

PENERAPAN PENDEKATAN *HYBRID NAÏVE BAYES* DAN *INDOBERT* UNTUK ANALISIS SENTIMEN ULASAN GOOGLE MAPS PADA RUMAH SAKIT JIH SOLO

Wahyu Agung Adi Nugroho, Dwi Hartanti, Afu Ichsan Pradana

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara No.55, tipes, Serengan, Surakarta, Jawa Tengah, Indonesia

220103109@mhs.udb.ac.id

ABSTRAK

Perkembangan platform digital seperti Google Maps memungkinkan masyarakat memberikan ulasan terhadap layanan rumah sakit secara daring. Ulasan tersebut dapat dimanfaatkan untuk mengetahui persepsi masyarakat terhadap kualitas layanan kesehatan. Namun, jumlah ulasan yang besar, data yang tidak terstruktur, serta adanya ketidakseimbangan distribusi kelas sentimen menyebabkan proses analisis menjadi lebih kompleks. Penelitian ini bertujuan membangun model analisis sentimen pada ulasan Google Maps Rumah Sakit JIH Solo menggunakan pendekatan *hybrid IndoBERT* dan *Multinomial Naïve Bayes*. Data penelitian berupa 2.432 ulasan yang dikumpulkan melalui teknik *web scraping* pada periode 2022–2026. *IndoBERT* digunakan untuk melabeli ulasan yang ambigu, sedangkan *Multinomial Naïve Bayes* digunakan sebagai model klasifikasi. Data kemudian melalui tahap *preprocessing* dan transformasi TF-IDF. Untuk mengatasi ketidakseimbangan data, diterapkan teknik *Random Over Sampling* (ROS) dan *Synthetic Minority Oversampling Technique* (SMOTE). Hasil pengujian menunjukkan bahwa model *hybrid* memperoleh akurasi sebesar 94% dengan nilai *recall* sentimen negatif mencapai 0,89. Nilai tersebut lebih tinggi dibandingkan model *Naïve Bayes* tanpa pendekatan *hybrid* yang memperoleh *recall* sebesar 0,83. Hasil penelitian menunjukkan bahwa kombinasi *IndoBERT*, *Naïve Bayes*, dan teknik *balancing* data mampu meningkatkan kinerja klasifikasi sentimen pada data yang tidak seimbang.

Kata kunci : Analisis Sentimen, *IndoBERT*, *Naïve Bayes*, Google Maps, SMOTE, ROS

1. PENDAHULUAN

Transformasi digital yang berkembang pesat telah mengubah pola masyarakat dalam menyampaikan pengalaman dan penilaian terhadap berbagai layanan, termasuk sektor kesehatan. Salah satu platform yang banyak dimanfaatkan untuk tujuan tersebut adalah Google Maps. Melalui fitur ulasan yang tersedia, pengguna dapat membagikan pengalaman mereka setelah menerima pelayanan di rumah sakit. Informasi yang terkandung dalam ulasan tersebut dapat menjadi sumber masukan yang berharga karena mencerminkan pengalaman pasien terkait mutu pelayanan, kelengkapan fasilitas, maupun kualitas interaksi dengan tenaga kesehatan. Meskipun demikian, peningkatan jumlah ulasan dari waktu ke waktu menyebabkan proses pengkajian secara konvensional menjadi kurang efisien. Sejumlah penelitian menunjukkan bahwa ulasan daring memiliki nilai strategis sebagai bahan evaluasi kualitas layanan kesehatan apabila diolah menggunakan pendekatan komputasional yang tepat [1]. Selain itu, ulasan yang diberikan pengguna juga dapat menggambarkan persepsi publik terhadap layanan rumah sakit secara lebih aktual [2].

Pengolahan sentimen pada data ulasan menghadapi berbagai tantangan. Karakteristik teks yang bebas, panjang, dan sering kali mengandung makna ganda menyulitkan proses interpretasi secara otomatis. Kondisi tersebut banyak ditemukan pada ulasan dengan rating menengah yang tidak secara jelas menunjukkan kecenderungan sentimen positif maupun negatif. Di samping itu, isi ulasan tidak selalu selaras dengan skor rating yang diberikan sehingga pelabelan

berdasarkan rating saja berisiko menghasilkan kategori sentimen yang kurang tepat. Beberapa studi mengungkapkan bahwa kemampuan memahami konteks bahasa merupakan aspek penting dalam meningkatkan kualitas klasifikasi sentimen [3]. Permasalahan lain yang umum dijumpai pada teks berbahasa Indonesia meliputi penggunaan kata tidak baku, singkatan, bahasa sehari-hari, serta keberadaan opini yang bercampur dalam satu ulasan [4]. Sementara itu, algoritma *Naïve Bayes* dikenal memiliki keterbatasan dalam memahami keterkaitan makna antar kata, sedangkan model berbasis transformer seperti *IndoBERT* membutuhkan sumber daya komputasi yang lebih tinggi. Oleh karena itu, diperlukan strategi yang dapat memanfaatkan kelebihan kedua metode tersebut secara bersamaan [5].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan menerapkan pendekatan *hybrid* antara *IndoBERT* dan *Multinomial Naïve Bayes* untuk analisis sentimen ulasan Google Maps pada Rumah Sakit JIH Solo. Pendekatan ini diharapkan mampu meningkatkan kualitas klasifikasi sentimen melalui pemanfaatan *IndoBERT* dalam memahami konteks bahasa yang ambigu, sekaligus mempertahankan efisiensi komputasi melalui penggunaan *Multinomial Naïve Bayes* sebagai model klasifikasi utama. Selain itu, penelitian ini bertujuan memberikan gambaran yang lebih komprehensif mengenai persepsi masyarakat terhadap kualitas layanan Rumah Sakit JIH Solo berdasarkan ulasan yang tersedia pada Google Maps. Pendekatan *hybrid* dipilih karena dapat menggabungkan kemampuan pemahaman konteks

dari model transformer dengan efisiensi komputasi yang dimiliki metode klasifikasi tradisional [6].

Untuk mencapai tujuan tersebut, penelitian dilakukan melalui beberapa tahapan yang terstruktur. Data ulasan dikumpulkan melalui proses *scraping* pada Google Maps, kemudian dilakukan pelabelan awal berdasarkan nilai rating. Ulasan yang memiliki kecenderungan ambigu diproses lebih lanjut menggunakan *IndoBERT* untuk menentukan label sentimennya. Data yang telah dilabeli selanjutnya melalui tahap *preprocessing* yang meliputi *case folding*, *cleansing*, *normalisasi*, *tokenizing*, *stopword removal*, dan *stemming*. Setelah itu dilakukan pembobotan fitur menggunakan TF-IDF sebelum proses klasifikasi menggunakan algoritma *Multinomial Naïve Bayes*. Penelitian ini juga menerapkan teknik *Random Over Sampling* (ROS) dan *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan distribusi data. Kinerja model dievaluasi menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* sehingga dapat diperoleh gambaran mengenai efektivitas pendekatan *hybrid* yang diusulkan dalam menganalisis sentimen ulasan pengguna.

2. TINJAUAN PUSTAKA

Kajian literatur dilakukan untuk memperoleh gambaran mengenai perkembangan metode analisis sentimen sekaligus mengidentifikasi pendekatan yang paling sesuai untuk diterapkan dalam penelitian ini.

Beberapa penelitian terdahulu telah memanfaatkan algoritma *Naïve Bayes* dalam pengolahan sentimen pada ulasan layanan kesehatan. Salah satunya adalah penelitian yang dilakukan oleh Rizki, Prihartono, dan Fathurrohman pada data ulasan Google Maps Rumah Sakit Khalishah Cirebon. Hasil evaluasi menunjukkan bahwa model yang digunakan mampu mencapai tingkat kinerja yang cukup baik, ditunjukkan oleh nilai akurasi 84%, *precision* 89%, *recall* 91%, serta *F1-score* 90%. Meskipun demikian, proses pemberian label masih bergantung pada rating pengguna sehingga interpretasi terhadap ulasan yang memiliki makna ganda belum dapat dilakukan secara optimal [7].

Penelitian lain yang dilakukan oleh Dzulkarnain, Ratnawati, dan Rahayudi juga menerapkan pendekatan *Naïve Bayes* untuk mengelompokkan opini masyarakat mengenai layanan rumah sakit di Kota Malang. Metode tersebut berhasil menghasilkan tingkat akurasi hingga 90% serta mampu mengidentifikasi sentimen berdasarkan aspek pelayanan tertentu, seperti fasilitas, administrasi, dan biaya. Namun demikian, model yang digunakan masih mengalami kendala dalam menangkap hubungan makna dan konteks yang terkandung dalam kalimat [8].

Perkembangan metode berbasis transformer kemudian menunjukkan hasil yang lebih menjanjikan. Mareta menggunakan model *Fine-Tuned IndoBERT*

untuk melakukan *aspect-based sentiment analysis* pada ulasan rumah sakit. Temuan penelitian tersebut memperlihatkan tingkat performa yang tinggi dengan nilai akurasi mencapai 96% dan *F1-score* sebesar 90%. Walaupun memberikan hasil klasifikasi yang sangat baik, penerapan *IndoBERT* pada keseluruhan data memerlukan kapasitas komputasi yang relatif besar sehingga kurang efisien untuk beberapa kondisi penelitian [9].

Keunggulan model *transformer* juga ditunjukkan dalam penelitian Apriyadi dan Styawati yang membandingkan algoritma *Support Vector Machine* (SVM) dengan *IndoBERT* pada kasus analisis sentimen mengenai serangan siber terhadap Bank Syariah Indonesia. Berdasarkan hasil pengujian, *IndoBERT* menghasilkan performa yang lebih unggul dengan akurasi sebesar 85% dan *F1-score* 82%. Penelitian tersebut turut memanfaatkan metode *Latent Dirichlet Allocation* (LDA) untuk mengekstraksi dan mengelompokkan tema-tema utama yang muncul dalam sentimen negatif [10].

Berdasarkan berbagai penelitian tersebut, dapat disimpulkan bahwa *Naïve Bayes* menawarkan kelebihan dari sisi kecepatan pemrosesan dan efisiensi penggunaan sumber daya, sedangkan *IndoBERT* lebih unggul dalam memahami konteks linguistik dan makna semantik bahasa Indonesia. Meskipun demikian, sebagian besar penelitian masih menerapkan kedua metode tersebut secara terpisah. Oleh sebab itu, penelitian ini mengembangkan strategi kombinasi antara *IndoBERT* dan *Multinomial Naïve Bayes* dengan tujuan meningkatkan ketepatan klasifikasi pada ulasan yang mengandung ambiguitas sekaligus mempertahankan efisiensi proses komputasi.

2.1. Analisis Sentimen

Analisis sentimen merupakan salah satu cabang kajian dalam bidang *Natural Language Processing* (NLP) yang berfokus pada proses pengenalan serta pengelompokan pendapat atau ekspresi yang terdapat dalam data teks. Melalui pendekatan ini, suatu pernyataan dapat dikategorikan ke dalam kelas tertentu, misalnya sentimen positif, negatif, maupun netral. Pada sektor layanan kesehatan, penerapan analisis sentimen dapat dimanfaatkan untuk menggali pandangan dan tingkat kepuasan masyarakat terhadap pelayanan yang diterima berdasarkan ulasan yang dipublikasikan secara online. Meskipun demikian, keberadaan kata tidak baku, singkatan, maupun kalimat yang memiliki lebih dari satu interpretasi sering kali menjadi tantangan dalam proses klasifikasi sentimen [2] [4].

2.2. Text Preprocessing

Sebelum data teks digunakan dalam proses analisis, diperlukan serangkaian tahapan persiapan data yang dikenal sebagai *text preprocessing*. Tahap ini bertujuan meningkatkan kualitas data dengan menghilangkan unsur-unsur yang tidak diperlukan serta menyeragamkan bentuk teks agar lebih mudah

diolah oleh sistem. Pada penelitian ini, proses yang dilakukan mencakup konversi seluruh karakter menjadi huruf kecil (*case folding*), pembersihan simbol, angka, dan karakter yang tidak relevan (*cleansing*), serta normalisasi kata untuk mengubah istilah tidak baku menjadi bentuk yang sesuai dengan kaidah bahasa. Setelah itu dilakukan pemisahan kalimat menjadi unit-unit kata (*tokenisasi*), penghapusan kata-kata umum yang tidak memberikan informasi penting (*stopword removal*), dan reduksi kata ke bentuk dasarnya melalui *stemming*. Keberhasilan tahapan ini sangat menentukan representasi fitur yang dihasilkan dan berdampak pada tingkat kinerja model klasifikasi yang digunakan [11].

2.3. Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen. Dalam klasifikasi teks, algoritma ini menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu berdasarkan frekuensi kemunculan kata. Keunggulan utama metode ini terletak pada kesederhanaan dan efisiensi komputasinya, sehingga cocok digunakan untuk data berdimensi tinggi. Namun, asumsi independensi antar kata menjadi keterbatasan karena tidak mampu menangkap hubungan konteks secara menyeluruh [12] [13].

2.4. IndoBERT

IndoBERT adalah model representasi bahasa Indonesia yang dikembangkan menggunakan teknologi transformer dan dilatih pada kumpulan data teks berbahasa Indonesia dalam jumlah besar. Berbeda dengan pendekatan tradisional yang memproses kata secara terpisah, *IndoBERT* mampu mempertimbangkan informasi dari kata-kata sebelum dan sesudahnya secara bersamaan ketika memahami suatu kalimat. Kemampuan tersebut memungkinkan model memperoleh pemahaman yang lebih baik terhadap konteks dan makna yang terkandung dalam teks. Oleh karena itu, *IndoBERT* banyak dimanfaatkan pada berbagai tugas NLP, termasuk analisis sentimen, terutama ketika berhadapan dengan kalimat yang kompleks atau mengandung ambiguitas. Sejumlah penelitian menunjukkan bahwa model ini mampu menghasilkan performa klasifikasi yang lebih tinggi dibandingkan berbagai pendekatan konvensional [6] [14].

2.5. Machine Learning

Machine learning merupakan salah satu bidang dalam kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan sistem yang mampu memperoleh pengetahuan dari data tanpa harus diprogram secara eksplisit untuk setiap tugas. Dengan memanfaatkan data historis, model dapat mempelajari pola tertentu dan menggunakannya sebagai dasar dalam menghasilkan prediksi maupun pengambilan keputusan secara otomatis.

Pada penelitian analisis sentimen, pendekatan yang banyak digunakan adalah *supervised learning*. Dalam metode ini, model dibangun menggunakan data pelatihan yang telah memiliki label kelas sehingga sistem dapat mempelajari hubungan antara karakteristik data dan kategori yang sesuai. Setelah proses pelatihan selesai, model dapat diterapkan untuk menentukan kecenderungan sentimen pada data yang belum pernah dianalisis sebelumnya dengan lebih cepat dan konsisten [15] [16].

2.6. Evaluasi Model

Tahap evaluasi dilakukan untuk mengetahui tingkat keberhasilan model dalam melakukan klasifikasi sentimen. Pengukuran performa pada penelitian ini menggunakan *confusion matrix* yang kemudian menghasilkan beberapa metrik penilaian, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion matrix* berfungsi sebagai alat untuk membandingkan hasil prediksi model dengan label aktual sehingga tingkat kesalahan maupun ketepatan klasifikasi dapat dianalisis secara lebih rinci [17].

Accuracy menggambarkan proporsi prediksi yang benar terhadap seluruh data yang diuji. *Precision* menunjukkan seberapa tepat model ketika memberikan prediksi pada suatu kelas tertentu. Sementara itu, *recall* digunakan untuk melihat kemampuan model dalam menemukan seluruh data yang benar-benar termasuk ke dalam suatu kelas. Adapun *F1-score* merupakan kombinasi antara *precision* dan *recall* yang dihitung menggunakan rata-rata harmonik, sehingga lebih representatif dalam menilai model pada data yang distribusi kelasnya tidak seimbang [18].

Persamaan yang digunakan untuk menghitung setiap metrik evaluasi ditunjukkan sebagai berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Keterangan :

- TP (*True Positive*) : jumlah data yang termasuk kelas positif dan berhasil diprediksi sebagai positif oleh model.
- TN (*True Negative*) : jumlah data yang termasuk kelas negatif dan berhasil dikenali sebagai negatif.
- FP (*False Positive*) : menunjukkan data yang sebenarnya negatif tetapi diprediksi sebagai positif.
- FN (*False Negative*) : menunjukkan data yang sebenarnya positif namun diklasifikasikan sebagai negatif

3. METODE PENELITIAN

3.1. Jenis dan Sumber Data

Penelitian ini memanfaatkan dua kelompok data, yaitu data utama dan data pendukung. Data utama berasal dari ulasan pengguna terhadap layanan Rumah Sakit JIH Solo yang tersedia pada platform Google Maps. Informasi yang dikumpulkan mencakup isi komentar pengguna beserta skor penilaian yang diberikan dalam rentang satu hingga lima bintang. Data tersebut menjadi sumber utama dalam proses pengolahan dan klasifikasi sentimen.

Selain itu, penelitian juga menggunakan data pendukung yang diperoleh dari berbagai referensi ilmiah, seperti artikel jurnal, prosiding konferensi, buku, dan sumber akademik lainnya. Referensi tersebut digunakan sebagai dasar konseptual serta acuan dalam penyusunan metode penelitian.

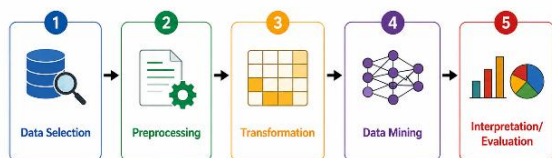
3.2. Metode Pengumpulan Data

Data penelitian diperoleh melalui kombinasi pendekatan studi pustaka dan observasi. Studi pustaka dilakukan dengan menelaah berbagai publikasi ilmiah yang membahas analisis sentimen, khususnya penelitian yang menerapkan algoritma *Naïve Bayes* maupun model berbasis *transformer* seperti *IndoBERT*.

Sementara itu, observasi dilakukan dengan mengambil data ulasan secara langsung dari Google Maps melalui teknik web *scraping*. Teknik ini memungkinkan proses pengambilan data dilakukan secara otomatis dalam jumlah besar. Informasi yang berhasil dikumpulkan meliputi isi ulasan, nilai bintang, serta atribut pendukung lainnya. Seluruh data hasil ekstraksi kemudian disimpan dalam format terstruktur agar lebih mudah digunakan pada tahapan pengolahan berikutnya.

3.3. Tahapan Penelitian

Penelitian ini mengadopsi kerangka kerja *Knowledge Discovery in Databases* (KDD) sebagai pedoman dalam proses pengolahan data. Kerangka tersebut terdiri atas beberapa tahapan yang saling berkaitan mulai dari pengumpulan data hingga evaluasi hasil model. Alur tahapan penelitian ditunjukkan pada Gambar 1



Gambar 1. Tahapan penelitian berbasis KDD

Berikut penjabaran dari metode yang akan digunakan :

a. Data Selection (Pemilihan Data)

Tahap pertama difokuskan pada pengambilan data ulasan pengguna Rumah Sakit JIH Solo dari

Google Maps. Data yang berhasil dikumpulkan berupa komentar pengguna dan nilai rating yang menyertainya. Setelah proses pengumpulan selesai, dilakukan pemberian label awal berdasarkan skor rating. Ulasan dengan rating rendah dimasukkan ke kategori negatif, sedangkan rating tinggi dikelompokkan sebagai sentimen positif. Adapun ulasan dengan rating menengah dianggap memiliki kecenderungan ambigu sehingga memerlukan proses identifikasi lanjutan menggunakan *IndoBERT*.

b. Preprocessing (Pra-pemrosesan Data)

Data mentah yang diperoleh dari proses scraping umumnya masih mengandung berbagai elemen yang dapat mengganggu proses analisis. Oleh karena itu, dilakukan serangkaian tahapan pembersihan dan penyeragaman data. Tahapan tersebut meliputi konversi huruf ke format lowercase, penghapusan karakter yang tidak relevan, perbaikan kata tidak baku menjadi bentuk standar, tokenisasi, penghilangan kata-kata umum yang tidak memberikan informasi penting, serta reduksi kata ke bentuk dasarnya. Melalui proses ini, data teks menjadi lebih terstruktur dan siap digunakan pada tahap ekstraksi fitur.

c. Transformation (Transformasi Data)

Setelah melalui tahap pra-pemrosesan, data teks dikonversi ke dalam representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Transformasi ini bertujuan menghasilkan bobot fitur yang mampu menggambarkan tingkat kepentingan suatu kata dalam dokumen sehingga dapat diproses oleh model machine learning.

d. Data Mining (Pemodelan)

Pada tahap ini dilakukan pembangunan model klasifikasi menggunakan algoritma *Multinomial Naïve Bayes*. Sebelum proses pelatihan dimulai, dataset dibagi menjadi data pelatihan dan data pengujian dengan proporsi 70% dan 30%. Data pelatihan digunakan untuk membentuk model, sedangkan data pengujian dimanfaatkan untuk mengukur kemampuan model dalam mengklasifikasikan sentimen.

Pengujian dilakukan pada tiga kondisi dataset yang berbeda, yaitu data asli tanpa penyeimbangan, data yang telah diproses menggunakan *Random Over Sampling* (ROS), serta data yang diseimbangkan menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). Perbandingan ketiga skenario tersebut digunakan untuk mengetahui pengaruh teknik balancing terhadap hasil klasifikasi.

e. Interpretation/Evaluation (Evaluasi dan Interpretasi)

Tahap akhir difokuskan pada pengukuran kinerja model yang telah dibangun. Penilaian dilakukan menggunakan confusion matrix dengan indikator *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai yang diperoleh dari setiap metrik digunakan untuk

menilai tingkat keberhasilan model dalam melakukan klasifikasi sentimen.

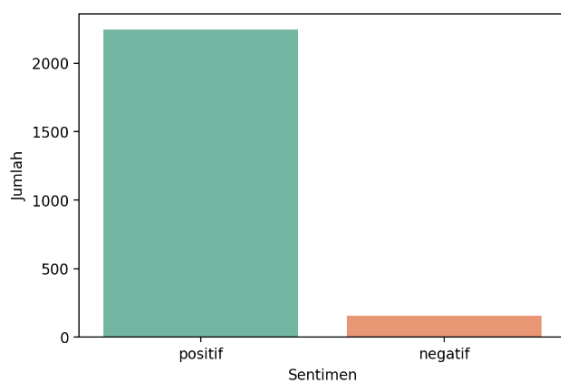
Selanjutnya, hasil pengujian diolah dalam bentuk analisis dan visualisasi sehingga pola distribusi sentimen dapat diamati dengan lebih mudah. Melalui tahap ini, dapat diperoleh gambaran mengenai kecenderungan opini masyarakat terhadap kualitas layanan yang diberikan oleh Rumah Sakit JIH Solo.

4. HASIL DAN PEMBAHASAN

4.1. Dataset dan Distribusi Data

Dataset yang digunakan pada penelitian ini diperoleh melalui proses *scraping* ulasan Google Maps Rumah Sakit JIH Solo menggunakan platform Apify pada rentang tahun 2022–2026. Data yang diperoleh berupa teks ulasan dan rating pengguna dengan total 2432 data. Selanjutnya, data diproses untuk menghasilkan dua kelas sentimen, yaitu positif dan negatif.

Berdasarkan hasil eksplorasi dataset, terlihat adanya perbedaan proporsi yang cukup mencolok antara kedua kategori sentimen. Jumlah ulasan dengan label positif jauh lebih banyak dibandingkan ulasan berlabel negatif. Ketimpangan distribusi tersebut berpotensi menyebabkan model lebih sering memprediksi kelas yang memiliki jumlah data terbesar sehingga kemampuan identifikasi terhadap kelas dengan jumlah data lebih sedikit menjadi berkurang. Untuk mengatasi kondisi tersebut, diterapkan dua metode penyeimbangan data, yaitu *Random Over Sampling* (ROS) dan *Synthetic Minority Oversampling Technique* (SMOTE), sebelum tahapan pelatihan model dilaksanakan.



Gambar 2. Distribusi dataset sentimen sebelum proses *balancing*

4.2. Hasil Preprocessing Data

Tahap *preprocessing* dilakukan untuk membersihkan dan menormalisasi data teks agar siap diproses pada tahap klasifikasi. Proses yang dilakukan meliputi *case folding*, *cleansing*, normalisasi, *tokenizing*, *stopword*, dan *stemming*. Tahapan ini bertujuan untuk mengurangi *noise* pada data serta meningkatkan kualitas representasi teks.

Hasil *preprocessing* menunjukkan bahwa data teks menjadi lebih konsisten dan terstruktur, sehingga

dapat meningkatkan efektivitas proses transformasi fitur menggunakan TF-IDF.

Tabel 1. Contoh Hasil Preprocessing Data

Tahap	Sebelum	Sesudah
<i>Case Folding</i>	Petugas ramah, ruang tunggu nyaman,	petugas ramah, ruang tunggu nyaman,
<i>Cleansing</i>	petugas ramah, ruang tunggu nyaman,	petugas ramah ruang tunggu nyaman
<i>Normalisasi</i>	petugas ramah ruang tunggu nyaman	petugas ramah ruang tunggu nyaman
<i>Tokenizing</i>	petugas ramah ruang tunggu nyaman	[petugas, ramah, ruang, tunggu, nyaman]
<i>Stopword Removal</i>	[petugas, ramah, ruang, tunggu, nyaman,]	[petugas, ramah, ruang, tunggu, nyaman,]
<i>Stemming</i>	[petugas, ramah, ruang, tunggu, nyaman]	[tugas, ramah, ruang, tunggu, nyaman]

4.3. Pelabelan Data Menggunakan IndoBERT

Pada penelitian ini, ulasan dengan rating bintang 3 dikategorikan sebagai data ambigu karena tidak secara langsung menunjukkan sentimen positif maupun negatif. Untuk mengatasi hal tersebut, *IndoBERT* digunakan untuk melakukan pelabelan ulang berdasarkan konteks kalimat pada ulasan.

Pendekatan ini menjadi bagian utama dari metode *hybrid* yang diterapkan karena *IndoBERT* mampu memahami konteks semantik bahasa Indonesia secara lebih baik dibandingkan pelabelan berbasis rating. Hasil pelabelan kemudian digabungkan kembali ke dalam dataset sebelum dilakukan proses klasifikasi menggunakan *Multinomial Naïve Bayes*.

ulasan	sentimen_baru
5 Alasan utama berobat ke rs JIH karna Dokternya. Kami sekeluarga adalah pasien Dokter Di	negatif
6 Saya operasi tgl 3 april 2026, biasanya setelah rawat inap ada WA pemberitahuan untuk ko	negatif
66 kaget banget di seprai kenapa ada serangga yaa (semut terbang) padahal harusnya kebersi	negatif
102 Pembayaran di farmasi lama pol.	negatif
147 Saya dan mantu (baru tadi malam pulang). Kekurangan jih solo y! saat patien opname sud	negatif

Gambar 3. Proses pelabelan data ambigu menggunakan *IndoBERT*

Tabel 2. Hasil Pelabelan Data Ambigu

Label sentimen	Jumlah data
Positif	0
Negatif	34

Berdasarkan hasil pelabelan pada Tabel 2, data ulasan ambigu cenderung mengandung konteks keluhan atau ketidakpuasan terhadap layanan sehingga seluruh data ambigu diklasifikasikan ke dalam sentimen negatif. Hasil tersebut menunjukkan bahwa pendekatan berbasis konteks menggunakan *IndoBERT* mampu mengidentifikasi kecenderungan sentimen secara lebih mendalam dibandingkan pelabelan berbasis rating saja.

4.4. Hasil Transformasi TF-IDF

Setelah seluruh tahapan prapemrosesan selesai dilakukan, data teks direpresentasikan ke dalam bentuk numerik menggunakan pendekatan TF-IDF (*Term Frequency–Inverse Document Frequency*). Teknik ini memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam dokumen dan tingkat keunikannya pada keseluruhan kumpulan dokumen. Dengan demikian, kata yang dianggap lebih informatif akan memperoleh bobot yang lebih tinggi dibandingkan kata yang sering muncul pada banyak dokumen.

Perhitungan bobot TF-IDF menggunakan persamaan berikut:

$$TF - IDF(t, d) = TF(t, d) \times \log \frac{N}{df(t)} \quad (5)$$

Keterangan :

- TF(t,d) : frekuensi kemunculan suatu istilah pada dokumen tertentu.
- N : total dokumen yang digunakan dalam dataset.
- df(t) : jumlah dokumen yang memuat istilah tersebut.

Representasi numerik yang dihasilkan dari proses TF-IDF selanjutnya digunakan sebagai variabel masukan pada proses pembentukan model klasifikasi menggunakan algoritma *Multinomial Naïve Bayes*.

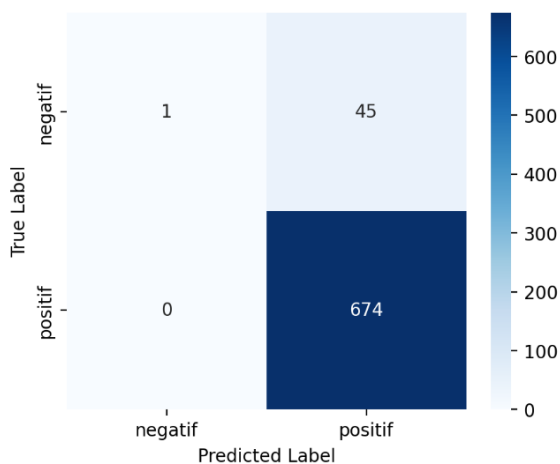
4.5. Hasil Pengujian Model

Pada skenario awal yang hanya menggunakan *Multinomial Naïve Bayes*, model menghasilkan tingkat akurasi sebesar 94%. Namun demikian, kemampuan model dalam mengidentifikasi sentimen negatif masih tergolong rendah yang ditunjukkan oleh nilai *recall* sebesar 0,02. Kondisi ini mengindikasikan bahwa prediksi model lebih banyak mengarah pada kelas dengan jumlah data dominan sehingga pengenalan terhadap kategori negatif belum berjalan secara optimal.

Tabel 3. Hasil perbandingan pengujian model

Model	Dataset	Sentimen	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	Data Real	Negatif	0.94	1	0.02	0.04
		Positif	0.94	0.94	1	0.97
Naïve Bayes	ROS	Negatif	0.94	0.52	0.83	0.64
		Positif	0.94	0.99	0.95	0.97
Naïve Bayes	SMOTE	Negatif	0.94	0.53	0.83	0.64
		Positif	0.94	0.99	0.95	0.97
Hybrid IndoBERT + NB	Data Real	Negatif	0.92	1	0.02	0.04
		Positif	0.92	0.92	1	0.96
Hybrid IndoBERT + NB	ROS	Negatif	0.94	0.57	0.89	0.68
		Positif	0.94	0.99	0.94	0.97
Hybrid IndoBERT + NB	SMOTE	Negatif	0.94	0.56	0.89	0.68
		Positif	0.94	0.99	0.94	0.96

4.6. Naïve Bayes Tanpa Hybrid

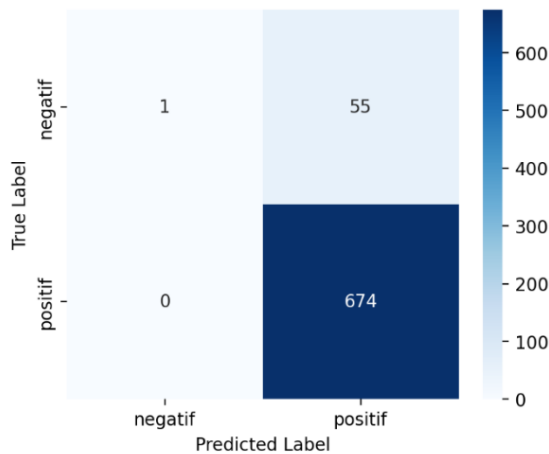


Gambar 4. Confusion matrix Naïve Bayes pada data real

Pengujian awal dilakukan menggunakan algoritma *Multinomial Naïve Bayes* tanpa pendekatan hybrid. Berdasarkan hasil pengujian pada data real, model memperoleh akurasi sebesar 94%. Namun, model memiliki kelemahan dalam mengenali kelas negatif dengan nilai *recall* hanya sebesar 0.02. Hal ini menunjukkan bahwa model cenderung bias terhadap kelas mayoritas akibat ketidakseimbangan dataset.

4.7. Hybrid IndoBERT + Naïve Bayes

Integrasi IndoBERT pada tahap pelabelan ulasan ambigu memberikan dampak positif terhadap proses klasifikasi. Hasil evaluasi menunjukkan bahwa kombinasi kedua metode mampu menghasilkan pemahaman konteks yang lebih baik, terutama pada ulasan yang memiliki makna tidak eksplisit. Dampaknya terlihat pada distribusi hasil klasifikasi yang lebih konsisten ketika model diterapkan pada dataset yang telah melalui proses *balancing*.

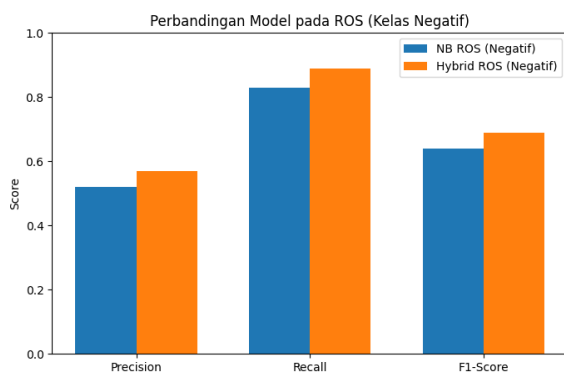


Gambar 5. Confusion matrix Hybrid IndoBERT + Naïve Bayes

4.8. Pengujian ROS

Penerapan *Random Over Sampling* dilakukan dengan menambah jumlah data pada kategori minoritas hingga distribusi kelas menjadi lebih proporsional. Berdasarkan hasil evaluasi, pendekatan ini meningkatkan kemampuan sistem dalam mendeteksi ulasan negatif dibandingkan penggunaan data asli tanpa *balancing*.

Pada skenario ROS, model *Naïve Bayes* memperoleh nilai *recall* sentimen negatif sebesar 0,83. Sementara itu, kombinasi *IndoBERT* dan *Naïve Bayes* menghasilkan *recall* sebesar 0,89. Hasil tersebut menunjukkan bahwa penambahan data pada kelas minoritas membantu model mempelajari karakteristik sentimen negatif dengan lebih baik. Kehadiran *IndoBERT* pada proses pelabelan juga memberikan kontribusi terhadap peningkatan kualitas klasifikasi karena konteks ulasan dapat dipahami secara lebih akurat.

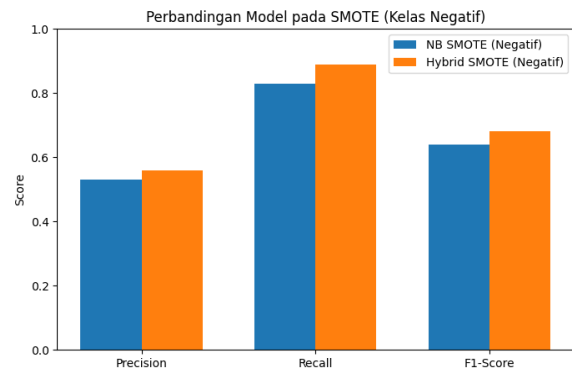


Gambar 6. Perbandingan performa model menggunakan ROS

4.9. Pengujian Menggunakan SMOTE

Berbeda dengan ROS yang melakukan penambahan data melalui duplikasi sampel, SMOTE menghasilkan sampel baru secara sintesis berdasarkan karakteristik data yang telah ada sebelumnya. Pendekatan ini menghasilkan variasi data yang lebih beragam pada kelas minoritas.

Hasil pengujian memperlihatkan bahwa model hybrid kembali memberikan performa yang lebih baik dibandingkan model *Naïve Bayes* tanpa integrasi *IndoBERT*. Nilai *recall* sentimen negatif pada model *hybrid* mencapai 0,89, sedangkan model konvensional memperoleh nilai 0,83. Walaupun tingkat akurasi keseluruhan relatif sama, peningkatan *recall* menunjukkan bahwa penggunaan SMOTE berhasil memperkuat kemampuan sistem dalam mendeteksi data pada kategori yang sebelumnya kurang terwakili.



Gambar 7. Perbandingan performa model menggunakan SMOTE

4.10. Pembahasan

Hasil penelitian menunjukkan bahwa distribusi kelas yang tidak seimbang memiliki dampak yang cukup besar terhadap kualitas prediksi model. Pada dataset asli, model lebih sering menghasilkan prediksi pada kelas positif sehingga kemampuan deteksi sentimen negatif menjadi rendah. Setelah dilakukan proses *balancing* menggunakan ROS maupun SMOTE, kemampuan identifikasi terhadap sentimen negatif mengalami peningkatan yang signifikan.

Selain faktor distribusi data, penggunaan pendekatan gabungan antara *IndoBERT* dan *Multinomial Naïve Bayes* juga memberikan kontribusi terhadap peningkatan kinerja model. *IndoBERT* berperan dalam memahami konteks bahasa pada ulasan yang ambigu, sedangkan *Naïve Bayes* memberikan keunggulan dari sisi efisiensi komputasi. Kombinasi keduanya menghasilkan model yang tidak hanya mempertahankan tingkat akurasi yang tinggi, tetapi juga lebih adaptif dalam menangani ulasan yang memiliki konteks kompleks dan distribusi data yang tidak proporsional.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa kombinasi metode *IndoBERT* dan *Multinomial Naïve Bayes* dapat dimanfaatkan secara efektif untuk melakukan klasifikasi sentimen pada ulasan Google Maps Rumah Sakit JIH Solo. Hasil eksperimen memperlihatkan bahwa ketidakseimbangan distribusi data menjadi salah satu faktor utama yang memengaruhi kemampuan model dalam mengidentifikasi sentimen negatif. Pada dataset asli, baik model konvensional maupun model kombinasi masih mengalami kesulitan

dalam mendeteksi kategori dengan jumlah data yang lebih sedikit. Penerapan teknik penyeimbangan data melalui *Random Over Sampling* (ROS) dan *Synthetic Minority Oversampling Technique* (SMOTE) terbukti memberikan perbaikan terhadap hasil klasifikasi. Setelah proses balancing dilakukan, model yang mengintegrasikan *IndoBERT* dan *Naïve Bayes* mampu mencapai tingkat akurasi sebesar 94% serta memperoleh nilai *recall* sentimen negatif sebesar 0,89. Nilai tersebut lebih tinggi dibandingkan model *Naïve Bayes* yang digunakan tanpa integrasi *IndoBERT* dengan *recall* sebesar 0,83. Temuan ini mengindikasikan bahwa pemanfaatan *IndoBERT* pada proses identifikasi ulasan yang ambigu dapat meningkatkan kualitas pelabelan data sehingga konteks sentimen lebih mudah dikenali oleh model klasifikasi. Di sisi lain, penggunaan metode ROS dan SMOTE berkontribusi dalam mengurangi dampak ketimpangan distribusi kelas. Kombinasi kedua pendekatan tersebut menghasilkan sistem klasifikasi yang lebih konsisten sekaligus mampu memberikan gambaran mengenai kecenderungan opini pengguna terhadap kualitas layanan Rumah Sakit JIH Solo yang tercermin dalam ulasan Google Maps.

Penelitian selanjutnya dapat memanfaatkan dataset dengan cakupan yang lebih luas, baik dari sisi jumlah data maupun variasi sumber ulasan, sehingga model memperoleh representasi data yang lebih beragam. Penggunaan data yang lebih besar juga berpotensi meningkatkan kemampuan generalisasi model pada berbagai kondisi sentimen.

Selain itu, pengembangan penelitian dapat diarahkan pada penerapan model deep learning secara end-to-end tanpa melibatkan algoritma klasifikasi tradisional. Alternatif lain yang dapat dipertimbangkan adalah penggunaan pendekatan *aspect-based sentiment analysis* (ABSA) sehingga hasil analisis tidak hanya menunjukkan polaritas sentimen secara umum, tetapi juga mampu mengidentifikasi opini pengguna terhadap aspek tertentu, seperti fasilitas, pelayanan tenaga medis, administrasi, maupun biaya layanan.

DAFTAR PUSTAKA

- [1] V. W. Ningsih, "Analisis Sentimen Ulasan Pengguna Google Maps terhadap Rumah Sakit di Bojonegoro Menggunakan Metode K-Nearest Neighbor," Universitas Nahdlatul Ulama Sunan Giri, 2024. [Online]. Available: <https://repository.unugiri.ac.id:8443/id/eprint/6066/>
- [2] A. C. T. Angel, V. H. Pranatawijaya, and Widiatry, "Analisis Sentimen dan Emosi dari Ulasan Google Maps untuk Layanan Rumah Sakit di Palangka Raya Menggunakan Machine Learning," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 35–49, 2024, doi: 10.24002/konstelasi.v4i1.8924.
- [3] M. Riski and M. Fikry, "Klasifikasi Sentimen Ulasan Aplikasi WhatsApp di Play Store Menggunakan Metode K-Nearest Neighbor," *KLIK Kaji. Ilm. Inform. dan Kompute*, vol. 4, no. 1, pp. 438–444, 2023, doi: 10.30865/klik.v4i1.1050.
- [4] A. Wirayudha, Murniyati, and Rosdiana, "Analisis Sentimen Terhadap Ulasan Access By KAI Pada Google Play Store Menggunakan Metode Indobert," *PRINSIP Portal Ris. Inov. Sist. Perangkat Lunak*, vol. 3, pp. 9–20, 2025, doi: 10.59696/prinsip.v3i1.69.
- [5] A. A. Qolbu, N. Fitriyati, and N. Inayah, "Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang," *J. FOURIER*, vol. 814, no. 1, pp. 29–44, 2025, doi: 10.14421/fourier.2025.141.29-44.
- [6] R. Merdiansah, Siska, and A. A. Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, pp. 221–228, 2024, doi: 10.55338/jikomsi.v7i1.2895.
- [7] A. F. Rizki, W. Prihartono, and Fathurrohman, "Analisis Sentimen Ulasan Google Maps Rumah Sakit Khalishah Di Cirebon dengan Algoritma Naive Bayes," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 2, pp. 728–738, 2025, doi: 10.23960/jitet.v13i2.6309.
- [8] T. Dzulkarnain, D. E. Ratnawati, and B. Rahayudi, "Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Penilaian Masyarakat Terhadap Pelayanan Rumah Sakit di Malang," *JTIK (Jurnal Teknol. Inf. dan Ilmu Kompute)*, vol. 11, no. 5, pp. 993–1000, 2024, doi: 10.25126/jtiik.2024117979.
- [9] A. P. Mareta, "Implementasi Indobert Untuk Analisis Sentimen Berbasis Aspek Terhadap Ulasan Rumah Sakit," Universitas Sriwijaya, 2025. [Online]. Available: https://repository.unsri.ac.id/189586/4/RAMA_57201_09031182227020_0013058302_01_front_ref.pdf
- [10] C. Apriyadi and Styawati, "Sentiment Analysis of Cyber Attacks in Bank Syariah Indonesia Using SVM and Indobert Method," *J. Tek. Inform.*, vol. 6, no. 2, pp. 819–838, 2025, doi: 10.52436/1.jutif.2025.6.2.2636 Sentiment.
- [11] F. A. Indriyani, A. Fauzi, and S. Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine Tiktok application sentiment analysis using naïve bayes algorithm and support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
- [12] I. G. S. D. Putra and I. N. T. A. Putra, "Implementasi Metode Naïve Bayes Pada Analisis Sentimen Pengguna Aplikasi Mobile Kita Bisa," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6423.

- [13] N. F. Rozam and T. N. Sari, "Analisis Sentimen Penilaian Masyarakat Terhadap Rumah Sakit Berdasarkan Ulasan Pada Google Maps Menggunakan Naive Bayes," *JurTI (Jurnal Teknol. Informasi)*, vol. 8, no. 2, 2024, doi: 10.36294/jurti.v8i2.4681.
- [14] A. L. Maysara and Muljono, "Optimalisasi Akurasi dan Stabilitas Analisis Sentimen Ulasan E-Commerce Indonesia melalui Fine-Tuning Transformer IndoBERT," *Infotekmesin*, vol. 17, no. 01, pp. 9–17, 2026, doi: 10.35970/infotekmesin.v17i1.3037.
- [15] M. R. Prasetyo and A. Fahrurrozi, "Analisa Sentimen Pada Ulasan Google Untuk Hotel Gran Mahakam Jakarta Menggunakan Pendekatan Machine Learning," *J. Ilm. Inform. Komput.*, vol. 28, pp. 203–217, 2023, doi: 10.35760/ik.2023.v28i3.9761.
- [16] Y. MZ, J. E. B, and A. A. Prasetyo, "Penerapan Algoritma Supervised Learning Untuk Analisis Sentimen Kebijakan Ekonomi Pada Masa Pemerintahan Prabowo Melalui Aplikasi X Application of Supervised Learning Algorithms for Sentiment Analysis of Economic Policies During Prabowo ' s Administration," *JIKO (JURNAL Inform. DANKOMPUTER)*, vol. 10, no. 1, pp. 180–192, 2026, doi: 10.26798/jiko.v10i1.2373.
- [17] S. Sathyanarayanan and B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, vol. 27, no. 4, 2024, doi: 10.53555/AJBR.v27i4S.4345.
- [18] P. Silitonga, D. El, R. Purba, F. Ojak, and I. Pardede, "Analisis Sentimen Ulasan Pengguna LinkedIn di Google Playstore Menggunakan Metode Support Vecotre Machine (SVM)," *J. Tek. Inform. Unika ST. Thomas*, vol. 09, pp. 137–142, 2024.