Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3976~3990
e-ISSN: 2548-6861

Comparative Analysis of Machine Learning Algorithms with SMOTE for Imbalanced Sentiment Classification of IndiHome on Platform X

Rizky Adisaputra [1]*, Muhammad Arifin [2], Soni Adiyono [3]
[1,2,3] Sistem Informasi, Universitas Muria Kudus
202253121@std.umk.ac.id [1], arifin.m@umk.ac.id [2], soni.adiyono@umk.ac.id [3]

Article Info

Article history:

Received 2026-06-09
Revised 2026-07-25
Accepted 2026-08-12

Keyword:

Imbalanced Data,
Machine Learning,
Sentiment Analysis,
SMOTE,
TF-IDF.

ABSTRACT

Sentiment analysis of IndiHome users on social media X faces a severe class imbalance, with negative tweets dominating 88.26% of the dataset. This study compares four machine learning algorithms, Support Vector Machine (SVM), Naive Bayes, Decision Tree, and Random Forest, for sentiment classification using SMOTE to address the imbalance. Initially, 20,001 Indonesian tweets were scraped using Tweet Harvest with the keyword "indihome". After duplicate removal and preprocessing, 7,199 tweets were retained. Each tweet was manually annotated into positive, negative, and neutral categories. TF-IDF was applied for feature extraction, and Stratified 5-Fold Cross Validation was used for evaluation. Algorithms were tested under two conditions: without and with SMOTE. Before SMOTE, SVM achieved the highest accuracy (94.55%) and F1-score (94.05%). After SMOTE, Random Forest outperformed others with 94.14% accuracy and 93.83% F1-score, as the only algorithm showing consistent improvement across all metrics, including balanced accuracy and MCC. Although Wilcoxon tests showed no statistically significant differences between algorithms, Random Forest demonstrated the most stable and consistent performance. These findings confirm that Random Forest with SMOTE is the most effective strategy for imbalanced sentiment classification in this context.

I. PENDAHULUAN

Platform media sosial X merupakan salah satu wadah yang kerap dimanfaatkan oleh masyarakat sebagai sarana untuk mengungkapkan pendapat, kritik, maupun pengalaman pribadi mereka terkait suatu layanan digital [1]. Pengguna memanfaatkan X untuk menyampaikan kepuasan atau keluhan terhadap kualitas jaringan, stabilitas koneksi, dan layanan pelanggan IndiHome [2]. IndiHome dipilih sebagai studi kasus karena merupakan salah satu penyedia layanan internet rumah dengan basis pelanggan terbesar di Indonesia, sehingga opini publik mengenai layanan ini memiliki cakupan dan dampak yang luas terhadap persepsi masyarakat terhadap kualitas layanan telekomunikasi nasional. Analisis sentimen merupakan salah satu metode yang lazim diterapkan guna mengenali pandangan publik berdasarkan data teks yang bersifat tidak terstruktur [3]. Penelitian sebelumnya menyebutkan bahwa analisis sentimen mampu digunakan untuk memahami opini dan emosi pengguna pada media sosial secara otomatis menggunakan pendekatan *machine learning* [4]. Perkembangan Artificial Intelligence (AI) dan digital transformation juga mendorong pemanfaatan machine learning dalam pengolahan data teks tidak terstruktur untuk mendukung pengambilan keputusan berbasis data secara lebih efektif [5]

Data media sosial memiliki karakteristik kompleks karena ditulis secara bebas oleh pengguna [6]. Permasalahan lain yang kerap dijumpai adalah ketimpangan proporsi antar kelas, yakni volume data pada salah satu kategori sentimen jauh lebih besar dibandingkan kategori lainnya sehingga komposisi dataset menjadi tidak berimbang [7]. Ketidakseimbangan distribusi tersebut berpotensi membuat model lebih banyak mempelajari pola dari kelas dominan dan mengurangi kemampuannya dalam mengenali kelas yang jumlah datanya lebih sedikit [8]. Penelitian terdahulu menyatakan bahwa ketidakseimbangan data dapat

memengaruhi hasil klasifikasi serta berpotensi menghasilkan interpretasi yang menyimpang apabila permasalahan tersebut dibiarkan tanpa penanganan yang memadai. [9].

Berbagai algoritma *machine learning* telah banyak diterapkan pada penelitian klasifikasi sentimen, di antaranya Naive Bayes, *Support Vector Machine* (SVM), Decision Tree, dan Random Forest [10]. Algoritma *machine learning* mampu mempelajari pola kompleks dari data berdimensi tinggi serta banyak diterapkan dalam proses klasifikasi karena unggul dalam menghasilkan prediksi yang andal pada data bertipe teks [11]. Studi yang dilakukan Ardiyansah membandingkan metode Naive Bayes dan SVM dalam analisis sentimen di platform media sosial X, dengan hasil yang memperlihatkan bahwa SVM mampu melampaui performa Naive Bayes setelah diterapkannya SMOTE [1]. Sementara itu, penelitian lain turut mengungkapkan bahwa algoritma SVM terbukti efektif dalam mengklasifikasikan teks pendek di lingkungan media sosial [12]. Selain itu, Random Forest dilaporkan mampu memberikan performa yang stabil pada proses klasifikasi karena menggunakan mekanisme ensemble learning dan majority voting yang mampu meningkatkan konsistensi prediksi model [13][14].

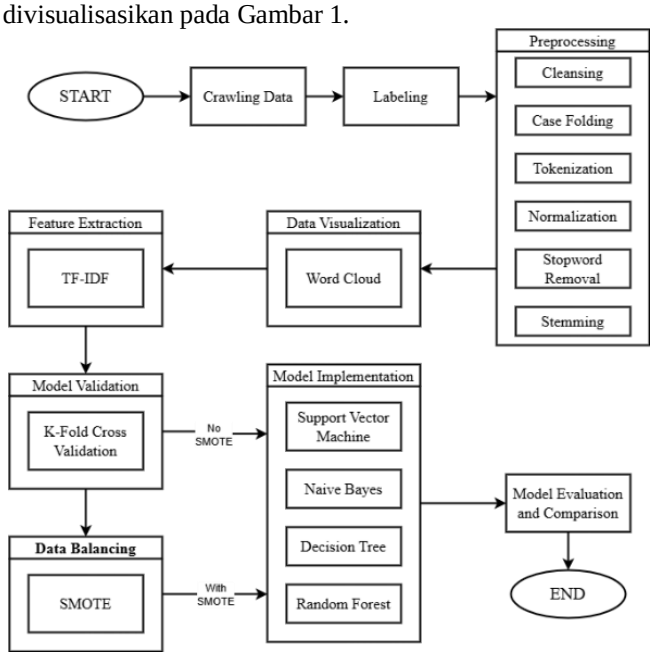
Guna meminimalkan dampak ketimpangan distribusi kategori kelas, studi ini menerapkan metode *Synthetic Minority Oversampling Technique* (SMOTE). Ketidakseimbangan distribusi data berpotensi membuat model condong berpihak pada kelas mayoritas, sehingga dibutuhkan suatu metode penyeimbangan data agar kapasitas generalisasi model dapat ditingkatkan secara optimal [15]. Pendekatan tersebut menambah sampel baru secara sintetis pada kelas yang jumlah datanya sedikit melalui proses interpolasi antar data yang memiliki karakteristik serupa, sehingga proporsi antar kelas menjadi lebih merata [16]. Sejumlah penelitian mengungkapkan bahwa penggunaan SMOTE terbukti mampu mendongkrak performa pengelompokan pada data yang tidak proporsional, khususnya dalam konteks analisis sentimen berbasis *machine learning* [17]. Namun demikian, pengaruh SMOTE terhadap setiap algoritma tidak selalu sama karena karakteristik masing-masing model klasifikasi berbeda [18]. Oleh sebab itu, diperlukan evaluasi lebih lanjut untuk mengetahui bagaimana pengaruh SMOTE terhadap performa beberapa algoritma *machine learning* dalam mengelompokkan sentimen pengguna IndiHome di platform media sosial X. Kebaruan penelitian ini terletak pada analisis mekanistik mengenai mengapa SMOTE berdampak berbeda pada masing-masing algoritma, suatu aspek yang masih jarang dibahas dalam literatur analisis sentimen berbahasa Indonesia.

Penelitian mengenai analisis sentimen layanan IndiHome pada platform media sosial sebelumnya telah banyak dilakukan, namun umumnya terbatas pada pengujian satu algoritma klasifikasi tanpa pembandingan menyeluruh antar metode [2][19]. Selain itu, sebagian penelitian yang menerapkan teknik oversampling untuk menangani ketidakseimbangan data melaporkan dampak yang minim atau tidak konsisten tanpa analisis lebih lanjut mengenai penyebabnya [20][21], serta cenderung mengandalkan accuracy dan F1-score konvensional sebagai metrik evaluasi utama tanpa mempertimbangkan metrik yang lebih representatif pada data tidak seimbang seperti balanced accuracy maupun Matthews Correlation Coefficient (MCC). Penelitian ini memberikan kontribusi dengan mengevaluasi secara komprehensif pengaruh SMOTE pada empat algoritma klasifikasi sekaligus (SVM, Naive Bayes, Decision Tree, dan Random Forest), serta menganalisis secara mendalam mengapa dampak SMOTE bervariasi antar algoritma, suatu aspek yang belum banyak dibahas pada penelitian-penelitian terdahulu mengenai sentimen layanan IndiHome.

Berangkat dari permasalahan ketidakseimbangan kelas pada data sentimen, penelitian ini mengevaluasi sejauh mana penerapan SMOTE memengaruhi kinerja beberapa algoritma machine learning dalam mengklasifikasikan sentimen pengguna IndiHome di platform X [22]. Algoritma yang digunakan dalam penelitian ini meliputi *Naive Bayes*, *Support Vector Machine*, *Decision Tree*, dan *Random Forest*. Evaluasi kinerja model diukur menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* guna memperoleh gambaran yang lebih komprehensif mengenai kemampuan masing-masing model dalam menangani data sentimen yang bersifat tidak seimbang. Evaluasi menggunakan beberapa metrik diperlukan agar performa model tidak hanya dinilai dari akurasi, tetapi juga kemampuan model dalam mengenali setiap kelas secara proporsional [23]

II. METODE

Studi ini mengimplementasikan analisis sentimen guna mengkategorikan pendapat pengguna platform media sosial X terkait layanan IndiHome. Keseluruhan alur penelitian divisualisasikan pada Gambar 1.



Gambar 1. Skema Alur Penelitian

A. Crawling Data

Pengumpulan data dilakukan menggunakan teknik *web scraping* pada Platform X (sebelumnya Twitter) dengan memanfaatkan Tweet Harvest berbasis Python [24]. Data dikumpulkan menggunakan kata kunci "indihome" dengan filter tweet berbahasa Indonesia (*lang*) pada periode April 2025 hingga Maret 2026. Karena adanya keterbatasan akses (*rate limit*) pada Platform X, proses *scraping* dilakukan secara bertahap setiap hari dengan batas pengambilan maksimal 100 tweet per proses [25]. Pengambilan data dilakukan secara berulang (*looping*) hingga mencakup seluruh hari dalam setiap bulan sehingga menghasilkan sejumlah berkas data harian sesuai jumlah hari pada bulan tersebut. Seluruh berkas harian kemudian digabungkan menjadi satu dataset bulanan, kemudian seluruh dataset bulanan digabungkan kembali menjadi satu dataset yang digunakan dalam penelitian.

B. Pelabelan Data

Data hasil scraping belum memiliki kategori sentimen sehingga perlu dilakukan proses anotasi terlebih dahulu agar dapat digunakan pada tahap pembelajaran model klasifikasi. Pelabelan dilakukan dalam dua tahap, yaitu pelabelan awal secara otomatis menggunakan kamus sentimen InSet Lexicon, kemudian diverifikasi secara manual oleh peneliti untuk memastikan kesesuaian label terhadap konteks kalimat. Data dikelompokkan ke dalam tiga kategori sentimen, yakni positif, negatif, dan netral. Untuk mengevaluasi konsistensi hasil pelabelan, dilakukan pengujian reliabilitas antar-anotator menggunakan Cohen's Kappa pada sampel data yang dilabeli secara independen oleh anotator kedua, dengan hasil yang dijelaskan lebih lanjut pada bagian Hasil dan Pembahasan.

C. Preprocessing

Preprocessing mentransformasi data mentah menjadi format terstruktur untuk meningkatkan efektivitas pemrosesan model dan mereduksi informasi tidak relevan [26]. Adapun tahapan *preprocessing* yang diterapkan dalam studi ini mencakup beberapa proses berikut.

1. *Cleansing:* Proses *cleansing* dijalankan guna membersihkan teks dari berbagai elemen yang tidak memberikan kontribusi berarti terhadap analisis sentimen. Tahap ini meliputi penghapusan mention, hashtag, tautan (URL), serta berbagai simbol dan karakter khusus yang berpotensi menghasilkan *noise* pada data. Hasilnya, teks yang diperoleh menjadi lebih bersih dan terfokus pada informasi yang memuat makna sentimen.
2. *Case Folding:* Tahap ini mentransformasikan seluruh teks menjadi huruf kecil (*lowercase*) untuk menstandarisasi format penulisan. Hal ini penting mengingat penggunaan huruf kapital pada unggahan di platform X cenderung tidak seragam.
3. *Tokenization:* Tahap ini memecah kalimat menjadi unit-unit terkecil yang disebut token, berupa kata maupun karakter sesuai kebutuhan, sehingga setiap token merepresentasikan satu satuan makna tertentu.
4. *Normalization:* Proses normalisasi bertujuan menstandarisasi kata-kata tidak baku maupun yang memiliki keragaman ejaan agar diperoleh representasi teks yang konsisten, misalnya mengubah "gak" menjadi "tidak" atau "udah" menjadi "sudah".
5. *Stopword Removal:* Tahap ini menyaring kata-kata yang kerap muncul namun tidak memberikan nilai informasi yang berarti bagi analisis teks, seperti "yang", "tapi", "masih", dan sejenisnya.
6. *Stemming:* Proses stemming mengubah kata menjadi wujud asalnya dengan cara menanggalkan afiks, prefiks, maupun sufiks, tanpa menggeser makna inti dari kata tersebut.

Penerapan teknik *preprocessing* yang tepat dan menyeluruh memiliki peran krusial dalam mendongkrak kinerja model [27], terutama pada konteks klasifikasi teks.

D. Data Visualization

Tahap visualisasi data dilaksanakan menggunakan *word cloud* guna memetakan kosakata dengan tingkat kemunculan paling tinggi pada setiap kategori sentimen. Visualisasi ini memudahkan identifikasi karakteristik kata yang dominan dari masing-masing kelas secara intuitif, di mana tinggi rendahnya ukuran kata mencerminkan seberapa tinggi intensitas kemunculan kata tersebut dalam dataset [28].

E. Feature Extraction

Proses ekstraksi fitur dilakukan untuk mentransformasikan data teks hasil *preprocessing* menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Dalam penelitian ini, metode TF-IDF (*Term Frequency–Inverse Document Frequency*) diterapkan sebagai pendekatan untuk menghitung bobot setiap kata yang terdapat dalam dokumen.Pembobotan ini mempertimbangkan tingkat kemunculan suatu kata pada dokumen tertentu sekaligus tingkat distribusinya di seluruh dataset [29]. Dengan demikian, kata yang dinilai lebih merepresentasikan isi dokumen akan mendapatkan nilai bobot lebih besar dibandingkan kata yang lazim ditemukan di banyak dokumen.

1. *Term Frequency (TF):* Komponen TF berfungsi sebagai alat untuk menghitung jumlah kehadiran sebuah kata di dalam suatu dokumen. Besaran nilai TF akan meningkat seiring bertambahnya tingkat keberulangan kata tersebut, sehingga kata yang lebih kerap hadir memperoleh bobot yang lebih tinggi. Persamaan TF yang digunakan dalam penelitian ini disajikan melalui formula berikut.

$$tf\left(t,d\right) = f_{t,d} \tag{1}$$

Keterangan :
$f_{(t,d)}$: frekuensi kemunculan kata $t$dalam dokumen d

2. *Inverse Document Frequency (IDF)*: IDF difungsikan untuk menghitung tingkat keunikan suatu kata dalam keseluruhan korpus dokumen [30]. Semakin sedikit sebuah kata ditemukan pada dokumen lain, semakin besar nilai IDF yang diperoleh. Sebaliknya, kata yang hadir pada hampir semua dokumen akan mendapatkan bobot yang lebih kecil karena dinilai kurang mampu membedakan karakteristik antar dokumen. Persamaan IDF yang diterapkan dalam penelitian ini disajikan melalui formula berikut.

$$idf(t,D) = \ln \frac{1+N}{1+df(t)} + 1 \quad (2)$$

Keterangan :
N : jumlah total dokumen dalam dataset.
df : jumlah dokumen yang mengandung kata tersebut.
ln : logaritma natural.

Nilai TF-IDF diperoleh dengan mengalikan kedua nilai tersebut:

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \quad (3)$$

F. Model Validation

Penelitian ini mengimplementasikan Stratified 5-Fold Cross Validation sebagai metode validasi model. Pendekatan tersebut digunakan guna mendapatkan gambaran performa yang lebih akurat dan menyeluruh melalui pembagian data ke dalam beberapa fold dengan proporsi kelas yang tetap terjaga [11]. Dalam metode ini, dataset dibagi ke dalam lima fold dengan tetap mempertahankan proporsi masing-masing label pada tiap fold. Pada tiap putaran pengujian, satu *fold* dialokasikan sebagai data uji, sedangkan empat *fold* yang tersisa digunakan sebagai bahan pelatihan model. Tahapan tersebut dijalankan secara bergiliran sampai keseluruhan fold pernah difungsikan sebagai data evaluasi. Hasil evaluasi yang dilaporkan merupakan rata-rata dari kelima iterasi tersebut. Pendekatan stratified dipilih mengingat distribusi kelas dalam dataset yang tidak seimbang, sehingga proporsi kelas pada setiap fold tetap terjaga dan hasil evaluasi lebih objektif.

G. Justifikasi Pemilihan Metode Penyeimbangan Data

Sebelum menetapkan SMOTE sebagai metode penyeimbangan data utama, penelitian ini melakukan pengujian pembanding menggunakan skema Stratified 5-Fold Cross Validation yang sama dengan evaluasi utama. Pendekatan class weighting diujikan pada algoritma SVM, Decision Tree, dan Random Forest (tidak diterapkan pada Naive Bayes karena keterbatasan dukungan pustaka scikit-learn). Hasil pengujian menunjukkan bahwa SMOTE secara konsisten menghasilkan weighted F1-score yang sedikit lebih tinggi dibandingkan class weighting pada ketiga algoritma tersebut (SVM 92,40% berbanding 92,17%; Decision Tree 91,79% berbanding 91,17%; Random Forest 93,84% berbanding 93,51%). Selain itu, SMOTE turut diujicobakan pada dua algoritma di luar cakupan utama penelitian, yaitu Logistic Regression dan K-Nearest Neighbor (KNN), yang menunjukkan pengaruh SMOTE bervariasi antar algoritma. Pada Logistic Regression, SMOTE tidak menurunkan weighted F1-score secara berarti (92,91% menjadi 92,89%) dan justru meningkatkan balanced accuracy dari 0,6654 menjadi 0,7974, sementara pada KNN, SMOTE menurunkan performa secara signifikan pada seluruh metrik evaluasi, termasuk weighted F1-score (91,88% menjadi 74,64%). Temuan ini memperkuat dasar pemilihan SMOTE sebagai metode utama serta pemilihan SVM, Naive Bayes, Decision Tree, dan Random Forest sebagai fokus algoritma penelitian, mengingat konsistensi keunggulan SMOTE pada algoritma tersebut sejalan dengan tujuan penelitian dalam menganalisis dampaknya secara mendalam pada algoritma klasik.

H. Handling Imbalance Data (SMOTE)

Kesenjangan jumlah data pada tiap kelas berpotensi mempengaruhi proses pelatihan model dan memicu kecenderungan prediksi yang condong ke arah kelas mayoritas. Oleh karena itu, penelitian ini mengimplementasikan SMOTE (*Synthetic Minority Over-sampling Technique*) sebagai solusi penyeimbang distribusi antar kelas. Melalui pendekatan ini, sampel sintetis dihasilkan dari data minoritas yang sudah tersedia sehingga komposisi dataset menjadi lebih proporsional dan representatif untuk proses pelatihan model [31]. Penerapan teknik ini bertujuan untuk meminimalkan bias pada model machine learning sekaligus meningkatkan akurasi dan efektivitas klasifikasi.

I. Modeling

Tahap modeling merupakan proses konstruksi model klasifikasi untuk mengkategorikan sentimen ke dalam tiga kelompok, yakni positif, negatif, dan netral [32]. Penelitian ini mengevaluasi performa empat algoritma klasifikasi, yakni Support Vector Machine (SVM), Naive Bayes, Decision Tree, dan Random Forest, dalam menjalankan tugas pengelompokan sentimen pengguna IndiHome di platform media sosial X. Seluruh algoritma diimplementasikan menggunakan pustaka *scikit-learn* dengan *random_state* sebesar 42 untuk menjamin reproduktifitas hasil eksperimen.

1. *Support Vector Machine (SVM)*: Algoritma klasifikasi yang bekerja dengan mengidentifikasi hyperplane terbaik guna mempartisi data antar kategori kelas dengan jarak pemisah terlebar. SVM dikenal mampu bekerja optimal pada data berdimensi tinggi seperti vektor teks yang dihasilkan dari proses ekstraksi TF-IDF. Pada penelitian ini, SVM dikonfigurasi menggunakan kernel linear dengan parameter regularisasi (C) sebesar 1,0.
2. *Naive Bayes*: Algoritma yang berlandaskan teori probabilitas dengan menerapkan Teorema Bayes dan mengasumsikan bahwa setiap fitur tidak saling mempengaruhi satu sama lain. Algoritma ini kerap dipilih dalam klasifikasi teks karena mekanismenya yang

sederhana dan tidak membutuhkan sumber daya komputasi yang besar. Penelitian ini menggunakan varian Multinomial Naive Bayes dengan parameter smoothing (alpha) sebesar 1,0.

3. *Decision Tree*: Metode klasifikasi yang membentuk struktur pohon keputusan secara bertahap berdasarkan atribut yang paling informatif dengan memanfaatkan nilai entropy dan information gain, hingga setiap simpul daun menghasilkan kelas akhir. Algoritma ini dikonfigurasi menggunakan kriteria pemecahan *gini impurity* tanpa pembatasan kedalaman pohon (max_depth).
4. *Random Forest*: Metode ensemble learning yang mengombinasikan beberapa Decision Tree melalui skema majority voting untuk menghasilkan prediksi yang lebih stabil sekaligus meminimalkan risiko overfitting. Algoritma ini dikonfigurasi menggunakan 100 pohon (n_estimators) dengan kriteria pemecahan *gini impurity* tanpa pembatasan kedalaman pohon.

J. Model Evaluasi

Penilaian model dalam penelitian ini memanfaatkan *confusion matrix* sebagai landasan perhitungan nilai *precision*, *recall*, dan *accuracy*. Tabel I memaparkan struktur *confusion matrix* yang diterapkan.

TABEL I
CONFUSION MATRIX

Kelas	**Klasifikasi**	
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

Keterangan :

- TP : Data pada suatu kelas yang berhasil diprediksi benar sesuai kelasnya.
- TN : Data bukan pada kelas yang diamati dan berhasil diprediksi bukan sebagai kelas tersebut.
- FP : Data bukan pada kelas yang diamati tetapi diprediksi sebagai kelas tersebut.
- FN : Data pada kelas yang diamati tetapi diprediksi sebagai kelas lain.

Pada penelitian dengan tiga kelas (positif, negatif, dan netral), perhitungan dilakukan secara *one-vs-rest*, yaitu setiap kelas diperlakukan sebagai kelas positif sementara kelas lainnya dianggap negatif. Rumus metrik evaluasi yang digunakan adalah sebagai berikut.

1. *Accuracy*: merupakan persentase prediksi yang benar dari keseluruhan data yang diuji.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

2. *Precision*: menggambarkan seberapa tepat model dalam memprediksi suatu kelas dari keseluruhan data yang diprediksi sebagai kelas tersebut.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

3. *Recall*: menggambarkan kemampuan model dalam menemukan seluruh data yang sebenarnya termasuk dalam suatu kelas.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

4. *F1-Score*: merupakan rata-rata harmonik dari precision dan recall yang memberikan keseimbangan antara keduanya.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

K. Uji Signifikansi Statistik

Selain evaluasi menggunakan accuracy, precision, recall, F1-score, balanced accuracy, dan Matthews Correlation Coefficient (MCC), penelitian ini juga melakukan uji signifikansi statistik untuk membandingkan performa antar algoritma. Uji Shapiro–Wilk terlebih dahulu digunakan untuk mengevaluasi normalitas nilai hasil Stratified 5-Fold Cross Validation. Meskipun hasil uji normalitas tidak menunjukkan penyimpangan yang signifikan, Wilcoxon Signed-Rank Test dipilih karena jumlah observasi hanya terdiri atas lima fold sehingga pendekatan nonparametrik dinilai lebih sesuai untuk ukuran sampel yang kecil. Tingkat signifikansi yang digunakan adalah $\alpha = 0{,}05$.

III. HASIL DAN PEMBAHASAN

A. Dataset

Proses pengumpulan data menghasilkan sebanyak 20.001 tweet yang diperoleh dari Platform X selama periode April 2025 hingga Maret 2026 menggunakan kata kunci "indihome". Data yang berhasil dikumpulkan merupakan hasil penggabungan seluruh berkas hasil *scraping* yang diperoleh secara bertahap setiap hari selama periode pengambilan data. Dataset mentah tersebut terdiri atas 15 atribut yang mencakup identitas tweet, waktu publikasi, isi tweet (*full_text*), informasi akun, serta metadata interaksi. Meskipun demikian, penelitian ini hanya memanfaatkan atribut *full_text* sebagai data utama karena atribut tersebut berisi isi tweet yang digunakan dalam proses pelabelan

sentimen, *preprocessing*, ekstraksi fitur TF-IDF, dan klasifikasi.

Sebelum digunakan pada tahap analisis, dataset terlebih dahulu melalui proses penyaringan (*data filtering*) untuk memastikan kualitas data. Pada tahap ini dilakukan penghapusan tweet duplikat sehingga jumlah data berkurang dari 20.001 tweet menjadi 7.268 tweet. Penghapusan data duplikat dilakukan agar setiap tweet hanya direpresentasikan satu kali sehingga tidak menimbulkan bias pada proses pelatihan model klasifikasi.

Dataset hasil scraping selanjutnya diproses melalui tahapan preprocessing yang meliputi *cleansing*, *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*. Selama proses tersebut, ditemukan sejumlah tweet yang menghasilkan teks kosong setelah dilakukan pembersihan sehingga tidak lagi mengandung informasi tekstual yang dapat dianalisis. Oleh karena itu, tweet tersebut tidak digunakan dalam penelitian. Setelah seluruh tahapan preprocessing selesai dilakukan, diperoleh sebanyak 7.199 tweet yang digunakan sebagai dataset akhir untuk proses pelabelan sentimen, ekstraksi fitur, serta pelatihan dan evaluasi model klasifikasi. Contoh data hasil scraping yang digunakan dalam penelitian ditampilkan pada Tabel II.

TABEL II
CONTOH DATA MENTAH

No	Cuitan
1	Indihome sekarang makin aneh udh restart modem berkali kali tapi internet tetep putus putus
...	.....
7188	sudah aman admin makasih 😇
7199	tolong cdm ya kak @IndiHome @IndiHomeCare

B. Pelabelan Data

Pelabelan data dilakukan untuk mengelompokkan setiap tweet ke dalam kategori sentimen sehingga dapat digunakan sebagai data pada proses klasifikasi. Penelitian ini menggunakan tiga kategori sentimen, yaitu positif, negatif, dan netral. Pelabelan dilaksanakan dalam dua tahap. Tahap pertama menggunakan kamus sentimen InSet Lexicon untuk menghasilkan label awal secara otomatis. Selanjutnya, seluruh hasil pelabelan diverifikasi secara manual oleh peneliti dengan meninjau setiap tweet satu per satu guna memastikan kesesuaian label terhadap konteks kalimat yang sebenarnya.

Tweet berlabel positif mencerminkan opini yang menunjukkan apresiasi, kepuasan, atau pengalaman baik terhadap layanan IndiHome. Tweet berlabel negatif berisi keluhan, kritik, maupun ungkapan ketidakpuasan terhadap layanan IndiHome. Sementara itu, tweet berlabel netral merupakan tweet yang tidak menunjukkan kecenderungan sentimen positif maupun negatif, seperti penyampaian informasi, pertanyaan, atau tanggapan yang bersifat informatif. Proses verifikasi manual dilakukan untuk memperbaiki kesalahan pelabelan yang disebabkan oleh penggunaan bahasa tidak baku, singkatan, maupun konteks kalimat yang belum dapat diinterpretasikan secara tepat oleh pendekatan berbasis leksikon. Seluruh proses pelabelan dilakukan oleh satu anotator utama.

Untuk mengevaluasi tingkat konsistensi dan objektivitas hasil pelabelan, dilakukan pengujian reliabilitas antar-anotator menggunakan Cohen's Kappa pada sampel sebanyak 500 tweet yang diambil secara stratified dengan proporsi kelas minoritas (netral dan positif) yang diperbanyak untuk memastikan pengujian mencakup area yang paling rawan subjektivitas. Sampel tersebut dilabeli secara independen oleh anotator kedua tanpa mengetahui hasil pelabelan sebelumnya. Hasil pengujian menunjukkan nilai Cohen's Kappa sebesar 0,8088 dengan tingkat kesesuaian mentah sebesar 88,00%, yang berada pada ambang batas kategori *substantial* menuju *almost perfect* menurut skala interpretasi Landis dan Koch (1977) [32]. Ketidaksesuaian yang teramati paling banyak terjadi antara kelas negatif dan netral, mengindikasikan bahwa batas interpretasi antara ungkapan keluhan bernada datar dan pernyataan informatif netral merupakan area yang secara inheren rawan perbedaan penilaian dalam tugas anotasi sentimen berbahasa Indonesia. Hasil ini memberikan indikasi awal bahwa kriteria pelabelan yang digunakan cukup konsisten meski diterapkan oleh anotator berbeda, meskipun pengujian reliabilitas menyeluruh pada seluruh dataset menggunakan anotator kedua tetap disarankan untuk penelitian selanjutnya guna memperkuat validitas hasil.

Setelah proses pelabelan selesai dilakukan, setiap tweet dikelompokkan ke dalam tiga kategori sentimen, yaitu negatif, positif, dan netral. Proses anotasi menghasilkan sebanyak 6.354 tweet berlabel negatif, 513 tweet berlabel positif, dan 332 tweet berlabel netral dari total 7.199 tweet yang digunakan dalam penelitian. Contoh hasil anotasi data ditampilkan pada Tabel III.

TABEL III
CONTOH HASIL ANOTASI DATA

No	Cuitan	Kategori
1	Indihome sekarang makin aneh udh restart modem berkali kali tapi internet tetep putus putus	negatif
...	.....	.....
7188	sudah aman admin makasih 😇	positif
7199	tolong cdm ya kak @IndiHome @IndiHomeCare	netral

C. Preprocessing Data

Tahap *preprocessing* merupakan langkah penting dalam analisis sentimen untuk meningkatkan kualitas data teks sebelum dilakukan ekstraksi fitur dan proses klasifikasi. Tweet hasil *scraping* umumnya masih mengandung berbagai elemen yang tidak berkontribusi terhadap analisis, seperti URL, *mention*, *hashtag*, tanda baca, angka, maupun karakter khusus. Oleh karena itu, dilakukan serangkaian tahapan *preprocessing* yang meliputi *cleansing*, *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*.

Tahap *cleansing* bertujuan menghapus elemen-elemen yang tidak diperlukan sehingga hanya menyisakan informasi

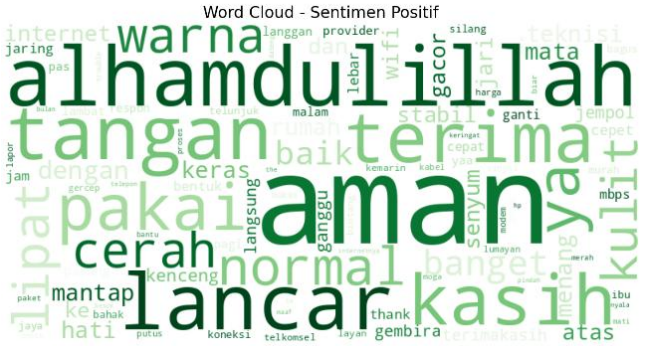
tekstual yang relevan. Selanjutnya, *case folding* mengubah seluruh huruf menjadi huruf kecil (*lowercase*) agar penulisan kata menjadi seragam. Tahap *tokenization* memisahkan kalimat menjadi kumpulan token, kemudian *normalization* mengubah kata tidak baku menjadi bentuk baku sesuai kamus yang digunakan. Setelah itu, *stopword removal* menghilangkan kata-kata umum yang memiliki kontribusi rendah terhadap proses klasifikasi, sedangkan *stemming* mengubah setiap kata menjadi bentuk kata dasar. Rangkaian tahapan tersebut menghasilkan data teks yang lebih bersih dan terstruktur sehingga dapat meningkatkan kualitas representasi fitur pada proses klasifikasi. Contoh hasil *preprocessing* ditampilkan pada Tabel IV.

TABEL IV
CONTOH HASIL PREPROCESSING DATA

Tahapan	Hasil Pemrosesan
Text awal	Indihome sekarang makin aneh udh restart modem berkali kali tapi internet tetep putus putus
Cleansing	indihome sekarang makin aneh udh restart modem berkali kali tapi internet tetep putus putus
Case folding	indihome sekarang makin aneh udh restart modem berkali kali tapi internet tetep putus putus
Tokenization	['indihome', 'sekarang', 'makin', 'aneh', 'udh', 'restart', 'modem', 'berkali', 'kali', 'tapi', 'internet', 'tetep', 'putus', 'putus']
Normalization	['indihome', 'sekarang', 'makin', 'aneh', 'sudah', 'restart', 'modem', 'berkali', 'kali', 'tapi', 'internet', 'tetap', 'putus', 'putus']
Stopword Removal	['indihome', 'aneh', 'sudah', 'restart', 'modem', 'berkali', 'kali', 'internet', 'tetap', 'putus', 'putus']
Stemming	['indihome', 'aneh', 'sudah', 'restart', 'modem', 'kali', 'kali', 'internet', 'tetap', 'putus', 'putus']
Hasil akhir	indihome aneh sudah restart modem kali kali internet tetap putus putus

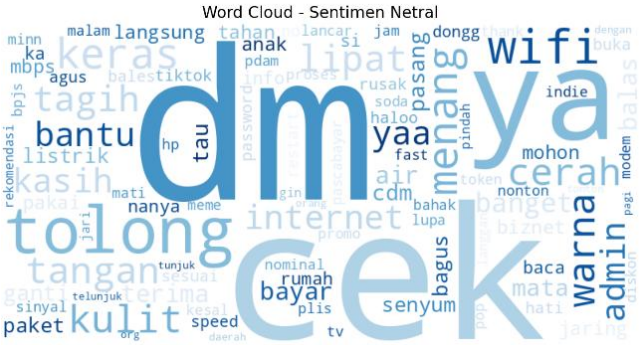
D. Word Cloud

Setelah proses *preprocessing*, dilakukan visualisasi data menggunakan *word cloud* untuk menggambarkan distribusi kata yang paling sering muncul pada setiap kategori sentimen. Ukuran setiap kata pada visualisasi merepresentasikan frekuensi kemunculannya dalam dataset, sehingga semakin besar ukuran suatu kata maka semakin sering kata tersebut muncul. Visualisasi ini digunakan sebagai analisis eksploratif untuk membantu mengidentifikasi karakteristik kosakata yang dominan pada masing-masing kategori sentimen sebelum dilakukan ekstraksi fitur menggunakan TF-IDF.



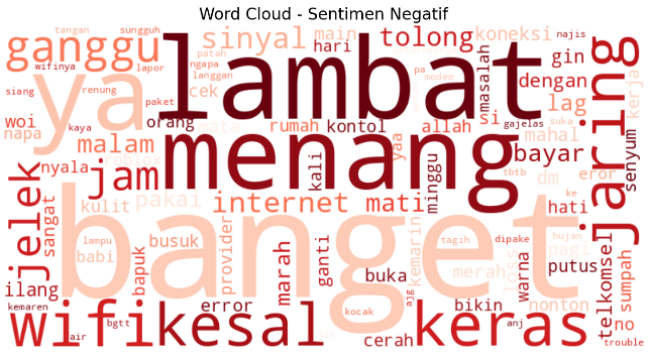
Gambar 2. Visualisasi Frekuensi Kata Sentimen Positif

Gambar 2 menampilkan *word cloud* sentimen positif yang memperlihatkan kosakata yang paling sering muncul pada tweet bermuatan positif terhadap layanan IndiHome. Kemunculan kata *alhamdulillah*, *aman*, *lancar*, *terima*, dan *kasih* mengindikasikan bahwa sebagian besar tweet pada kategori ini berisi ungkapan rasa syukur, apresiasi, dan kepuasan terhadap layanan yang diterima. Selain itu, kata *teknisi*, *gacor*, dan *kencang* juga muncul dengan frekuensi yang cukup tinggi, menunjukkan bahwa pengguna memberikan penilaian positif terhadap kualitas jaringan maupun respons teknisi dalam menangani permasalahan layanan.



Gambar 3. Visualisasi Frekuensi Kata Sentimen Netral

Gambar 3 menampilkan *word cloud* sentimen netral yang memperlihatkan kosakata yang sering muncul tanpa menunjukkan kecenderungan sentimen positif maupun negatif. Kata-kata seperti *cek*, *dm*, *tolong*, *wifi*, dan *admin* mendominasi visualisasi, yang mengindikasikan bahwa sebagian besar tweet pada kategori ini berisi permintaan informasi, tindak lanjut layanan, maupun komunikasi administratif antara pengguna dan pihak IndiHome. Selain itu, kemunculan kata *internet*, *bayar*, *password*, dan *modem* menunjukkan bahwa pembahasan pada kelas netral lebih banyak berkaitan dengan informasi teknis maupun administratif dibandingkan ekspresi kepuasan atau ketidakpuasan pengguna.



Gambar 4. Visualisasi Frekuensi Kata Sentimen Negatif

Gambar 4 menampilkan *word cloud* sentimen negatif yang memperlihatkan kosakata yang paling sering muncul pada tweet bernada negatif terhadap layanan IndiHome. Kemunculan kata *lambat*, *ganggu*, *mati*, *putus*, *lag*, *error*, *sinyal*, dan *wifi* mengindikasikan bahwa sebagian besar keluhan pengguna berkaitan dengan kualitas koneksi serta kestabilan jaringan internet. Selain itu, kata *kesal* menunjukkan adanya ekspresi ketidakpuasan pengguna terhadap layanan yang diterima. Temuan ini mengindikasikan bahwa persepsi negatif terhadap layanan IndiHome selama periode pengambilan data lebih banyak dipengaruhi oleh permasalahan teknis yang menghambat aktivitas pengguna.

Secara keseluruhan, visualisasi *word cloud* memberikan gambaran awal mengenai karakteristik kosakata pada masing-masing kategori sentimen. Meskipun visualisasi ini hanya menunjukkan frekuensi kemunculan kata dan belum mempertimbangkan bobot maupun hubungan antarkata, hasilnya mampu memberikan informasi awal mengenai pola opini pengguna terhadap layanan IndiHome. Analisis tersebut selanjutnya diperkuat melalui proses ekstraksi fitur menggunakan metode TF-IDF.

E. Feature Extraction

Setelah proses *preprocessing*, data teks ditransformasikan ke dalam bentuk numerik menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*). Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen (*term frequency*) serta tingkat kelangkaan kata tersebut pada seluruh kumpulan dokumen (*inverse document frequency*). Dengan demikian, kata yang sering muncul pada suatu tweet tetapi relatif jarang muncul pada tweet lainnya akan memperoleh bobot yang lebih tinggi karena dianggap memiliki kemampuan yang lebih baik dalam membedakan karakteristik suatu dokumen.

Representasi numerik yang dihasilkan oleh TF-IDF selanjutnya digunakan sebagai vektor fitur pada proses klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM), Naive Bayes, Decision Tree, dan Random Forest. Sepuluh kata dengan skor TF-IDF tertinggi ditampilkan pada Tabel V.

TABEL V
KATA-KATA REPRESENTATIF BERDASARKAN SKOR TF-IDF

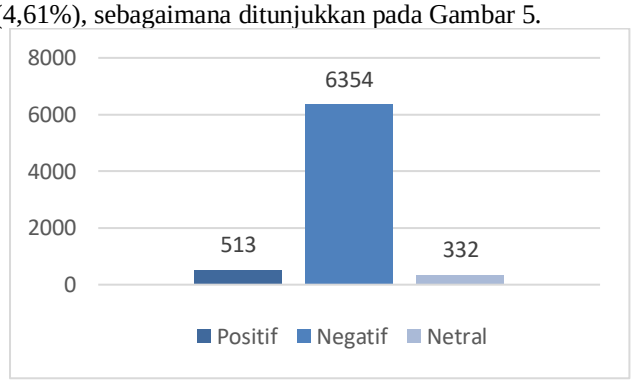
Kata	TF-IDF Score
banget	0.050759
lambat	0.033002
kesal	0.028417
ganggu	0.025715
wifi	0.022447
jelek	0.021970
jaring	0.021487
sinyal	0.018776
mati	0.013815
lag	0.013362

Berdasarkan hasil pembobotan TF-IDF, kata *banget* memperoleh skor tertinggi sebesar 0,050759, diikuti oleh kata *lambat* sebesar 0,033002, *kesal* sebesar 0,028417, *ganggu* sebesar 0,025715, *wifi* sebesar 0,022447, *jaring* sebesar 0,021487, *sinyal* sebesar 0,018776, *mati* sebesar 0,013815, dan *lag* sebesar 0,013362. Kata-kata tersebut didominasi oleh istilah yang berkaitan dengan kualitas jaringan internet, gangguan layanan, serta ekspresi pengguna ketika mengalami kendala dalam menggunakan layanan IndiHome. Sementara itu, kata *banget* berfungsi sebagai kata penguat (*intensifier*) yang mempertegas intensitas opini yang disampaikan pengguna sehingga turut memberikan kontribusi dalam membedakan karakteristik dokumen.

Hasil tersebut menunjukkan bahwa metode TF-IDF mampu memberikan bobot yang lebih besar pada kata-kata yang memiliki daya pembeda tinggi antardokumen. Dengan demikian, fitur yang dihasilkan tidak hanya merepresentasikan frekuensi kemunculan kata, tetapi juga mampu menangkap istilah-istilah yang relevan dalam membedakan sentimen pengguna terhadap layanan IndiHome sebelum digunakan pada proses klasifikasi.

F. Distribusi Data dan Penerapan SMOTE

Sebelum dilakukan proses pelatihan model, distribusi kelas pada dataset dianalisis untuk mengetahui tingkat ketidakseimbangan data. Berdasarkan hasil pelabelan, dataset terdiri atas 6.354 tweet berlabel negatif (88,26%), 513 tweet berlabel positif (7,13%), dan 332 tweet berlabel netral (4,61%), sebagaimana ditunjukkan pada Gambar 5.

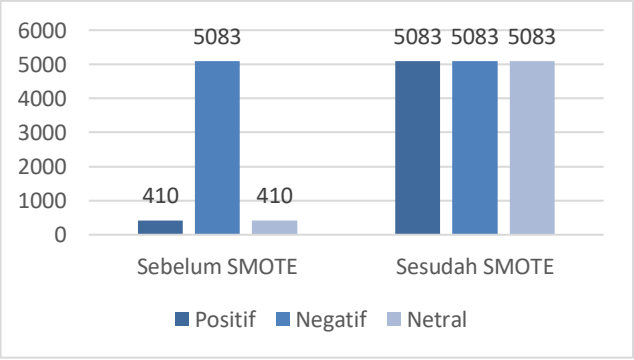


Gambar 5. Distribusi Data Setelah Dilakukan Proses Labeling

Dominasi sentimen negatif pada penelitian ini tidak serta-merta menggambarkan keseluruhan persepsi pelanggan IndiHome terhadap layanan yang diberikan. Distribusi yang sangat timpang ini lebih mencerminkan karakteristik Platform X sebagai kanal pengaduan dan keluhan, dibandingkan representasi proporsional dari seluruh opini pelanggan. Oleh karena itu, hasil penelitian ini tidak dapat digeneralisasi sebagai gambaran keseluruhan kepuasan pelanggan IndiHome secara absolut. Kondisi tersebut dipengaruhi oleh karakteristik Platform X yang lebih banyak dimanfaatkan pengguna untuk menyampaikan keluhan maupun laporan gangguan layanan. Selain itu, pengumpulan data menggunakan kata kunci "indihome" pada periode April 2025 hingga Maret 2026 juga berpotensi menghasilkan lebih banyak tweet yang berkaitan dengan permasalahan layanan dibandingkan pengalaman positif pengguna. Ketidakseimbangan distribusi kelas tersebut berpotensi menyebabkan model klasifikasi lebih cenderung mempelajari pola pada kelas mayoritas sehingga kemampuan model dalam mengenali kelas minoritas menjadi kurang optimal.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan metode SMOTE (Synthetic Minority Over-sampling Technique). Penerapan SMOTE dilakukan hanya pada data latih (training set) pada setiap fold dalam proses Stratified 5-Fold Cross Validation, sedangkan data uji (testing set) tetap menggunakan distribusi data asli. Pendekatan ini bertujuan untuk menghindari terjadinya data leakage sehingga hasil evaluasi model tetap merepresentasikan kemampuan model dalam mengklasifikasikan data yang belum pernah digunakan selama proses pelatihan.

Sebagai ilustrasi, distribusi data latih pada Fold 1 sebelum penerapan SMOTE terdiri atas 5.083 tweet berlabel negatif, 410 tweet berlabel positif, dan 266 tweet berlabel netral. Setelah proses oversampling menggunakan SMOTE, jumlah data pada kelas positif dan netral ditingkatkan hingga sama dengan jumlah kelas mayoritas, sehingga masing-masing kelas memiliki 5.083 data latih. Ilustrasi distribusi data latih sebelum dan sesudah penerapan SMOTE ditunjukkan pada Gambar 6.



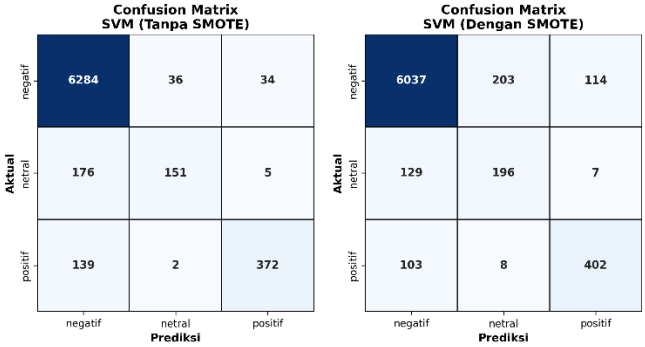
Gambar 6. Distribusi Data Latih Sebelum dan Sesudah SMOTE

Melalui proses oversampling tersebut, distribusi kelas pada data latih menjadi lebih seimbang sehingga setiap algoritma memperoleh kesempatan yang lebih proporsional untuk mempelajari karakteristik masing-masing kelas. Selanjutnya, data hasil oversampling digunakan pada proses pelatihan model menggunakan 4 algoritma yang telah direncanakan. sedangkan proses pengujian tetap dilakukan menggunakan data uji yang tidak mengalami proses oversampling.

Penerapan SMOTE hanya pada data latih pada setiap fold juga bertujuan untuk meminimalkan potensi overfitting yang dapat muncul akibat pembentukan sampel sintetis. Dengan mempertahankan distribusi asli pada data uji, hasil evaluasi tetap merepresentasikan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Hasil penelitian menunjukkan bahwa penerapan SMOTE tidak selalu meningkatkan accuracy pada seluruh algoritma. SVM, Naive Bayes, dan Decision Tree justru mengalami sedikit penurunan accuracy setelah proses oversampling, sedangkan peningkatan yang diperoleh lebih terlihat pada balanced accuracy, macro F1-score, dan kemampuan model dalam mengenali kelas minoritas. Temuan tersebut mengindikasikan bahwa pada konfigurasi eksperimen yang digunakan, penerapan SMOTE lebih berperan dalam mengurangi bias terhadap kelas mayoritas dibandingkan menunjukkan indikasi peningkatan performa yang disebabkan oleh overfitting.

G. Penerapan Algoritma Support Vector Machine (SVM)

Kemampuan Support Vector Machine (SVM) dalam memaksimalkan margin antarkelas pada ruang fitur berdimensi tinggi diuji untuk klasifikasi sentimen tweet (negatif, netral, positif) tanpa dan dengan SMOTE pada validasi 5-fold (Gambar 7).



Gambar 7. Confusion Matrix Algoritma SVM

Berdasarkan Gambar 7, model SVM tanpa SMOTE sangat baik dalam mengenali kelas negatif (mayoritas). Dari 6.354 tweet negatif, 6.284 berhasil diklasifikasikan dengan benar. Sebaliknya, kemampuan mengenali kelas minoritas masih rendah: hanya 151 dari 332 tweet netral dan 372 dari 513 tweet positif yang terprediksi benar, sebagian besar kesalahan klasifikasi terjadi ketika tweet netral maupun positif diprediksi sebagai sentimen negatif. Distribusi data yang tidak seimbang menyebabkan model lebih banyak mempelajari kelas mayoritas sehingga batas keputusan cenderung mengarah ke kelas negatif.

Hasil *classification report* memperkuat temuan tersebut. Kelas negatif memperoleh precision 95,23%, recall 98,90%, dan F1-score 97,03%. Sebaliknya, kelas netral memperoleh precision 79,89% namun recall hanya 45,48%, artinya ketika model memprediksi netral, prediksi tersebut umumnya benar,

tetapi masih banyak tweet netral yang tidak terdeteksi. Pola serupa terlihat pada kelas positif dengan precision 90,51% dan recall 72,51%. Dengan demikian, performa tinggi SVM tanpa SMOTE masih didominasi oleh kelas mayoritas.

Setelah SMOTE, kemampuan model dalam mengenali kelas minoritas meningkat. Tweet netral yang terprediksi benar meningkat menjadi 196, dan tweet positif menjadi 402. Namun, prediksi benar pada kelas negatif menurun menjadi 6.037 tweet karena peningkatan sensitivitas terhadap kelas minoritas.

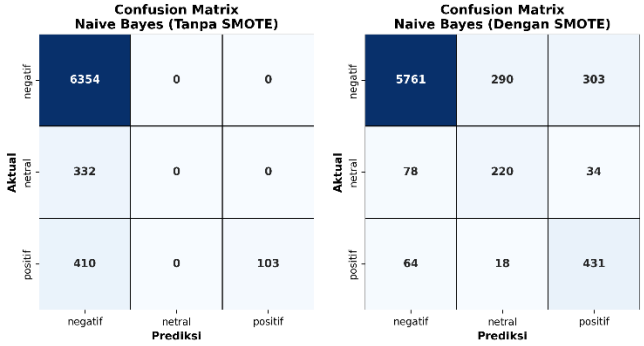
Perubahan tercermin pada *classification report*: recall netral meningkat dari 45,48% menjadi 59,04%, recall positif dari 72,51% menjadi 78,36%, diikuti penurunan precision pada kedua kelas. Ini menunjukkan *trade-off* antara kemampuan menemukan data minoritas dengan meningkatnya *false positive*. Setelah distribusi data lebih seimbang, model menjadi lebih sensitif terhadap kelas netral dan positif, meskipun sebagian tweet negatif ikut terprediksi sebagai kelas minoritas.

Evaluasi 5-fold menunjukkan accuracy menurun dari 94,55% menjadi 92,17% setelah SMOTE. Namun, balanced accuracy meningkat dari 0,7230 menjadi 0,7747, dan macro F1-score berubah dari 0,7850 menjadi 0,7543, sementara weighted F1-score menurun dari 94,05% menjadi 92,40%. Metrik seperti balanced accuracy dan macro F1-score lebih representatif untuk data tidak seimbang dibandingkan accuracy saja.

Secara keseluruhan, penerapan SMOTE pada SVM bertujuan meningkatkan kemampuan mengenali kelas minoritas, bukan meningkatkan accuracy. Evaluasi model pada data tidak seimbang perlu mempertimbangkan berbagai metrik secara simultan agar kualitas klasifikasi setiap kelas dapat dinilai lebih objektif.

H. Penerapan Algoritma Naive Bayes

Naive Bayes, yang mengandalkan teorema probabilitas dengan asumsi independensi antar fitur, diterapkan untuk mengklasifikasikan sentimen tweet (negatif, netral, positif) dalam dua skenario: tanpa dan dengan SMOTE pada validasi 5-fold (Gambar 8).



Gambar 8. Confusion Matrix Algoritma Naive Bayes

Berdasarkan Gambar 8, Naive Bayes tanpa SMOTE menunjukkan kecenderungan sangat kuat mengklasifikasikan tweet ke kelas negatif (mayoritas). Seluruh 6.354 tweet negatif berhasil dikenali, namun kemampuan mengenali kelas minoritas sangat rendah: tidak ada satu pun dari 332 tweet netral yang terklasifikasi benar, dan hanya 103 dari 513 tweet positif yang dikenali sesuai kelas. Sebagian besar tweet netral dan positif diprediksi sebagai negatif. Distribusi data yang sangat timpang menyebabkan model memprioritaskan kelas mayoritas sehingga variasi karakteristik kelas minoritas tidak dapat dipelajari secara memadai.

Hasil *classification report* memperkuat kecenderungan tersebut. Kelas negatif memperoleh precision 89,54%, recall 100%, dan F1-score 94,48%, menandakan hampir seluruh prediksi berpusat pada kelas negatif. Sebaliknya, kelas netral memiliki precision, recall, dan F1-score 0%, artinya model gagal total mengenali kelas tersebut. Pada kelas positif, precision mencapai 100% namun recall hanya 20,08%, artinya model hanya memprediksi positif pada tweet dengan karakteristik sangat kuat, sementara sebagian besar tweet positif lainnya tetap diklasifikasikan sebagai negatif. Performa Naive Bayes tanpa SMOTE lebih didominasi kelas mayoritas daripada kemampuan membedakan ketiga kelas secara seimbang.

Setelah SMOTE, kemampuan model mengenali kelas minoritas meningkat signifikan. Tweet netral yang terdeteksi naik menjadi 220, tweet positif menjadi 431, sedangkan prediksi benar pada kelas negatif turun menjadi 5.760. Penambahan sampel sintetis memperkaya representasi kelas minoritas sehingga model tidak lagi terlalu bergantung pada distribusi kelas negatif.

Perubahan tercermin pada *classification report*: recall netral melonjak dari 0% menjadi 66,27%, recall positif dari 20,08% menjadi 84,02%, menunjukkan SMOTE meningkatkan sensitivitas terhadap kelas minoritas. Namun, precision netral turun menjadi 41,67% dan precision positif menjadi 56,12%, menandakan model menjadi lebih agresif memprediksi kelas minoritas sehingga *false positive* meningkat. Penerapan SMOTE berhasil memperbaiki deteksi kelas minoritas dengan konsekuensi penurunan ketepatan prediksi.

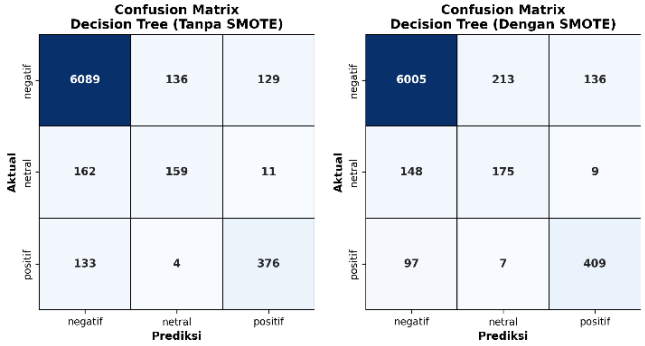
Evaluasi 5-fold menunjukkan accuracy turun tipis dari 89,69% menjadi 89,07% setelah SMOTE. Sebaliknya, macro F1-score meningkat dari 0,4264 menjadi 0,7082 dan balanced accuracy dari 0,4003 menjadi 0,8032, mengindikasikan perbaikan keseimbangan klasifikasi antar kelas. Weighted F1-score naik dari 85,78% menjadi 90,12%, dan MCC dari 0,3338 menjadi 0,6050, menunjukkan kualitas klasifikasi keseluruhan lebih baik setelah data diseimbangkan.

Naive Bayes merupakan algoritma paling sensitif terhadap ketidakseimbangan data dibandingkan algoritma lain dalam penelitian ini. Tanpa SMOTE, model hampir sepenuhnya berfokus pada kelas mayoritas dan gagal mengenali kelas netral. Setelah data diseimbangkan, kemampuan mengenali kelas minoritas meningkat drastis, meskipun diikuti penurunan precision. SMOTE memberikan manfaat lebih nyata pada Naive Bayes dibandingkan kondisi tanpa penyeimbangan. Hasil ini selanjutnya dibandingkan dengan Decision Tree dan Random Forest untuk memperoleh

gambaran komprehensif pengaruh SMOTE pada masing-masing metode.

I. Penerapan Algoritma Decision Tree

Decision Tree, yang membentuk aturan klasifikasi secara hierarkis berdasarkan atribut paling informatif, digunakan untuk mengklasifikasikan sentimen tweet (negatif, netral, positif) tanpa dan dengan SMOTE pada validasi 5-fold (Gambar 9).



Gambar 9. Confusion Matrix Algoritma Decision Tree

Berdasarkan Gambar 9, Decision Tree menunjukkan kemampuan klasifikasi relatif lebih seimbang dibandingkan Naive Bayes meskipun menggunakan distribusi data asli. Dari 6.354 tweet negatif, 6.089 berhasil diklasifikasikan benar. Pada kelas netral, 159 dari 332 tweet teridentifikasi sesuai label, dan pada kelas positif 376 dari 513 tweet terprediksi benar. Meskipun sebagian tweet netral dan positif masih terklasifikasi sebagai negatif, jumlah prediksi benar pada kedua kelas minoritas menunjukkan Decision Tree mampu membentuk aturan klasifikasi lebih baik terhadap kelas minoritas.

Hasil *classification report* memperkuat temuan. Kelas negatif memperoleh precision 95,38%, recall 95,83%, dan F1-score 95,60%. Kelas netral memperoleh precision 53,18%, recall 47,89%, dan F1-score 50,40%, sedangkan kelas positif memperoleh precision 72,87%, recall 73,29%, dan F1-score 73,08%. Nilai precision dan recall yang relatif berdekatan pada setiap kelas menunjukkan Decision Tree menghasilkan klasifikasi lebih seimbang daripada Naive Bayes, meskipun performa kelas minoritas masih dipengaruhi ketidakseimbangan data.

Setelah SMOTE, kemampuan model mengenali kelas minoritas meningkat. Tweet netral yang terklasifikasi benar naik menjadi 175, dan tweet positif menjadi 409, sementara prediksi benar pada kelas negatif turun menjadi 6.005. Penambahan sampel sintetis membantu model mempelajari pola kelas minoritas tanpa mengubah karakteristik Decision Tree secara drastis.

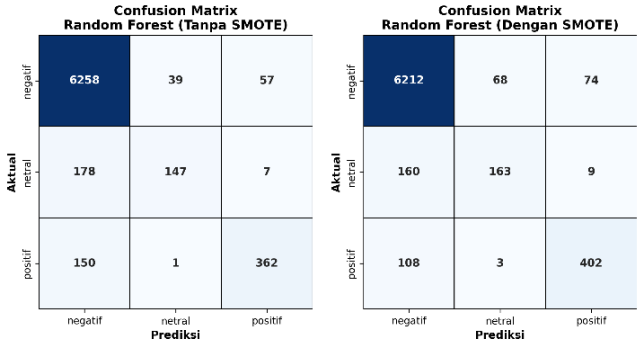
Perubahan tercermin pada *classification report*: recall netral meningkat dari 47,89% menjadi 52,71%, recall positif dari 73,29% menjadi 79,73%, sementara precision netral turun menjadi 44,30% dan precision positif relatif stabil pada 73,83%. Penerapan SMOTE meningkatkan sensitivitas terhadap kelas minoritas dengan konsekuensi peningkatan prediksi kurang tepat pada kelas netral. Dibandingkan Naive Bayes, perubahan performa Decision Tree lebih moderat sehingga keseimbangan precision dan recall tetap terjaga.

Evaluasi 5-fold menunjukkan accuracy turun tipis dari 92,01% menjadi 91,53% setelah SMOTE. Balanced accuracy meningkat dari 0,7234 menjadi 0,7565, macro F1-score dari 0,7303 menjadi 0,7336, weighted F1-score hanya berubah kecil dari 91,91% menjadi 91,79%, dan MCC meningkat dari 0,6203 menjadi 0,6259. SMOTE tidak meningkatkan accuracy secara berarti, tetapi memperbaiki keseimbangan klasifikasi antar kelas dengan tetap mempertahankan performa keseluruhan.

Decision Tree menunjukkan performa relatif stabil sebelum maupun sesudah SMOTE. Berbeda dengan Naive Bayes yang sangat dipengaruhi ketidakseimbangan data, Decision Tree telah mampu mengenali sebagian besar kelas minoritas bahkan sebelum penyeimbangan. SMOTE tetap memberikan manfaat peningkatan klasifikasi pada kelas netral dan positif, namun peningkatannya tidak sedrastis Naive Bayes karena Decision Tree sejak awal memiliki kemampuan lebih seimbang. Karakteristik ini menjadi dasar perbandingan dengan Random Forest pada pembahasan berikutnya.

J. Penerapan Algoritma Random Forest

Random Forest, yang menggabungkan berbagai pohon keputusan melalui mekanisme *ensemble*, diterapkan untuk klasifikasi sentimen tweet (negatif, netral, positif) tanpa dan dengan SMOTE pada validasi 5-fold (Gambar 10).



Gambar 10. Confusion Matrix Algoritma Random Forest

Berdasarkan Gambar 10, Random Forest tanpa SMOTE telah menunjukkan kemampuan klasifikasi tinggi pada ketiga kelas. Dari 6.354 tweet negatif, 6.258 berhasil diklasifikasikan benar. Pada kelas netral, 147 dari 332 tweet teridentifikasi sesuai label, dan pada kelas positif 362 dari 513 tweet terprediksi benar. Meskipun sebagian besar kesalahan masih mengarah ke kelas negatif, model mampu mempertahankan kemampuan baik dalam mengenali kelas positif tanpa mengorbankan performa mayoritas. Random Forest telah membentuk pola klasifikasi relatif stabil meskipun dilatih dengan data tidak seimbang.

Hasil *classification report* memperkuat temuan. Kelas negatif memperoleh precision 95,02%, recall 98,49%, dan F1-score 96,72%. Kelas netral memperoleh precision 78,61%

namun recall hanya 44,28%, artinya model cukup akurat saat memprediksi netral tetapi masih belum mampu menemukan sebagian besar tweet kelas tersebut. Kelas positif memperoleh precision 84,98%, recall 70,57%, dan F1-score 77,10%. Random Forest memiliki kemampuan klasifikasi baik terhadap kelas minoritas, meskipun dominasi kelas negatif masih terlihat terutama pada recall netral.

Setelah SMOTE, kemampuan model mengenali kelas minoritas meningkat tanpa mengurangi performa keseluruhan secara berarti. Tweet positif yang terklasifikasi benar naik menjadi 402, dan tweet netral menjadi 163, sementara prediksi benar pada kelas negatif hanya turun sedikit menjadi 6.212. Penyeimbangan data meningkatkan sensitivitas terhadap kelas minoritas dengan tetap mempertahankan kemampuan klasifikasi pada kelas mayoritas.

Perubahan tercermin pada *classification report*: recall positif meningkat dari 70,57% menjadi 78,36%, diikuti perubahan precision dari 84,98% menjadi 82,89%. Penurunan precision kecil ini merupakan konsekuensi wajar dari peningkatan sensitivitas terhadap kelas minoritas. Pada kelas netral, recall meningkat dari 44,28% menjadi 49,10% sementara precision turun menjadi 69,66%. Dibandingkan algoritma lain, Random Forest mampu meningkatkan deteksi kelas minoritas tanpa penurunan precision terlalu besar, sehingga *oversampling* menghasilkan peningkatan performa lebih seimbang.

Evaluasi 5-fold menunjukkan accuracy meningkat dari 94,00% menjadi 94,14% setelah SMOTE. Macro F1-score naik dari 0,7682 menjadi 0,7832, balanced accuracy dari 0,7111 menjadi 0,7507, weighted F1-score dari 93,48% menjadi 93,84%, dan MCC dari 0,6897 menjadi 0,7086. Berbeda dengan algoritma lain yang cenderung menurun accuracy-nya, Random Forest justru mempertahankan bahkan sedikit meningkatkan performa pada hampir seluruh metrik.

Random Forest memperoleh manfaat paling konsisten dari SMOTE. Peningkatan tidak hanya pada kemampuan mengenali kelas minoritas, tetapi juga tercermin pada berbagai metrik tanpa mengorbankan accuracy. Mekanisme *ensemble* yang menggabungkan banyak pohon keputusan memungkinkan Random Forest memanfaatkan data seimbang secara lebih efektif, menghasilkan performa lebih stabil pada dataset tidak seimbang. Hasil ini menjadi dasar perbandingan seluruh algoritma untuk menentukan model paling sesuai pada konfigurasi penelitian ini.

K. Perbandingan Algoritma

Perbandingan performa dilakukan untuk mengevaluasi pengaruh penerapan SMOTE terhadap setiap algoritma klasifikasi pada dataset yang memiliki distribusi kelas tidak seimbang. Evaluasi tidak hanya didasarkan pada accuracy, tetapi juga mempertimbangkan weighted F1-score, macro F1-score, balanced accuracy, dan Matthews Correlation Coefficient (MCC) agar kemampuan model dalam mengenali seluruh kelas sentimen dapat dievaluasi secara lebih komprehensif. Ringkasan hasil evaluasi sebelum dan sesudah penerapan SMOTE beserta perubahan (Δ) setiap metrik disajikan pada Tabel VI.

TABEL VI
PERBANDINGAN SEBELUM VS SETELAH SMOTE

Model	Metrik	Sebelum SMOTE	Setelah SMOTE	Delta (Δ)
Support Vector Machine (SVM)	Accuracy (%)	94,55%	92,17%	-2,38%
	Weighted F1 (%)	94,05%	92,40%	-1,65%
	Macro F1	0,7850	0,7543	-0,0307
	Balanced Accuracy	0,7230	0,7747	0,0517
	MCC	0,7190	0,6506	-0,0684
Naive Bayes	Accuracy (%)	89,69%	89,07%	-0,62%
	Weighted F1 (%)	85,78%	90,12%	4,34%
	Macro F1	0,4264	0,7082	0,2818
	Balanced Accuracy	0,4003	0,8032	0,4029
	MCC	0,3338	0,6050	0,2712
Decision Tree	Accuracy (%)	92,01%	91,53%	-0,48%
	Weighted F1 (%)	91,91%	91,79%	-0,12%
	Macro F1	0,7303	0,7336	0,0033
	Balanced Accuracy	0,7234	0,7565	0,0331
	MCC	0,6203	0,6259	0,0056
Random Forest	Accuracy (%)	94,00%	94,14%	0,14%
	Weighted F1 (%)	93,48%	93,84%	0,36%
	Macro F1	0,7682	0,7832	0,0150
	Balanced Accuracy	0,7111	0,7507	0,0396
	MCC	0,6897	0,7086	0,0189

Berdasarkan Tabel VI, penerapan SMOTE memberikan pengaruh yang berbeda pada setiap algoritma. SVM memperoleh accuracy tertinggi sebelum penerapan SMOTE sebesar 94,55%, namun mengalami penurunan sebesar 2,38% setelah data diseimbangkan. Meskipun demikian, balanced accuracy meningkat dari 0,7230 menjadi 0,7747, menunjukkan bahwa kemampuan model dalam mengenali kelas minoritas menjadi lebih baik. Hasil ini menunjukkan bahwa penurunan accuracy tidak selalu menunjukkan penurunan kualitas model, karena proses penyeimbangan data berhasil mengurangi bias terhadap kelas mayoritas.

Naive Bayes merupakan algoritma yang memperoleh dampak paling besar dari penerapan SMOTE. Meskipun accuracy hanya berubah sebesar −0,62%, macro F1-score meningkat sebesar 0,2818, balanced accuracy meningkat sebesar 0,4029, dan MCC meningkat sebesar 0,2712. Peningkatan tersebut menunjukkan bahwa SMOTE secara signifikan meningkatkan kemampuan Naive Bayes dalam mengenali kelas minoritas yang sebelumnya sangat dipengaruhi oleh dominasi kelas negatif. Sementara itu,

Decision Tree menunjukkan perubahan performa yang relatif kecil pada seluruh metrik evaluasi, sehingga mengindikasikan bahwa algoritma ini telah memiliki kemampuan klasifikasi yang cukup stabil meskipun sebelum proses penyeimbangan data dilakukan.

Berbeda dengan algoritma lainnya, Random Forest merupakan satu-satunya algoritma yang mengalami peningkatan pada seluruh metrik evaluasi setelah penerapan SMOTE. Accuracy meningkat sebesar 0,14%, weighted F1-score sebesar 0,36%, macro F1-score sebesar 0,0150, balanced accuracy sebesar 0,0396, serta MCC sebesar 0,0189. Konsistensi peningkatan tersebut menunjukkan bahwa Random Forest mampu memanfaatkan distribusi data yang telah diseimbangkan secara lebih efektif dibandingkan algoritma lainnya. Secara nominal, accuracy Random Forest dengan SMOTE (94,14%) masih sedikit lebih rendah dibandingkan SVM tanpa SMOTE (94,55%), sehingga pemilihan model terbaik tidak didasarkan pada accuracy tertinggi secara absolut, melainkan pada konsistensi peningkatan di seluruh metrik evaluasi. Berdasarkan kombinasi seluruh metrik, Random Forest dengan SMOTE dipilih sebagai model dengan performa paling konsisten pada konfigurasi penelitian ini.

Perbedaan respons terhadap SMOTE antar algoritma disebabkan oleh karakteristik masing-masing model. Naive Bayes, yang mengasumsikan independensi antar fitur, lebih rentan terhadap *noise* yang diperkenalkan oleh sampel sintetis karena interpolasi fitur dapat melanggar asumsi independensi tersebut, sehingga peningkatan performa paling drastis justru terjadi pada algoritma ini. Sebaliknya, Random Forest memanfaatkan mekanisme *ensemble* yang menggabungkan berbagai pohon keputusan, sehingga reduksi variansi yang dihasilkan membuatnya lebih *robust* terhadap fluktuasi distribusi data akibat *oversampling*. Hal ini menjelaskan mengapa Random Forest merupakan satu-satunya algoritma yang menunjukkan peningkatan pada seluruh metrik evaluasi.

Temuan tersebut terbatas pada dataset, periode pengambilan data, serta konfigurasi eksperimen yang digunakan, sehingga masih memerlukan validasi lebih lanjut pada dataset dan domain yang berbeda. Tidak ditemukan indikasi overfitting pada model yang diuji karena SMOTE diterapkan hanya pada data latih di setiap fold, sementara evaluasi tetap menggunakan data uji dengan distribusi asli. Konsistensi performa pada data uji menunjukkan bahwa peningkatan kemampuan klasifikasi kelas minoritas tidak semata-mata disebabkan oleh overfitting terhadap sampel sintetis.

TABEL VII
HASIL UJI SIGNIFIKANSI WILCOXON SIGNED-RANK TEST

Pasangan Algoritma	p-value Tanpa SMOTE	p-value Dengan SMOTE	Keputusan
SVM vs Naive Bayes	0,0625	0,0625	Tidak signifikan
SVM vs Decision Tree	0,0625	0,3125	Tidak signifikan
SVM vs Random Forest	0,1250	0,0625	Tidak signifikan
Naive Bayes vs Decision Tree	0,0625	0,0625	Tidak signifikan
Naive Bayes vs Random Forest	0,0625	0,0625	Tidak signifikan
Decision Tree vs Random Forest	0,0625	0,0625	Tidak signifikan

Keterangan: $\alpha = 0{,}05$.

Untuk mengevaluasi signifikansi statistik perbedaan performa antar algoritma, dilakukan uji Wilcoxon Signed-Rank Test berdasarkan hasil Stratified 5-Fold Cross Validation. Hasil pengujian disajikan pada Tabel VII. Berdasarkan hasil uji pada Tabel VII, seluruh pasangan algoritma menghasilkan nilai *p* lebih besar dari 0,05, baik sebelum maupun setelah penerapan SMOTE. Hasil tersebut menunjukkan bahwa tidak terdapat perbedaan performa yang signifikan secara statistik antar algoritma pada tingkat signifikansi 5%. Oleh karena itu, pemilihan Random Forest sebagai model terbaik dalam penelitian ini tidak didasarkan pada adanya perbedaan performa yang signifikan secara statistik, melainkan pada konsistensi peningkatan seluruh metrik evaluasi setelah penerapan SMOTE, yaitu accuracy, weighted F1-score, macro F1-score, balanced accuracy, dan Matthews Correlation Coefficient (MCC) pada konfigurasi eksperimen yang digunakan.

Secara praktis, temuan ini dapat dimanfaatkan oleh penyedia layanan IndiHome dalam beberapa hal. Dominasi kata seperti lambat, ganggu, mati, putus, dan sinyal pada kelas sentimen negatif (lihat Tabel V dan Gambar 4) mengindikasikan keluhan pelanggan sebagian besar berkaitan dengan stabilitas koneksi dan kualitas jaringan. Informasi tersebut dapat digunakan sebagai dasar prioritas perbaikan infrastruktur maupun evaluasi kualitas layanan di wilayah yang sering menerima keluhan. Selain itu, kemampuan Random Forest dengan SMOTE dalam mengenali kelas minoritas secara lebih seimbang berpotensi dimanfaatkan untuk membangun sistem pemantauan sentimen pelanggan otomatis, sehingga keluhan maupun apresiasi terdeteksi lebih cepat, mendukung pengambilan keputusan berbasis data, dan membantu tim layanan pelanggan menentukan prioritas penanganan masalah.

IV. KESIMPULAN

Penelitian ini berhasil membandingkan performa empat algoritma *machine learning*, yaitu Support Vector Machine (SVM), Naive Bayes, Decision Tree, dan Random Forest, dalam klasifikasi sentimen pengguna IndiHome pada

Platform X menggunakan representasi fitur TF-IDF dengan dan tanpa penerapan SMOTE. Hasil penelitian menunjukkan bahwa ketidakseimbangan distribusi data berpengaruh terhadap kemampuan model dalam mengenali kelas minoritas. Penerapan SMOTE terbukti meningkatkan kemampuan klasifikasi pada kelas minoritas yang ditunjukkan melalui peningkatan nilai balanced accuracy, macro F1-score, dan Matthews Correlation Coefficient (MCC) pada sebagian besar algoritma, meskipun tidak selalu diikuti oleh peningkatan accuracy.

SVM memperoleh accuracy tertinggi pada kondisi tanpa SMOTE, namun peningkatan kemampuan klasifikasi setelah penerapan SMOTE tidak terjadi secara konsisten pada seluruh metrik evaluasi. Sebaliknya, Random Forest merupakan satu-satunya algoritma yang menunjukkan peningkatan pada seluruh metrik evaluasi, sehingga menghasilkan performa yang paling konsisten dibandingkan algoritma lainnya pada konfigurasi penelitian ini. Meskipun uji Wilcoxon tidak menunjukkan perbedaan performa yang signifikan secara statistik antar algoritma, konsistensi Random Forest dalam meningkatkan seluruh metrik evaluasi (accuracy, weighted F1, macro F1, balanced accuracy, dan MCC) menjadikannya model paling stabil dan direkomendasikan secara praktis untuk penanganan data tidak seimbang pada konteks ini. Temuan tersebut menunjukkan bahwa pemilihan model pada dataset tidak seimbang sebaiknya tidak hanya didasarkan pada accuracy, tetapi juga mempertimbangkan metrik evaluasi lain yang mampu menggambarkan performa klasifikasi pada seluruh kelas secara lebih objektif.

Penelitian ini memberikan kontribusi berupa evaluasi komprehensif terhadap pengaruh SMOTE pada beberapa algoritma machine learning untuk klasifikasi sentimen berbahasa Indonesia, sehingga dapat menjadi referensi dalam pemilihan model klasifikasi pada dataset dengan distribusi kelas yang tidak seimbang. Namun demikian, penelitian ini memiliki beberapa keterbatasan. Representasi fitur TF-IDF yang digunakan hanya mempertimbangkan frekuensi kemunculan kata tanpa menangkap konteks semantik maupun hubungan antar kata dalam kalimat, sehingga model berpotensi kurang sensitif terhadap nuansa makna seperti sarkasme atau struktur kalimat yang kompleks. Algoritma klasifikasi yang digunakan juga masih bersifat konvensional dan belum dibandingkan dengan pendekatan berbasis transformer yang lebih mutakhir. Selain itu, hasil penelitian ini terbatas pada dataset tweet yang diperoleh menggunakan satu kata kunci pencarian ("IndiHome") pada Platform X selama periode pengambilan data serta konfigurasi eksperimen yang digunakan, sehingga belum sepenuhnya merepresentasikan keseluruhan opini pelanggan yang disampaikan tanpa menyertakan kata kunci tersebut secara eksplisit. Oleh karena itu, penelitian selanjutnya dapat mengevaluasi teknik penyeimbangan data lainnya, memanfaatkan representasi fitur berbasis transformer seperti IndoBERT, serta melakukan validasi pada dataset dan periode pengambilan data yang berbeda untuk memperoleh hasil yang lebih general.

DAFTAR PUSTAKA

[1] Ardiyansah and Parjito, "Perbandingan Metode Naïve Bayes dan Support Vector Machine Dalam Analisis Sentimen Terhadap Tokoh Publik," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 6, pp. 2813–2821, Jun. 2024.

[2] Diana Puspitasari and Tata Sutabri, "Analisis Sentimen Berdasarkan pada Twitter (X) terhadap Layanan Indihome Menggunakan Algoritma Support Vector Machine (SVM)," *JUMINTAL: Jurnal Manajemen Informatika dan Bisnis Digital*, vol. 3, no. 2, 2024, doi: 10.55123/jumintal.v3i2.4449.

[3] H. H. Limbong and Norhikmah, "Optimasi Analisis Sentimen Ulasan Aplikasi Amikom One Menggunakan SMOTE pada Algoritma Artificial Neural Network," *Sistematis : Jurnal Sistem Informasi*, vol. 13, no. 5, 2024.

[4] D. Kurniawan and M. N. D. Satria, "Analisis Sentimen Opini Publik Tentang Gempa Megathrust di Indonesia Menggunakan Metode Support Vector Machine dan Naïve Bayes," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, 2024, doi: 10.47065/bits.v6i3.6213.

[5] S. Adiyono and M. Arifin, "AI and Digital Transformation Trends: A Systematic Review with Multi-Criteria Analysis," *Journal of Information Systems and Informatics*, vol. 8, no. 1, pp. 1011–1044, Mar. 2026, doi: 10.63158/journalisi.v8i1.1462.

[6] N. Malasari and M. Ramli, "Analisis Sentimen Media Sosial Menggunakan Algoritma BERT dan LSTM," *Journal of Computer Science and Information Technology*, vol. 1, no. 3, 2025.

[7] L. Cahyaningrum, A. Luthfiarta, and M. Rahayu, "Sentiment Analysis on the Impact of MBKM on Student Organizations Using Supervised Learning with Smote to Handle Data Imbalance," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, 2024, doi: 10.25139/inform.v9i1.7484.

[8] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, 2022, doi: 10.30812/matrik.v21i3.1726.

[9] S. E. Prianto, B. Berlilana, and R. E. Saputro, "Analisis Sentimen Program Makan Bergizi Gratis Menggunakan Random Forest dan Support Vector Machine dengan Penyeimbangan Data Berbasis SMOTE," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 6, no. 2, pp. 353–365, Mar. 2026, doi: 10.52436/1.jpti.1491.

[10] F. Panjaitan *et al.*, "Studi Komparatif Algoritma Machine Learning Pada Analisis Sentimen Media Sosial," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, 2025, doi: 10.36040/jati.v9i2.13277.

[11] M. Arifin and S. Adiyono, "Hyperparameter Tuning in Machine Learning to Predicting Student Academic Achievement," *International Journal of Artificial Intelegence Research*, vol. 8, no. 1, 2024, doi: https://doi.org/10.29099/ijair.v8i1.1.1214.

[12] N. Fajriyah, N. T. Lapatta, D. W. Nugraha, and R. Laila, "Implementasi Svm Dan Smote Pada Analisis Sentimen Media Sosial X Terhadap Pelantikan Agus Harimurti Yudhoyono," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, 2025, doi: 10.29100/jipi.v10i2.6246.

[13] A. F. Anjani, D. Anggraeni, and I. M. Tirta, "Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, 2023, doi: 10.25077/teknosi.v9i2.2023.163-172.

[14] S. Adiyono and T. Hidayat, "Predictive Health Monitoring of a Marine Propulsion Plant Using Multi-Output Random Forest Regression," *International Journal of Marine Engineering Innovation and Research*, vol. 11, no. 1, pp. 248–258, 2026.

[15] M. Arifin, S. Adiyono, and A. Kurniawan, "Advanced Machine Learning Approaches to Combat Rice Challenges: A Comparative Study of SMO and PSO-ACO," *International Journal on Technical and Physical Problems of Engineering*, vol. 17, no. 4, pp. 40–50, 2025.

[16] C. Mulia and A. Kurniasih, "Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Bank Customer Churn Menggunakan Algoritma Naïve bayes dan Logistic Regression," *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, vol. 4, no. 2, 2023.

[17] P. P. Putra, M. K. Anam, A. S. Chan, A. Hadi, N. Hendri, and A. Masnur, "Optimizing Sentiment Analysis on Imbalanced Hotel Review Data Using SMOTE and Ensemble Machine Learning Techniques," *Journal of Applied Data Sciences*, vol. 6, no. 2, 2025, doi: 10.47738/jads.v6i2.618.

[18] A. Annur Rohman, G. Alfa Trisnapradika, and K. Kunci, "Perbandingan Algoritma NBC, SVM, Logistic Regression untuk Analisis Sentimen Terhadap Wacana KaburAjaDulu di Media Sosial X," *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 1, 2025.

[19] M. Nur Akbar, N. A. S. Yusuf, N. Nasrullah, and M. Mubarak, "Analisis Sentimen Pengguna Indihome dengan Metode Klasifikasi Support Vector Machine (SVM)," *Journal Software, Hardware and Information Technology*, vol. 2, no. 1, pp. 13–21, Jan. 2022, doi: 10.24252/shift.v2i1.18.

[20] A. R. Muhammad Fikri, J. Jondri, and W. Astuti, "Sentiment Analysis Against IndiHome and First Media Internet Providers Using Ensemble Stacking Method," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, Sep. 2022, doi: 10.47065/bits.v4i2.1969.

[21] A. N. Aini, J. Jondri, and W. Astuti, "Sentiment Analysis on Twitter Against IndiHome Providers Using Chi-Square and Ensemble Bagging Methods," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, pp. 516–521, Sep. 2022, doi: 10.47065/bits.v4i2.1967.

[22] Alwin Marcellino, Yosefa Camilia Moniung, and Rusbandi, "Penerapan Teknik SMOTE Pada Analisis Sentimen Bea Cukai Menggunakan Algoritma Naïve Bayes," *Jurnal Algoritme*, vol. 4, no. 2, 2022.

[23] S. Adiyono, N. Latifah, and D. Laily Fithri, "Analysis of Digital Readiness in the Social Assistance Distribution System with the Unified Theory of Acceptance and Use of Technology (UTAUT)," *Journal of Applied Informatics and Computing*, vol. 9, no. 2, pp. 483–489, Mar. 2025, doi: 10.30871/jaic.v9i2.9070.

[24] I. Abdul Ghofur and A. D. Putri, "Analisis Perbandingan Svm, Knn Dan Naïve Bayes Pada Sentiment Analysis Tweet Makan Bergizi Gratis," *Computer Based Information System Journal*, vol. 14, no. 1, pp. 100–117, Mar. 2026, doi: 10.33884/cbis.v14i1.10994.

[25] S. Dermawan and A. T. Ayunda, "Sentiment Analysis of Coretax on Social Media X Using Naive Bayes, SVM, and LSTM for Service Improvement," *Journal of Applied Informatics and Computing*, vol. 9, no. 6, 2025, doi: 10.30871/jaic.v9i6.11063.

[26] D. Saputra and M. R. Pribadi, "Analisis Sentimen Masyarakat terhadap Layanan Provider Internet di Indonesia menggunakan SVM," *MDP Student Conference*, vol. 2, no. 1, pp. 32–41, Apr. 2023, doi: 10.35957/mdp-sc.v2i1.4554.

[27] A. Alim Murtopo, M. Aditdya, P. Septiana Ananda, and G. Gunawan, "Penerapan Computer Vision Untuk Mendeteksi Kelengkapan Atribut Siswa Menggunakan Metode CNN," *PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, vol. 11, no. 2, pp. 247–258, Sep. 2024, doi: 10.30656/prosisko.v11i2.8752.

[28] N. F. Octavia and Berlilana, "Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Ulasan Produk Elektronik pada Platform E-Commerce," *Jurnal Algoritma*, vol. 22, no. 2, Nov. 2025, doi: 10.33364/algoritma/v.22-2.3083.

[29] A. Setiawan and H. Yuliansyah, "Analisis Sentimen Berbasis Aspek Terhadap Ulasan Pengguna pada Game Honkai: Star Rail Menggunakan Naïve Bayes Classifier," *SISTEMASI: Jurnal Sistem Informasi*, vol. 13, no. 5, 2024.

[30] D. Septiani and I. Isabela, "Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks," *Sintesia*, vol. 1, 2022.

[31] I. K. Dharmendra, I. M. Agus, W. Putra, and Y. P. Atmojo, "Evaluasi Efektivitas SMOTE dan Random Under Sampling pada Klasifikasi Emosi Tweet," *Informatics for Educators And Professionals : Journal of Informatics*, vol. 9, no. 2, 2024.

[32] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, p. 159, Mar. 1977, doi: 10.2307/2529310.