

Analisis Sentimen Layanan Kesehatan Klinik Yamet Komentar Instagram Dengan *Naïve Bayes*

¹Reina Salsa Kinanty, ²Ari Wedhasmara, ³Rizka Dhini Kurnia

^{1,2,3}Universitas Sriwijaya, Indonesia

¹reinasalsak@gmail.com; ²a_wedhasmara@unsri.ac.id; ³rizkadhini@gmail.com;

Article Info

Article history:

Received, 2025-11-09

Revised, 2025-11-12

Accepted, 2025-11-18

Kata Kunci:

Analisis Sentimen
Media Sosial
Klinik Yamet Palembang
Naïve Bayes
TF-IDF
SMOTE
Layanan Kesehatan Digital

ABSTRAK

Kemajuan teknologi informasi yang pesat mendorong sektor kesehatan untuk bertransformasi menuju pemanfaatan platform digital demi meningkatkan efisiensi dan kemudahan akses layanan. Klinik Yamet Palembang, sebagai pusat terapi bagi anak berkebutuhan khusus, telah menggunakan media sosial sebagai kanal komunikasi utama, meskipun situs web resminya masih belum dioptimalkan secara maksimal. Permasalahan utama yang diangkat dalam penelitian ini adalah belum teridentifikasinya persepsi publik terhadap layanan klinik tersebut di ranah digital. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap layanan Klinik Yamet Palembang melalui komentar yang terdapat pada akun Instagram resminya. Algoritma *Naïve Bayes* dipilih karena memiliki performa yang baik dalam klasifikasi teks dengan jumlah data yang relatif terbatas serta tingkat kompleksitas yang rendah. Prosedur penelitian mencakup tahap pra-pemrosesan teks, pembobotan menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), dan penerapan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan proporsi data antar kategori sentimen. Berdasarkan hasil pengujian, model yang dibangun mampu mengklasifikasikan opini masyarakat ke dalam tiga kategori utama, yaitu positif, netral, dan negatif, dengan tingkat akurasi sebesar 86,56% sebelum penerapan SMOTE. Namun, setelah dilakukan penyeimbangan data menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE), tingkat akurasi model mengalami penurunan menjadi 70,00%. Kombinasi TF-IDF, Naïve Bayes, dan SMOTE terbukti efektif dalam menggambarkan persepsi masyarakat terhadap layanan kesehatan berbasis digital. Secara aplikatif, hasil penelitian ini memberikan rekomendasi bagi Klinik Yamet Palembang dalam mengembangkan strategi komunikasi digital yang lebih adaptif, serta memberikan kontribusi teoretis bagi penelitian analisis sentimen di bidang kesehatan di Indonesia.

ABSTRACT

Keywords:

Sentiment Analysis
Social Media
Yamet Clinic Palembang
Naïve Bayes
TF-IDF
SMOTE
Digital Health Services

*The rapid advancement of information technology has driven the healthcare sector to adopt digital platforms in order to enhance service efficiency and accessibility. Yamet Clinic Palembang, as a therapy center for children with special needs, has utilized social media as its primary communication channel, although its official website has not yet been fully optimized. The main issue addressed in this study is the lack of understanding regarding public perception of the clinic's digital services. Therefore, this research aims to analyze public sentiment toward Yamet Clinic Palembang's services through comments posted on its official Instagram account. The Naïve Bayes algorithm was selected due to its strong performance in text classification tasks involving limited datasets and low computational complexity. The research process includes text preprocessing, feature weighting using *Term Frequency–Inverse Document Frequency* (TF-IDF), and the application of the *Synthetic Minority Over-sampling Technique* (SMOTE) to balance data distribution among sentiment categories. The experimental results indicate that the developed model successfully classified public opinions into three categories—positive, neutral, and negative—with an accuracy rate of 70%. The combination of TF-IDF, Naïve Bayes, and SMOTE proved effective in capturing public perceptions of digital health services. Practically, the findings provide valuable insights for Yamet Clinic Palembang in developing a more adaptive digital communication strategy and theoretically contribute to the*

advancement of sentiment analysis research within the healthcare service context in Indonesia.

This is an open access article under the [CC BY-SA](#) license.



Penulis Korespondensi:

Ari Wedhasmara,
Program Studi Sistem Informasi,
Universitas Sriwijaya,
Email: a_wedhasmara@unsri.ac.id

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong sektor kesehatan untuk bertransformasi melalui platform digital guna meningkatkan efisiensi layanan dan memperluas akses informasi kesehatan. Dalam era digital saat ini, masyarakat semakin bergantung pada media sosial dan situs web sebagai sumber utama informasi terkait kesehatan [1]. Berdasarkan laporan Asosiasi Penyelenggara Jasa Internet Indonesia [2], jumlah pengguna internet di Indonesia pada tahun 2024 mencapai 221,5 juta jiwa atau sekitar 77% dari total populasi nasional [3]. Kondisi ini menunjukkan peran penting media sosial sebagai sarana komunikasi interaktif antara penyedia layanan kesehatan dan masyarakat.

Klinik Yamet Palembang merupakan lembaga terapi bagi anak berkebutuhan khusus yang menyediakan layanan terapi wicara, okupasi, dan perilaku. Meskipun telah memiliki situs web resmi, penggunaannya belum dioptimalkan sebagai media komunikasi dua arah. Berdasarkan wawancara dengan pihak klinik, diketahui bahwa interaksi publik justru lebih aktif terjadi melalui akun Instagram resmi yang digunakan untuk berbagi informasi edukatif dan promosi layanan. Namun, hingga saat ini belum ada kajian ilmiah yang meneliti bagaimana opini publik terbentuk terhadap layanan Klinik Yamet melalui komentar di media sosial. Analisis terhadap opini publik tersebut penting untuk memahami persepsi, tingkat kepuasan, dan harapan masyarakat terhadap kualitas layanan yang diberikan.

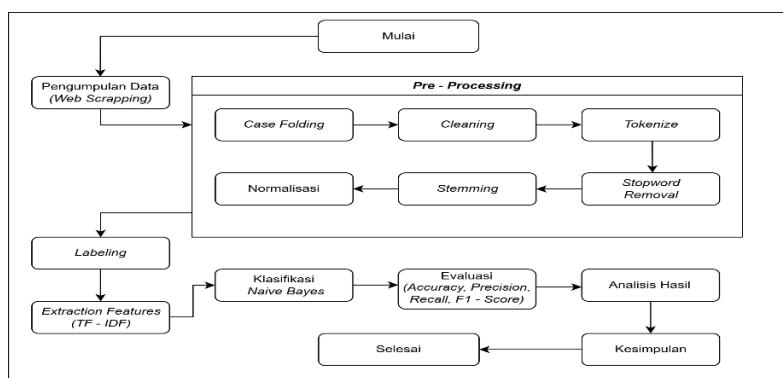
Berbagai penelitian terdahulu telah menerapkan analisis sentimen dalam konteks layanan kesehatan digital, seperti pada aplikasi *Halodoc* [4], *MySiloam* [5], serta layanan kesehatan di platform Twitter [6]. Namun, sebagian besar penelitian tersebut berfokus pada rumah sakit atau aplikasi daring, sementara kajian mengenai klinik terapi anak melalui platform Instagram masih sangat terbatas. Selain itu, beberapa penelitian terdahulu menghadapi permasalahan ketidakseimbangan data yang berdampak pada akurasi model klasifikasi. Untuk mengatasi hal tersebut, penelitian ini mengombinasikan *Naïve Bayes* dengan *Term Frequency–Inverse Document Frequency* (TF-IDF) serta menerapkan *Synthetic Minority Over-sampling Technique* (SMOTE) sebagai upaya penyeimbangan data.

Meskipun SMOTE secara umum digunakan untuk meningkatkan performa model klasifikasi, hasil penelitian ini menunjukkan bahwa penerapannya tidak selalu memberikan peningkatan akurasi pada algoritma *Naïve Bayes*. Hal tersebut diduga dipengaruhi oleh karakteristik data dan asumsi independensi fitur pada model probabilistik. Dengan demikian, penelitian ini juga memberikan temuan empiris bahwa efektivitas SMOTE perlu disesuaikan dengan konteks dan sifat dataset yang digunakan.

Evaluasi hasil klasifikasi dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur tingkat performa dan keandalan model. Secara teoretis, penelitian ini berkontribusi pada pengembangan kajian analisis sentimen dan pemrosesan bahasa alami (*Natural Language Processing*) dalam konteks layanan kesehatan digital. Secara praktis, hasil penelitian ini diharapkan dapat menjadi masukan bagi Klinik Yamet Palembang dalam merancang strategi komunikasi digital yang lebih adaptif terhadap opini publik serta memperkuat citra layanan kesehatan berbasis data.

2. METODE PENELITIAN

Alur penelitian ini dirancang secara sistematis agar setiap tahap dapat saling mendukung dan menghasilkan output yang sesuai tujuan penelitian. Secara keseluruhan, tahapan penelitian ditunjukkan melalui *flowchart* pada Gambar 1 yang meliputi proses awal pengumpulan data hingga tahap akhir penarikan kesimpulan.

Gambar 1 *Flowchart* Tahapan Penelitian

PENGUMPULAN DATA

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu komentar publik yang telah diunggah oleh pengguna pada akun resmi Instagram Klinik Yamet Palembang. Pengumpulan data dilakukan tanpa melakukan interaksi langsung dengan responden, melainkan melalui pemanfaatan data digital yang tersedia secara terbuka (public data).

Proses pengambilan data dilakukan menggunakan ekstensi peramban Chrome “*IG Comment Export Tool*”, yang memungkinkan peneliti mengekspor komentar dari unggahan Instagram ke dalam format CSV/Excel [7]. Pengumpulan dilakukan pada bulan september 2025, dengan menentukan rentang waktu unggahan yang relevan, yaitu posting yang berkaitan dengan layanan terapi anak berkebutuhan khusus.

Dari proses tersebut diperoleh sebanyak 1.846 komentar mentah. Setelah melalui tahap seleksi awal dan pembersihan data (menghapus komentar ganda, spam, emoji, atau komentar yang tidak relevan), jumlah data yang digunakan dalam penelitian berkurang menjadi 1.831 komentar, atau sekitar 99,2% dari total data awal. Data yang bersifat pribadi, mengandung informasi sensitif, atau tidak berkaitan langsung dengan layanan Klinik Yamet tidak disertakan dalam analisis.

Seluruh proses pengumpulan data dilakukan dengan memperhatikan etika penelitian serta ketentuan privasi pengguna. Data yang dikumpulkan bersifat publik dan digunakan semata-mata untuk kepentingan akademik.

Penggunaan alat berbasis *web scraping* untuk mengumpulkan data media sosial seperti Instagram juga telah diterapkan dalam berbagai penelitian sebelumnya. Misalnya, penelitian oleh [4] dan [6] yang menunjukkan bahwa metode ini efektif dan efisien untuk memperoleh data autentik dari media sosial sebagai bahan analisis sentimen.

PRE - PROCESSING

Proses *Pre - Processing* data dilakukan melalui beberapa tahapan yang digambarkan pada *flowchart* di Gambar 1. Tahapan tersebut meliputi *case folding*, *data cleaning*, *tokenize*, *stopword removal*, *stemming*, hingga normalisasi sehingga data teks yang dihasilkan lebih bersih, seragam, dan siap dianalisis. Seluruh tahapan *pre - processing* menggunakan bahasa pemrograman *Python* dengan memanfaatkan fungsi bawaan.

Pada tahap *case folding*, seluruh huruf diubah menjadi huruf kecil agar bentuk penulisan seragam dan sistem tidak membedakan kata berdasarkan kapitalisasi [8]. Selanjutnya dilakukan *cleaning* dengan menghapus karakter non-alfabet seperti angka, tanda baca, emoji, tautan, dan simbol lain yang tidak memiliki makna semantik. Proses ini bertujuan untuk mengurangi *noise* pada data komentar media sosial yang umumnya tidak terstruktur dengan baik [9].

Tahap berikutnya adalah *tokenize* dan *stopword removal*, yaitu memecah kalimat menjadi potongan kata kemudian menghapus kata-kata umum seperti “yang”, “dan”, “di”, atau “ke” yang tidak memiliki pengaruh terhadap analisis sentimen. Penghapusan stopword terbukti dapat meningkatkan kinerja model klasifikasi dengan mengurangi kata-kata yang tidak relevan terhadap konteks opini yang dianalisis [10]. Setelah itu dilakukan *stemming*, yaitu proses mengembalikan setiap kata ke bentuk dasar sehingga kata seperti “bermain”, “dimainkan”, dan “bermainan” akan dianggap sebagai satu bentuk kata, yaitu “main”. Tahap ini penting untuk memastikan model tidak menganggap variasi kata dengan makna serupa sebagai entitas berbeda [11].

Terakhir, dilakukan normalisasi, yaitu penyeragaman bentuk penulisan atau singkatan yang sering digunakan dalam komentar media sosial, seperti mengubah “bgt” menjadi “banget” atau “gk” menjadi “enggak” [12].

LABELING

Pada tahap penyaringan data, setiap ulasan yang terkumpul diberi label sesuai konteks analisis sentimen. Label ini dapat berupa kategori atau kelas tertentu, seperti positif, negatif, atau netral. Proses *labeling* dilakukan dengan mengklasifikasikan setiap data ulasan ke dalam tiga kelas utama, yaitu positif, netral, dan negatif untuk mempermudah proses analisis sentimen dan pelatihan model [13]. Sebagai bentuk validasi tambahan, hasil pelabelan juga dikonsultasikan dengan dosen pembimbing agar tetap selaras dengan konteks penelitian.

EXTRACTION FEATURES

Tahap ekstraksi fitur dalam penelitian ini dilakukan menggunakan metode TF – IDF (*Term Frequency – Inverse Document Frequency*). Dalam penelitian ini, tahap ekstraksi fitur digunakan untuk mengubah data teks yang semula berupa komentar/ulasan menjadi representasi numerik (vektor fitur) yang dapat dipahami dan diproses oleh algoritma klasifikasi seperti *Naïve Bayes*.

Studi terbaru menunjukkan bahwa pembobotan melalui metode TF-IDF merupakan teknik yang efektif untuk merepresentasikan tingkatan kepentingan tiap term dalam korpus sebagai nilai numerik, sehingga memungkinkan model klasifikasi melakukan perhitungan probabilistik [15], [16], [17].

KLASIFIKASI NAÏVE BAYES

Tahap klasifikasi dalam penelitian ini menggunakan algoritma *Naïve Bayes*, salah satu metode *machine learning* berbasis probabilistik yang banyak digunakan dalam pemrosesan teks karena kesederhanaannya dan efektivitasnya dalam pengklasifikasian data [17]. Algoritma ini bekerja berdasarkan prinsip *Teorema Bayes*, yaitu menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu dengan mempertimbangkan distribusi kata dalam dokumen [14].

Dalam penelitian ini, proses perhitungan probabilitas dilakukan terhadap fitur teks yang telah diberi bobot menggunakan metode TF – IDF untuk memperoleh representasi numerik yang relevan terhadap konteks dokumen [18]. Kelas dengan probabilitas tertinggi kemudian dipilih sebagai hasil klasifikasi, sesuai dengan prinsip utama algoritma *Naïve Bayes* yang memilih kelas dengan peluang paling besar [19].

EVALUASI MODEL

Tahap evaluasi model dilakukan setelah proses klasifikasi selesai dengan tujuan untuk mengetahui sejauh mana algoritma yang digunakan mampu bekerja secara akurat dalam mengklasifikasikan data. Pada penelitian ini, evaluasi dilakukan terhadap model *Naïve Bayes* yang telah dilatih menggunakan representasi data TF – IDF. Proses evaluasi biasanya dilakukan dengan membandingkan hasil prediksi model dengan label sebenarnya pada data uji.

Dari perbandingan tersebut dihitung berbagai metrik performa seperti akurasi, *precision*, *recall*, dan *F1-Score*, yang merupakan ukuran umum untuk menilai kinerja model klasifikasi dalam pembelajaran mesin [20]. Dengan menggunakan metrik-metrik tersebut, dapat diketahui apakah model *Naïve Bayes* sudah memberikan hasil yang memuaskan atau masih perlu dilakukan perbaikan, misalnya dengan penyesuaian parameter atau penggunaan teknik tambahan seperti SMOTE untuk menyeimbangkan data [21].

ANALISIS HASIL

Analisis hasil bertujuan untuk menafsirkan kinerja model setelah dievaluasi dengan metrik-metrik seperti akurasi, *precision*, *recall*, dan *F1-Score* sehingga dapat ditentukan sejauh mana model bekerja secara efektif, dari hasil evaluasi tersebut dapat dilihat kemampuan algoritma *Naïve Bayes* yang menggunakan representasi TF – IDF dalam mengklasifikasikan teks secara tepat, nilai akurasi memberikan gambaran performa keseluruhan model, sedangkan *precision*, *recall*, dan *F1-Score* memberikan informasi detail mengenai ketepatan dan kelengkapan prediksi pada tiap kelas [20].

Analisis mendalam terhadap metrik-metrik ini menjadi dasar untuk menilai apakah model telah mencapai tingkat performa yang memadai atau masih memerlukan perbaikan, misalnya melalui tuning parameter, penerapan teknik penyeimbangan data seperti SMOTE, atau pengujian algoritma banding [21].

3. HASIL DAN ANALISIS

Berdasarkan hasil penelitian yang telah dilakukan berikut pemaparan hasil dan analisis dimulai dari pengumpulan data. Data yang berhasil dikumpulkan melalui komentar instagram resmi Klinik Yamet Palembang pada periode 2023 – 2025 berjumlah 1.846 komentar mentah yang terdiri dari tanggapan masyarakat terhadap berbagai ungkapan promosi, kegiatan terapi, serta informasi layanan klinik. Beberapa data komentar hasil pengumpulan data ditampilkan pada Tabel 1.

Tabel 1. Hasil Pengumpulan Data

Komentar
@t_gunadi ok baik, paham dok   trima kasih banyak dok jawaban nya 
Pengen bawa anak ke yamet tp trkendala di biaya  
@t_gunadi terimakasih banyak dok atas jawabannya. Terjawab sudah selama ini masalah sy 
ADHD itu contohnya apa ya dok?
@t_gunadi dok kisaran biaya untuk konsultasi kira2 brpa ya dok

Pada tahap *pre – processing* terdiri dari, *case folding*, *cleaning*, *tokenize*, *stopword removal*, *stemming*, dan normalisasi. Berikut hasil *pre – processing*. Tiap hasil dari *Pre – Processing* ditampilkan pada tabel 2, 3, 4, 5 dan 6.

Tabel 2. Hasil *Case Folding*

Case Folding
@t_gunadi ok baik, paham dok   trima kasih banyak dok jawaban nya 
pengen bawa anak ke yamet tp trkendala di biaya  
@t_gunadi terimakasih banyak dok atas jawabannya. Terjawab sudah selama ini masalah sy 
adhd itu contohnya apa ya dok?
@t_gunadi dok kisaran biaya untuk konsultasi kira2 brpa ya dok

Tabel 3. Hasil *Cleaning*

Cleaning
ok baik, paham dok trima kasih banyak dok jawaban nya
pengen bawa anak ke yamet tp trkendala di biaya
terimakasih banyak dok atas jawabannya. Terjawab sudah selama ini masalah sy
adhd itu contohnya apa ya dok?
dok kisaran biaya untuk konsultasi kira brpa ya dok

Tabel 4. Hasil *Tokenize* dan *Stopword Removal*

Case Folding	Stopword Removal
['ok', 'baik', 'paham', 'dok', 'trima', 'kasih', 'banyak', 'dok', 'jawaban', 'nya']	['baik', 'paham', 'dok', 'trima', 'kasih', 'banyak', 'dok', 'jawaban', 'nya']
['pengen', 'bawa', 'anak', 'ke', 'yamet', 'tp', 'trkendala', 'di', 'biaya']	['pengen', 'bawa', 'anak', 'yamet', 'tp', 'trkendala', 'biaya']
['terimakasih', 'banyak', 'dok', 'atas', 'jawabannya', 'terjawab', 'sudah', 'selama', 'ini', 'masalah', 'sy']	['terimakasih', 'banyak', 'dok', 'atas', 'jawabannya', 'terjawab', 'selama', 'masalah', 'sy']
['adhd', 'itu', 'contohnya', 'apa', 'ya', 'dok']	['adhd', 'contohnya', 'apa', 'dok']
['dok', 'kisaran', 'biaya', 'untuk', 'konsultasi', 'kira', 'brpa', 'ya', 'dok']	['dok', 'kisaran', 'biaya', 'konsultasi', 'kira', 'brpa', 'dok']

Tabel 5. Hasil *Stemming*

Stemming
baik paham dok trima kasih banyak dok jawab nya
ken bawa anak yamet tp trkendala biaya
terimakasih banyak dok atas jawab jawab lama masalah sy
adhd contoh apa dok
dok kisar biaya konsultasi kira brpa dok

Tabel 6. Hasil Normalisasi

Normalisasi

baik paham dokter trima kasih banyak dokter jawab nya
ken bawa anak yamet tapi kendala biaya
terimakasih banyak dokter atas jawab jawab lama masalah
saya
adhd contoh apa dokter
dokter kisar biaya konsultasi kira brpa dokter

Setelah melewati beberapa tahapan *pre – processing* pada data mentah pengumpulan data. Jumlah data komentar yang semula berjumlah 1.846 mengalami pengurangan menjadi 1.831. Selanjutnya proses *labeling* dengan mengklasifikasikan setiap data komentar ke dalam tiga kelas utama, yaitu positif, netral, dan negatif. Hasil *labeling* ditampilkan pada Tabel 7.

Tabel 7. Hasil *Labeling*

Normalisasi	Label
baik paham dokter trima kasih banyak dokter jawab nya	Positif
ken bawa anak yamet tapi kendala biaya	Netral
terimakasih banyak dokter atas jawab jawab lama masalah saya	Negatif
adhd contoh apa dokter	Netral
dokter kisar biaya konsultasi kira brpa dokter	Netral

Tahapan selanjutnya ekstraksi fitur yang dilakukan dengan metode TF – IDF proses ini menghasilkan matriks berukuran 1.831×1.000 , yang merepresentasikan 1.831 komentar valid dan 1.000 fitur unik. Cuplikan hasil bobot TF-IDF ditampilkan pada Tabel 8.

Tabel 8. Hasil *Extraction Features*

No.	Kata Kunci	Komentar ke - 1	Komentar ke - 2	Komentar ke - 3
1.	terapi	0.214	0.185	0.267
2.	anak	0.187	0.204	0.156
3.	yamet	0.225	0.198	0.210
4.	bagus	0.167	0.000	0.174
5.	bicara	0.000	0.195	0.000

Tahap klasifikasi menggunakan *Multinomial Naïve Bayes* dengan dataset hasil ekstraksi fitur TF – IDF yang dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan *train-test split*. Pada tahap ini hasil menunjukkan bahwa model cenderung memprediksi komentar dalam kelas mayoritas, yaitu netral, sementara kelas positif dan negatif tidak terdeteksi sama sekali hal ini terjadi karena adanya ketidakseimbangan data (*imbalanced dataset*). Nilai akurasi yang diperoleh sebesar 86,56%, namun tidak mempresentasikan performa sebenarnya, lebih detail bisa dilihat pada hasil klasifikasi yang ditampilkan pada Tabel 9.

Tabel 9. Hasil Klasifikasi *Naïve Bayes*

Label Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.00	0.00	0.00	15
Netral	0.87	1.00	0.93	277
Positif	0.00	0.00	0.00	28
Macro avg	0.29	0.33	0.31	320
Akurasi Model			86.56%	320

Untuk mengatasi permasalahan tersebut, dilakukan penyeimbangan data menggunakan metode SMOTE (*Synthetic Minority Oversampling Technique*). Setelah penerapan SMOTE, jumlah data pada masing-masing kelas menjadi seimbang, yaitu 1.100 data untuk setiap kategori sentimen, seperti ditunjukkan pada Tabel 10.

Tabel 10. Hasil Perbandingan penyeimbangan data

Label Sentimen	Jumlah Sebelum SMOTE	Jumlah Sesudah SMOTE
Netral	1.375	1.100
Positif	137	1.100
Negatif	86	1.100

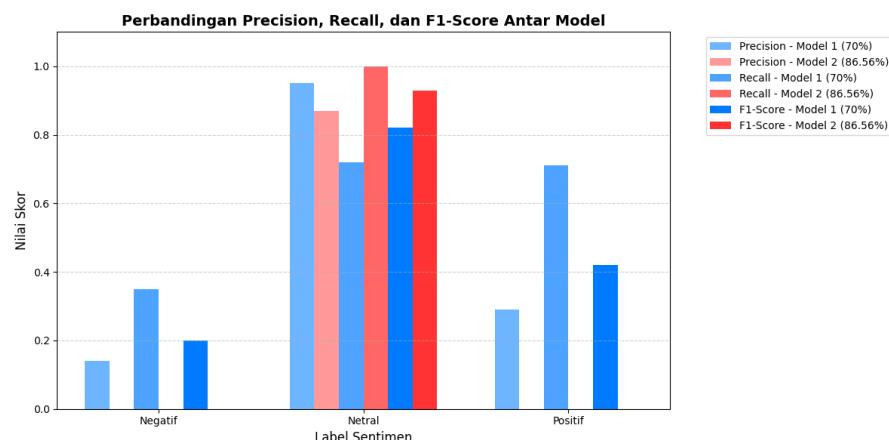
Setelah dilakukan penyeimbangan data, model dilatih ulang dengan algoritma yang sama. Hasil pengujian menunjukkan peningkatan performa terutama pada kelas minoritas, sebagaimana ditampilkan pada Tabel 11.

Tabel 11. Hasil Klasifikasi *Naïve Bayes* setelah penerapan SMOTE

Label Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.14	0.35	0.20	17
Netral	0.95	0.72	0.82	275
Positif	0.29	0.71	0.42	28
Macro avg	0.46	0.60	0.48	320
Akurasi Model		70.00%		320

Perubahan signifikan terjadi pada kelas positif dan negatif, di mana nilai *recall* meningkat secara drastis menjadi 0.71 dan 0.35. Meskipun akurasi model menurun ini menunjukkan performa yang lebih proporsional antar kelas. Penurunan akurasi ini merupakan konsekuensi dari peningkatan kemampuan model dalam mengenali variasi kelas minoritas, bukan karena penurunan kualitas prediksi.

Evaluasi dilakukan untuk menilai kinerja algoritma *Multinomial Naïve Bayes* sebelum dan sesudah penerapan SMOTE. Sebelum penyeimbangan data, model memperoleh akurasi 86,56%. Setelah diterapkan SMOTE, distribusi tiap kelas menjadi seimbang (1.100 data per kelas) dan performa model terhadap kelas minoritas meningkat seperti yang ditampilkan pada Tabel 11, meskipun hasil akurasinya menurun menjadi 70%. Grafik visualisasi perbandingan klasifikasi sebelum dan sesudah penerapan SMOTE ditampilkan pada Gambar 2.



Gambar 2 Grafik Perbandingan Klasifikasi

Hasil menunjukkan bahwa penerapan SMOTE berhasil meningkatkan sensitivitas model terhadap kelas minoritas dan menghasilkan klasifikasi sentimen yang lebih proporsional. Penurunan akurasi bukan menandakan penurunan performa, melainkan peningkatan kemampuan model dalam mengenali variasi sentimen secara seimbang. Secara keseluruhan, kombinasi TF-IDF, *Naïve Bayes*, dan SMOTE efektif dalam menganalisis sentimen publik terhadap layanan Klinik Yamet Palembang di media sosial Instagram.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa analisis sentimen terhadap komentar publik pada akun Instagram Klinik Yamet Palembang dapat dilakukan secara efektif menggunakan pendekatan *Naïve Bayes* dengan proses prapemrosesan teks dan pembobotan TF-IDF. Model yang dibangun mampu mengklasifikasikan opini masyarakat ke dalam tiga kategori utama, yaitu positif, netral, dan negatif, dengan tingkat akurasi sebesar 86,56% sebelum penerapan SMOTE. Namun, setelah dilakukan penyeimbangan data menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE), tingkat akurasi model mengalami penurunan menjadi 70,00%. Hal ini menunjukkan bahwa penerapan SMOTE tidak selalu meningkatkan performa algoritma *Naïve Bayes*, terutama pada data dengan karakteristik teks pendek dan distribusi fitur yang tidak linier. Temuan ini menjadi bukti empiris bahwa efektivitas metode *oversampling* perlu disesuaikan dengan karakteristik data dan jenis model klasifikasi yang digunakan. Dengan demikian, masalah penelitian telah terjawab: kombinasi *Naïve Bayes* dan TF-IDF mampu menganalisis sentimen publik secara akurat, namun penambahan SMOTE tidak meningkatkan performa model pada konteks dataset penelitian ini. Secara praktis, hasil penelitian ini dapat menjadi dasar bagi Klinik Yamet Palembang untuk memahami persepsi masyarakat di media sosial dan menyusun strategi komunikasi digital yang lebih adaptif serta berbasis data.

REFERENSI

- [1] Alviandy Ayu Rizka, Arvian Rully Shenia, and Wasir Riseandy, "LITERATURE REVIEW: ANALISIS TRANSFORMASI DIGITAL LAYANAN KESEHATAN DI INDONESIA TERHADAP PERAN SEKTOR PUBLIK DAN SWASTA," Jakarta, Jun. 2025.
- [2] Asosiasi Penyelenggaraan Jasa Internet Indonesia, "Laporan Survei Internet Indonesia 2024." Accessed: Sep. 24, 2025. [Online]. Available: <https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang>
- [3] We Are Social, "Digital 2024: 5 billion social media users." Accessed: Sep. 24, 2025. [Online]. Available: <https://wearesocial.com/id/blog/2024/01/digital-2024-5-billion-social-media-users/>
- [4] I. B. Setiawan, J. Maulindar, and N. Nurchim, "Penerapan Algoritma Naïve Bayes untuk Analisis Sentimen Pada Aplikasi Kesehatan Digital," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 4, pp. 2301–2312, Oct. 2024, doi: 10.70609/gtech.v8i4.5020.
- [5] A. Lia, A. Rahim, and T. A. Yoga Siswa, "ANALISIS SENTIMEN APLIKASI MYSILOAM MENGGUNAKAN METODE NAÏVE BAYES," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5997.
- [6] M. Usman *et al.*, "Analyzing patients satisfaction level for medical services using twitter data," *PeerJ Comput Sci*, vol. 10, 2024, doi: 10.7717/peerj-cs.1697.
- [7] IG Comment Export Tool, "IG Comment Export Tool - Export Instagram Comment to CVS/Excel."
- [8] S. Khomsah and A. S. Aribowo, "Terakreditasi SINTA Peringkat 2 Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia," *masa berlaku mulai*, vol. 1, no. 3, pp. 648–654, 2017.
- [9] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [10] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, "Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 4, p. 717, Aug. 2022, doi: 10.25126/jtiik.2022934707.
- [11] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00413-1.
- [12] A. N. A. Saputra, R. E. Saputro, and D. I. S. Saputra, "Enhancing Sentiment Analysis Accuracy Using SVM and Slang Word Normalization on YouTube Comments," *Sinkron*, vol. 9, no. 2, pp. 687–699, Apr. 2025, doi: 10.33395/sinkron.v9i2.14613.
- [13] A. Hasan, Y. R. Ramadhan, and M. Minarto, "Sentiment Analysis of Telemedicine Applications on Twitter Using Lexicon-Based and Naïve Bayes Classifier Methods," *Jurnal Riset Informatika*, vol. 5, no. 4, pp. 481–490, Sep. 2023, doi: 10.34288/jri.v5i4.244.
- [14] A. Ardi and Kurniawan, "Optimasi Metode Naïve Bayes Classifier Menggunakan Pendekatan Term Frequency-Inverse Document Frequency (TF-IDF) Pada Analisis Sentimen," *JSAI (Journal Scientific and Applied Informatics)*, vol. 7, no. 3, pp. 458–463, Nov. 2024, doi: 10.36085/jsai.v7i3.7153.

- [15] Y. Hapsari, S. Mujahidin, and N. Fadhlina, “Analisis Sentimen Isu Vaksinasi Covid-19 pada Twitter dengan Metode Naive Bayes dan Pembobotan TF-IDF Tokenisasi 1-2 Gram,” *SPECTA Journal of Technology*, vol. 7, no. 2, pp. 573–583, Aug. 2023, doi: 10.35718/specta.v7i2.812.
- [16] L. Hartimar, Y. Manza, and K. Putriani Siregar, “Text Classification Using TF-IDF and Naïve Bayes: Case Study of MyXL App User Review Data,” *Journal of Technology and Computer (JOTECHCOM)*, vol. 2, no. 2, pp. 100–108, 2025.
- [17] S. Khoerunnisa, D. F. Shiddiq, and D. Nurhayati, “Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 2, pp. 566–577, Mar. 2025, doi: 10.57152/malcom.v5i2.1852.
- [18] W. Trisnawati and A. Wibowo, “SENTIMENT ANALYSIS OF ICT SERVICE USER USING NAIVE BAYES CLASSIFIER AND SVM METHODS WITH TF-IDF TEXT WEIGHTING,” *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 3, pp. 709–719, May 2024, doi: 10.52436/1.jutif.2024.5.3.1784.
- [19] A. F. Subowo Moh Hadi, “Moneter: Jurnal Keuangan Dan Perbankan Akreditasi SINTA 3 and COPERNICUS,” Jul. 2024.
- [20] G. Naidu, T. Zuva, and E. M. Sibanda, “A Review of Evaluation Metrics in Machine Learning Algorithms,” 2023, pp. 15–25. doi: 10.1007/978-3-031-35314-7_2.
- [21] Y. Aprianti, A. Lia Hananto, S. Shofiah Hilabi, S. Informasi, and U. Buana Perjuangan Karawang, “Klasifikasi Sentimen Komentar Pengguna pada Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes,” vol. 9, p. 2025, Jun. 2025, doi: 10.47002/metik.v9i1.1023.