [Online]. Available: <https://api.semanticscholar.org/CorpusID:302123>
- [21] F. Arifadilah *et al.*, *Perbandingan Hyperparameter Optimization Population Based Training, Random Search, Bayesian Optimization Pada Analisis Sentimen Radikalisme*. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2023. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/76873>
- [22] M. Y. Ridho and E. Yulianti, “From Text to Truth : Leveraging IndoBERT and Machine Learning Models for Hoax Detection in Indonesian News,” vol. 10, no. 3, pp. 544–555, 2024, doi: 10.26555/jiteki.v10i3.29450.
- [23] N. K. Nissa and E. Yulianti, “Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model,” *Int. J. Electr. Comput. Eng.*, vol. 13, no. 5, pp. 5641–5652, 2023, doi: 10.11591/ijece.v13i5.pp5641-5652.
- [24] N. Shahid, M. U. Ilyas, J. S. Alowibdi, and N. R. Aljohani, “Word cloud segmentation for simplified exploration of trending topics on Twitter,” *IET Softw.*, vol. 11, no. 5, pp. 214–220, 2017, doi: 10.1049/iet-sen.2016.0307.
- [25] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, “Evaluating the quality of machine learning explanations: A survey on methods and metrics,” *Electron.*, vol. 10, no. 5, pp. 1–19, 2021, doi: 10.3390/electronics10050593.
- [26] R. Indransyah, Y. H. Chrisnanto, P. N. Sabrina, and S. Kom, “Klasifikasi Sentimen Pergelaran MotoGP di Indonesia Menggunakan Algoritma Correlated Naive Bayes Clasifier,” *INFOTECH J.*, vol. 8, no. 2, pp. 60–66, 2022, doi: 10.31949/infotech.v8i2.3103.
- [27] B. IANCU, “Web Crawler for Indexing Video e-Learning Resources: A YouTube Case Study,” *Inform. Econ.*, vol. 23, no. 2/2019, pp. 15–24, 2019, doi: 10.12948/issn14531305/23.2.2019.02.