JUKOMIKA (Jurnal Ilmu Komputer dan Informatika)
Volume 9, Nomor 1, July 2026 || Hal 82-89
p-ISSN **: 2655-7541 ; e-ISSN : 2745-5823**
DOI : 10.54650

Analisis Sentimen Masyarakat terhadap Layanan Publik pada Media Sosial X Menggunakan IndoBERT dan *Support Vector Machine*

Marissa Utami*[1], Erwin Dwika Putra[2]
Fakutlas Teknik, Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia[1,2]
marissautami@umb.ac.id[1], erwindwikap@gmail.com[2]
*Corresponding author : marissautami@umb.ac.id[1]

Abstrak—Media sosial X menjadi salah satu platform yang banyak digunakan masyarakat untuk menyampaikan opini terhadap berbagai layanan publik sehingga dapat dimanfaatkan sebagai sumber informasi dalam mengevaluasi kualitas pelayanan pemerintah. Penelitian ini bertujuan membandingkan kinerja algoritma *Support Vector Machine* (SVM) dan IndoBERT dalam mengklasifikasikan sentimen masyarakat terhadap layanan publik menggunakan *Figshare Twitter Dataset*. Dataset terlebih dahulu melalui proses filtering berdasarkan kata kunci layanan publik, kemudian dilakukan preprocessing yang meliputi *case folding, cleaning, stopword removal*, dan *tokenization*. Pada model SVM digunakan ekstraksi fitur TF-IDF, sedangkan IndoBERT menerapkan proses *fine-tuning* berbasis *Transformer*. Evaluasi dilakukan menggunakan metrik *accuracy, precision, recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model IndoBERT memperoleh performa terbaik dengan *accuracy* 91,89%, *precision* 92,51%, *recall* 92,56%, dan *F1-score* 92,54%, sedangkan SVM memperoleh *accuracy* 81,20%, *precision* 78,02%, *recall* 78,74%, dan *F1-score* 78,88%. Hasil tersebut menunjukkan bahwa IndoBERT lebih efektif dalam memahami konteks bahasa Indonesia pada media sosial dibandingkan SVM berbasis TF-IDF. Penelitian ini diharapkan menjadi referensi dalam pengembangan sistem analisis sentimen untuk mendukung evaluasi layanan publik berbasis media sosial.

Abstract—*Social media platform X has become one of the primary channels for the public to express opinions regarding public services, making it a valuable source for evaluating government service quality. This study aims to compare the performance of Support Vector Machine (SVM) and IndoBERT in classifying public sentiment toward public services using the Figshare Twitter Dataset. The dataset was first filtered using public service-related keywords, followed by preprocessing steps including case folding, text cleaning, stopword removal, and tokenization. The SVM model employed the Term Frequency–Inverse Document Frequency (TF-IDF) feature extraction method, while the IndoBERT model was fine-tuned using a Transformer-based architecture. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results show that IndoBERT outperformed SVM, achieving an accuracy of 91.89%, precision of 92.51%, recall of 92.56%, and F1-score of 92.54%. In comparison, SVM achieved an accuracy of 81.20%, precision of 78.02%, recall of 78.74%, and F1-score of 78.88%. These findings indicate that IndoBERT is more effective in capturing the contextual semantics of Indonesian-language social media texts than the TF-IDF-based SVM approach. The proposed approach can support the development of intelligent sentiment analysis systems for monitoring and evaluating public services through social media.*

Keywords—*Sentiment Analysis, Public Services, Social Media X, IndoBERT, Support Vector Machine, Natural Language Processing.*

1. Pendahuluan

Perkembangan teknologi informasi telah mendorong perubahan pola komunikasi masyarakat dalam menyampaikan opini, kritik, maupun apresiasi terhadap berbagai layanan publik. Media sosial X (sebelumnya Twitter) menjadi salah satu platform digital yang paling aktif digunakan masyarakat untuk menyampaikan pengalaman mereka terhadap kebijakan pemerintah maupun kualitas pelayanan publik secara langsung dan real-time. Karakteristik media sosial yang terbuka, cepat, dan menghasilkan data dalam jumlah besar menjadikan platform ini sebagai sumber informasi yang potensial untuk mengevaluasi persepsi masyarakat terhadap berbagai layanan pemerintah. Oleh karena itu, analisis terhadap opini publik pada media sosial menjadi salah satu pendekatan yang efektif dalam mendukung proses pengambilan keputusan berbasis data (*data-driven decision making*) bagi instansi pemerintah [1].
Seiring meningkatnya jumlah pengguna media sosial, volume data teks yang dihasilkan juga semakin besar dan beragam. Data tersebut mengandung informasi penting mengenai tingkat kepuasan masyarakat, kritik terhadap kualitas pelayanan, serta harapan terhadap peningkatan layanan publik. Namun demikian, karakteristik teks pada media sosial yang tidak terstruktur, mengandung singkatan, bahasa tidak baku,

emoji, maupun kesalahan penulisan menyebabkan proses analisis secara manual menjadi kurang efektif. Oleh karena itu, diperlukan suatu pendekatan otomatis yang mampu mengolah data teks secara efisien. Salah satu cabang ilmu yang berkembang pesat dalam mengatasi permasalahan tersebut adalah *Natural Language Processing* (NLP), yang memungkinkan komputer memahami, memproses, dan menginterpretasikan bahasa manusia secara otomatis [2].

Salah satu implementasi NLP yang paling banyak diterapkan adalah analisis sentimen, yaitu proses mengidentifikasi polaritas suatu opini ke dalam kategori positif, negatif, maupun netral berdasarkan isi teks yang ditulis oleh pengguna. Analisis sentimen telah dimanfaatkan secara luas dalam berbagai bidang, seperti pendidikan, kesehatan, *e-commerce*, telekomunikasi, hingga evaluasi kebijakan pemerintah. Pada sektor pelayanan publik, analisis sentimen memberikan informasi yang berguna dalam mengidentifikasi tingkat kepuasan masyarakat terhadap suatu layanan sehingga dapat digunakan sebagai dasar evaluasi maupun penyusunan kebijakan yang lebih responsif terhadap kebutuhan masyarakat [3].

Metode klasifikasi berbasis machine learning masih menjadi pendekatan yang banyak digunakan dalam analisis sentimen. Salah satu algoritma yang memiliki performa tinggi dalam klasifikasi data teks adalah *Support Vector Machine* (SVM). Algoritma ini bekerja dengan membangun *hyperplane* optimal yang mampu memisahkan data ke dalam kelas-kelas tertentu sehingga menghasilkan kemampuan generalisasi yang baik, terutama pada data berdimensi tinggi. Keunggulan tersebut menjadikan SVM sebagai salah satu algoritma yang banyak digunakan dalam penelitian analisis sentimen karena mampu memberikan hasil klasifikasi yang stabil dengan kompleksitas komputasi yang relatif rendah [4]. Representasi fitur yang umum digunakan pada metode ini adalah *Term Frequency–Inverse Document Frequency* (TF-IDF) yang mampu memberikan bobot terhadap setiap kata berdasarkan tingkat kepentingannya dalam suatu dokumen sehingga meningkatkan kualitas proses klasifikasi teks [5].

Perkembangan teknologi kecerdasan buatan mendorong munculnya pendekatan berbasis *deep learning* yang mampu memahami konteks bahasa secara lebih baik dibandingkan metode konvensional. Salah satu perkembangan paling berpengaruh adalah diperkenalkannya arsitektur *Transformer* yang menggunakan mekanisme *self-attention* untuk mempelajari hubungan antarkata dalam sebuah kalimat secara kontekstual [6]. Berdasarkan arsitektur tersebut, dikembangkan model *Bidirectional Encoder Representations from Transformers* (BERT) yang mampu mempelajari representasi bahasa secara dua arah (*bidirectional*) sehingga menghasilkan performa yang jauh lebih baik pada berbagai tugas NLP, termasuk klasifikasi teks dan analisis sentimen [7].

Untuk mendukung pemrosesan bahasa Indonesia, dikembangkan model IndoBERT yang dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar sehingga mampu memahami karakteristik linguistik bahasa Indonesia secara lebih baik dibandingkan model multilingual. Berbagai penelitian menunjukkan bahwa IndoBERT mampu menghasilkan tingkat akurasi yang lebih tinggi dibandingkan algoritma machine learning konvensional pada berbagai tugas klasifikasi teks. Penelitian yang dilakukan oleh Simanihuruk et al. menunjukkan bahwa IndoBERT memberikan performa yang lebih baik dibandingkan SVM pada analisis sentimen data pasar saham Indonesia [8]. Hasil serupa juga ditunjukkan oleh penelitian Br. Sembiring et al. pada analisis sentimen *platform* pendidikan digital yang memperlihatkan bahwa IndoBERT mampu meningkatkan nilai *F1-score* secara signifikan dibandingkan metode klasifikasi tradisional [9], [10].

Penerapan IndoBERT juga telah dilakukan pada berbagai domain lainnya. Kurniawati et al. membandingkan performa SVM dan IndoBERT pada analisis sentimen masyarakat terhadap implementasi Undang-Undang Perlindungan Data Pribadi di Indonesia dan menunjukkan bahwa IndoBERT memiliki kemampuan yang lebih baik dalam memahami konteks bahasa pada media sosial X [1]. Penelitian Widayanti dan Kasih juga menunjukkan bahwa pendekatan *Class-Weighted* IndoBERT mampu meningkatkan performa klasifikasi pada dataset yang tidak seimbang dibandingkan model SVM konvensional [11]. Selain itu, penelitian Abriananta et al. memperlihatkan bahwa IndoBERT menghasilkan akurasi tertinggi dibandingkan *Naïve Bayes* dan SVM pada analisis sentimen terhadap Program Makan Bergizi Gratis [12].

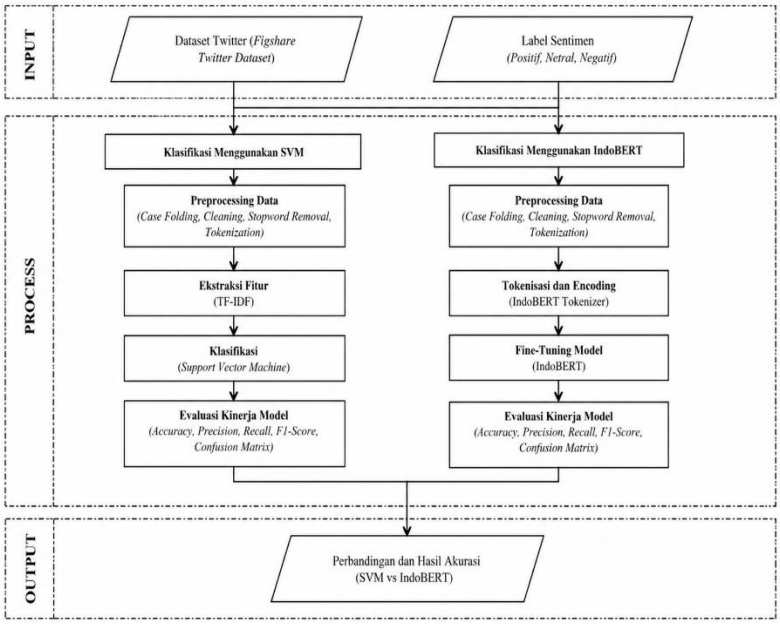
Penelitian mengenai perbandingan IndoBERT dan SVM juga dilakukan pada berbagai domain lain, seperti layanan akademik [13], isu migrasi tenaga kerja Indonesia [14], layanan telekomunikasi [15], serta berbagai isu sosial dan ekonomi [16]. Secara umum, penelitian-penelitian tersebut menunjukkan bahwa model berbasis Transformer mampu menghasilkan representasi semantik yang lebih baik dibandingkan metode berbasis fitur statistik. Meskipun demikian, SVM masih memiliki keunggulan dalam hal efisiensi komputasi, kebutuhan data pelatihan yang lebih kecil, serta waktu pelatihan yang lebih singkat sehingga tetap relevan digunakan sebagai model pembanding dalam penelitian analisis sentimen [17].

Berdasarkan hasil telaah literatur, sebagian besar penelitian sebelumnya menggunakan dataset yang diperoleh dari proses crawling pada objek tertentu, seperti layanan akademik, pendidikan, maupun

kebijakan pemerintah. Penelitian yang membandingkan performa *Support Vector Machine* dan IndoBERT dengan memanfaatkan dataset global yang telah melalui proses pelabelan sentimen masih relatif terbatas. Oleh karena itu, penelitian ini memanfaatkan *Figshare Twitter Dataset* sebagai dataset utama yang kemudian dilakukan proses filtering berdasarkan kata kunci layanan publik sehingga diperoleh data yang relevan dengan topik penelitian. Selanjutnya, dilakukan tahapan preprocessing, ekstraksi fitur menggunakan TF-IDF untuk model SVM, serta proses *fine-tuning* IndoBERT untuk memperoleh representasi kontekstual pada model *Transformer*. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision, recall, F1-score*, dan *confusion matrix* untuk membandingkan performa kedua metode. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem analisis sentimen pada sektor pelayanan publik sekaligus menjadi referensi bagi pemerintah dalam memanfaatkan data media sosial sebagai instrumen evaluasi kualitas layanan publik.

2. Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan kinerja algoritma *Support Vector Machine* (SVM) dan IndoBERT dalam melakukan klasifikasi sentimen masyarakat terhadap layanan publik pada media sosial X.



Gambar 1. Alur Penelitian

Alur penelitian diawali dengan pengumpulan data menggunakan *Figshare Twitter Dataset*, yaitu dataset global yang telah memiliki label sentimen berupa positif, netral, dan negatif. Dataset tersebut dipilih karena telah melalui proses anotasi sehingga dapat langsung digunakan sebagai data pelatihan maupun pengujian. Agar sesuai dengan tujuan penelitian, dilakukan proses *filtering* berdasarkan kata kunci yang berkaitan dengan layanan publik, seperti *government*, *public service*, *healthcare*, *education*, *transportation*, *tax*, *passport*, *municipality*, dan berbagai istilah lain yang merepresentasikan pelayanan publik. Tahapan penyaringan ini bertujuan untuk memperoleh data yang benar-benar relevan sehingga proses klasifikasi dapat menghasilkan representasi sentimen masyarakat terhadap layanan publik secara lebih akurat.

Setelah proses penyaringan selesai, seluruh data diproses melalui tahap *preprocessing* untuk meningkatkan kualitas teks sebelum digunakan dalam proses klasifikasi. Tahapan ini meliputi *case folding* dengan mengubah seluruh karakter menjadi huruf kecil, pembersihan data (*cleaning*) untuk menghilangkan URL, mention, hashtag, emoji, angka, tanda baca, dan karakter khusus yang tidak memiliki makna terhadap proses klasifikasi. Selanjutnya dilakukan stopword removal untuk menghapus kata-kata umum yang tidak memberikan kontribusi terhadap makna kalimat, kemudian dilakukan tokenization sehingga setiap kalimat dipecah menjadi kumpulan token yang siap diproses oleh model klasifikasi. Proses *preprocessing* bertujuan menghasilkan data yang lebih bersih, konsisten, dan mampu meningkatkan kualitas representasi fitur pada tahap berikutnya.

Data hasil preprocessing selanjutnya diproses menggunakan dua pendekatan klasifikasi yang berbeda. Pada pendekatan pertama digunakan algoritma *Support Vector Machine* (SVM) dengan representasi fitur menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode TF-IDF memberikan bobot terhadap setiap kata berdasarkan frekuensi kemunculannya pada dokumen dan tingkat kepentingannya terhadap keseluruhan koleksi dokumen sehingga mampu menghasilkan representasi numerik yang efektif untuk proses klasifikasi teks. Representasi tersebut kemudian digunakan sebagai masukan bagi algoritma SVM yang bekerja dengan membangun *hyperplane optimal* untuk memisahkan data ke dalam kelas sentimen positif, netral, dan negatif. Pada penelitian ini digunakan kernel linear karena memiliki kemampuan yang baik dalam menangani data teks berdimensi tinggi serta memberikan proses pelatihan yang relatif efisien.

Pendekatan kedua menggunakan model IndoBERT, yaitu model bahasa berbasis arsitektur *Transformer* yang telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar. Berbeda dengan SVM yang memanfaatkan representasi TF-IDF, IndoBERT menggunakan proses tokenisasi dan encoding melalui IndoBERT Tokenizer sehingga setiap kalimat dikonversi menjadi representasi numerik kontekstual yang mampu mempertimbangkan hubungan antar kata dalam sebuah kalimat. Selanjutnya dilakukan proses fine-tuning menggunakan dataset hasil *preprocessing* sehingga model dapat menyesuaikan parameter yang telah dipelajari sebelumnya dengan karakteristik data layanan publik pada media sosial X. Pendekatan ini memungkinkan model memahami konteks kalimat secara lebih baik, termasuk penggunaan bahasa informal, singkatan, maupun variasi penulisan yang umum ditemukan pada media sosial.

Setelah kedua model menghasilkan prediksi sentimen, dilakukan proses evaluasi untuk mengetahui tingkat performa masing-masing metode. Evaluasi dilakukan menggunakan beberapa metrik pengukuran, yaitu *accuracy, precision, recall, F1-score*, dan *confusion matrix*. *Accuracy* digunakan untuk mengukur tingkat ketepatan klasifikasi secara keseluruhan, sedangkan *precision* dan *recall* digunakan untuk mengevaluasi kemampuan model dalam mengidentifikasi setiap kelas sentimen. Nilai *F1-score* digunakan sebagai ukuran keseimbangan antara *precision* dan *recall*, sementara *confusion matrix* dimanfaatkan untuk menganalisis distribusi prediksi yang benar maupun salah pada setiap kategori sentimen. Seluruh proses evaluasi dilakukan menggunakan data pengujian yang sama sehingga hasil yang diperoleh dapat dibandingkan secara objektif.

Tahap akhir penelitian adalah melakukan analisis komparatif terhadap performa *Support Vector Machine* dan IndoBERT berdasarkan seluruh metrik evaluasi yang diperoleh. Perbandingan ini bertujuan untuk mengetahui metode yang memiliki kemampuan terbaik dalam mengklasifikasikan sentimen masyarakat terhadap layanan publik pada media sosial X. Selain membandingkan tingkat akurasi, penelitian ini juga menganalisis kelebihan dan keterbatasan masing-masing metode dari aspek kemampuan memahami konteks bahasa, efisiensi komputasi, serta kualitas klasifikasi pada data media sosial. Hasil analisis tersebut diharapkan dapat memberikan rekomendasi mengenai metode yang paling efektif untuk diterapkan pada sistem analisis sentimen layanan publik berbasis media sosial serta menjadi referensi bagi penelitian selanjutnya dalam pengembangan teknologi *Natural Language Processing* untuk mendukung evaluasi pelayanan publik secara otomatis.

3. Hasil dan Pembahasan

Penelitian ini bertujuan membandingkan performa algoritma *Support Vector Machine* (SVM) dan IndoBERT dalam mengklasifikasikan sentimen masyarakat terhadap layanan publik pada media sosial X. Dataset yang digunakan berasal dari *Figshare Twitter Dataset* yang telah memiliki label sentimen berupa positif, negatif, dan netral. Sebelum proses klasifikasi dilakukan, dataset melalui tahapan penyaringan (*filtering*) menggunakan kata kunci yang berkaitan dengan layanan publik sehingga data yang digunakan benar-benar merepresentasikan opini masyarakat terhadap pelayanan pemerintah. Tahap ini penting karena dataset Figshare memiliki cakupan topik yang luas, sedangkan penelitian difokuskan pada sentimen mengenai layanan publik. Hasil penyaringan kemudian diproses lebih lanjut melalui tahapan preprocessing untuk menghasilkan data yang bersih dan siap digunakan pada proses klasifikasi.

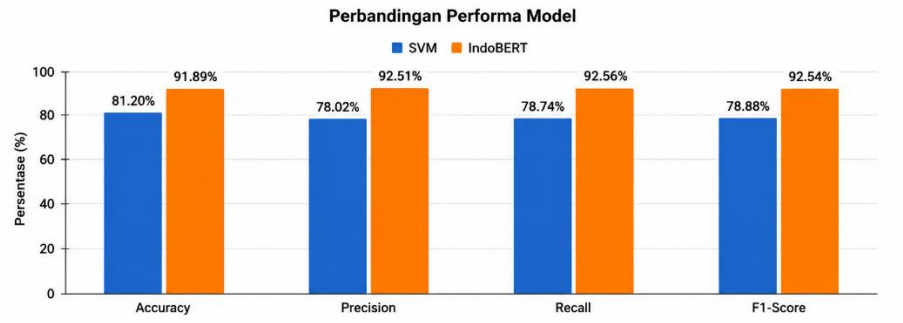
Tahapan preprocessing dilakukan untuk meningkatkan kualitas data teks dengan menghilangkan berbagai unsur yang tidak memiliki kontribusi terhadap proses klasifikasi, seperti URL, *mention, hashtag, emoji*, angka, tanda baca, dan karakter khusus lainnya. Selanjutnya dilakukan proses *case folding*, stopword *removal*, dan *tokenization* sehingga setiap tweet memiliki struktur yang lebih konsisten. Tahapan ini berperan penting dalam mengurangi *noise* pada data media sosial yang umumnya memiliki karakteristik bahasa tidak baku, singkatan, maupun variasi penulisan yang tinggi. Data hasil preprocessing kemudian dibagi menjadi data pelatihan dan data pengujian sebelum diproses menggunakan dua pendekatan klasifikasi yang berbeda, yaitu *Support Vector Machine* dan IndoBERT.

Pada pendekatan pertama, algoritma *Support Vector Machine* memanfaatkan representasi fitur menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Representasi tersebut mengubah setiap dokumen menjadi vektor numerik berdasarkan bobot kemunculan setiap kata sehingga dapat digunakan sebagai masukan bagi algoritma SVM. Pendekatan ini dipilih karena telah banyak digunakan pada penelitian analisis sentimen dan dikenal memiliki kemampuan yang baik dalam menangani data teks berdimensi tinggi. Sementara itu, pendekatan kedua menggunakan model IndoBERT yang memanfaatkan representasi bahasa berbasis Transformer. Berbeda dengan TF-IDF yang hanya mempertimbangkan frekuensi kata, IndoBERT mampu memahami hubungan kontekstual antar kata melalui mekanisme self-attention sehingga informasi semantik pada setiap kalimat dapat dipertahankan selama proses klasifikasi.
Evaluasi terhadap kedua model dilakukan menggunakan metrik *Accuracy, Precision, Recall,* dan *F1-Score* untuk mengetahui kemampuan masing-masing model dalam mengklasifikasikan sentimen masyarakat terhadap layanan publik. Hasil pengujian menunjukkan bahwa model IndoBERT memberikan performa yang lebih baik dibandingkan *Support Vector Machine* pada seluruh metrik evaluasi. Ringkasan hasil pengujian ditunjukkan pada Tabel 1.

Tabel 1. Perbandingan Kinerja Model Klasifikasi

Model	***Accuracy* (%)**	***Precision* (%)**	***Recall* (%)**	***F1-Score* (%)**
Support Vector Machine	81.20	78.02	78.74	78.88
IndoBERT	**91.89**	**92.51**	**92.56**	**92.54**

Berdasarkan Tabel 1, model *Support Vector Machine* memperoleh nilai *accuracy* sebesar 81,20%, dengan *precision* sebesar 78,02%, *recall* sebesar 78,74%, dan *F1-score* sebesar 78,88%. Hasil tersebut menunjukkan bahwa SVM mampu melakukan klasifikasi sentimen dengan tingkat akurasi yang cukup baik. Penggunaan TF-IDF sebagai representasi fitur mampu menangkap kata-kata penting yang muncul pada tweet sehingga model dapat membedakan kelas sentimen positif, negatif, maupun netral. Akan tetapi, pendekatan ini masih memiliki keterbatasan dalam memahami hubungan antar kata maupun konteks kalimat karena setiap kata diperlakukan secara independen berdasarkan frekuensi kemunculannya. Akibatnya, SVM cenderung mengalami kesalahan klasifikasi ketika menghadapi kalimat yang mengandung negasi, ironi, atau penggunaan bahasa informal yang umum ditemukan pada media sosial X.



Gambar 2. Grafik perbandingan model

Sebaliknya, model IndoBERT menunjukkan performa yang jauh lebih tinggi dengan *accuracy* sebesar 91,89%, *precision* sebesar 92,51%, *recall* sebesar 92,56%, dan *F1-score* sebesar 92,54%. Peningkatan akurasi sebesar 10,69% dibandingkan SVM menunjukkan bahwa representasi kontekstual yang dimiliki IndoBERT mampu memahami makna kalimat secara lebih baik dibandingkan representasi berbasis TF-IDF. Selain itu, nilai *precision*, *recall*, dan *F1-score* yang seluruhnya berada di atas 92% menunjukkan bahwa model tidak hanya mampu memberikan prediksi yang tepat, tetapi juga mampu mengidentifikasi sebagian besar data pada setiap kelas sentimen secara konsisten. Kemampuan tersebut diperoleh karena IndoBERT memanfaatkan mekanisme *bidirectional encoding* sehingga hubungan semantik antar kata dapat dipelajari secara menyeluruh, termasuk pada kalimat yang menggunakan bahasa sehari-hari maupun struktur kalimat yang kompleks.
Perbedaan performa kedua model menunjukkan bahwa representasi fitur memiliki pengaruh yang signifikan terhadap hasil klasifikasi sentimen. Pada SVM, kualitas prediksi sangat bergantung pada representasi statistik yang dihasilkan oleh TF-IDF sehingga hubungan semantik antar kata belum dapat dipahami secara optimal. Sebaliknya, IndoBERT menghasilkan representasi berbasis konteks yang mampu

mempertahankan makna setiap kata sesuai dengan posisi dan hubungan terhadap kata lain di dalam kalimat. Hal ini menjadi faktor utama yang menyebabkan IndoBERT menghasilkan nilai evaluasi yang lebih tinggi dibandingkan SVM. Temuan ini juga menunjukkan bahwa model berbasis *Transformer* lebih sesuai digunakan pada data media sosial yang memiliki karakteristik bahasa yang dinamis, penggunaan singkatan, serta variasi penulisan yang tinggi.
Hasil penelitian ini sejalan dengan penelitian Kurniawati et al. [1] yang melaporkan bahwa IndoBERT memberikan performa lebih baik dibandingkan *Support Vector Machine* pada analisis sentimen masyarakat terhadap implementasi Undang-Undang Perlindungan Data Pribadi di platform X. Penelitian Simanihuruk et al. [8] juga menunjukkan bahwa IndoBERT menghasilkan akurasi yang lebih tinggi dibandingkan SVM pada analisis sentimen pasar saham Indonesia. Temuan serupa dilaporkan oleh Br. Sembiring et al. [9], Widayanti dan Kasih [11], serta Abriananta et al. [12], yang menyimpulkan bahwa model berbasis *Transformer* memiliki kemampuan lebih baik dalam memahami konteks bahasa Indonesia dibandingkan pendekatan machine learning konvensional. Dengan demikian, hasil penelitian ini memperkuat temuan penelitian terdahulu bahwa penggunaan IndoBERT merupakan pendekatan yang efektif untuk analisis sentimen berbahasa Indonesia, khususnya pada data media sosial.

Meskipun IndoBERT memberikan performa terbaik, penggunaan model ini memerlukan sumber daya komputasi yang lebih tinggi dibandingkan SVM. Proses fine-tuning membutuhkan waktu pelatihan yang lebih lama serta perangkat keras yang mendukung komputasi berbasis GPU agar proses pelatihan berlangsung secara optimal. Sebaliknya, *Support Vector Machine* masih memiliki keunggulan dari sisi efisiensi komputasi, kebutuhan memori yang lebih kecil, dan waktu pelatihan yang lebih singkat. Oleh karena itu, pemilihan model dapat disesuaikan dengan kebutuhan implementasi. Jika prioritas utama adalah akurasi klasifikasi, maka IndoBERT menjadi pilihan yang lebih tepat. Namun, apabila sistem akan diterapkan pada lingkungan dengan keterbatasan sumber daya komputasi atau membutuhkan proses pelatihan yang cepat, *Support Vector Machine* masih dapat menjadi alternatif yang layak.
Secara keseluruhan, hasil penelitian menunjukkan bahwa IndoBERT merupakan model yang lebih efektif dalam melakukan analisis sentimen masyarakat terhadap layanan publik pada media sosial X dibandingkan *Support Vector Machine*. Peningkatan performa yang diperoleh pada seluruh metrik evaluasi menunjukkan bahwa pemanfaatan representasi bahasa berbasis *Transformer* mampu meningkatkan kualitas klasifikasi sentimen pada data berbahasa Indonesia. Hasil ini diharapkan dapat menjadi referensi bagi pengembangan sistem pemantauan opini publik berbasis kecerdasan buatan, sekaligus memberikan masukan bagi instansi pemerintah dalam memanfaatkan media sosial sebagai salah satu sumber evaluasi terhadap kualitas layanan publik secara berkelanjutan.

4. Kesimpulan

Penelitian ini berhasil membandingkan performa *Support Vector Machine* (SVM) dan IndoBERT dalam melakukan analisis sentimen masyarakat terhadap layanan publik pada media sosial X menggunakan *Figshare Twitter Dataset*. Berdasarkan hasil evaluasi, model IndoBERT menunjukkan kinerja yang lebih unggul dibandingkan SVM dengan memperoleh *accuracy* sebesar 91,89%, *precision* 92,51%, *recall* 92,56%, dan *F1-score* 92,54%, sedangkan SVM memperoleh *accuracy* 81,20%, *precision* 78,02%, *recall* 78,74%, dan *F1-score* 78,88%. Peningkatan akurasi sebesar 10,69% menunjukkan bahwa kemampuan representasi kontekstual pada IndoBERT lebih efektif dalam memahami karakteristik bahasa Indonesia yang digunakan pada media sosial dibandingkan pendekatan berbasis TF-IDF pada SVM. Meskipun demikian, SVM masih memiliki keunggulan pada aspek efisiensi komputasi dan waktu pelatihan sehingga tetap menjadi alternatif yang baik untuk implementasi dengan sumber daya terbatas. Hasil penelitian ini menunjukkan bahwa IndoBERT lebih direkomendasikan untuk pengembangan sistem analisis sentimen layanan publik yang membutuhkan tingkat akurasi tinggi. Penelitian selanjutnya dapat memanfaatkan dataset yang lebih besar, menerapkan teknik penyeimbangan kelas, serta membandingkan IndoBERT dengan model *Transformer* terbaru seperti RoBERTa, XLM-RoBERTa, atau ModernBERT untuk memperoleh performa klasifikasi yang lebih optimal.

5. Daftar Pustaka

[1] Y. Kurniawati, R. B. Hamid, D. I. Sensuse, S. Lusa, P. A. W. Putro, and S. Indriasari, “Analysis of Public Sentiment Indonesia’s Personal Data Protection Law: A Comparison of SVM and IndoBERT on X Platform,” *J. Tek. Inform.*, vol. 7, no. 2, pp. 1007–1027, Apr. 2026, doi: 10.52436/1.jutif.2026.7.2.5415.

[2] R. Simanihuruk, E. Yulianti, K. Azizah, and W. Jatmiko, "Sentiment Analysis on Indonesian Stock Market Texts: A Comparative Study of Support Vector Machine (SVM) and IndoBERT," in *Proceedings of 2025 IEEE International Conference on Data and Software Engineering, ICoDSE 2025*, IEEE, Oct. 2025, pp. 455–460. doi: 10.1109/ICoDSE68111.2025.11351619.

[3] S. Shafwah and H. Al Azies, "Komparasi SVM dan IndoBERT dalam Klasifikasi Sentimen Program Makanan Bergizi Gratis," *JTERA (Jurnal Teknol. Rekayasa)*, vol. 10, no. 2, p. 105, Jan. 2026, doi: 10.31544/jtera.v10.i2.2025.105-112.

[4] R. A. Siregar, F. Fitriyani, and L. S. Darfiansa, "Sentiment Analysis Of Indihome Service Based On Geo Location Using The Bert Model On Platform X," *J. Tek. Inform.*, vol. 6, no. 5, pp. 2975–2990, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.4993.

[5] Nida Nur Aini Aryanti and Ozzi Suria, "ANALISIS SENTIMEN TERHADAP PEMUTUSAN HUBUNGAN KERJA DI INDONESIA : KOMPARASI INDOBERT DENGAN SVM, RANDOM FOREST, DAN DECISION TREE DENGAN OPTIMASI TF - IDF," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1158–1176, Jul. 2025, doi: 10.36341/rabit.v10i2.6364.

[6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 4171–4186, May 2019, [Online]. Available: http://arxiv.org/abs/1810.04805

[7] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 2023, [Online]. Available: http://arxiv.org/abs/1706.03762

[8] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513–523, Jan. 1988, doi: 10.1016/0306-4573(88)90021-0.

[9] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations ofwords and phrases and their compositionality," *Adv. Neural Inf. Process. Syst.*, Oct. 2013, [Online]. Available: http://arxiv.org/abs/1310.4546

[10] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Am. Med. Informatics Assoc.*, vol. 18, no. 5, pp. 544–551, Sep. 2011, doi: 10.1136/amiajnl-2011-000464.

[11] R. Widayanti and F. C. Kasih, "Comparative Analysis of Baseline IndoBERT , Class-Weighted IndoBERT , and SMOTE with Support Vector Machine for Handling Imbalanced Sentiment Classification in Indonesian," vol. 7, no. 3, pp. 2857–2875, 2026.

[12] H. Abriananta *et al.*, "Comparison of Naive Bayes , Support Vector Machine , and Indobert Methods for Classifying Public Sentiment towards the MBG Program on," vol. 10, no. 3, pp. 2806–2815, 2026.

[13] M. Ibnu, L. Muflikhah, and F. A. Bachtiar, "Perbandingan SVM dan IndoBERT untuk Analisis Sentimen Layanan Akademik Mahasiswa," vol. 7, no. 2, pp. 265–277, 2026, doi: 10.47065/bit.v5i2.2913.

[14] A. Amrullah, "Analisis Sentimen Publik terkait Migrasi Tenaga Kerja Indonesia di Platform X menggunakan SVM-IndoBERT," *J. UJMC*, vol. 11, no. 1, pp. 53–63, 2025.

[15] A. B. S. Br Sembiring, R. Robet, and L. Hoki, "Comparison of IndoBERT and SVM Performance in Sentiment Analysis of Digital Education Platforms," *Sinkron*, vol. 10, no. 1, pp. 64–74, Jan. 2026, doi: 10.33395/sinkron.v10i1.15472.

[16] H. Pal and B. Bhushan, "Sentiment Analysis on Twitter Dataset using Voting Classifier," in *2024 International Conference on Electrical, Electronics and Computing Technologies, ICEECT 2024*, IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ICEECT61758.2024.10739316.

[17] N. A. P. Masaling, T. Lubis, A. Amalia, A. R. Lubis, and M. S. Lydia, "Category Based

Sentiment Analysis for Basic Skin-Cares Using LDA and SVM Approach," in *Proceeding - ELTICOM 2022: 6th International Conference on Electrical, Telecommunication and Computer Engineering 2022*, IEEE, Nov. 2022, pp. 182–189. doi: 10.1109/ELTICOM57747.2022.10037876.

6. Penulis

Marissa Utami
Fakultas Teknik, Universitas Muhammadiyah Bengkulu,
Bengkulu, Indonesia
Penulis merupakan tenaga pendidik di Program Studi Sistem Informasi Universitas Muhammadiyah Bengkulu

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